Efficient and Accurate Spectroscopic Simulations with Symmetry-Preserving Neural Network Models for Tensorial Properties

Yaolong Zhang, Sheng Ye, Jinxiao Zhang, Jun Jiang, Bin Jiang*

*Hefei National Laboratory for Physical Science at the Microscale, Department of Chemical Physics, Key Laboratory of Surface and Interface Chemistry and Energy Catalysis of Anhui Higher Education Institutes, University of Science and Technology of China, Hefei, Anhui 230026, China

*: corresponding author: bjiangch@ustc.edu.cn
Abstract

Machine learning (ML) methods have revolutionized the high-dimensional representations for scalar molecular properties such as potential energy. However, there have been scarce ML models targeting tensorial properties, which suffer from intrinsic covariant transformation with regard to rotation. Here, we generalize an atomistic neural network (NN) model to learn both tensorial response and transition properties, in which atomic position vectors are multiplied with scalar NN outputs or their derivatives to preserve the rotationally covariant symmetry. This strategy keeps these geometric descriptors symmetry invariant so that the resulting tensorial NN models are as efficient as their scalar counterparts. We validate the performance and universality of this approach by learning response properties of water oligomers and liquid water. More impressively, by learning transition dipole moments of a model structural unit of proteins, the ab initio based ultraviolet spectra of real proteins are faithfully reproduced with ~10^6 times speedup. This concept can be readily adapted to any atomistic NN framework to predict tensorial properties, promising efficient and accurate spectroscopic simulations for biomolecules and materials.
Introduction

Machine learning (ML) techniques have shown great promise in solving challenging problems in physics, chemistry and material science\textsuperscript{1,2}. Although various ML methods are well-established in computer science, one common problem is that they do not automatically recognize any intrinsic symmetry of any data. To describe a given physical quantity precisely, it is therefore necessary to symmetrize the ML representation in terms of translation, rotation, and permutation operation. By designing various symmetry-invariant descriptors as the input of ML methods\textsuperscript{3-12}, there have been quite successful ML applications in representing scalar property, e.g. the potential energy, which is invariant with respect to symmetry operations on the molecular configuration\textsuperscript{13}. These ML-based potential energy surfaces (PESs) provide analytical and continuous atomic forces, thus making molecular dynamics simulations orders of magnitude faster than first-principles calculations.

However, much fewer studies have focused on the ML representation of tensorial molecular properties, e.g. permanent/transient dipole moment (PDM/TDM) and polarizability tensor. These tensorial properties are generally more difficult to learn, as they contain multiple coordinate-dependent components that are covariant when the system is rotated. One may bypass this challenge by manually aligning the molecules to a local reference frame\textsuperscript{14-19}. This common practice is however less well-defined in heavily-distorted structures or dissociable systems and could cause discontinuity in the boundary of the reference frame. More rigorously, symmetry adaption in kernel-based regression has been realized by using covariant tensor
kernels adapted to the rotational symmetry of response properties\textsuperscript{20-23}, or acting the response operator act on the kernels explicitly considering the external field and internal degrees of freedom of the molecule\textsuperscript{24}. Unlike kernel-based regression, the nonlinearity of neural networks (NNs) would scramble the covariant symmetry embedded in geometric descriptors and one has to use a symmetry-invariant NN output as a multiplier for tensorial properties\textsuperscript{22}.

For permeant dipoles, for example, it is well-known that one can simply learn the atomic charge multiplied with the corresponding coordinate vector, reducing the ML model to an efficient scalar analogue\textsuperscript{25-28}. However, it is less well-noticed that this simple approach is unable to account for transition dipoles, which describe the transition (or change) of charge distributions in two states, as shown below. It is not applicable to the polarizability tensor either. In this work, we extend this simple NN model to describe the TDM and polarizability tensor in an efficient way. The core idea is to construct the desired tensorial form by multiplying virtual NN outputs and/or their partial derivatives with atomic coordinate vectors, while keeping original geometrical descriptors symmetry-invariant. We demonstrate that the tensorial NN models yield accurate predictions for a large number of \textit{ab initio} PDM, TDM and polarizability data in various benchmark systems. Furthermore, the NN-predicted TDM surfaces for a peptide fragment are used to faithfully reproduce the ab initio based electronic spectra of two real proteins with about six orders of magnitude speedup.
Methods

Let us go through the PDM ($\mathbf{\mu}$) case quickly, which corresponds classically to the separated charge ($q$) multiplied by the vector pointing from the negative to the positive charge ($\mathbf{\bar{r}}$), i.e. $\mathbf{\bar{\mu}} = q\mathbf{\bar{r}}$. For a given system with $N$ atoms, PDM can be readily obtained by the sum of atomic contributions,

$$ \mathbf{\bar{\mu}} = \sum_{i=1}^{N} q_i \mathbf{\bar{r}}_i, $$

(1)

where $q_i$ is the atomic effective charge of $i$th atom (not physically meaningful) and $\mathbf{\bar{r}}_i = (x_i, y_i, z_i)^T$ is the corresponding atomic coordinate vector originated from the center of mass. Note that $q_i$ is a scalar quantity with the same symmetry of atomic energy that can be easily fitted in the same manner, e.g. in the atomistic NN framework, which immediately warrants the translational and permutational invariance and the rotational covariance of PDM (see Fig. 1).

