Electronic Supplementary Information: Active Sparse Bayesian Committee Machine Potential for Isothermal-Isobaric Molecular Dynamics Simulations

Soohaeng Yoo Willow,[1] Dong Geon Kim,[1] R. Sundheep,[1]
Amir Hajibabaei,[2] Kwang S. Kim,[3] and Chang Woo Myung[1,*]

[1] *Department of Energy Science, Sungkyunkwan University, Seobu-ro 2066, Suwon, 16419, Korea*
[2] *Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom*
[3] *Center for Superfunctional Materials, Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea*

(Dated: July 30, 2024)

CONTENTS

We provide additional supporting data as well as contextual information to the main text here. All input and output files are provided on GitHub, which contains a Jupyter Notebook file that analyses the data.

S1. ACTIVE BCM MD SIMULATION

The goal of on-the-fly active learning MD simulations is to automatically construct both the dataset and the inducing set. Using these sets, the weight parameter vector is determined by fitting the predicted values of energy, forces, and virial pressures to their target values. During the MD simulations, we update the inducing set $z_P = \{\chi_j\}$ and dataset $X_P = \{R_n\}$ according to the SGPR technique. The SGPR technique works as follows.

1. A new configuration x^* is generated from the MD simulation.

2. The local chemical environments (LCEs) ρ of atoms are generated.

3. The kernel matrices for energy $\mathcal{K}(\rho_i, \chi_j)$, forces $\dot{\mathcal{K}}_i^{\mu}(R, \chi_j)$, and virial pressures $\dot{\mathcal{K}}_i^{\mu}(R, \chi_j) r_i^{\nu}$ are calculated as per Eq (6), (7), and (9).

4. Energy, forces, and virial pressures are predicted using Eq. (11), (13), and (14) for the BCM potential.

5. Check for new LCEs by calculating the covariance loss $\sigma^2(\rho)$ between an LCE and the inducing sets of all local experts:

$$\sigma^2(\rho) = (1 - \mathbf{K}_{\rho m} \mathbf{K}_{mm}^{-1} \mathbf{K}_{\rho m}^T). \tag{S1}$$

6. Update the inducing set when new LCEs are detected, with the condition $\sigma^2(\rho) > \sigma_{\text{cutoff}}$. Here, $\sigma_{\text{cutoff}} = 0.05$.

7. The updated inducing set $z = \{\chi_j\}$ changes the kernel matrices for energy, forces, and virial pressures, which in turn affects the predicted energy, forces, and virial pressures. If the change in the predicted energy ΔE is significant ($\Delta E > \Delta E_{\text{cutoff}}$), we update the dataset by adding the new configuration x^* along with its energy, forces, and virial pressures. In our calculation, $\Delta E_{\text{cutoff}} = 0.05$ eV.

