Supplementary Information:

NEP-MB-pol: A unified machine-learned framework for fast and accurate prediction of water's thermodynamic and transport properties

Ke Xu[1,2], Ting Liang[2], Nan Xu[3], Penghua Ying[4], Shunda Chen[*5], Ning Wei[6], Jianbin Xu[†2], and Zheyong Fan[‡1]

[1] *College of Physical Science and Technology, Bohai University, Jinzhou 121013, P. R. China*
[2] *Department of Electronic Engineering and Materials Science and Technology Research Center, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, 999077, P. R. China*
[3] *College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310058, P. R. China*
[4] *Department of Physical Chemistry, School of Chemistry, Tel Aviv University, Tel Aviv, 6997801, Israel*
[5] *Department of Civil and Environmental Engineering, George Washington University, Washington, DC 20052, USA*
[6] *Jiangsu Key Laboratory of Advanced Food Manufacturing Equipment and Technology, Jiangnan University, Wuxi, 214122, China*

November 22, 2024

[*]Email: phychensd@gmail.com
[†]Email: jbxu@ee.cuhk.edu.hk
[‡]Email: brucenju@gmail.com

Supplementary Figures

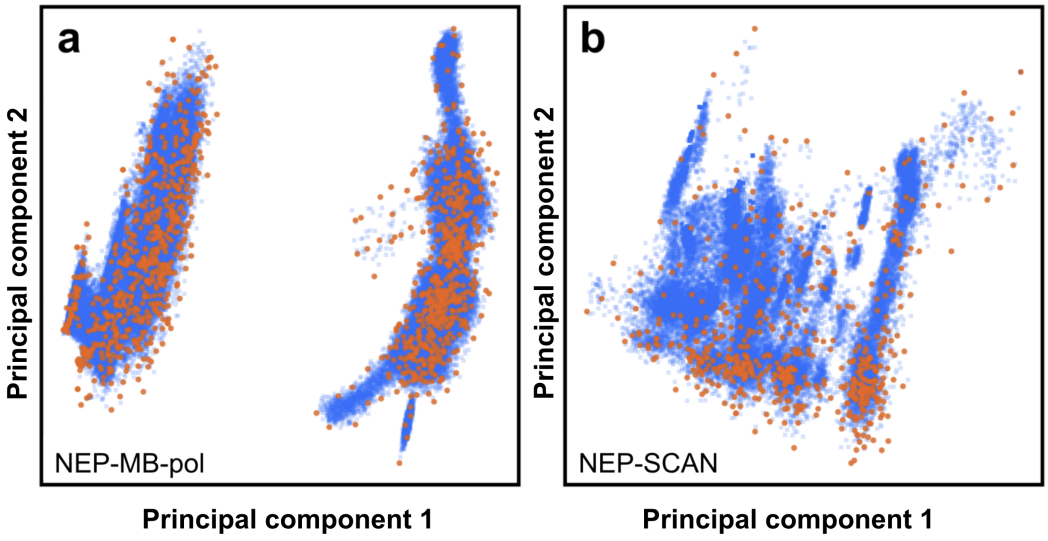

Figure S1: **Training data subsampling.** Representative training datasets were selected from the original extensive datasets [1, 2] using farthest point sampling in descriptor space and visualized via principal component analysis for (a) the neuroevolution potential (NEP) trained on many-body polarization (MB-pol) reference dataset and (b) the NEP trained on the strongly constrained and appropriately normed (SCAN) functional reference dataset. Blue points represent the original full datasets, while orange points indicate the selected representative datasets, comprising about 1% of the original full datasets.