It may not be so obvious that this approach is inapplicable to TDM. The quantum description of PDM corresponds to the expectation of dipole operator in terms of a given electronic wavefunction $\psi_i$, namely, $\mathbf{\bar{\mu}} = \langle \psi_i | q\mathbf{\bar{r}} | \psi_i \rangle$. In this regard, PDM relies only on the charge distribution of this electronic state. However, TDM associates with the transition between the two different states, $\mathbf{\bar{\mu}}_T = \langle \psi_f | q\mathbf{\bar{r}} | \psi_i \rangle$, which is well-known to be affected by the relative phase of the initial and final electronic states as well as the transition type (i.e. the change of charge distribution upon transition). The phase problem can be mitigated by a phase correction procedure comparing the overlaps between wavefunctions of neighboring configurations when generating training data. However, the orientation of the TDM vector is dependent
on the molecular orbitals involved in the transition, which is not at all taken into account in Eq. (1). For example, for a triatomic molecule lying in the $yz$ plane with a perpendicular transition, the TDM vector is nonzero only in its $\mu_z$ component. However, the $x_i$ component in each atomic coordinate vector is zero, vanishing the right side of Eq. (1). As a consequence, the aforementioned atomic charge model would completely fail in such a simple case, giving rise to unavoidably large fitting errors. To solve the problem, we introduce two vectors in the same way as the PDM vector, namely,

$$\mu_T^i = \sum_{j=1}^{N} q_i R_j \quad (j=1, 2),$$  

(2)

where $q_i^j$ $(j=1, 2)$ can be obtained by two different outputs of the same atomic NN (see Fig. 1b). As long as $q_i^1$ and $q_i^2$ are not accidentally identical, $\mu_T^1$ and $\mu_T^2$ will define a specific plane (e.g., the molecular plane for a planar geometry) and their cross product will give rise to a third vector perpendicular to this plane,

$$\mu_T^3 = \sum_{i=1}^{N} q_i^3 (\mu_T^1 \times \mu_T^2),$$  

(3)

where $q_i^3$ is given by another output of the same atomic NN that determines the magnitude of $\mu_T^3$. These three fundamental vectors can be then linearly combined, namely $\mu_T^{NN} = \mu_T^1 + \mu_T^2 + \mu_T^3$, to mimic a TDM vector that is not necessarily restricted in the molecular plane, with the correct rotational covariance (see Fig. 1). We keep using the symbol $q_i^j$ for consistency, which contains no physical meaning here. Since Eqs. (2-3) take advantage of the same atomistic NN structure with multiple outputs, the NN models for TDM and PDM are comparably efficient as their counterpart for PES.

Next, we will follow this concept to design the tensorial NN model for molecular
polarizability \( (\alpha) \). Let us recall that \( \alpha \) is the second-order response of potential energy \( (V) \) to an electric field \( \hat{E} \),

\[
\alpha = \frac{\partial^2 \mu_{\text{ind}}}{\partial \hat{E}^2} = \frac{\partial^2 V}{\partial \hat{E} \partial \hat{E}}. \tag{4}
\]

where \( \mu_{\text{ind}} \) is the induced dipole moment. Apparently, \( \alpha \) is a second-rank tensor and a \( 3 \times 3 \) symmetric matrix, which is translationally and permutationally invariant but subject to the following transformation under rotation,

\[
\alpha^b = U \alpha^a U^T. \tag{5}
\]

where \( U \) is the transformation matrix, \( \alpha^a \) and \( \alpha^b \) are polarizability tensors at symmetry-equivalent molecular configurations, respectively.

A convenient way to construct a \( 3 \times 3 \) symmetric matrix is taking the product of a \( 3 \times M \) matrix \( (M \geq 3) \) and its transpose. To this end, we can design an effective induced dipole-like NN structure similar to that in Eq. (2) with multiple \( (M) \) outputs,

\[
\mu_{\text{ind}}^j = \sum_{i=1}^{N} q_i \tilde{r}_i \quad (j = 1, 2, ..., M), \tag{6}
\]

which has the same rotational covariance as the dipole moment. Alternatively, we can also calculate the partial derivatives of the virtual quantity generated by the NN model with respect to atomic coordinates, leading to a \( 3 \times N \) matrix,

\[
D^j = \sum_{i=1}^{N} \frac{\partial q_i}{\partial \tilde{r}_j} \quad (j = 1, 2, ..., N), \tag{7}
\]

Multiplying either \( R \) or \( D \) matrix by its transpose gives us the required \( 3 \times 3 \) matrix,

\[
\alpha^{\text{NNI}} = D(D)^T \text{ or } \alpha^{\text{NNI}} = \mu_{\text{ind}} \mu_{\text{ind}}^T \tag{8}
\]

We choose to work with the \( D \) matrix in this work, which is found to produce more accurate results. However, with either choice, \( \alpha^{\text{NNI}} \) is a semidefinite matrix by
construction, while molecular polarizability itself is not necessarily semidefinite. We thus create another symmetric matrix \( \alpha^{NN2} \) in the following way,

\[
\alpha^{NN2} = r(D)^T + Dr^T,
\]

which is obviously not semidefinite. Furthermore, it is important to note that both \( \alpha^{NN1} \) and \( \alpha^{NN2} \) become a rank-deficient matrix when the molecular geometry is planar, while the molecular polarizability tensor is not. The simplest way to correct this is to incorporate a scalar matrix \( \alpha^{NN3} \), whose diagonal element can be optimized by a very simple NN. Combining these three terms yields the full representation of the NN-based polarizability tensor,

\[
\alpha^{NN} = \alpha^{NN1} + \alpha^{NN2} + \alpha^{NN3},
\]

that fulfills the symmetry of \( \alpha \) (see Fig. 1).

In practice, any atomistic NN methods proposed to represent scalar quantities can be readily adapted within the formulism discussed above, and importantly, one needs no modification of the symmetry-invariant descriptors. In this work, we generalize our recently proposed embedded atom neural network (EANN) model to representing these tensorial properties. The accuracy and efficiency of the scalar EANN model have been demonstrated in our recent publication. More details on the implementation of the tensorial EANN (T-EANN) model are given in the Supplementary Methods.