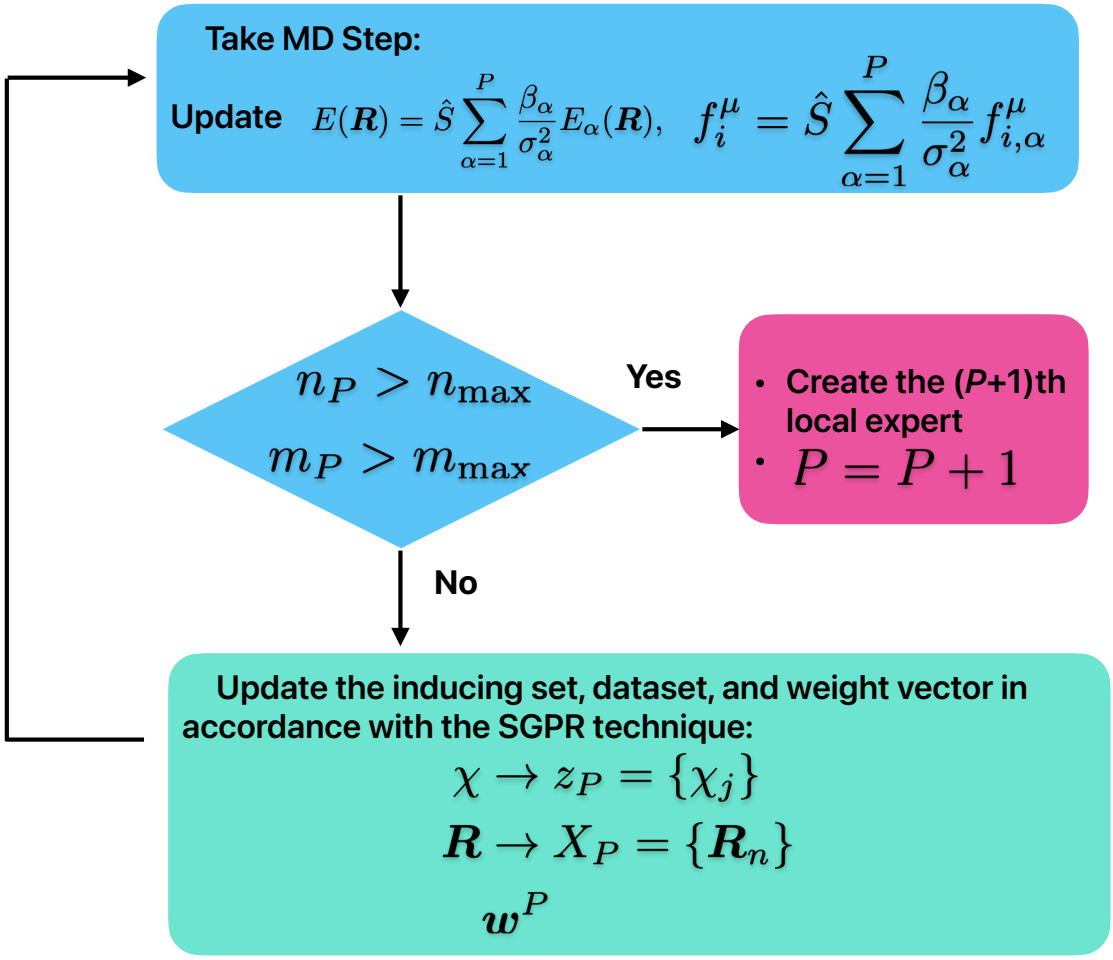

FIG. S1. On-the-fly active learning of Bayesian committee machine (BCM) potential during MD simulations. During each time-step of the MD simulation, the BCM potential calculates energy, forces, and stresses. If the number of data in $X_P = \{R_n\}$ or inducing set $z_P = \{\chi_j\}$ exceeds the threshold of a predefined kernel size (for example, $n_{\max} = 50$, $m_{\max} = 200$), a new local expert sparse Gaussian process regression potential model is trained.

S2. LIQUID BORON NITRIDE NVT MD SIMULATION

NVT MD simulations on liquid boron nitride were carried out at $T = 7500$ K. The simulations were conducted on a unit cell (with a lattice parameter of 7.252Å) consisting of 64 atoms as shown in Figure S2.

The MD simulations were carried out using the MD engine from the Atomic Simulation Environment (ASE) python library.[1] The simulations ran for 8 ps using Nosé-Hoover dynamics with a target temperature of 7500 K, a time step of 1 fs and time constant of 25 fs. For these ASE MD simulations, the forces were calculated using the Vienna Ab initio Simulation Package (VASP).[2,3]

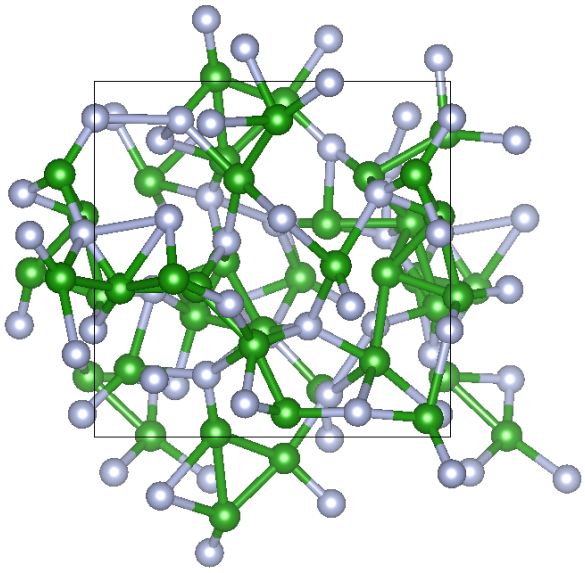

FIG. S2. High-density liquid-phase BN Unit cell where green represents Boron atoms, and white represents Nitrogen atoms, respectively.

