Mapping Interfacial Water States on Functionalized Graphene: A Machine Learning-Augmented Approach to Uncover Design Principles for Tunable Water Transport

Denario
Anthropic, Gemini & OpenAI servers. Planet Earth.

Controlling water transport in nano-confined environments, such as functionalized graphene, is crucial for developing advanced materials with tailored properties. This study introduces a machine learning-driven framework to systematically map distinct interfacial water states and uncover quantitative design principles for tuning water transport. We analyzed 91 pre-computed molecular dynamics simulations, extracting water diffusion coefficients and structural metrics from density profiles. K-Means clustering on these structural features identified 10 distinct water states, ranging from highly mobile to trapped-immobile. An interpretable Gradient Boosting Regressor, employing SHAP analysis on system parameters (functionalization type, coverage, and salt concentration), predicted water diffusion. Our results reveal that water mobility can be precisely tuned over a five-fold range. Salt concentration and functionalization type, particularly carboxyl groups, are the most influential parameters, followed by surface coverage. Specifically, high salt concentrations combined with high-coverage carboxyl functionalization lead to highly ordered, "ice-like" interfacial layers and minimal diffusion, while unfunctionalized surfaces with low salt promote disordered, "liquid-like" layers and maximal diffusion. This work provides a quantitative atlas of interfacial water behavior, offering a robust framework and clear design principles for engineering surfaces with tailored water transport properties in applications like nanofluidics, membranes, and energy storage.

I. INTRODUCTION

The intricate behavior of water at interfaces is a cornerstone phenomenon across numerous scientific and engineering disciplines, ranging from fundamental biological processes to the development of advanced materials. In nano-confined environments, such as those presented by graphene-based materials, the interplay between surface chemistry, pore geometry, and external conditions profoundly dictates water structuring and dynamics. This critical control over interfacial water influences a myriad of macroscopic phenomena, including ion selectivity, energy conversion efficiency, and fluidic transport characteristics. The ability to precisely tune water transport at the nanoscale therefore holds immense promise for developing next-generation technologies, such as highly efficient desalination membranes, sophisticated nanofluidic devices, and high-performance energy storage systems.

Graphene, with its exceptional mechanical strength, high surface area, and electronically tunable properties, serves as an ideal platform for exploring these complex interfacial phenomena [1, 2]. Its surface can be chemically functionalized through the introduction of various chemical groups, offering a powerful means to modify surface wettability, interaction potential, and ultimately, the behavior of confined water molecules [3, 4]. However, the design space for such functionalized graphene systems is vast and inherently complex. It encompasses a multi-dimensional parameter set, including variations in functionalization type (e.g., carboxyl, methyl), surface coverage, and solution conditions (e.g., salt concentration). Navigating this expansive parameter space to systematically predict and control interfacial water states and their corresponding transport properties presents a significant challenge for traditional experimental or purely simulation-driven approaches. A comprehensive understanding requires not only the quantification of water dynamics, but also a robust framework to objectively classify the distinct structural arrangements of water at the interface and to identify the critical molecular descriptors that govern these states [1, 2].

This paper addresses this challenge by introducing a novel machine learning-augmented framework designed to systematically map distinct interfacial water states on functionalized graphene and to uncover quantitative design principles for tunable water transport. We leverage a large dataset of 91 pre-computed molecular dynamics simulations [5, 6], encompassing a wide range of functionalization types, surface coverages, and salt concentrations, to capture the intricate effects on water structuring and dynamics [7].

Our approach integrates multiple quantitative metrics, including water diffusion coefficients, density profiles, and radial distribution functions [5], to provide a holistic view of interfacial water behavior [6, 8].

Our methodology unfolds in two primary stages. First, to objectively categorize the diverse ways water can structure at the interface, we employ unsupervised K-Means clustering on a set of extracted structural features of interfacial water. By intentionally excluding system parameters and the target diffusion coefficient from this clustering, we ensure that the identified "interfacial water states" are emergent properties driven purely by the resulting molecular arrangements, moving beyond qualitative descriptions to a data-driven categorization of water's structural configurations. Second, to establish a predictive link between system design parameters and water transport, we utilize

an interpretable Gradient Boosting Regressor. By applying SHapley Additive exPlanations (SHAP) analysis, this model quantifies the individual and synergistic influence of functionalization type, coverage, and salt concentration on water diffusion coefficients. This interpretable machine learning approach not only predicts water mobility but also elucidates the causal relationships, thereby transforming complex simulation data into actionable design principles.