**Results and Discussion**

**A. Tensorial Response Properties**

We first apply the T-EANN model to water-related systems, for which there have
been ab initio data available for comparison in the work of the symmetry-adapted Gaussian approximation potential (SA-GAP) method developed by Ceriotti and coworkers\textsuperscript{20}. For each system, half of the total 1000 ab initio data of $\bar{\mu}$ and $\alpha$ were randomly selected for training and the rest for prediction. The prediction errors are compared with those obtained by SA-GAP\textsuperscript{20} in Table I. Due to the widespread numerical ranges of ab initio $\bar{\mu}$ ($\alpha$) values, the prediction root-mean-squared-errors (RMSEs) hereafter are always relative to intrinsic standard deviation of the testing samples. The T-EANN models represent both quantities in all systems very accurately, yielding RMSEs of $\bar{\mu}$ ($\alpha$) as 0.019\% (0.040\%), 6.6\% (4.2\%), 1.3\% (0.42\%), and 17\% (3.5\%) for H$_2$O, (H$_2$O)$_2$, and H$_5$O$_{2+}$, and liquid water, respectively. It is known that kernel-based models typically require fewer data than NN-based ones to reach the same level of accuracy, but the former have often higher computational cost and worse scaling with respect to the number of data points\textsuperscript{29}, due to their local interpolation nature. It is therefore encouraging that, despite its simplicity, our T-EANN model seems comparably accurate as the SA-GAP model with the same small amount of data. Moreover, the T-EANN model does not increase much the cost of evaluating atomic density-like descriptors, which scales linearly with respect to the number of neighboring atoms in the cutoff radius\textsuperscript{12}.

Our T-EANN model for polarizability differs in spirit from the SA-GAP model which accounts for the covariant symmetry in terms of the tensorial smooth overlap of atomic positions ($\lambda$-SOAP) kernels\textsuperscript{20} and the regression is linear. Ceriotti and coworkers\textsuperscript{22} also discussed the framework of combining the $\lambda$-SOAP representation...
with NNs. Because NNs primarily consist of nonlinear mapping functions, the tensorial λ-SOAP features have to be incorporated to the linear output layer of a scalar SOAP-based NNs where the output serves as a multiplier for the tensorial features. In other words, the hidden layers of NNs virtually output linear expansion coefficients for the tensorial λ-SOAP descriptors only. This is completely different from the T-EANN representation implemented here, in which the virtual outputs of EANN are differentiated with respect to nuclear coordinates, similar in spirit to the ML representation for the electronic friction tensor of an adsorbate at a metal surface and the nonadiabatic coupling vector between two electronic states.

It is worthwhile to check the necessity of each of the three terms in Eq. (10) in precisely representing the polarizability tensor. To this end, we compare the performance of incomplete T-EANN models, i.e. fitting the ab initio $\alpha$ values with $\alpha_{NN1}$, $\alpha_{NN1} + \alpha_{NN2}$, $\alpha_{NN1} + \alpha_{NN2} + \alpha_{NN3}$, respectively, and the converged T-EANN model in Table II. As expected, using $\alpha_{NN1}$ only results in huge prediction errors, especially for water and water dimer. This is understandable as they contain either completely or mostly planar configurations. Similar results are observed when using $\alpha_{NN2}$ only (not shown). An accompanying issue is that “over-fitting” severely takes place when training this single term. Including $\alpha_{NN2}$ leads to a marginal improvement but largely remedies the over-fitting issue, making the training process more stable. Interestingly, adding $\alpha_{NN3}$ substantially lowers the prediction errors, suggesting that the singularity of $\alpha_{NN1}$ and $\alpha_{NN2}$ at planar configurations is problematic. Incorporating all three terms fix this problem and provide sufficient repeatability of the T-EANN model,
which further considerably improves the results. This problem still exists for a more complex system (H2O2\(^{+}\)), though less severe because fewer configurations are planar. We note in passing that the SA-GAP model transforms the tensor from the Cartesian coordinate space to the irreducible spherical representation\(^{32}\), which could eliminate the singularity problem in planar configurations.

**B. Transition Dipole Moment**

There were few ML models representing global TDM surfaces\(^{17,18}\), however, to our best knowledge, none of them rigorously considered the covariant symmetry of the TDM vector. To show the critical role of the symmetry, we train a T-EANN model with the TDM data of the N-methylacetamide (NMA) molecule (see Fig. 2a). The nπ* and ππ* excitations of NMA (see Fig. 2b) have been extensively used to model the ultraviolet (UV) spectra of the amide group of the protein backbone\(^{17,33}\). Our training set consists of the TDM data for nπ* and ππ* transitions of NMA at the time-dependent density functional theory (TDDFT) and PBE0/cc-pVDZ level with the Gaussian 16 package\(^{34}\), for ~50000 NMA configurations extracted from 1000 different protein backbones at room temperature from the RCSB Protein Data Bank (PDB)\(^{35}\). Phase corrections have been carefully done by evaluating the wavefunction overlaps of neighboring configurations along the trajectories\(^{18}\).

In Figs. 2c-f shows the correlation diagrams between the ab initio TDM values of NMA upon nπ* and ππ* excitations and corresponding T-EANN predictions. It is worth noting that the incomplete T-EANN model based merely on Eq. (1) performs poorly for the nπ* transition but quite well for the ππ* transition, as evidenced by the
huge relative RMSE (202.60%) for nπ* versus that (5.99%) for ππ*. Interestingly, the disparate representability of Eq. (1) for nπ* and ππ* excitations turns out be a natural consequence of their transition characters. Indeed, the UV absorption of the NMA molecule is dominated by the peptide bond, namely H-N-C=O group. To be more specific, as illustrated in Fig. 2b, the nπ* transition corresponds to a perpendicular transition from a lone pair p orbital of oxygen to the anti-bonding π* orbital of the N-C=O group, while the ππ* transition takes place parallel to the N-C=O plane between the nonbonding π to the anti-bonding π* orbital. As most configurations in the data set were moderately deviated from the equilibrium geometry where the H-N-C=O group is nearly planar, the in-plane ππ* TDM can be well described by the PDM expression in Eq. (1). In contrast, the nπ* TDM vectors are almost orthogonal to the atomic coordinate vectors and largely incompatible with the implementation of Eq. (1), leading to inevitable errors. Within the proper symmetrization scheme as described in Eqs. (2-3), the T-EANN representation achieves a much better accuracy for the TDM of both nπ* and ππ* excitations, with no prior information required about the transition type. As shown in Figs. 2e-f, the relative RMSE for the nπ* transition is now as small as ~1.62% and the description for the ππ* TDM is also slightly improved. This level of accuracy is much better than that of the NN model using the regular Coulomb matrix as descriptors and aligning the molecule to a reference frame, especially for the nπ* transition. This system thus represents an excellent showcase for the importance of preserving the desired symmetry of the transition tensorial property.
We further test the quality of the T-EANN TDM model by computing the UV spectra of two real proteins, namely 2bmm and 5h34, which were not included in the training set. These proteins contain so many atoms that it is prohibitively difficult to compute their spectra from atomistic molecular dynamics simulations. Instead, their spectra have been approximated by the Frenkel exciton model\textsuperscript{36,37} in which the system is divided into many molecular excitons (\textit{i.e.} NMA fragments) and their couplings are estimated by dipole-dipole approximation\textsuperscript{38} (see Supplementary Methods for details). Impressively, as shown in Fig. 3, the T-EANN derived UV spectra of both proteins are in excellent agreement with the TDDFT counterparts, not only for the main peak positions/intensities and absorption band widths, but also for many fine structures due to exciton splitting energies that are very sensitive to the TDM vectors. Moreover, the T-EANN model is over 6 orders of magnitude faster than TDDFT. Specifically, the former takes only $\sim 2.39$ s CPU time per core for calculating both $n\pi^*$ and $\pi\pi^*$ TDMs of 10000 NMA frames, compared to $3.62 \times 10^6$ s by TDDFT. These results provide convincing evidence of the high accuracy and efficiency of the T-EANN TDM model.

