Supplementary Materials for Deep Neural Network for Phonon-Assisted Optical Spectra of Semiconductors at Finite Temperatures

Shengguo Yang,[1,2,3] Shishir Kumar Pandey,[4,5]
Zhanghao Zhouyin,[6] and Qiangqiang Gu[1,2,3,*]

[1]*School of Artificial Intelligence and Data Science, University of Science and Technology of China, Hefei 230026, China*
[2]*Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China*
[3]*Suzhou Big Data & AI Research and Engineering Center, Suzhou 215123, China*
[4]*Department of General Sciences, Birla Institute of Technology and Science, Pilani-Dubai Campus, Dubai International Academic City, Dubai 345055, UAEE*
[5]*Department of Physics, Birla Institute of Technology and Science, Pilani, Hyderabad Campus, Jawahar Nagar, Kapra Mandal, Medchal District, Telangana 500078, India*
[6]*Department of Physics, McGill University, Montreal, Quebec, Canada H3A2T8*

[*] guqq@ustc.edu.cn

I. VALIDATION OF DEEPTB AND DEEPMD MODELS

A. Validation of DeePMD models

The trained DeePMD [1] models are evaluated on the testing dataset for the mean absolute errors (MAE) of various properties. For Si, the MAE of energy, force, and virial predictions were 6.06 meV/atom, 136.44 meV/Å, and 21.54 meV/atom, respectively. In the case of GaAs, the corresponding MAE values were 4.66 meV/atom for energy, 129.02 meV/Åfor force, and 21.02 meV/atom for virial. Fig. S1 and S2 illustrate the comparison between the DP model predictions and the DFT reference data for energies, forces, and virials for Si and GaAs, respectively. These figures provide a visual representation of the model's performance and accuracy in predicting these key properties for both semiconductor materials.

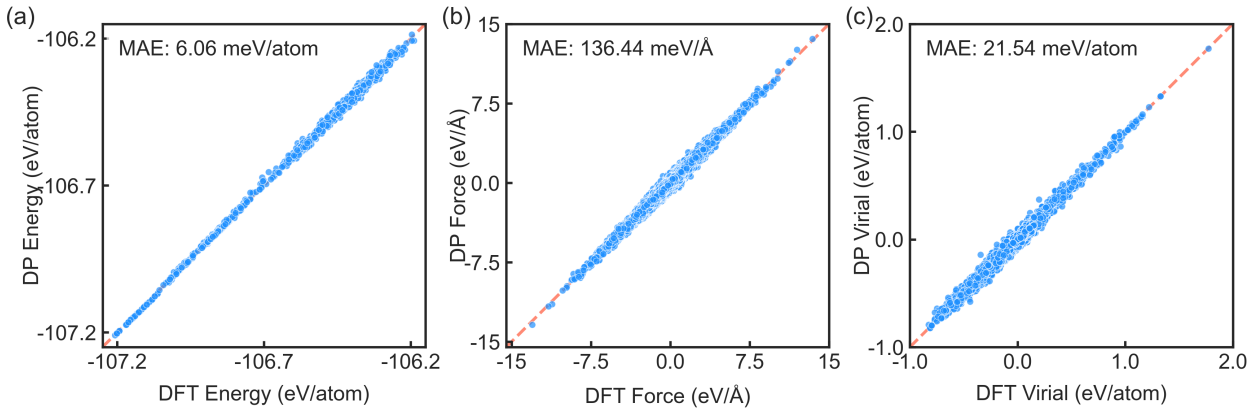

FIG. S1. The comparison of the DeePMD model predicted energies, forces, and virials with the DFT reference data for Si. (a) The energy comparison. (b) The force comparison. (c) The virial comparison.

