Entropy of Liquid Water as Predicted by the Two-Phase Thermodynamic Model and Data-Driven Many-Body Potentials

Ching-Hwa Ho*,† and Francesco Paesani*,†,‡,¶,§

†*Department of Chemistry and Biochemistry, University of California San Diego,*
La Jolla, California 92093, United States
‡*Materials Science and Engineering, University of California San Diego,*
La Jolla, California 92093, United States
¶*Halicioğlu Data Science Institute, University of California San Diego,*
La Jolla, California 92093, United States
§*San Diego Supercomputer Center, University of California San Diego,*
La Jolla, California 92093, United States

E-mail: c9ho@ucsd.edu; fpaesani@ucsd.edu

Abstract

We investigate the entropy of liquid water at ambient conditions using the two-phase thermodynamic (2PT) model, applied to both common pairwise-additive water models and the MB-pol and MB-pol(2023) data-driven many-body potentials. Our simulations demonstrate that the 2PT model yields entropy values in semiquantitative agreement with experimental data when using MB-pol and MB-pol(2023). Additionally, our analyses indicate that the entropy values predicted by pairwise-additive water models may benefit from error compensation between the inherent approximations of the 2PT model and the known limitations of these water models in describing many-body interactions. Despite its approximate nature, the simplicity of the 2PT model makes it a valuable tool for estimating relative entropy changes of liquid water across various environments, especially when combined with water models that provide a consistently robust representation of the structural, thermodynamic, and dynamical properties of liquid water.

Introduction

Entropy is a central thermodynamic property across various fields, having broad implications not only in chemistry and physics but also in biology, information theory, and social sciences.[?][?][?][?] The second law of thermodynamics states that the total entropy of the universe cannot decrease, thus determining the spontaneity of all transformations.[?] For example, entropy changes play a crucial role in determining reaction dynamics,[?][?][?] understanding protein folding,[?] exploring phase transitions,[?] and designing high-entropy materials.[?] Although Boltzmann's statistical interpretation of entropy laid the groundwork over a century ago,[?] accurately calculating entropy changes for complex molecular systems continues to represent a significant challenge.

The two-phase thermodynamic (2PT) model provides a straightforward approach for estimating the entropy of a liquid by decomposing the total density of states into solid-like and gas-like components.[? ? ?] Within the 2PT framework, the solid-like component is approximated by normal modes describing an ideal solid, while the gas-like component is approximated by the diffusive motions of a hard-sphere gas. Since the diffusion of a liquid can be fully attributed to the gas-like component, the decomposition of the total density of states can be uniquely determined. More specifically, the fluidicity, f, describes the fraction of the system that behaves like the gas-like component. It is defined by the ratio of the system's self-diffusivity to the hard-sphere diffusivity at the same density and zero pressure limit. By incorporating the Enskog theory[?] and Carnahan–Starling equation of state,[?] the fluidicities for translational and rotational degrees of freedom can be solved using the following equation:[? ? ?]

$$2\Delta_{\mathrm{mot}}^{-9/2} f_{\mathrm{mot}}^{15/2} - 6\Delta_{\mathrm{mot}}^{-3} f_{\mathrm{mot}}^{5} - \Delta_{\mathrm{mot}}^{-3/2} f_{\mathrm{mot}}^{7/2} + 6\Delta_{\mathrm{mot}}^{-3/2} f_{\mathrm{mot}}^{5/2} + 2f_{\mathrm{mot}} - 2 = 0 \tag{1}$$

where f_{mot} represents the fluidicity of translational or rotational motion, and Δ_{mot} is the normalized diffusivity constant defined as:

$$\Delta_{\mathrm{mot}} = \frac{2s_{\mathrm{mot},0}}{9N} \left(\frac{\pi k_B T}{m} \right)^{1/2} \rho^{1/3} \left(\frac{6}{\pi} \right)^{2/3} \tag{2}$$

Here, $s_{\mathrm{mot},0}$ is the translational or rotational density of states of the system at zero frequency, N is the number of molecules in the system, m is the mass of the molecule, and ρ is the density of the system. All vibrational motions of the system are considered to be harmonic.