**Conclusions**

Tensorial response and transition properties of chemical systems are crucial in spectroscopic simulations and essentially determine the spectroscopic transition rules. The major difficulty of symmetrizing NNs to represent tensorial molecular properties is that the nonlinearity of NNs would scramble any covariant symmetry in the geometric descriptors. To meet this challenge, in this work, we construct the tensorial model by the product of atomic coordinate vectors and virtual NN outputs (for dipole
moment) or their partial derivatives with respect to atomic coordinates (for polarizability tensor). This strategy accounts for the intrinsic covariant symmetry of tensorial properties and takes advantage of the great representability of NNs. In particular, the directional property of a TDM vector is taken into account by introducing a cross-product of two NN-based vectors, which is essential to accurately predict TDMs without prior knowledge of the transition type. Implemented with the EANN model, numerical tests on PDM, TDM, and molecular polarizability tensor in various benchmark systems demonstrate the high fidelity and efficiency of this strategy and its potential in simulating infrared, Raman, and electronic spectra of large biological complexes and condensed phase systems. This approach can be straightforwardly adapted to represent other important tensorial properties like magnetic dipole moment, or even the electronic Hamiltonian which is subject to a similar covariant symmetry\textsuperscript{39}, and combined with any conventional atomistic NN framework.

**Acknowledgements:** This work was supported by National Key R&D Program of China (2017YFA0303500 and 2018YFA0208603), National Natural Science Foundation of China (91645202, 21722306, and 21633006), and Anhui Initiative in Quantum Information Technologies (AHY090200). We appreciate the Supercomputing Center of USTC and AM-HPC for high-performance computing services. We thank Prof Zhenggang Lan for some helpful discussion on phase correction.
**Author Contributions:** B.J. and J.J. conceived the project. Y.Z. developed the EANN program and performed numerical tests. S.Y. and J.Z. performed ab initio calculations for NMA and protein spectra. Y.Z. and B.J. took the lead in writing the manuscript. All authors read and commented on the manuscript.

**Data availability.** The data that support the findings of this study are available from the corresponding author upon request