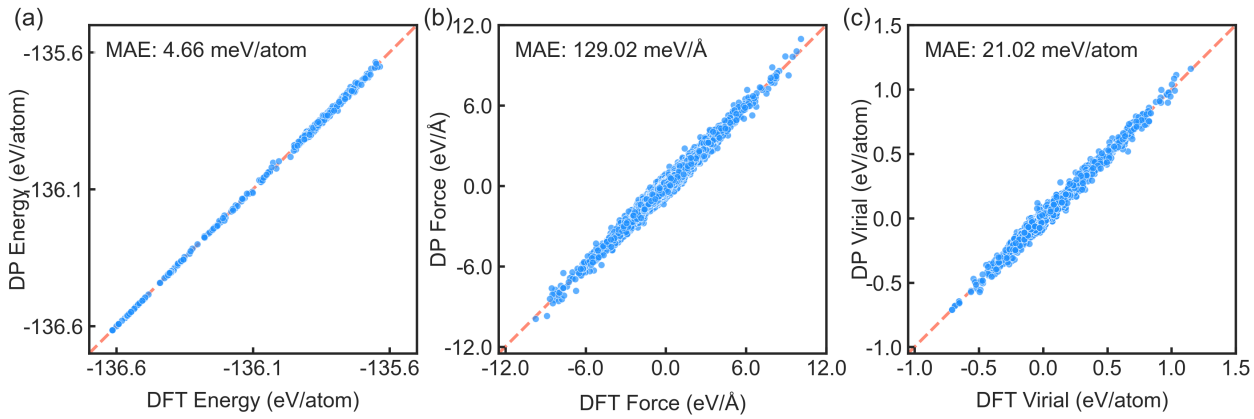

FIG. S2. The comparison of the DeePMD model predicted energies, forces, and virials with the DFT reference data for GaAs. (a) The energy comparison. (b) The force comparison. (c) The virial comparison.

B. Validation of DeePTB models on convention units cells

The DeePTB [2] model was tested using the test dataset with conventional unit cells, the same size as the training datasets. The MAE in the eigenvalues was found to be 44.39 meV for Si, 24.86 meV for GaAs, and 30.7 meV for GaP. These results demonstrate the model's capability in predicting electronic properties with reasonable accuracy for both semiconductor materials. To visually represent the performance of the TB model, we present comparisons between the model predictions and the DFT reference data for Si, GaAs, and GaP in Fig. S3 (a), (b), and (c), respectively. These figures offer a comprehensive view of the model's predictive accuracy across the range of calculated eigenvalues, providing insight into its reliability for electronic structure predictions in these semiconductor systems.

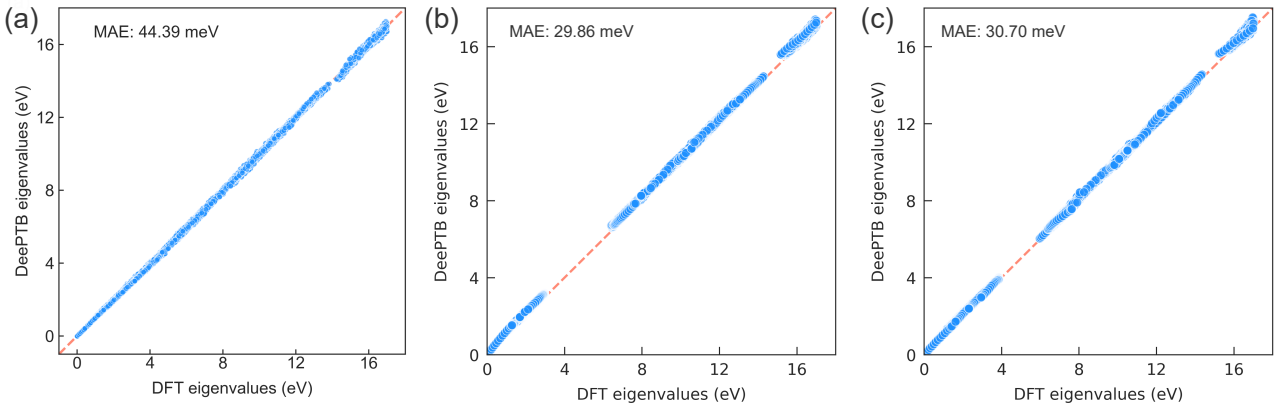

FIG. S3. The comparison between the DeePTB model predicted eigenvalues and the DFT reference data for Si (a), GaAs (b), and GaP (c).