Water is arguably the most important liquid on Earth, directly mediating chemical, physical, and biological processes.[?] It is thus not surprising that many molecular models[? ? ? ?] have been developed since the first Monte Carlo (MC)[?] and molecular dynamics (MD)[?] simulations of liquid water. The most common models describe the water molecules as rigid objects and represent the underlying interactions as a sum of physics-inspired, pairwise-additive contributions. These models attempt to capture actual many-body interactions through effective pairwise terms that are empirical parameterized to reproduce a subset of experimental properties (e.g., density, freezing point, enthalpy of vaporization, etc.).[?] Examples of empirical pairwise-additive models include SPC,[?] SPC/E,[?] TIP3P,[?] TIP4P,[?] TIP4P-Ew,[?] TIP4P/2005,[?] and TIP5P.[?] Empirical pairwise-additive models with flexible monomers (e.g., TIP4P/2005f,[?] and SPC/Fw[?]) have also been developed. Although empirical pairwise-additive models have been widely used in computer simulations of water and various aqueous solutions they suffer from intrinsic shortcomings that limit their predictive power and transferability across different phases.[?]

Recent advances in correlated electronic structure methods and machine learning algorithms have enabled the development of water potentials rigorously derived from the many-body expansion of the energy (MBE) calculated at the coupled cluster level of theory, including single, double, and perturbative triple excitations, *i.e.*, CCSD(T). CCSD(T) is currently considered the "gold standard" of chemical accuracy for molecular interactions.[?, ?] Among these data-driven many-body potentials, MB-pol[?, ?, ?] has been shown to accurately reproduce the properties of water from gas-phase clusters to liquid water to ice.[?, ?] Notably, MB-pol is the first and, to date, only molecular model capable of accurately predicting the phase diagram of water.[?] In addition, MB-pol has also enabled accurate simulations of the hydration properties of halide and alkali metal ions from small clusters[?, ?, ?, ?, ?, ?, ?, ?] to aqueous solutions,[?, ?, ?, ?, ?] as well as of water adsorbed in metal-organic frameworks at different relative humidity values.[?, ?, ?, ?, ?, ?, ?, ?, ?] More recently, an updated version of MB-pol, MB-pol(2023), trained on larger training sets of CCSD(T) many-body energies, has been shown to achieve even higher predictive accuracy for simulations of water in both gas and liquid phases.[?]

In this study, we report the entropy values of liquid water at ambient conditions for the MB-pol and MB-pol(2023) potentials as predicted by the 2PT model and compare them with values calculated with several commmon pairwise additive water models. Since the 2PT model involves the calculation of the vibrational density of states, which depends sensitively on the water model used in the simulations, we also investigate the ability of empirical pairwise-additive models, as well as the MB-pol and MB-pol(2023) potentials, to consistently describe both the entropy and the structural and dynamical properties of liquid water at ambient conditions. Our analyses indicate that the entropy values calculated with both MB-pol and MB-pol(2023) are about 10% smaller than the corresponding experimental value. Since these data-driven many-body potentials provide remarkable agreement with experimental data across all water's phases, the discrepancy found for the entropy of liquid water might primarily be due to the approximate nature of the 2PT model. Similar discrepancies are also found for empirical pairwise-additive models (e.g., TIP4P/2005 and TIP4P/Ew) that are known to provide a robust description of the properties of water at ambient conditions, supporting the hypothesis that these discrepancies reflect inherent limitations of the 2PT model. This is further corroborated by the good performance in reproducing the experimental entropy value exhibited by pairwise-additive water models (e.g., TIP3P and SPC) that provide a poor description of both structural and dynamical properties of liquid water. This apparent agreement thus hinges on fortuitous error compensation between the approximations inherent to the 2PT model and the known inaccuracies of these pairwise-additive water models in correctly representing many-body interactions. Nevertheless, given its simplicity, the 2PT model can still serve as a valuable tool for estimating relative changes in the entropy of liquid water across various environments, especially when applied to simulations carried out with models that provide a reliable representation of the structural, thermodynamic, and dynamical properties of water across all its phases.