**Competing interests:** The authors declare no competing financial interests.
References

1. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547-555 (2018).
2. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**, 360-365 (2018).
3. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
4. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
5. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
6. Jiang, B. & Guo, H. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. *J. Chem. Phys.* **139**, 054112 (2013).
7. Shao, K., Chen, J., Zhao, Z. & Zhang, D. H. Communication: Fitting potential energy surfaces with fundamental invariant neural network. *J. Chem. Phys.* **145**, 071101 (2016).
8. Chmiela, S., Tkatchenko, A., Sauceda, H. E., Poltavsky, I., Schutt, K. T. & Muller, K. R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015 (2017).
9. Huang, S. D., Shang, C., Zhang, X. J. & Liu, Z. P. Material discovery by combining stochastic surface walking global optimization with a neural network. *Chem. Sci.* **8**, 6327-6337 (2017).
10. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
11. Zhang, L., Han, J., Wang, H., Car, R. & E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
12. Zhang, Y., Hu, C. & Jiang, B. Embedded atom neural network potentials: Efficient and accurate machine learning with a physically inspired representation. *J. Phys. Chem. Lett.* **10**, 4962-4967 (2019).
13. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
14. Bereau, T., Andrienko, D. & von Lilienfeld, O. A. Transferable Atomic Multipole Machine Learning Models for Small Organic Molecules. *J. Chem. Theory Comput.* **11**, 3225-3233 (2015).
15. Liang, C., Tocci, G., Wilkins, D. M., Grisafi, A., Roke, S. & Ceriotti, M. Solvent fluctuations and nuclear quantum effects modulate the molecular hyperpolarizability of water. *Phys. Rev. B* **96**, 041407 (2017).
16. Zhang, Y., Maurer, R. J., Guo, H. & Jiang, B. Hot-electron effects during reactive scattering of H\(_2\) from Ag(111): the interplay between mode-specific electronic friction and the potential energy landscape. *Chem. Sci.* **10**, 1089-1097 (2019).
17. Ye, S., et al. A neural network protocol for electronic excitations of N-methylacetamide. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11612-11617 (2019).
18. Westermayr, J., Gastegger, M., Menger, M., Mai, S., Gonzalez, L. & Marquetand, P. Machine learning enables long time scale molecular photodynamics simulations. *Chem. Sci.* **10**, 8100-8107 (2019).
19. Guan, Y., Guo, H. & Yarkony, D. R. Extending the Representation of Multistate Coupled Potential Energy Surfaces To Include Properties Operators Using Neural Networks: Application to the 1,21A States of Ammonia. J. Chem. Theory Comput. 16, 302-313 (2020).
20. Grisafi, A., Wilkins, D. M., Csányi, G. & Ceriotti, M. Symmetry-adapted machine learning for tensorial properties of atomistic systems. Phys. Rev. Lett. 120, 036002 (2018).
21. Wilkins, D. M., Grisafi, A., Yang, Y., Lao, K. U., DiStasio, R. A. & Ceriotti, M. Accurate molecular polarizabilities with coupled cluster theory and machine learning. Proc. Natl. Acad. Sci. U. S. A. 116, 3401-3406 (2019).
22. Willatt, M. J., Musil, F. & Ceriotti, M. Atom-density representations for machine learning. J. Chem. Phys. 150, 154110 (2019).
23. Raimbault, N., Grisafi, A., Ceriotti, M. & Rossi, M. Using Gaussian process regression to simulate the vibrational Raman spectra of molecular crystals. New. J. Phys. 21, 105001 (2019).
24. Christensen, A. S., Faber, F. A. & Lilienfeld, O. A. v. Operators in quantum machine learning: Response properties in chemical space. J. Chem. Phys. 150, 064105 (2019).
25. Huang, X., Braams, B. J. & Bowman, J. M. Ab initio potential energy and dipole moment surfaces for H$_2$O$_2$*. J. Chem. Phys. 122, 044308 (2005).
26. Gastegger, M., Behler, J. & Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. Chem. Sci. 8, 6924-6935 (2017).
27. Sifain, A. E., et al. Discovering a Transferable Charge Assignment Model Using Machine Learning. J. Phys. Chem. Lett. 9, 4495-4501 (2018).
28. Zhang, L., Chen, M., Wu, X., Wang, H., E, W. & Car, R. Deep neural network for the dielectric response of insulators. Preprint at https://ui.adsabs.harvard.edu/abs/2019arXiv190611434Z (2019).
29. Zuo, Y., et al. Performance and Cost Assessment of Machine Learning Interatomic Potentials. J. Phys. Chem. A 124, 731-745 (2020).
30. Zhang, Y., Maurer, R. J. & Jiang, B. Symmetry-Adapted High Dimensional Neural Network Representation of Electronic Friction Tensor of Adsorbates on Metals. J. Phys. Chem. C 124, 186-195 (2020).
31. Westermayr, J., Gastegger, M. & Marquetand, P. Combining SchNet and SHARC: The SchNarc machine learning approach for excited-state dynamics. Preprint at https://ui.adsabs.harvard.edu/abs/2020arXiv200207264W (2020).
32. Stone, A. J. Transformation between cartesian and spherical tensors. Mol. Phys. 29, 1461-1471 (1975).
33. Wang, L., Middleton, C. T., Zanni, M. T. & Skinner, J. L. Development and Validation of Transferable Amide I Vibrational Frequency Maps for Peptides. J. Phys. Chem. B 115, 3713-3724 (2011).
34. Frisch, M. J., et al. Gaussian 16 Rev. B.01. (ed*$^*$eds) (2016).
35. Berman, H. M., et al. The Protein Data Bank. Nucleic Acids Res. 28, 235-242 (2000).
36. Abramavicius, D., Palmieri, B. & Mukamel, S. Extracting single and two-exciton couplings in photosynthetic complexes by coherent two-dimensional electronic spectra. Chem. Phys. 357, 79-84 (2009).
37. Abramavicius, D., Jiang, J., Bulheller, B. M., Hirst, J. D. & Mukamel, S. Simulation Study of Chiral Two-Dimensional Ultraviolet Spectroscopy of the Protein Backbone. J. Am. Chem. Soc.
38. Kasha, M., Rawls, H. R. & El-Bayoumi, M. A. The exciton model in molecular spectroscopy. *Pure Appl. Chem.*, 11, 371-392 (1965).

39. Schütz, K. T., Gastegger, M., Tkatchenko, A., Müller, K. R. & Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* 10, 5024 (2019).
Table I. Comparison of relative RMSEs for permanent dipole moment and polarizability (in atomic unit) in several water-related systems with the T-EANN and SA-GAP\textsuperscript{20} models.

| System          | $\mu$ (a.u.) | $\alpha$ (a.u.) |
|-----------------|--------------|-----------------|
|                 | T-EANN(%)    | SA-GAP(%)   $\lambda=1^*$ | T-EANN(%) | $\lambda$-SOAP-GAP(%) $\lambda=0/\lambda=2^*$ |
| H$_2$O          | 0.019        | ~0.11         | 0.040      | ~0.021/0.12 |
| (H$_2$O)$_2$     | 6.6          | ~5.3          | 4.2        | ~6.4/7.8    |
| (H$_3$O$_2$)$^+$ | 1.3          | ~2.4          | 0.42       | ~3.8/0.97   |
| (H$_2$O)$_{32}$  | 17           | \             | 3.5        | ~5.8/19**   |

* The SA-GAP method combines the transformation of the tensor into its irreducible spherical tensor (IST) representation with the covariant $\lambda$-SOAP kernels. There are three IST components for dipoles ($\lambda=1$) and six IST components for polarizabilities (one for $\lambda=0$ and five for $\lambda=2$).

** In Ref. 20, RMSEs were reported for dielectric response tensors by indirect learning of molecular polarizability, which are used here for qualitative comparison only.
Table II. RMSEs for predicting polarizability tensor (in atomic unit) in several molecular systems with different implementations (see text for details).

| System      | $\alpha_{NN1}$ | $\alpha_{NN1}^{NN2}$ | $\alpha_{NN1}^{NN2} + \alpha_{NN3}$ | $\alpha_{NN1}^{NN3} + \alpha_{NN2}$ | $\alpha_{NN1} + \alpha_{NN2} + \alpha_{NN3}$ |
|-------------|----------------|----------------------|-------------------------------------|---------------------------------|---------------------------------------------|
| H$_2$O      | 9.6            | 9.6                  | 3.6x10^{-1}                         | 2.0x10^{-3}                     |                                             |
| (H$_2$O)$_2$| 8.4            | 7.8                  | 9.1x10^{-1}                         | 4.3x10^{-1}                     |                                             |
| (H$_5$O$_2$)$^+$ | 9.5x10^{-1} | 8.3x10^{-1}          | 7.9x10^{-2}                         | 3.5x10^{-2}                     |                                             |
Fig. 1 Schematic diagram of tensorial embedded atom neural network models for PDM, TDM, and polarizability tensor. Regular feedforward neural networks are shown in black while the red ones indicate that partial derivatives are evaluated.
Fig. 2 (a) Molecular geometry of NMA with the peptide bond placed in the $xy$ plane and (b) schematic diagram of its $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ electronic transitions. Correlation plots between the TDDFT results of $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transition dipole moments ($\mathbf{\mu}_T$) and the T-EANN predictions obtained from the incomplete model based on Eq. (1) only (c-d) and the correct one based on Eqs. (2-3) (e-f), respectively.
Fig. 3 UV adsorption spectra of protein (a) 5h34 and (b) 2bmm calculated by the T-EANN model for TDM (black) and TDDFT (red).
Supplementary Information

for

Efficient and Accurate Spectroscopic Simulations with Symmetry-Preserving Neural Network Models for Tensorial Properties