Before running the MD simulations, we performed convergence calculations for cutoff energy and the k-point grid (Figure S3) to achieve accurate DFT results. Based on these results, we employed an energy cutoff of 650 eV, and $(3 \times 3 \times 3)$ Monkhorst-pack k-grid at the r^2SCAN-D4 functional level.[4,5]

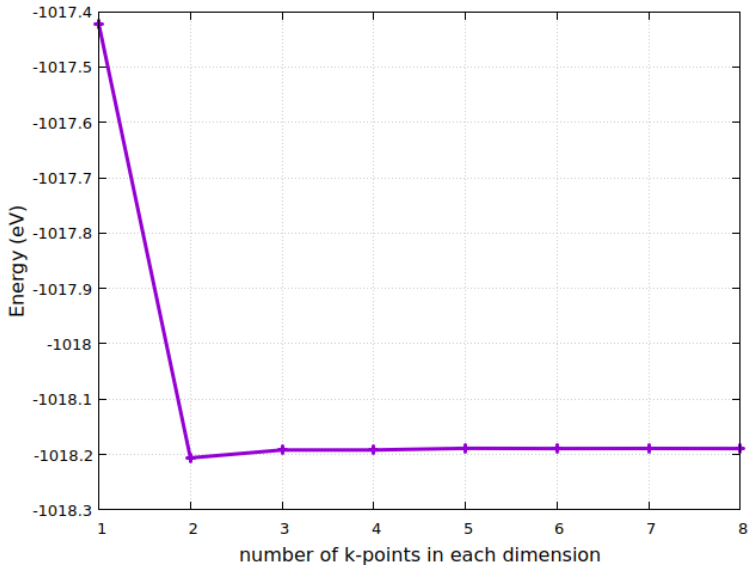

FIG. S3. K-point grid convergence calculations for high density liquid phase of BN.

S3. $Li_{10}GE(PS_6)_2$ NpT MD SIMULATION

All DFT calculations were performed using the Vienna Ab-initio Simulation Package(VASP)[2,3] with the projector augmented wave (PAW) method at the Perdew–Burke–Ernzerhof (PBE)[6] functional level to train the SGPR potential. The LGPS model, sourced from the Materials Project database (ID: mp-696128), consisted of 50 atoms (Li: 20, Ge: 2, P: 4, S: 24) within a tetragonal box measuring 8.8 Å × 8.8 Å × 12.7 Å. NpT MD simulations were performed under the atmospheric pressure (∼0.101 MPa) using a Nosé-Hoover thermostat and a Parrinello-Rahman barostat, with 1 fs time step, 500 eV energy cut-off, and $(1 \times 1 \times 1)$ Monkhorst-pack k-point mesh.

We calculated the mean square displacement(MSD) as

$$\mathrm{MSD}(t) = \frac{1}{N}\sum_{i=1}^{N} |\vec{r}_i(t) - \vec{r}_i(0)|^2 \tag{S2}$$

where N is the number of particles, $\vec{r}_i(t)$ and $\vec{r}_i(0)$ is the position of the i-th particle at time t and reference position. The diffusivity (D) is given as

$$D = \lim_{t\to\infty} \frac{1}{2dt}\langle \mathrm{MSD}(t)\rangle \tag{S3}$$

where d is number of dimensions and $\langle\cdot\rangle$ indicates the ensemble average. We estimated the diffusion activation energy by assuming an Arrhenius temperature dependence