Methods

We consider water models belonging to the SPCx (SPC,[?] SPC/E,[?] SPC/Fw[?]) and TIPnP (TIP3P,[?] TIP4P,[?] TIP4P/2005,[?] TIP4P/2005f,[?] TIP4P/Ew,[?] TIP5P[?]) families of pairwise-additive water models along with the MB-pol[? ? ?] and MB-pol(2023)[?] data-driven many-body potentials. All MD simulations were performed using the Large-scale Atomic/

Molecular Massively Parallel Simulator (LAMMPS)? for a cubic box of $N = 256$ water molecules in periodic boundary conditions. In the case of MB-pol and MB-pol(2023), the simulations were enabled by the MBX C++ library? ? interfaced with LAMMPS.

The equations of motion were propagated using the velocity-Verlet algorithm with a time step of 0.2 fs in the canonical (NVT: constant number of molecules N, volume V, and temperature T) ensemble,? with the temperature controlled using a global Nosé-Hoover chain consisting of four thermostats.? The nonbonded interactions were truncated at an atom-atom distance of 9.0 Å. The long-range electrostatic interactions for models belonging to the SPCx and TIPnP families were calculated using the particle-particle particle-mesh solver as implemented in LAMMPS. The long-range electrostatic interactions for MB-pol and MB-pol(2023) were calculated using the particle mesh Ewald method as implemented in MBX.? ? ? The entropy values of all water models were calculated using the 2PT model? by averaging over 20 independent NVT trajectories, each 50 ps long.

The tetrahedral order parameter, q_{tet}, was calculated according to?

$$q_{tet} = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos\psi_{ijk} + \frac{1}{3} \right)^2 , \tag{3}$$

where ψ_{ijk} is the angle between the oxygen atom of the central water molecule with index i and the oxygen atoms of two neighboring water molecules with index j and k lying within a distance smaller than 3.5 Å from the central molecule. $q_{tet} = 0$ corresponds to a completely disordered structural arrangement as in an ideal gas, while $q_{tet} = 1$ corresponds to a perfect tetrahedral arrangement. The orientational correlation function was calculated by averaging over the same 20 trajectories according to the following expression?

$$C_2(t) = \langle P_2[\vec{e}(0) \cdot \vec{e}(t)] \rangle . \tag{4}$$

Here, $\vec{e}$ represents a unit vector along one of the OH bonds of a water molecule, P_2 is the second-order Legendre polynomial, and the bracket indicates an ensemble average over all

OH bonds at a given time t.

Results and Discussions

The entropy values calculated with all water models considered in this study are listed in Table **??** along with the experimental value. The entropy values obtained for SPC, SPC/E, SPC/Fw, TIP3P, TIP4P/2005, and TIP4P/Ew agree with those reported in previous studies.[?] [?] SPC and TIP3P display the closest agreement with the experimental value within the SPCx and TIPnP families of models, respectively. All other SPC-type and TIPnP-type models, on the other hand, predict similar entropy values that are consistently smaller than the experimental value. In this regard, it should be noted that, despite water being empirical and described by effective two-body potential energy terms, TIP4P and TIP5P provide slightly better agreement with the experimental value than MB-pol and MB-pol(2023), which were developed based on a rigorous many-body formalism and CCSD(T) reference energies.