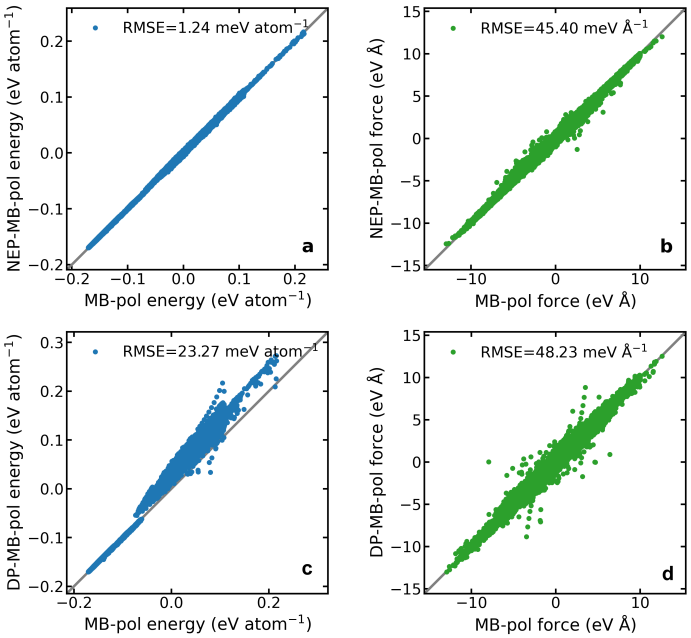

Figure S2: **Parity plots for MB-pol dataset.** Parity plots comparing energy and force predictions from (a-b) neuroevolution potential (NEP) and (c-d) deep potential (DP) trained on many-body polarization (MB-pol) reference datasets [1].

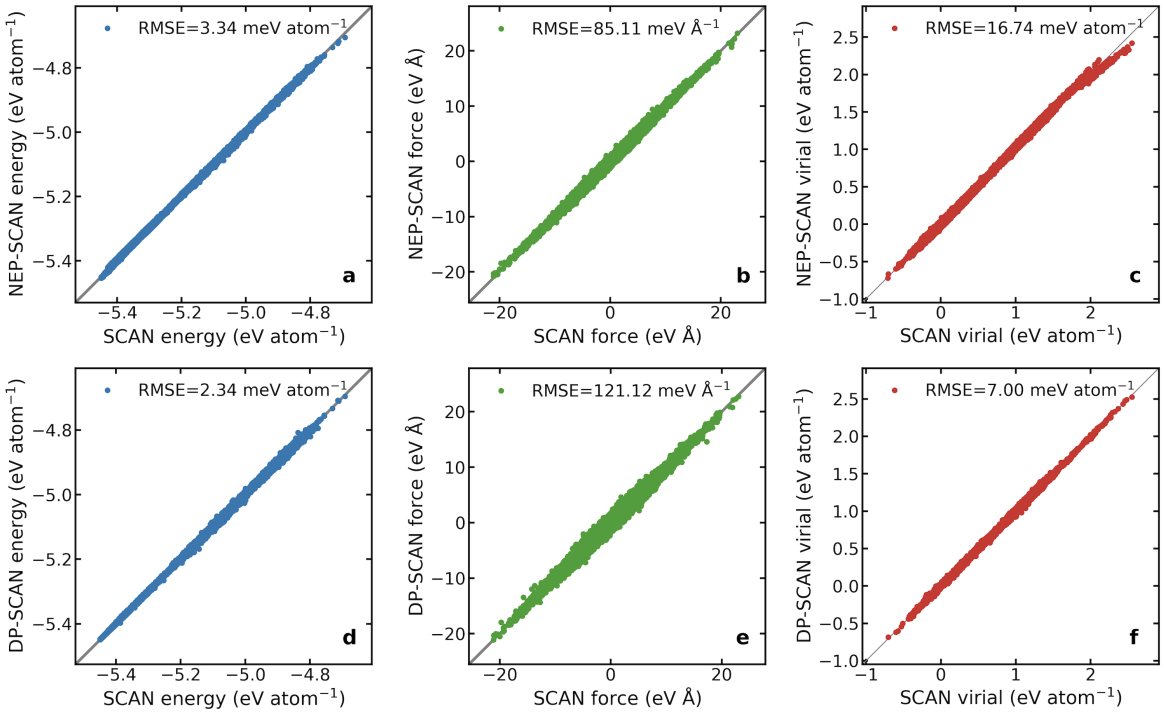

Figure S3: **Parity plots for SCAN dataset.** Parity plots comparing energy, force, and virial predictions from (a-b) neuroevolution potential (NEP) and (c-d) deep potential (DP) trained on the strongly constrained and appropriately normed (SCAN) functional datasets [2].