$$D = D_0 e^{-E_a/k_B T} \tag{S4}$$

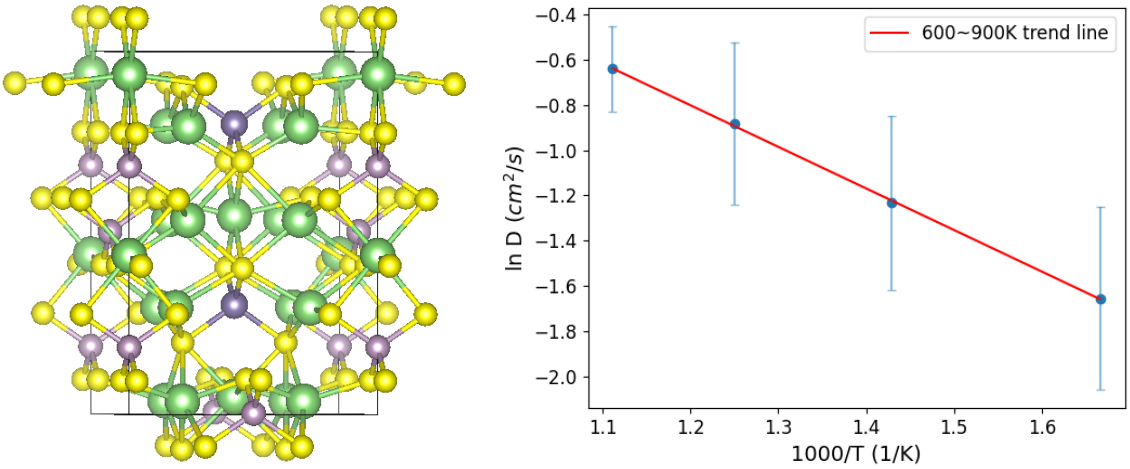

FIG. S4. The unit cell of LGPS (left) and Li-ion diffusivity (cm^2/s) with respect to temperature (right) (Li: green, S: yellow, P: purple, Ge: indigo). The red line represents the trend ranging from 600 K to 900 K. The error bars are estimated using the block average with block size of 10 ps.

S4. ICE-LIQUID NpH MD SIMULATIONS

TABLE S1. Ice-liquid coexisting systems. Total number of water molecules in the simulation box (N_{H_2O}) and initial cell dimensions ($\vec{a}, \vec{b}, \vec{c}, \alpha, \beta, \gamma$) of ice-liquid coexisting systems for NpH MD simulations.

	N_{H_2O}	($\lvert\vec{a}\rvert$, $\lvert\vec{b}\rvert$, $\lvert\vec{c}\rvert$, α, β, γ)
Ice II - Liquid	864	(23.415, 23.415, 43.910, 84.826, 84.826, 113.100)
Ice III- Liquid	648	(19.982, 20.349, 41.379, 90, 90, 90)
Ice V - Liquid	672	(18.553, 45.374, 20.833, 90, 109.215, 90)

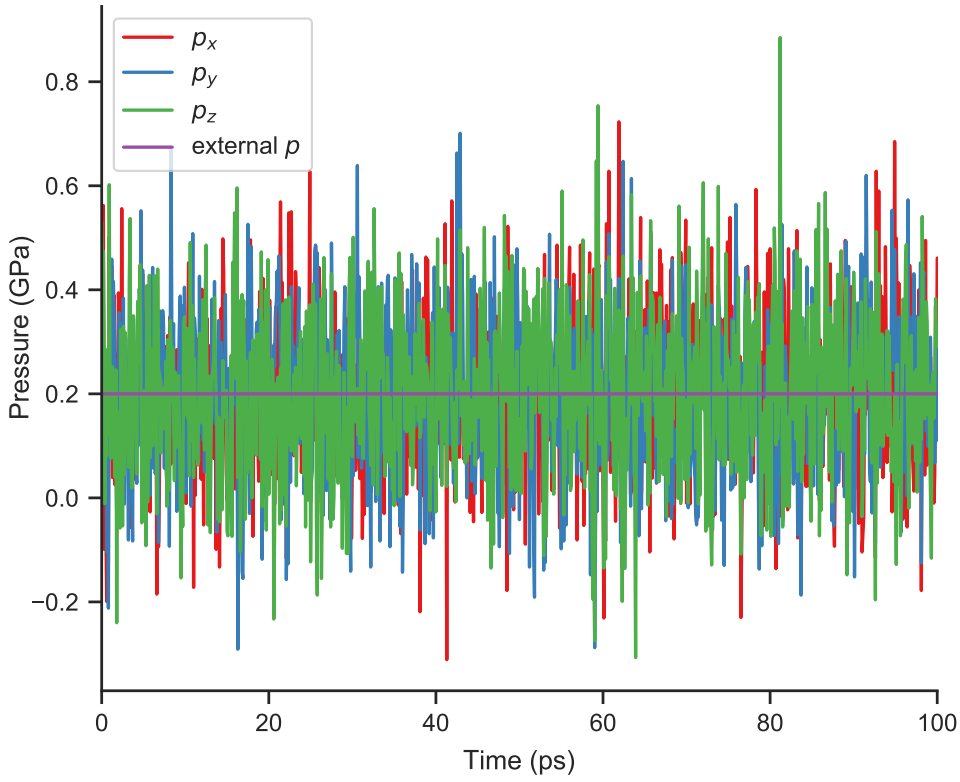

FIG. S5. Instantaneous pressures p_x, p_y, and p_z values in anisotropic NpH MD simulations of coexisting ice II-liquid system at pressure of $p = 0.2$ GPa. Instantaneous pressures are adjusted to specified target pressure during NpH MD simulations. Note that anisotropic NpH MD simulations show no significant pressure differences along different axes.