C. Validation of DeePTB models on 2×2×2 supercells

Our model's design is based on the principle of locality, meaning the properties of an atom are determined by its local environment. The conventional unit cells used for training were sufficiently large to contain the full local environment defined by our cutoff radius. Therefore, the model is expected to be directly transferable to larger supercells.

To explicitly assess the transferability of the proposed model, a validation test was conducted using larger supercells. The DFT eigenvalues were calculated for 2×2×2 supercells of Si, GaAs, and GaP. These results were then compared with predictions from the DeePTB

models, which had been trained exclusively on data from conventional unit cells using the HSE functional. As shown in Fig. S4(a), Fig. S5(a), and Fig. S6(a), the resulting MAE for the 2×2×2 supercells are 43.8 meV for Si, 12.8 meV for GaAs, and 16.3 meV for GaP, respectively. Furthermore, Fig. S4(b), Fig. S5(b), and Fig. S6(b) show a direct comparison of the band structures for a randomly selected snapshot, demonstrating excellent agreement across the entire Brillouin zone. This test quantitatively confirms the excellent transferability of our model and demonstrates that its accuracy is independent of the supercell size used for training.

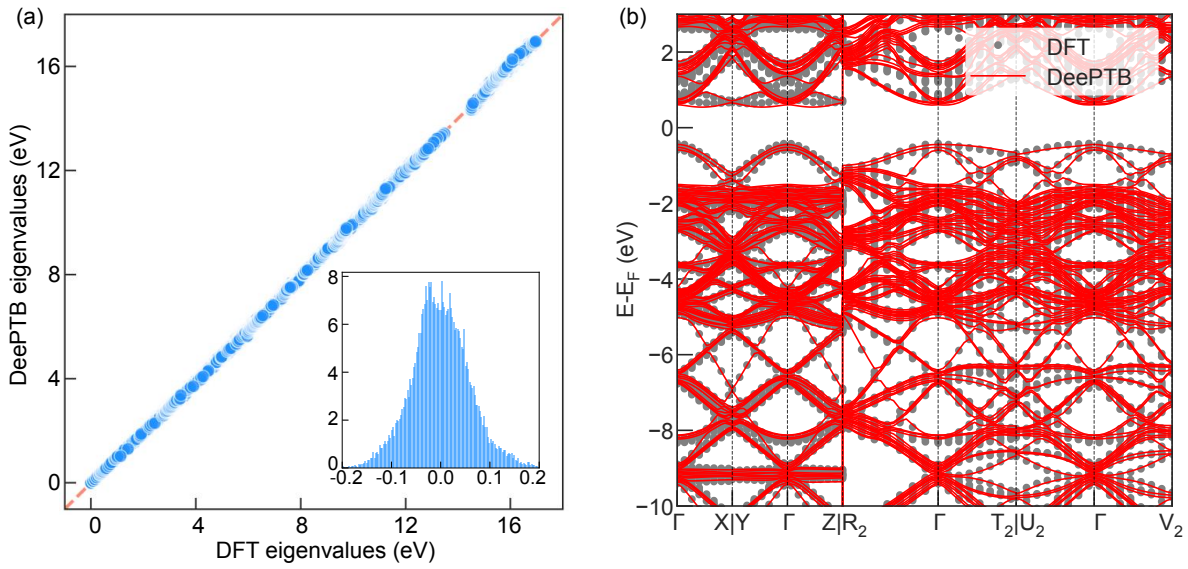

FIG. S4. Transferability test of the Si DeePTB model on a 2×2×2 supercell. (a) Comparison of predicted eigenvalues vs. DFT (HSE) eigenvalues. The MAE is 43.8 meV, with the error distribution shown in the inset. (b) Comparison of the band structure from a randomly selected configuration in the test set, showing excellent agreement between the DFT reference (gray dots) and the DeePTB prediction (red lines).