Yaolong Zhang, Sheng Ye, Jinxiao Zhang, Jun Jiang, Bin Jiang*

Hefei National Laboratory for Physical Science at the Microscale, Department of Chemical Physics, Key Laboratory of Surface and Interface Chemistry and Energy Catalysis of Anhui Higher Education Institutes, University of Science and Technology of China, Hefei, Anhui 230026, China
Supplementary Methods

A. Generalized embedded atom neural network representation

As discussed in the main text, any atomistic NN framework for scalar quantities can be readily adapted to represent tensorial properties with our proposed algorithm and one needs no modification of the symmetry-invariant descriptors. In this work, we employed our recently proposed embedded atom neural network (EANN) model\(^1\). The EANN model is inspired from the physically-derived embedded atom method (EAM)\(^2\) force field. The total energy of a system is the sum of the atomic energy, each of which is a function of electron density at the central atom position embedded in the surrounding environment. Atomic NNs thus represent the complex relationship between the embedded density and the atomic energy,

\[
E = \sum_{i=1}^{N} E_{i}^{\text{NN}}(\rho^i), \tag{S1}
\]

where \(N\) is the total number of atoms in the system and \(\rho^i\) is electron density vector at \(i\)th atom, which can be approximated by the square of the linear combination of different atomic orbitals,

\[
\rho^i_{L,a,c} = \sum_{l_x, l_y, l_z = l} \frac{L!}{l_x! l_y! l_z!} \left( \sum_{j=1}^{n_{\text{atom}}} c_j \phi_{l_x, l_y, l_z}^{a_j r_j} (r^i_j) \right)^2. \tag{S2}
\]

where \(c_j\) is the orbital coefficient of atom \(j\) and is optimized in the training process. \(n_{\text{atom}}\) is the number of neighboring atoms. \(\phi_{l_x, l_y, l_z}^{a_j r_j} (r^i_j)\) is Gaussian-type like orbitals

\[
\phi_{l_x, l_y, l_z}^{a_j r_j} (r^i_j) = x^{l_x} y^{l_y} z^{l_z} \exp(-a |r^i_j - r_j|^2), \tag{S3}
\]

where \(r=(x, y, z)\) is the coordinates vector of an electron relative to the corresponding center atom, \(r\) is the norm of the vector, \(a\) and \(r_s\) are parameters used for tuning radial distributions of atomic orbitals, \(L\) specifies the orbital angular momentum, which is the
summation of angular moment along each axis $l_x, l_y, l_z$. These hyper-parameters used to
define the density-like descriptors for each system in this work are listed in Table S1
and S2.

By construction, these embedded density-like descriptors can preserve the
translational, rotational and permutational symmetry of potential energy in an efficient
way\(^3\). Through a straightforward transformation\(^3\), these descriptors include the angular
information with a two-body computational cost in comparison to the commonly used
atom centered symmetry functions\(^4\). The EANN model for a scalar property can be
trained by minimizing the following cost function,

$$ S(\mathbf{w}) = \sum_{m=1}^{N_{\text{data}}} \left| E^{\text{NN}}_m - E^{\text{QC}}_m \right|^2 / N_{\text{data}}, $$

(S4)

where $\mathbf{w}$ is the collection of the NN weight parameters and $N_{\text{data}}$ is the size of training
dataset, $E^{\text{NN}}_m$ and $E^{\text{QC}}_m$ are the sum of atomic NN energies of the $m$th data point and
the corresponding target total energy obtained from quantum chemical (QC)
calculations. The formulas of the tensorial EANN (T-EANN) models for permanent
dipole moment (PDM), transition dipole moment (TDM), as well as molecular
polarizability have been detailed in the main text. Since the geometric descriptors
remain unchanged, one just needs to slightly modify the objective function to be
minimized. As a result, the T-EANN models for dipole moments can be trained in the
following way,

$$ S(\mathbf{w}) = \sum_{m=1}^{N_{\text{data}}} \left| \mathbf{\mu}^{\text{NN}}_m - \mathbf{\mu}^{\text{QC}}_m \right|^2 / N_{\text{data}} / 3. $$

(S5)

where $\mathbf{\mu}^{\text{NN}}_m$ is the sum of all atomic dipole moments via Eq. (1) for PDM or Eqs (2-3)
for TDM, and $\mathbf{\mu}^{\text{QC}}_m$ is corresponding QC-calculated PDM or TDM vector. For the
polarizability tensor, similarly, we minimize the objective function expressed as

\[
S(w) = \sum_{m=1}^{N_{\text{data}}} \left| a^{\text{NN}}_m - a^{\text{QC}}_m \right|^2 / N_{\text{data}} \cdot \theta .
\]  

(S6)

More details about the EANN model can be found in our previous work\(^1\).

To quantify the quality of the NN fit, given the wide numerical range of the tensorial properties, we present in this work root-mean-squared-error (RMSE) relative to intrinsic standard deviation (\(\sigma\)) of the testing samples\(^5\) instead of the commonly used absolute RMSE when fitting potential energy surface. Specifically,

\[
\text{RMSE} = \sqrt{\frac{1}{N_{\text{data}}} \sum_{m=1}^{N_{\text{data}}} \left( \Gamma^{\text{NN}} - \Gamma^{\text{QC}} \right)^2} / \sigma(\Gamma),
\]

(S7)

and the standard deviation is given by,

\[
\sigma(\Gamma) = \sqrt{\frac{1}{N_{\text{data}} - 1} \sum_{m=1}^{N_{\text{data}}} \left( \Gamma^{\text{QC}} - \langle \Gamma^{\text{QC}} \rangle \right)^2}.
\]

(S8)

Here \(\Gamma\) represent a tensorial property, namely \(\mathbf{\mu}, \mathbf{\mu}_T\) or \(\mathbf{a}\) in this work.