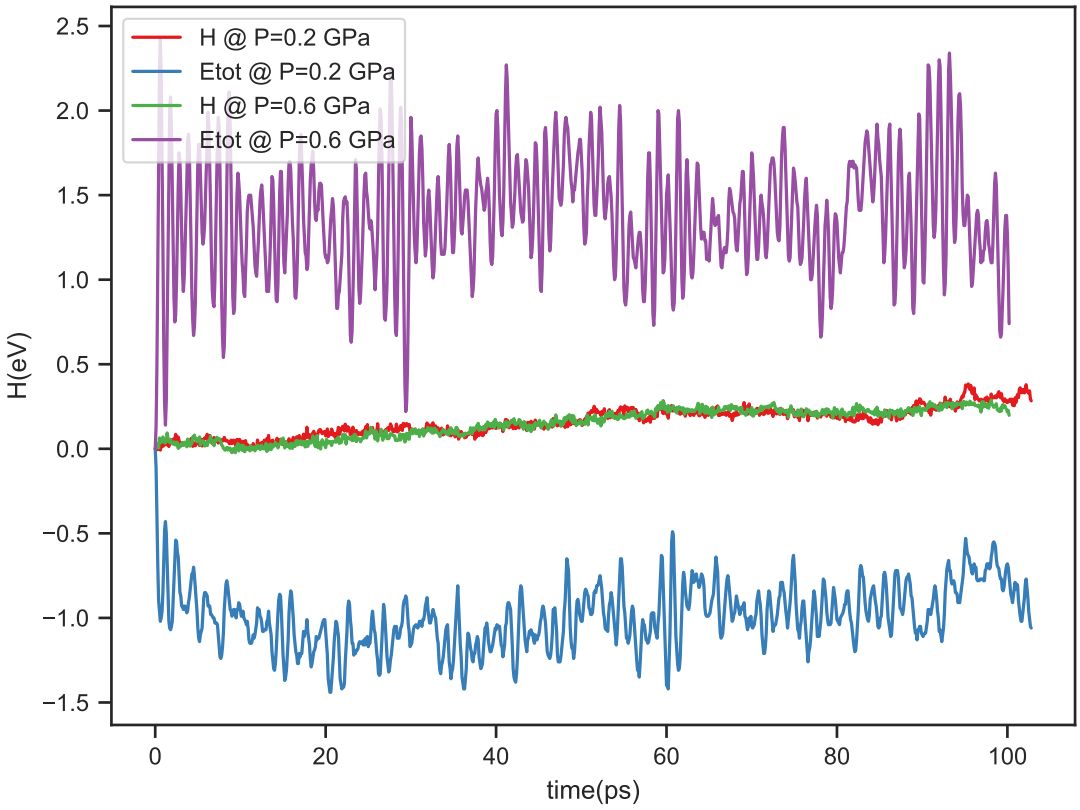

FIG. S6. Enthalpy (H) and total energy (E_{tot}) during NpH MD simulations of coexisting ice II-liquid systems at pressures of $p = 0.2$ and 0.6 GPa. H is conserved while the total energies fluctuate.