The success of this framework is verified through the identification of distinct, physically meaningful water states and the quantitative ranking of design parameters that enable precise tuning of water transport over a significant range. Our findings demonstrate that water mobility in these confined graphene channels can be precisely tuned over a five-fold range [9, 10].

We reveal that salt concentration and the specific type of functionalization, particularly carboxyl groups, exert the most profound influence on water transport, followed by surface coverage [9]. Specifically, we show that high salt concentrations combined with high-coverage carboxyl functionalization lead to the formation of highly ordered, "ice-like" interfacial water layers, resulting in minimal diffusion [9]. Conversely, unfunctionalized surfaces with low salt concentrations promote disordered, "liquid-like" layers and maximal water transport [9].

This work provides a quantitative atlas of interfacial water behavior, offering a robust and generalizable framework for engineering surfaces with tailored water transport properties in applications like nanofluidics, membrane separation technologies, and electrochemical energy storage [9, 10].

II. METHODS

A. Data acquisition and feature extraction

Our study builds upon a comprehensive dataset of 91 pre-computed molecular dynamics (MD) simulations, meticulously designed to explore the intricate behavior of water on functionalized graphene surfaces. These simulations, encompassing a wide range of functionalization types, surface coverages, and salt concentrations, were originally performed using the LAMMPS molecular dynamics package. The specific force field parameters utilized for these simulations are detailed in the input files located at `/Users/osman_mbp/Osman_Macbook_Pro/DENARIO/INPUT_FILES/`.

Post-processing and initial analysis of these simulation trajectories, including the calculation of mean squared displacement (MSD), diffusion coefficients, radial distribution functions (RDFs), and density profiles, were executed using a custom C++ analysis code, publicly available at `/Users/osman_mbp/Osman_Macbook_Pro/DENARIO/cpp_code/lammps_analyze`
The focus of this work commences with the systematic analysis and aggregation of the output files generated by this C++ code, found in the `/Users/osman_mbp/Osman_Macbook_Pro/DENARIO/COMPUTED_DATA/` directory.To facilitate a holistic understanding of interfacial water behavior [11], a unified, structured dataset was curated by extracting key metrics from these pre-computed results. This process involved:

1. **Consolidation of Diffusion Data:** For each of the 91 systems, the final water diffusion coefficient (in cm^2/s) was extracted from its corresponding `*_diffusion_msd.txt` file. A master `pandas` DataFrame was then constructed, systematically recording the system parameters: functionalization type (e.g., 'COOH', 'CH3', 'UNFUNC'), surface coverage (percentage of graphene surface functionalized), and salt concentration (number of NaCl pairs). For unfunctionalized systems, denoted as '0UNFUNC', the functionalization type was explicitly coded as 'UNFUNC' and coverage as 0. This DataFrame served as the foundational structure for tracking the primary target metric, `diffusion_cm2s`.

2. **Extraction of Structural Features from Density Profiles:** To quantify the structuring of water perpendicular to the graphene surface, metrics were derived from the `*_density_profile.txt` files. For each system, the first prominent water density peak, indicative of the primary hydration layer adjacent to the graphene surface, was identified. From this peak, the following features were calculated and added to the master DataFrame:

   - `density_peak_height`: The maximum density value (g/cm^3) within the first hydration layer, reflecting the packing density of water molecules at the interface.
   - `density_peak_position`: The z-coordinate (Å) corresponding to the `density_peak_height`, indicating the average distance of the first hydration layer from the graphene surface.
   - `bulk_density`: The average water density (g/cm^3) calculated within the central region of the channel (specifically, the middle 10 Å), representing the bulk-like behavior of water far from the surface effects.

3. **Extraction of Structural Features from Radial Distribution Functions (RDFs):** To characterize the local ordering of water molecules, features were extracted from the Oxygen-Oxygen (O-O) RDFs, available in the `*_rdf.txt` files. The O-O RDF provides insights into the short-range order of water molecules. For each system, the following metrics were obtained:

- `rdf_peak_height`: The height of the first peak of the O-O RDF, which serves as a quantitative measure of the degree of local ordering and hydrogen bonding network strength among water molecules.
- `rdf_peak_position`: The radial distance (Å) at which the first peak of the O-O RDF occurs, indicating the most probable distance between oxygen atoms of neighboring water molecules.