The differences in entropy values among the water models are related to the differences in structural and dynamical properties of liquid water predicted by the models themselves. To gain insights into the relationship between the structure of water and the calculated entropy values, Figure **??** shows the oxygen-oxygen radial distribution function (RDF) of liquid water calculated with each model at 298 K and experimental density. While the SPC and TIP3P models provide the closest agreement with the experimental data, their RDFs lack prominent long-range structural organization, which leads to an overall poor representation of the overall structure of liquid water. On the other hand, all other models predict similar RDFs, correctly predicting the presence of well-defined first and second solvation shells. In particular, MB-pol and MB-pol(2023), which predict similar entropy values, provide the closest agreement with the experimental oxygen-oxygen RDF.

It should be noted that the trend in entropy values predicted by the various water models qualitatively follows the structural ordering inferred from the analysis of the correspond-

Table 1: Entropy of liquid water at 298 K calculated using the 2PT model and MD simulations carried out at 298 K with the MB-pol and MB-pol(2023) data-driven many-body potentials, as well as various pairwise-additive models belonging to the SPCx and TIPnP families.

Model	Entropy (J/mol-K)
Experiment[?]	69.95 ± 0.03
Data-driven many-body potentials	
MB-pol	62.9 ± 0.3
MB-pol(2023)	62.0 ± 0.5
SPCx pairwise-additive models	
SPC	66.9 ± 0.3
SPC/E	61.7 ± 0.3
SPC/Fw	61.4 ± 0.4
TIPnP pairwise-additive models	
TIP3P	70.6 ± 0.2
TIP4P	64.3 ± 0.2
TIP4P/2005	58.3 ± 0.2
TIP4P/2005f	60.2 ± 0.2
TIP4P/Ew	59.9 ± 0.2
TIP5P	63.5 ± 0.3

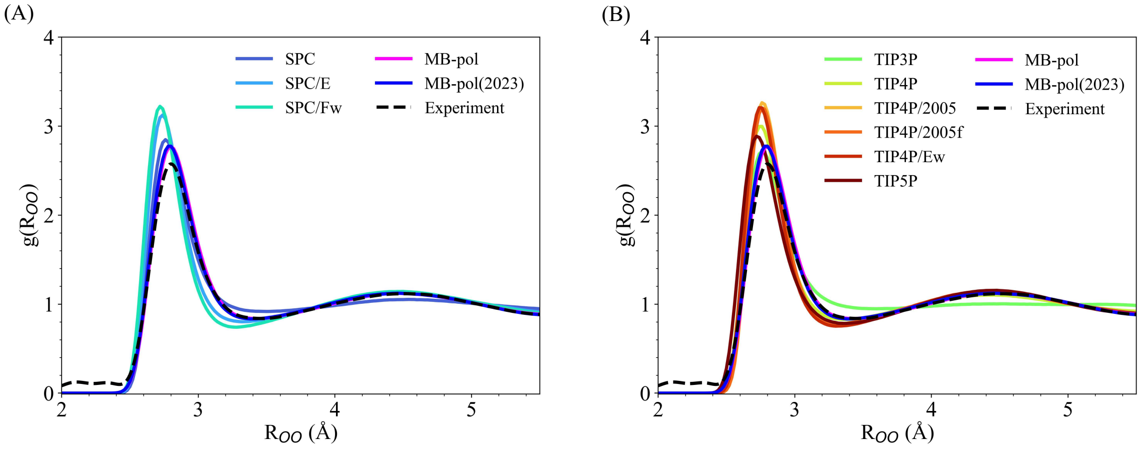

Figure 1: Oxygen-oxygen radial distribution function, $g(R_{OO})$, calculated from MD simulations of liquid water carried out at 298 K with the MB-pol and MB-pol(2023) data-driven many-body potentials, as well as various pairwise-additive models belonging to the SPCx (A) and TIPnP (B) families. The experimental oxygen-oxygen radial distribution function from Ref. **?** is also shown for comparison.