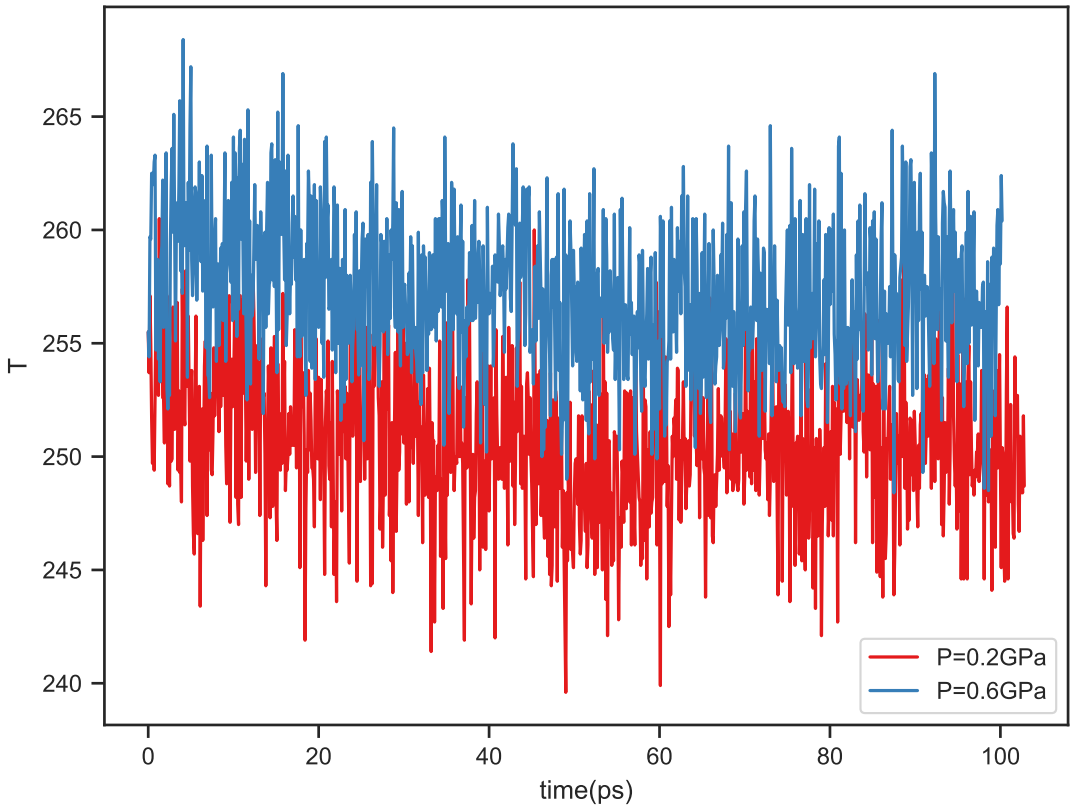

FIG. S7. Instantaneous temperature values in NpH MD simulations of coexisting ice II-liquid system at pressures of $p = 0.2$ and 0.6 GPa. Instantaneous temperatures are tuned to approach melting temperature, ensuring that they meet the condition $\mu_{\mathrm{ice}}(p,T)_{T=T_m} = \mu_{\mathrm{liq}}(p,T)_{T=T_m}$.

S5. MELTING POINTS OF ICE-WATER FOR MB-POL

In Figure S8, we compared the melting points of ice-water using the MB-pol model in our study (shown with circle lines) with those from Bore and Paesani's study[7] (indicated by square lines). While the melting lines of ice III are highly very similar, those of ice V are significant different.

To understand why ice V behaves differently, we aim to share the methods used in both studies. Bore and Paesani[7] described their approach in their supplementary Note3: "These calculations were carried out following a four-stage procedure ... (a) we determined the classical melting points from the chemical potential differences between liquid water and each ice polymorph using enhanced-coexistence simulations carried with the DNN@MB-pol potential, (b) starting from the classical melting points obtained with DNN@MB-pol, we used thermodynamic perturbation theory to determine the classical melting points of MB-pol, (c) ..."

$$\exp\left[-\beta\Delta G(p,T)\right] = \langle \exp[-\beta\Delta U]\rangle_{p,T,\mathrm{DNN@MB-pol}} \tag{S5}$$

$$\Delta U = U_{\mathrm{MB-pol}}(R_{\mathrm{DNN@MB-pol}}) - U_{\mathrm{DNN@MB-pol}}(R_{\mathrm{DNN@MB-pol}}) \tag{S6}$$

$$\Delta\mu = \mu_{\mathrm{MB-pol}} - \mu_{\mathrm{DNN@MB-pol}} = \frac{\Delta G}{N_{\mathrm{H_2O}}} \tag{S7}$$

To calculate the chemical potentials $\mu_{\mathrm{MB-pol}}$ of the MB-pol water model in Eq. (S7) using Eq. (S5), they utilized the coordinates $\{R_{\mathrm{DNN@MB-pol}}\}$ sampled within the DNN@MB-pol potential model to compute the potential energy $U_{\mathrm{MB-pol}}$ in Eq. (S6). Eq. (S5) represents the mathematical formulation of the free energy perturbation (FEP) identity introduced by Zwanzig in 1954[8] rather than thermodynamic perturbation theory. Generally, free energy perturbation is only valid when the coordinates sampled from DNN@MB-pol closely match those from MB-pol. Hence, the accuracy of the free energy difference calculated using Eq. (S5) strongly relies on how much two thermodynamic states overlap in configuration space. To mitigate this overlap requirement, a series of intermediate thermodynamic states are introduced between two thermodynamic end states $U_{\mathrm{MB-pol}}$ and $U_{\mathrm{DNN@MB-pol}}$ as $(1-\lambda)U_{\mathrm{DNN@MB-pol}} + \lambda U_{\mathrm{MB-pol}}$. Because Bore and Paesani[7] did not performed MD simulations with all intermediate and the final thermodynamic states, their estimated potential energies were inaccurate due to neglecting the requirement of the overlap between adjacent thermodynamic states.