**B. Electronic structure calculations**

The permanent dipole moment (\(\mathbf{\mu}\)) and polarizability (\(\mathbf{a}\)) data of H\(_2\)O, (H\(_2\)O\(_2\)), H\(_3\)O\(^+\) and liquid water have been calculated by Ceriotti and coworkers\(^5\). We refer the readers to Ref.\(^5\) for more computational details. Briefly, ab initial calculations for H\(_2\)O, (H\(_2\)O\(_2\)), and H\(_3\)O\(^+\) were performed at the CCSD/d-aug-cc-pvtz level using Dalton 2016\(^6\). Liquid water was described by a 32-water box and calculated at the DFT/PBE-USPP level, using a plane-wave basis set with a kinetic energy cutoff of 55 Ry. The first Brillouin zone of the periodic system has been sampled with 5 k-points along each reciprocal lattice direction\(^5\). Overall, 1000 configurations were collected for each system and half of them were randomly chosen for training, leaving the rest 500 points as test set.
The transition dipole moment data for $n\pi^*$ and $\pi\pi^*$ transitions of N-methylacetamide (NMA) were calculated by time-dependent density functional theory (TDDFT) implemented in the Gaussian 16 package$^7$, at the PBE0/cc-pVDZ level. We extracted 50000 NMA configurations from 1000 different protein backbones at room temperature from the RCSB Protein Data Bank (PDB)$^8$. Phase corrections have been carefully done by evaluating the wavefunction overlaps of neighboring configurations along the trajectories$^9$.

C. Frenkel exciton model

Proteins consist of peptides and amino acid residues. It is known that ultraviolet absorption (UV) spectra of proteins are largely due to the electronic excitation of peptides in protein backbones. In this work, we split the protein into different peptide fragments and amino acid residues. Frenkel exciton model$^{10,11}$ is employed to construct the exciton Hamiltonian derived from the interaction between different peptides and amino acid residues,

$$\hat{H} = \sum_{ma} \epsilon_{ma} \hat{B}_m^\dagger \hat{B}_m + \sum_{ma,nb} J_{ma,nb} \hat{B}_m^\dagger \hat{B}_{nb} ,$$  \hspace{1cm} (S9)

where $m$ and $n$ represent different peptide fragments, $a$ and $b$ correspond to different transition types (i.e., $n\pi^*$ (perpendicular) and $\pi\pi^*$ (parallel) transitions in this work), $\hat{B}_m^\dagger$ and $\hat{B}_m$ are the electron creation and annihilation operators of corresponding transitions, $\epsilon_{ma}$ is the excitation energy of the $mth$ peptide coupled with neighboring amino acid residues, and $J_{ma,nb}$ is the coupling between excited states of two peptide fragments. More details about this model can be found elsewhere$^{10,11}$.

In principle, an accurate depiction of these couplings require very time-consuming
two-electron integral and excited state calculations. Dipole-dipole approximation is regarded as a common and simple way to estimate the interaction between electrons\textsuperscript{12}. To this end, $\epsilon_{ma}$ and $J_{ma,nb}$ can be approximated as,

$$
\epsilon_{ma} = \epsilon_{ma}^0 + \sum_k \frac{1}{4\pi\epsilon_0} \left( \frac{\mathbf{\mu}_{T,ma} \cdot \mathbf{\mu}_k}{|\mathbf{r}_{mk}|^3} - 3 \left( \frac{(\mathbf{\mu}_{T,ma} \cdot \mathbf{r}_{mk})(\mathbf{\mu}_k \cdot \mathbf{r}_{mk})}{|\mathbf{r}_{mk}|^5} \right) \right),
$$

(S10)

$$
J_{ma,nb} = \sum_{m,n} \frac{1}{4\pi\epsilon_0} \left( \frac{\mathbf{\mu}_{T,ma} \cdot \mathbf{\mu}_{T,nb}}{|\mathbf{r}_{mn}|^3} - 3 \left( \frac{(\mathbf{\mu}_{T,ma} \cdot \mathbf{r}_{mn})(\mathbf{\mu}_{T,nb} \cdot \mathbf{r}_{mn})}{|\mathbf{r}_{mn}|^5} \right) \right).
$$

(S11)

Here, $\epsilon_{ma}^0$ is the excitation energy of an isolated peptide (i.e., the NMA molecule in this work), $\mathbf{\mu}_{T,ma}$ and $\mathbf{\mu}_{T,nb}$ are the TDMs of the corresponding peptide, $\mathbf{\mu}_k$ is the PDM of an amino acid residue, $\mathbf{r}_{mk} / |\mathbf{r}_{mk}|$ is the position vector of a peptide referred to an amino acid residue or another peptide. To further simplify the simulations, we have ignored the coupling between peptides and amino acid residues because of their negligible influence\textsuperscript{10}. In comparison of ab initio based and NN predicted UV spectra of two real proteins, namely 2bmm and 5h34, $\epsilon_{ma}^0$, $\mathbf{\mu}_{T,ma}$, and $\mathbf{\mu}_{T,nb}$ generated from TDDFT calculations are represented by scalar and tensorial EANN models, respectively. The scalar EANN model of $\epsilon_{ma}^0$ is essentially the same as that for fitting potential energy surfaces, yielding an average RMSE of 7.8 meV and 5.1 meV for $n\pi^*$ and $\pi\pi^*$ transitions, which are sufficiently accurate for calculating the UV spectra. The accuracy of the T-EANN model for TDM has been presented in the main text.