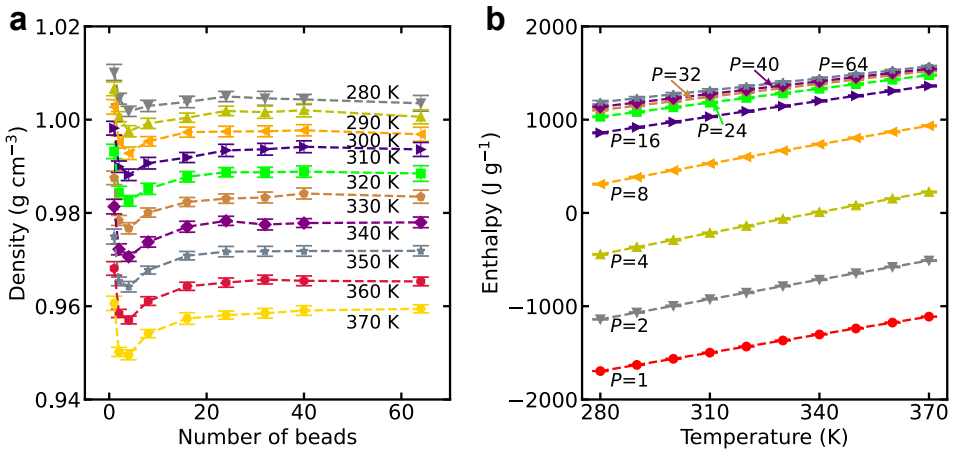

Figure S4: **Density and enthalpy of water.** (a) Density of water as a function of the number of beads at various temperatures, calculated using path-integral molecular dynamics (PIMD) simulations with the neuroevolution potential trained on many-body polarization reference dataset (NEP-MB-pol). (b) System enthalpy as a function of temperature for different numbers of beads, P, obtained using PIMD simulations with the same potential. The pressure is kept at 1 atm.

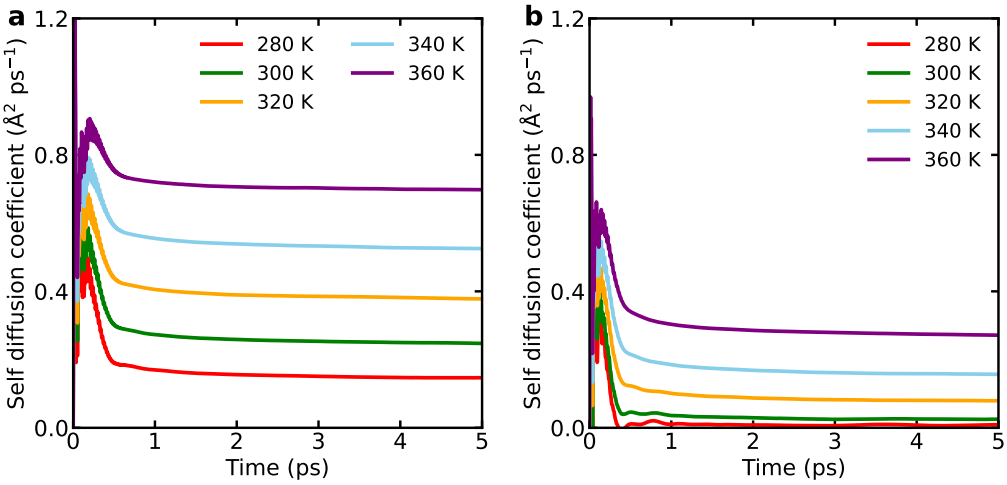

Figure S5: **Running self-diffusion coefficient of water.** Convergence of the self-diffusion coefficient of water as a function of correlation time, calculated via thermostatted ring-polymer molecular dynamics simulations with 32 beads, using the neuroevolution potential trained on (a) the many-body polarization (NEP-MB-pol) and (b) the strongly constrained and appropriately normed functional (NEP-SCAN) reference datasets at various temperatures and 1 atm.