In our study, on the other hand, we sampled all configurations $\{R_{\mathrm{MB-pol}}\}$ from NpH MD simulations with the MB-pol water model when we estimate the melting points for MB-pol. Similarly, the melting points for the BCM-MLP model were estimated with all configurations $\{R_{\mathrm{BCM-MLP}}\}$ sampled from NpH MD simulations with the BCM-MLP model. In short, we directly estimated

the melting points, which rely on their specific thermodynamic states and potential energy surface.

In summary, the main reason why our melting points and theirs don't match up, especially melting points of ice V, is because we used different sets of configurations. To estimate the melting points for the MB-pol model, they used the configurations from DNN@MB-pol, while we used ones from MB-pol directly. We predict that if they had calculated the melting points using the MB-pol potential instead of using DNN@MB-pol potential, the melting points of ice V would be much more similar.

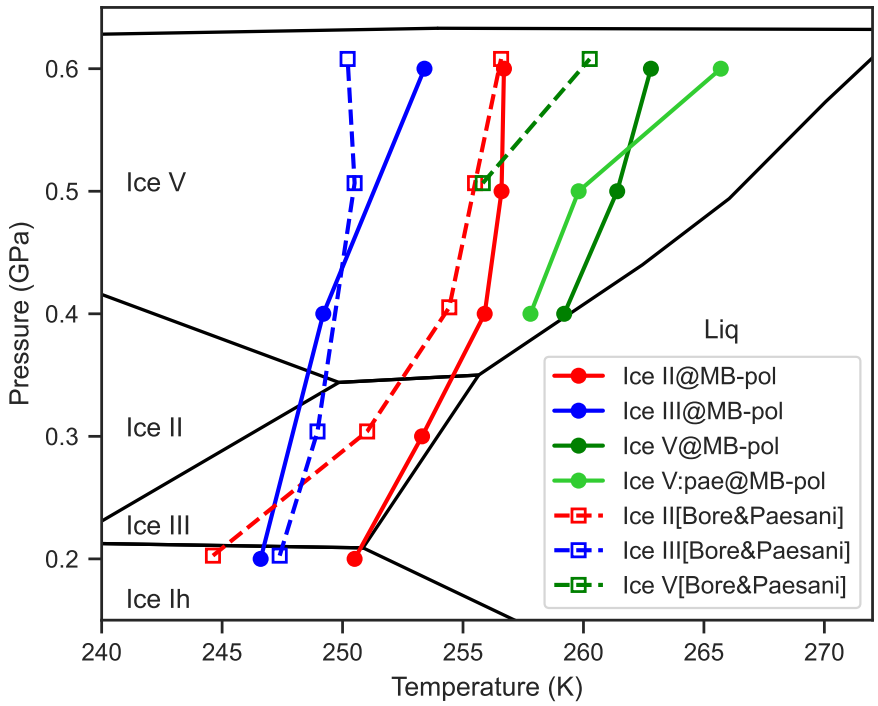

FIG. S8. The melting points of ice-water for MB-pol in our study (shown with circle lines) compared to those in Paesani's study (indicated by square lines). Melting points of Ice V:pae@MB-pol were obtained with the initial configuration that the Bore and Paesani used in their study.

* cwmyung@skku.edu

[1] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *J. Phys. Condens. Matter*, 2017, **29**, 273002.

[2] G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169–11186.

[3] G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15–50.

[4] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew and J. Sun, *J. Phys. Chem. Lett.*, 2020, **11**, 8208–8215.

[5] E. Caldeweyher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2017, **147**, 034112.

[6] J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.

[7] S. L. Bore and F. Paesani, *Nat. Commun.*, 2023, **14**, 3349.

[8] R. W. Zwanzig, *J. Chem. Phys.*, 1954, **22**, 1420–1426.