II. LOW-PRECISION CALCULATIONS

Figure S7 shows absorption spectra of GaAs and GaP predicted by DeePTB models trained on SCAN and PBE XC-functionals with scissor correction. The corresponding training datasets were constructed from 100 snapshots taken from MD trajectories of the conventional unit cell at 100 and 300 K. Each snapshot was labeled with DFT eigenvalues

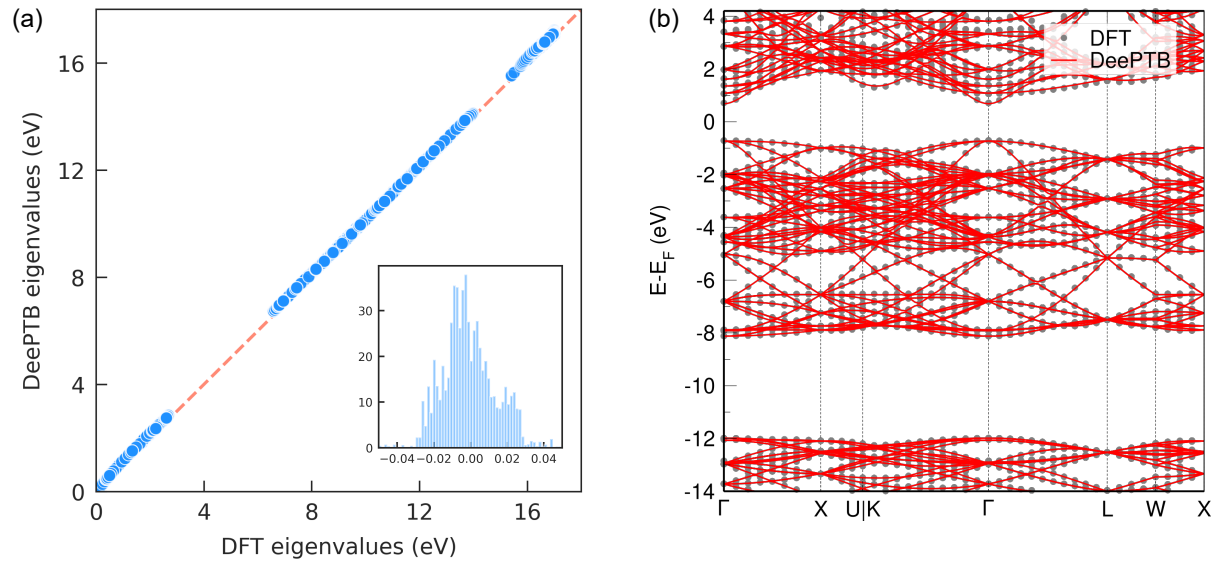

FIG. S5. Transferability test of the GaAs DeePTB model on a 2×2×2 supercell. (a) Comparison of predicted eigenvalues vs. DFT (HSE) eigenvalues. The MAE is 12.8 meV, with the error distribution shown in the inset. (b) Comparison of the band structure from a randomly selected configuration, showing excellent agreement between the DFT reference (gray dots) and the DeePTB prediction (red lines).

computed in ABACUS using either the SCAN or PBE functional and a double-zeta plus polarization (DZP) orbital basis set. The training parameters for the DeePTB model were the same as those used for Si, as detailed in the Methods section of the main text. After training, the model yields MAEs of 21.4 and 23.5 meV on the test set for GaAs and GaP, respectively. The well-trained models are then used to calculate the optical spectra with a rigid bandgap shift, i.e., scissor correction.

As shown in Fig. S7(a) and (b) for GaAs and GaP, the calculated spectra are in qualitative agreement with available experimental data [3]. Nevertheless, in comparison to the SCAN and PBE results corrected with scissors, the spectra predicted by the HSE-trained DeePTB model (in the main text) show significantly improved agreement with experiment in both the absorption onset and the overall spectral lineshapes. This indicates that a more accurate description of the electronic structure is crucial for quantitatively reproducing optical spectra.