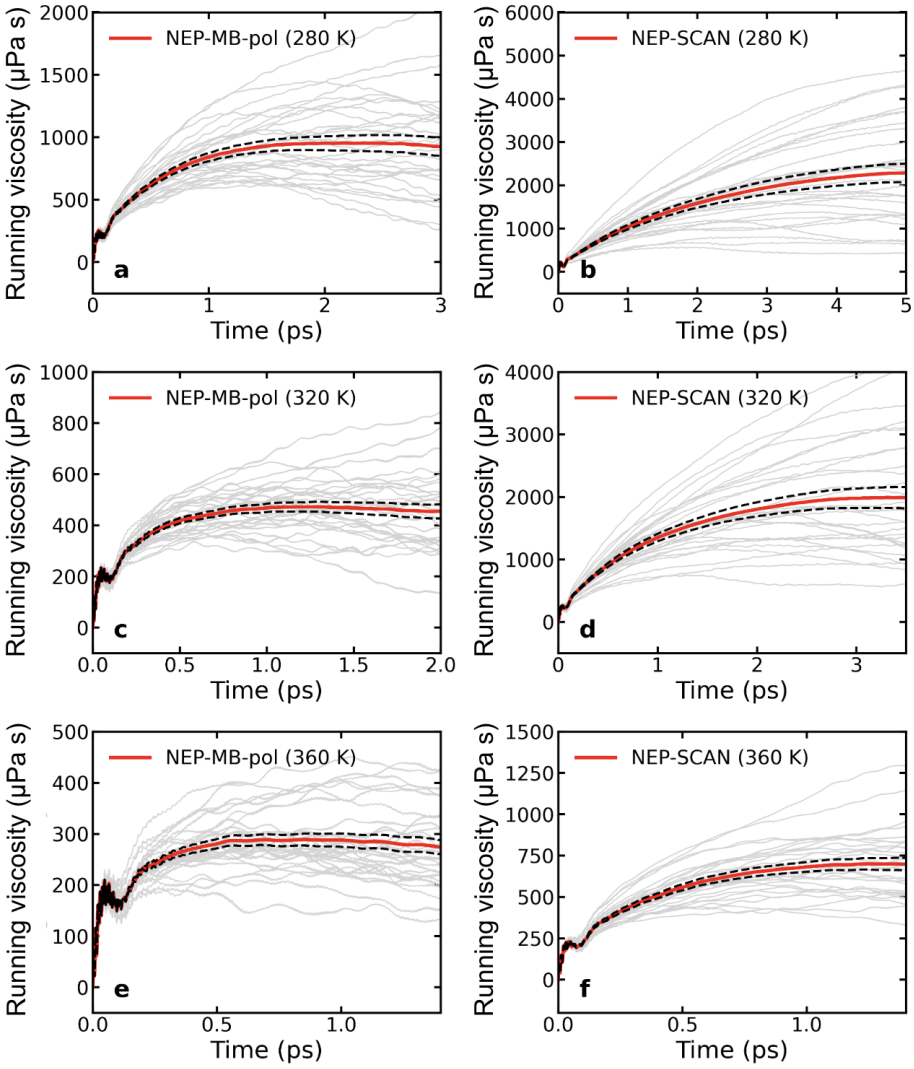

Figure S6: **Running viscosity of water.** Convergence of viscosity of water as a function of correlation time, calculated via thermostatted ring-polymer molecular dynamics (TRPMD) simulations with 32 beads, using neuroevolution potential trained on (a,c,e) the many-body polarization (NEP-MB-pol) and (b,d,f) the strongly constrained and appropriately normed functional (NEP-SCAN) reference datasets at various temperatures and 1 atm.

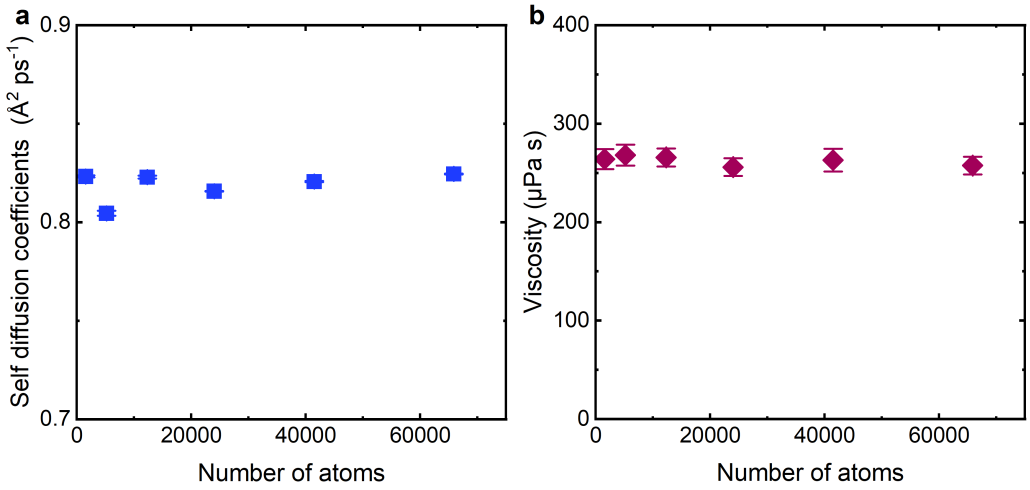

Figure S7: **Size convergence.** Convergence of (a) self-diffusion coefficient and (b) viscosity of water as a function of system size (number of atoms), calculated via thermostatted ring-polymer molecular dynamics (TRPMD) simulations with 32 beads, using NEP-MB-pol at 300 K and 1 atm.