**Supplementary Results and Discussion**

A. **Numerical verification of the T-EANN model for polarizability tensor**

In the T-EANN representation of $\mathbf{a}$, we need three terms $\alpha_{NN1}^{NN1}$, $\alpha_{NN2}^{NN2}$ and $\alpha_{NN3}^{NN3}$ to
guarantee the proper property of an arbitrary polarizability tensor. Taking water monomer as an illustrative example, we demonstrate below numerically how these terms satisfy the symmetry of polarizability. The molecule is placed in the $x$-$z$ plane and its molecular geometry can be expressed in terms of atomic Cartesian coordinate vectors as,

$$
\mathbf{r} = \begin{bmatrix}
  x_O & y_O & z_O \\
  x_{H_1} & y_{H_1} & z_{H_1} \\
  x_{H_2} & y_{H_2} & z_{H_2}
\end{bmatrix}.
$$

We take one arbitrary configuration from the training data set,

$$
\mathbf{r} = \begin{bmatrix}
  0.057536 & 0.000000 & -0.033219 \\
  -0.841234 & -0.000000 & -0.402656 \\
  -0.071907 & 0.000000 & 0.929858
\end{bmatrix}.
$$

and its ab initio polarizability tensor reads,

$$
\alpha^{QC} = \begin{bmatrix}
  9.931967 & 0.000000 & 0.205256 \\
  0.000000 & 9.459317 & -0.000000 \\
  0.205256 & -0.000000 & 10.168985
\end{bmatrix}.
$$

As clearly seen, $\alpha^{QC}$ is a full-rank symmetric matrix although H$_2$O is completely planar. However, by construction, $\alpha^{NN1}$ and $\alpha^{NN2}$ are both rank-deficient matrices. After successful training process, the T-EANN model contains the following $\alpha^{NN1}$ and $\alpha^{NN2}$ matrices,

$$
\alpha^{NN1} = \begin{bmatrix}
  0.744341 & 0.000000 & 0.024220 \\
  0.000000 & 0.000000 & 0.000000 \\
  0.024220 & 0.000000 & 0.772309
\end{bmatrix},
$$

$$
\alpha^{NN2} = \begin{bmatrix}
  -0.271611 & 0.000000 & 0.181350 \\
  0.000000 & 0.000000 & 0.000000 \\
  0.181350 & 0.000000 & -0.062198
\end{bmatrix},
$$

both with zero value for $\alpha_{yy}$. However, the scalar matrix $\alpha^{NN3}$ adequately solves this
issue,

\[
\alpha^{NN3} = \begin{bmatrix}
9.459202 & 0.000000 & 0.000000 \\
0.000000 & 9.459202 & 0.000000 \\
0.000000 & 0.000000 & 9.459202 \\
\end{bmatrix}.
\] (S17)

The overall T-EANN polarizability tensor is the sum of these three terms,

\[
\alpha^{NN} = \begin{bmatrix}
9.931932 & 0.000000 & 0.205570 \\
0.000000 & 9.459202 & 0.000000 \\
0.205570 & 0.000000 & 10.169314 \\
\end{bmatrix},
\] (S18)

which is in excellent agreement with the ab initio value \( \alpha^{QC} \).
Reference

1. Zhang, Y., Hu, C. & Jiang, B. Embedded atom neural network potentials: Efficient and accurate machine learning with a physically inspired representation. *J. Phys. Chem. Lett.* 10, 4962–4967 (2019).

2. Daw, M. S. & Baskes, M. I. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B* 29, 6443–6453 (1984).

3. Takahashi, A., Seko, A. & Tanaka, I. Conceptual and practical bases for the high accuracy of machine learning interatomic potentials: Application to elemental titanium. *Phys. Rev. Materials* 1, 063801 (2017).

4. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* 98, 146401 (2007).

5. Grisafi, A., Wilkins, D. M., Csányi, G. & Ceriotti, M. Symmetry-adapted machine learning for tensorial properties of atomistic systems. *Phys. Rev. Lett.* 120, 036002 (2018).

6. Aidas, K., et al. The Dalton quantum chemistry program system. *WIREs Computational Molecular Science* 4, 269–284 (2014).

7. Frisch, M. J., et al. Gaussian 16 Rev. B.01 (2016).

8. Berman, H. M., et al. The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242 (2000).

9. Westermayr, J., Gastegger, M., Menger, M., Mai, S., Gonzalez, L. & Marquetand, P. Machine learning enables long time scale molecular photo-dynamics simulations. *Chem. Sci.* 10, 8100-8107 (2019).

10. Abramavicius, D., Palmieri, B. & Mukamel, S. Extracting single and two-exciton couplings in photosynthetic complexes by coherent two-dimensional electronic spectra. *Chem. Phys.* 357, 79–84 (2009).

11. Abramavicius, D., Jiang, J., Bulheller, B. M., Hirst, J. D. & Mukamel, S. Simulation Study of Chiral Two-Dimensional Ultraviolet Spectroscopy of the Protein Backbone. *J. Am. Chem. Soc.* 132, 7769–7775 (2010).

12. Kasha, M., Rawls, H. R. & El-Bayoumi, M. A. The exciton model in molecular spectroscopy. *Pure Appl. Chem.* 11, 371–392 (1965).
Table S1: Hyperparameters for depicting embedded density used in the T-EANN models for PDMs and polarizability tensors of H₂O, (H₂O)₂, H₅O₂⁺, and liquid water, respectively.

| Hyper-parameters | L  | Δrᵢ (Å) | α (Å²) | Rₛ (Å) |
|------------------|----|---------|--------|--------|
| H₂O              | 1  | 0.38    | 1.4    | 3.5    |
| (H₂O)₂           | 1  | 0.54    | 0.70   | 5.0    |
| H₅O₂⁺            | 1  | 0.64    | 0.48   | 6.0    |
| Liquid water     | 1  | 0.54    | 0.70   | 5.0    |
Table S2: Hyper-parameters for depicting embedded density used in the T-EANN model for TDM of N-methylacetamide (NMA) molecule for nπ*/ππ* excitation.

| Hyper-parameters | $L$  | $\Delta r$ (Å) | $\alpha$ (Å$^2$) | $R_c$ (Å) |
|------------------|------|---------------|-----------------|----------|
| nπ*/ππ*          | 2    | 0.53         | 0.71            | 6.0      |