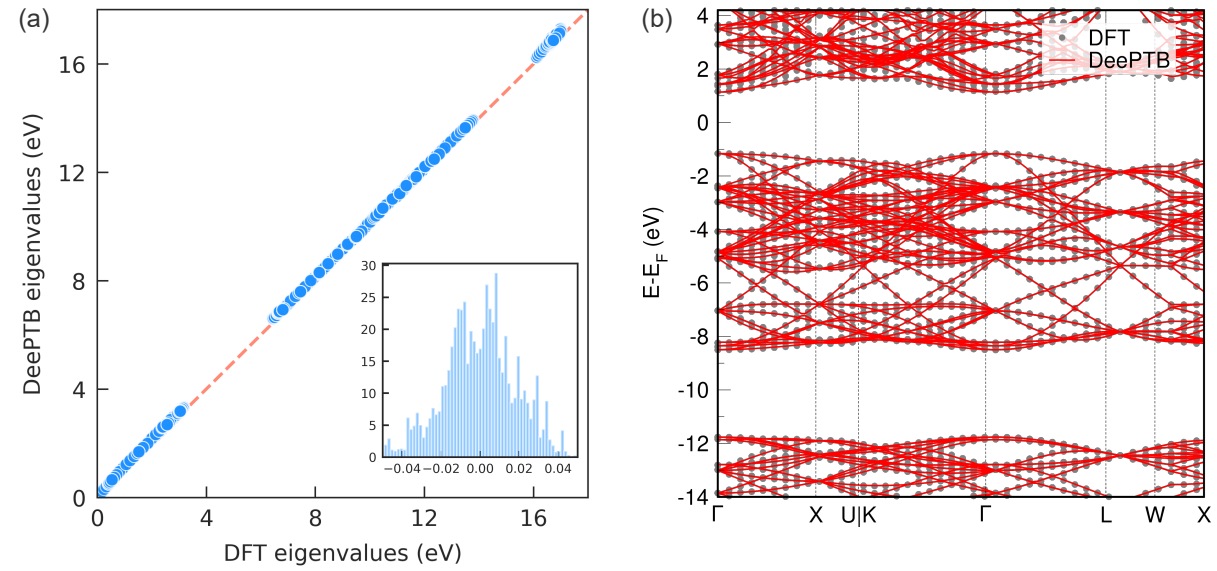

FIG. S6. Transferability test of the GaP DeePTB model on a 2×2×2 supercell. (a) Comparison of predicted eigenvalues vs. DFT (HSE) eigenvalues. The MAE is 16.3 meV, with the error distribution shown in the inset. (b) Comparison of the band structure from a randomly selected configuration, showing excellent agreement between the DFT reference (gray dots) and the DeePTB prediction (red lines).

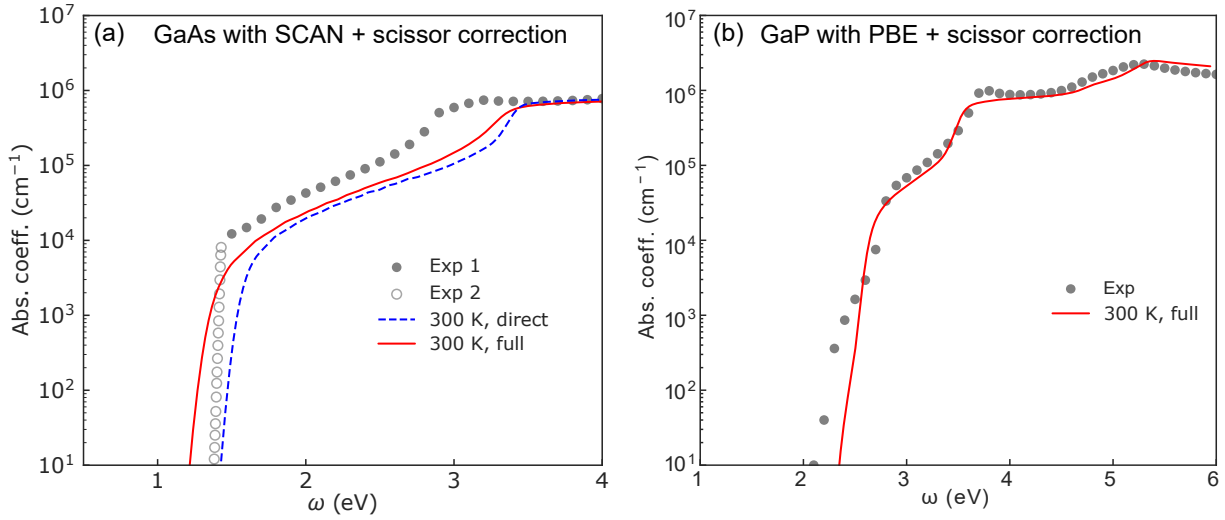

FIG. S7. Temperature-dependent absorption spectra of GaAs and GaP predicted by DeePTB models trained on different electronic-structure data, which are compared with experiment. (a) GaAs results obtained using a SCAN-trained DeePTB model with scissor correction. (b) GaP results obtained using a PBE-trained DeePTB model with scissor correction.