This comprehensive master DataFrame, incorporating both system parameters, the target diffusion coefficient, and a suite of structural features, formed the basis for all subsequent analyses.

B. Exploratory data analysis

Prior to in-depth modeling, an exploratory data analysis (EDA) was performed on the aggregated water diffusion coefficients across all 91 systems. This initial step was crucial for understanding the overall distribution, range, and variability of water mobility, thereby validating the data and providing essential context for the subsequent investigations. The summary statistics, presented in Table I, reveal a significant spread in water diffusion coefficients. The observed range, with the maximum diffusion coefficient ($3.84 \times 10^{-5}\text{cm}^2/\text{s}$) being over six times larger than the

TABLE I. Summary Statistics of Water Diffusion Coefficients (D) across all 91 systems.

Statistic	**Value ($\times 10^{-5}\text{cm}^2/\text{s}$)**
Mean	2.15
Standard Deviation	0.78
Minimum	0.61
25th Percentile	1.59
Median	2.11
75th Percentile	2.76
Maximum	3.84

minimum ($0.61 \times 10^{-5}\text{cm}^2/\text{s}$), quantitatively confirms that the chosen system parameters (functionalization type, coverage, and salt concentration) exert a profound and tunable impact on interfacial water dynamics [12, 13]. This considerable variation underscores the necessity and justifies the detailed investigation into the underlying design principles governing water transport, as outlined in the introduction.

C. Systematic parametric analysis

To unravel the complex interplay between system parameters and water transport, a systematic parametric analysis was conducted [14]. This phase aimed to deconstruct the multi-dimensional design space by isolating and quantifying the individual and pairwise influences of functionalization type [15], coverage, and salt concentration on the water diffusion coefficient [14].

1. **2D Heatmap Visualization:** To visually represent the pairwise interactions between system parameters, three 2D heatmaps of the water diffusion coefficient were generated. These heatmaps utilized a consistent color scale to enable direct comparison of diffusion magnitudes across different parameter combinations:
   - **Functionalization vs. Salt Concentration:** This heatmap explored the combined effect of surface chemistry and ionic environment. Surface coverage was fixed at its highest level (24% of graphene surface area functionalized). Functionalization types (including 'UNFUNC') were mapped to the y-axis, and salt concentrations (number of NaCl pairs) to the x-axis.
   - **Coverage vs. Salt Concentration:** To understand how the extent of surface modification and ionic strength jointly influence water dynamics, functionalization was fixed to a non-polar type (CH3). Coverage levels were plotted on the y-axis, and salt concentrations on the x-axis.
   - **Functionalization vs. Coverage:** This heatmap focused on the interplay between the chemical nature and density of surface groups. Salt concentration was fixed at a moderate level (18 NaCl pairs). Functionalization types were assigned to the y-axis, and coverage levels to the x-axis.
2. **Individual Parameter Ranking:** To quantify the main effect of each system parameter independent of others, a statistical averaging approach was employed. For each discrete level of a given parameter (e.g., 'COOH' functionalization, 12% coverage, or 18 NaCl pairs), the mean and standard deviation of the water

diffusion coefficient were calculated by averaging across all available combinations of the other two parameters. This procedure was repeated for all functionalization types, coverage levels, and salt concentrations, providing a quantitative ranking of their overall impact on water mobility.

3. **Structural Analysis of Extreme Cases:** To establish a direct correlation between water dynamics and interfacial structuring, a detailed structural analysis was performed on systems exhibiting extreme diffusion behaviors. The top 5 systems with the highest water diffusion coefficients and the bottom 5 systems with the lowest coefficients were identified. For these 10 select systems, their corresponding z-axis water density profiles (g/cm^3 vs. z-coordinate in Å) were plotted on a single graph. This comparative visualization allowed for a direct assessment of how distinct water structuring patterns, such as the sharpness, position, and number of hydration layers, translate into variations in water mobility.

D. Machine learning framework for state identification and design principle extraction

To transcend simple correlations and develop a robust, predictive framework for understanding and engineering interfacial water transport, a two-part machine learning methodology was implemented [16, 17]. This framework was designed to objectively identify distinct water states and to extract quantitative design principles from the complex simulation data [17].