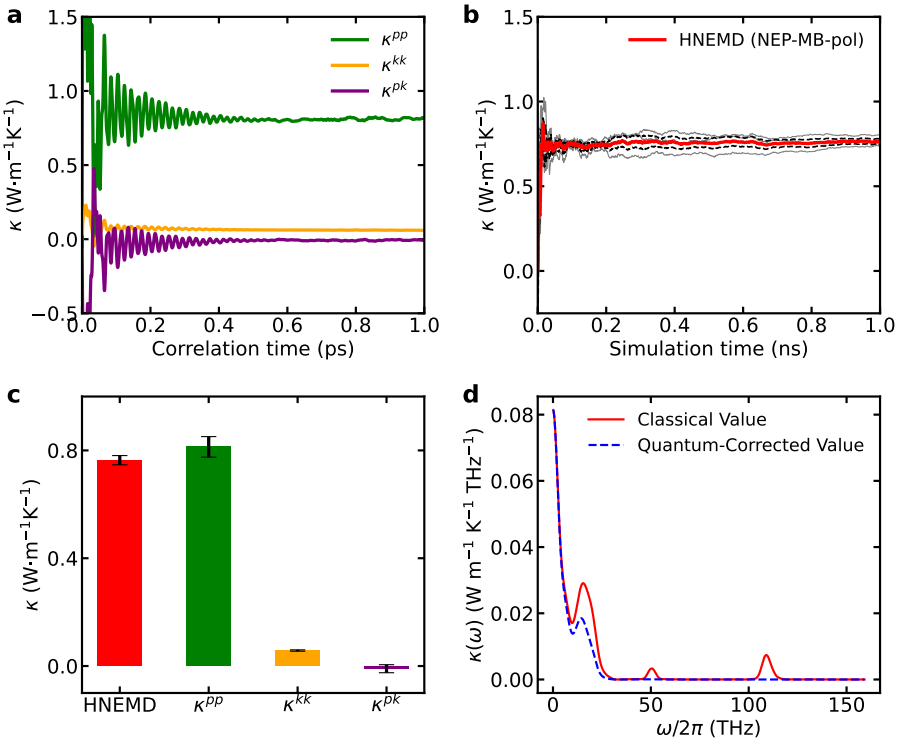

Figure S8: **Thermal conductivity.** (a) Convergence of water's thermal conductivity as a function of correlation time t, calculated using equilibrium molecular dynamics (EMD) with NEP-MB-pol at 300 K and 1 atm, showing contributions from the potential-potential term (κ^{pp}), the kinetic-kinetic term (κ^{kk}), and the cross term (κ^{pk}), respectively. (b) Thermal conductivity (potential-potential term) of water as a function of simulation time obtained using homogeneous nonequilibrium molecular dynamics (HNEMD) method. (c) Comparison of thermal conductivity values from EMD and HNEMD methods. (d) Classical and quantum-corrected spectral thermal conductivity (potential-potential term) for water.

Supplementary References

[1] Yaoguang Zhai, Alessandro Caruso, Sigbjørn Løland Bore, Zhishang Luo, and Francesco Paesani. A "short blanket" dilemma for a state-of-the-art neural network potential for water: Reproducing experimental properties or the physics of the underlying many-body interactions? *The Journal of Chemical Physics*, 158(8): 084111, 02 2023. ISSN 0021-9606. doi: 10.1063/5.0142843. URL https://doi.org/10.1063/5.0142843.

[2] Linfeng Zhang, Han Wang, Roberto Car, and Weinan E. Phase diagram of a deep potential water model. *Phys. Rev. Lett.*, 126:236001, Jun 2021. doi: 10.1103/PhysRevLett.126.236001. URL https://link.aps.org/doi/10.1103/PhysRevLett.126.236001.