III. SCALABILITY DEMONSTRATION WITH A LINEAR-SCALING METHOD

The 4096-atom supercell used for the high-precision calculations in the main text should not be viewed as a capability limit of our ML framework. The core components of our method, the DeePMD potential and the DeePTB Hamiltonian prediction, both scale linearly, $\mathcal{O}(N)$, with system size. The bottleneck for the high-precision optical spectra shown in the main text was the choice to use matrix diagonalization, a post-processing step that scales as $\mathcal{O}(N^3)$.

To demonstrate the true scalability of our framework, linear-scaling post-processing methods can be employed. As definitive proof of this capability, we performed an optical absorption spectrum calculation for a 512,000-atom Si supercell using the Tight-Binding Propagation Method (TBPM), an $\mathcal{O}(N)$ approach. As shown in Fig. S8, this massive-scale calculation successfully reproduces the key spectral features, such as the direct and indirect absorption regimes, confirming that our framework opens the door to simulations on scales that were previously intractable. We emphasize that this calculation serves primarily as a demonstration of the method's scalability. The statistical noise observed at low absorption coefficient values is a known feature of the stochastic TBPM approach and can be systematically reduced with increased computational sampling.

[1] L. Zhang, J. Han, H. Wang, R. Car, and W. E, Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics, Phys. Rev. Lett. **120**, 143001 (2018).

[2] Q. Gu, Z. Zhouyin, S. K. Pandey, P. Zhang, L. Zhang, and W. E, Deep learning tight-binding approach for large-scale electronic simulations at finite temperatures with ab initio accuracy, Nat. Commun. **15**, 6772 (2024).

[3] D. E. Aspnes and A. A. Studna, Dielectric functions and optical parameters of si, ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6.0 eV, Phys. Rev. B **27**, 985 (1983).

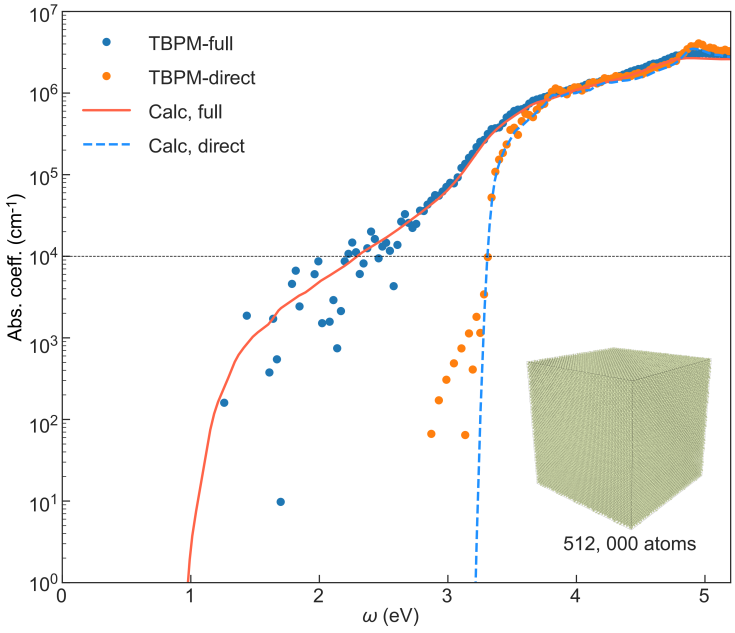

FIG. S8. Scalability demonstration of the ML framework with a 512,000-atom Si supercell. The optical absorption spectrum at 300 K (dots) was calculated using the $\mathcal{O}(N)$ Tight-Binding Propagation Method (TBPM) and is compared with the converged $\mathcal{O}(N^3)$ diagonalization results from the main text (lines). This calculation is presented to demonstrate the framework's capability to scale to massive system sizes.