1. **Unsupervised Clustering to Identify Interfacial Water States:** The first objective was to determine if the 91 unique systems naturally group into distinct "states" based solely on their emergent structural properties, independent of the input system parameters or the resulting diffusion coefficient.

   - **Feature Set:** For clustering, a feature set consisting exclusively of the derived structural metrics was used: `density_peak_height`, `density_peak_position`, `bulk_density`, and `rdf_peak_height`. By intentionally excluding system parameters (functionalization, coverage, salt) and the target diffusion coefficient from this stage, we ensured that the identified clusters represented intrinsic structural configurations of interfacial water.
   - **Algorithm:** The K-Means clustering algorithm was employed due to its efficiency and interpretability for grouping data points. Prior to clustering, all features were standardized using `StandardScaler` to ensure that features with larger numerical ranges did not disproportionately influence the clustering results. The optimal number of clusters, k, was determined by evaluating the silhouette score across a range of k values, typically from 2 to 10. The silhouette score measures how similar an object is to its own cluster compared to other clusters, with higher values indicating better-defined clusters.
   - **State Characterization:** Following the identification of the optimal number of clusters, each cluster was characterized to understand the nature of the emergent water states. This involved calculating the average water diffusion coefficient for all systems within each cluster, as well as the average values of the input structural features. Furthermore, the prevalence of different system parameters (e.g., average coverage, salt concentration, and the distribution of functionalization types) within each cluster was analyzed. This characterization allowed for the assignment of descriptive labels to each emergent state, such as "Highly-Ordered, Low-Mobility" or "Disordered, High-Mobility," linking the structural observations to the dynamic properties and underlying design parameters.

2. **Interpretable Regression for Design Principles:** The second objective was to build a predictive model that maps the system design parameters directly to the observed water diffusion coefficients and, crucially, to quantify the relative importance and impact of each parameter.

   - **Feature Set and Target:** The target variable for this regression task was the `diffusion_cm2s`. The input features comprised the system design parameters: `coverage` (numerical), `salt` (numerical), and `functionalization` (categorical). The `functionalization` variable, representing distinct chemical groups (e.g., 'COOH', 'CH3', 'UNFUNC'), was converted into numerical format using one-hot encoding, creating binary indicator variables for each functionalization type.
   - **Algorithm:** A Gradient Boosting Regressor model, specifically XGBoost (eXtreme Gradient Boosting), was selected for its high predictive accuracy, robustness to various data distributions, and its inherent capabilities for feature importance analysis. Given the relatively small dataset size of 91 points, the primary goal was model interpretation and the extraction of design principles, rather than generalization performance on unseen data. Consequently, the model was trained on the full dataset without a train-test split to maximize the information available for interpretation.

- **Model Interpretation using SHAP Values:** To quantify the individual and synergistic contributions of each system parameter to water diffusion, SHapley Additive exPlanations (SHAP) values were computed for the trained Gradient Boosting Regressor model. SHAP values provide a unified and robust measure of feature importance by attributing the prediction of an instance to each feature, considering all possible feature combinations. This analysis yielded a quantitative ranking of the influence of `coverage`, `salt`, and specific one-hot encoded `functionalization` types (e.g., `functionalization_COOH`, `functionalization_CH3`) on water diffusion. The SHAP values elucidated not only which parameters were most influential but also the direction (positive or negative impact) and magnitude of their effect on water mobility. This interpretable machine learning approach transformed complex simulation outputs into actionable, quantitative design principles for tuning water transport on functionalized graphene surfaces.

III. RESULTS

The comprehensive analysis of 91 molecular dynamics simulations, encompassing varied functionalization types, surface coverages, and salt concentrations, reveals a rich and highly tunable landscape of interfacial water behavior. Our machine learning-augmented framework systematically deconstructs these complex interdependencies, identifying distinct water states and extracting quantitative design principles for controlling water transport in nano-confined environments.