ing oxygen-oxygen RDFs. In particular, the high entropy values predicted by the SPC and TIP3P models are reflected in the absence of a second solvation shell in their oxygen-oxygen RDFs. As shown in Table **??**, the SPC/E, SPC/Fw, TIP4P/2005, TIP4P/2005f, and TIP4P/Ew models predict smaller entropy values than MB-pol and MB-pol(2023). This difference correlates with the first hydration shell in the oxygen-oxygen RDFs calculated with these models being overstructured compared to that predicted by MB-pol and MB-pol(2023). Overall, the trend in entropy values displayed by the various water models appears to be inversely correlated with their ability to predict the correct structure of liquid water. This analysis suggests that the better agreement of the SPC and TIP3P models with the experimental entropy value, compared to all other models, is likely due to error compensation between the approximations adopted by the 2PT model and the intrinsic limitations of these water models in correctly representing many-body interactions in liquid water. Additionally, based on the consistent agreement of MB-pol and MB-pol(2023) with experimental data for several structural, thermodynamic, and dynamical properties of water, it seems reasonable to assume that, due to its inherent approximate nature, the 2PT model likely underestimates the entropy of liquid water at ambient conditions predicted by these two models by approximately 10%. In the context of the 2PT model framework, this suggests that the predicted fluidicities for the translational and rotational degrees of freedom, as calculated using Eq. **??**, are underestimated. Backtracking the definition of fluidicity, this underestimation can potentially stem from the neglect of intermolecular interactions in the hard-sphere model. This observation also suggests that the 2PT model may be more accurate in predicting the entropy of molecular fluids where intermolecular interactions are weak, assuming the molecular model is sufficiently accurate.

The relationship between the structure and the entropy of water is further supported by the analysis of the tetrahedral order parameter (q_{tet}) for each water model, as shown in Figure **??**. Except for SPC and TIP3P, all water models exhibit a similar bimodal distribution, characterized by a shallow peak at $q_{tet} \approx 0.5$ and a pronounced peak at $q_{tet} \approx 0.8$. The

$P(q_{\text{tet}})$ distributions calculated with the SPC and TIP3P models instead display a single peak at $q_{tet} \approx 0.5$. The analyses of the RDFs (Figure **??**) and q_{tet} distributions (Figure **??**) demonstrate that the SPC and TIP3P models are unable to accurately predict the structure of both the first and second solvation shells, despite providing the closest agreement with the experimental entropy values. Moreover, the larger entropy values predicted by the SPC and TIP3P models compared to other models qualitatively align with their $P(q_{tet})$ distributions, since both models provide more pronounced distributions at smaller q_{tet} values, directly correlating with more disordered structural arrangements. In addition, the slightly higher entropy value predicted by TIP4P compared to MB-pol and MB-pol(2023) can be understood in terms of the differences in the corresponding $P(q_{tet})$ distributions. Specifically, the distribution calculated with TIP4P is noticeably higher at $q_{tet} \approx 0.5$ and lower at $q_{tet} \approx 0.8$ compared to the distributions calculated with MB-pol and MB-pol(2023), indicating slightly more disordered arrangements of water molecules residing within the first solvation shell.

Insights into the relationship between entropy values and dynamical properties predicted by the different water models can be gained by analyzing the calculated orientational correlation functions, as shown in Figure **??**. Following the same trends displayed by the

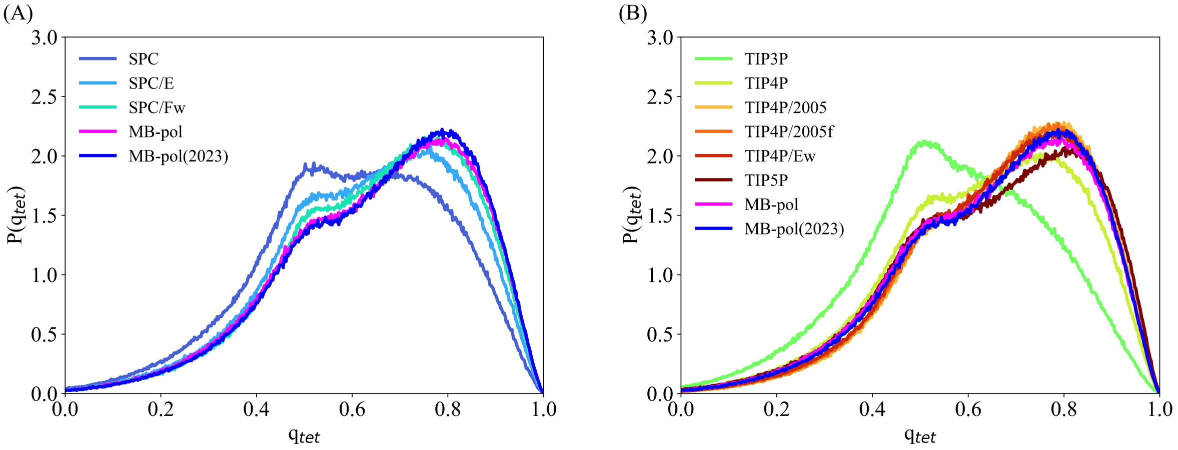

Figure 2: Probability distribution of the tetrahedral order parameter, $P(q_{tet})$, calculated from MD simulations of liquid water carried out at 298 K with the MB-pol and MB-pol(2023) data-driven many-body potentials, as well as various pairwise-additive models belonging to the SPCx (A) and TIPnP (B) families.

entropy values, RDFs, and $P(q_{tet})$ distributions, all water models, except for SPC, TIP3P, and TIP4P, predict a similar decay of the orientational correlation functions as a function of time. The orientational relaxation times for all water models, τ_2, which were determined by fitting the long-time decay of the corresponding orientational correlation functions to a single exponential function, are listed in Table **??**. Among all models considered in this study, MB-pol and SPC/E provide the closest agreement with the experimental value, followed by MB-pol(2023), SPC/Fw, TIP4P/2005f, TIP4P/Ew, and TIP5P. Interestingly, while MB-pol and MB-pol(2023) effectively display the same ability to predict the structural properties of liquid water, as shown in Figures **??** and **??**, they provide slightly different orientational correlation functions. In addition, it is noteworthy that models with smaller τ_2 values generally exhibit larger entropy values, as shown in Table **??**. Smaller τ_2 values, corresponding to rapid decays of the corresponding correlation functions, indicate that water molecules more quickly lose memory of their initial orientations. Consequently, information about the water structure is more rapidly lost over time. In this context, the trends in τ_2 and entropy values among all water models suggest that, despite being inherently approximate, the 2PT model qualitatively captures the essential features of entropy, directly correlating the entropy value

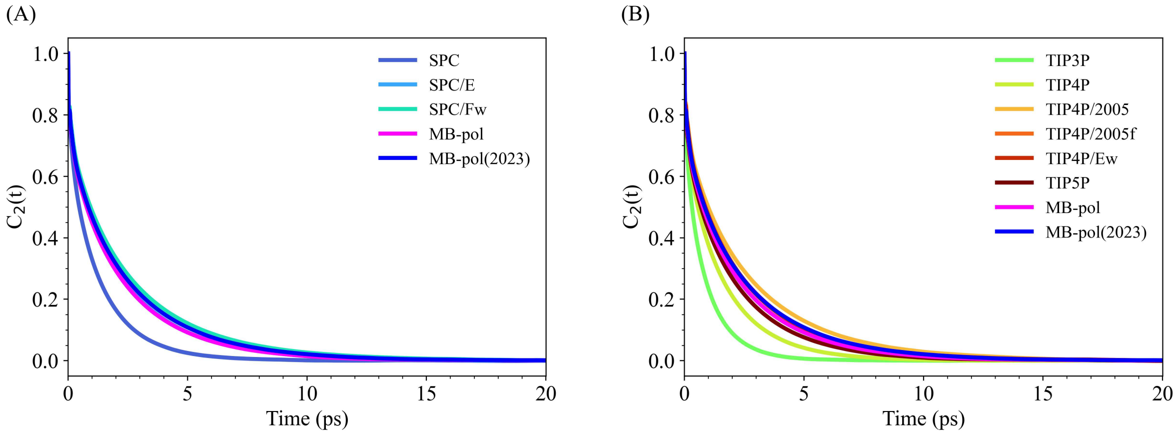

Figure 3: Orientational correlation function, $C_2(t)$, calculated from MD simulations of liquid water carried out at 298 K with the MB-pol and MB-pol(2023) data-driven many-body potentials, as well as various pairwise-additive models belonging to the SPC (A) and TIPnP (B) families.

Table 2: Orientational relaxation time, τ_2, calculated from MD simulations of liquid water carried out at 298 K with the MB-pol and MB-pol(2023) data-driven many-body potentials, as well as various pairwise-additive models belonging to the SPCx and TIPnP families. For of each water mode, τ_2 was calculated by fitting the long-time decay of the corresponding orientational correlation function to an exponential function: $C_2(t) = A\exp(-t/\tau_2)$.

Model	τ_2 (ps)
Experiment[?]	2.5
Data-driven many-body potentials	
MB-pol	2.6 ± 0.1
MB-pol(2023)	2.8 ± 0.1
SPCx pairwise-additive models	
SPC	1.5 ± 0.1
SPC/E	2.6 ± 0.1
SPC/Fw	3.0 ± 0.1
TIPnP pairwise-additive models	
TIP3P	1.1 ± 0.1
TIP4P	1.8 ± 0.1
TIP4P/2005	3.1 ± 0.1
TIP4P/2005f	2.7 ± 0.1
TIP4P/Ew	2.7 ± 0.1
TIP5P	2.3 ± 0.1

with the extent of structural order/disorder and information content.

Conclusions

In this study, we have investigated the entropy of liquid water as predicted by the 2PT model used in combination with various models, including the MB-pol and MB-pol(2023) data-driven many-body potentials and models belonging to the SPCx and TIPnP families. We have demonstrated that the entropy values calculated using the 2PT model qualitatively correlate with the extent of structural ordering and fast dynamics predicted by a given water model. However, our analyses indicate that the 2PT model does not consistently align the entropy value calculated with a given water model with the ability of the same model to reproduce other structural, thermodynamic, and dynamical properties of liquid water. This discrepancy can be attributed to the interplay between the approximations adopted by the 2PT model and the ability of a water model to accurately describe many-body interactions in water. This is evidenced by the ability of the SPC and TIP3P models to predict entropy values in remarkable agreement with the experimental value while providing a poor description of other structural and dynamical properties. On the other hand, the MB-pol and MB-pol(2023) data-driven many-body potentials consistently provide excellent agreement with experimental data for water across all phases but, when used in 2PT calculations, they underestimate the entropy value by approximately 10%. Overall, our study suggests that, despite its inherent approximate nature, the 2PT model can serve as an efficient tool to estimate relative entropy changes of water in different environments when it is used in combination with water models that provide a robust description of both the structure and dynamics of liquid water.

Acknowledgments

We are grateful to Dr. Tod Pascal and Alexandria Do for insightful discussions about the 2PT model. This research was supported by the Natural Science Foundation through award no. 2311260. All simulations were carried out on Expanse at the San Diego Supercomputer Center (SDSC) through allocation CHE230052 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by NSF grants nos. 2138259, 2138286, 2138307, 2137603, and 2138296, as well as Triton Shared Computing Cluster (TSCC) at SDSC.

Data availability

Any data generated and analyzed for this study that are not included in this article are available from the authors upon request.

Code availability

The molecular models used in the MD simulations are publicly available on GitHub (https://github.com/paesanilab/Data_Repository/tree/main/water_entropy) in the format for LAMMPS or the MBX[?] interface with LAMMPS.[?] All computer codes used in the analysis presented in this study are available from the authors upon request.

TOC Graphic