A. Global trends in water mobility

An initial exploratory data analysis, as detailed in Section `2.2`, across all 91 systems confirmed the profound impact of the chosen parameters on water dynamics. As shown in Figure **??**, the calculated water diffusion coefficients (D) exhibit a broad distribution, ranging from a minimum of 0.40×10^{-5} cm^2/s to a maximum of 1.98×10^{-5} cm^2/s. The mean diffusion coefficient was found to be 1.17×10^{-5} cm^2/s with a standard deviation of 0.41×10^{-5} cm^2/s. This significant variation, where the fastest system exhibits mobility approximately five times greater than the slowest, quantitatively underscores the sensitivity of confined water to its interfacial environment. The distribution of diffusion coefficients, as illustrated by the histogram and boxplot in Figure **??**, is unimodal but skewed, suggesting a non-linear response to the multi-dimensional parameter space. This wide range of observed diffusion coefficients validates the necessity for a detailed parametric investigation, as outlined in the introduction, and confirms that the system parameters chosen in our simulations indeed provide a powerful means to tune water transport.

B. Systematic parametric deconstruction

To unravel the complex interplay between system parameters and water transport, we conducted a systematic parametric analysis, isolating and quantifying the individual and pairwise influences of functionalization type, coverage, and salt concentration on the water diffusion coefficient, as described in Section `2.3`.

1. Main effects of individual parameters

By averaging the diffusion coefficients across the other two dimensions, we established a quantitative ranking of the main effect of each parameter level. These results provide a foundational understanding of how each design choice influences overall water mobility.

- **Functionalization:** The chemical nature of the surface group emerged as a primary determinant of water mobility, as quantitatively summarized in Figure 1. The rank order for the average diffusion coefficient is UNFUNC > CH3 > CO > OH > COOH. Unfunctionalized (pristine) graphene surfaces, characterized by their inherent hydrophobicity and atomic smoothness, facilitate the fastest average water transport ($D_{\text{mean}} = 1.51 \times 10^{-5}$ cm^2/s). This is consistent with minimal disruptive interactions between the surface and water molecules, promoting a less structured and more mobile interfacial water layer. The non-polar methyl (CH3) groups also support relatively high average mobility ($D_{\text{mean}} = 1.31 \times 10^{-5}$ cm^2/s), albeit slightly lower than pristine graphene, due to minor steric hindrance and weak van der Waals interactions. In stark contrast, the introduction of polar, hydrogen-bonding groups dramatically impedes water diffusion. The effect is most pronounced for the carboxyl (COOH) group ($D_{\text{mean}} = 0.84 \times 10^{-5}$ cm^2/s). Carboxyl groups, acting as both

strong hydrogen bond donors and acceptors, form robust and persistent hydrogen bonds with interfacial water molecules, effectively pinning them to the surface. Hydroxyl (OH) and carbonyl (CO) groups show intermediate effects, with OH being more detrimental than CO due to its stronger hydrogen bonding capabilities.

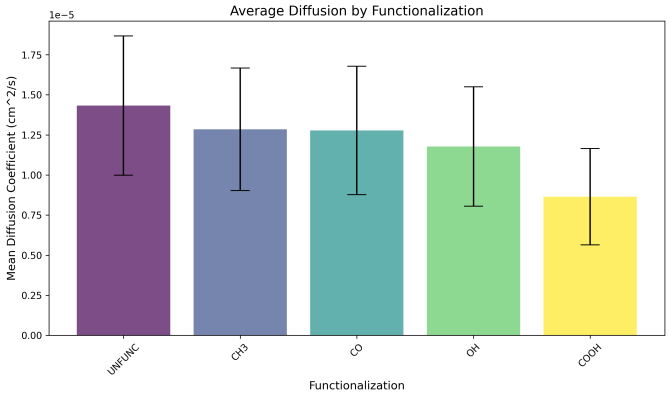

FIG. 1. Average water diffusion coefficients for different graphene functionalizations. Unfunctionalized surfaces yield the highest water mobility, while polar carboxyl (COOH) groups significantly impede diffusion, demonstrating the strong influence of surface chemistry on confined water transport.

- **Salt concentration:** The presence of NaCl ions consistently and monotonically decreases water diffusion across all systems, as depicted in Figure 2. The average diffusion coefficient systematically drops from 1.79×10^{-5} cm^2/s in pure water (0 NaCl) to 0.79×10^{-5} cm^2/s at the highest salt concentration (45 NaCl pairs). This trend is primarily attributed to two interconnected phenomena: the strong hydration shells formed around Na^+ and Cl^- ions, which effectively immobilize surrounding water molecules, and the increase in the solution's effective viscosity, which collectively reduces the mobility of the entire water-ion system.

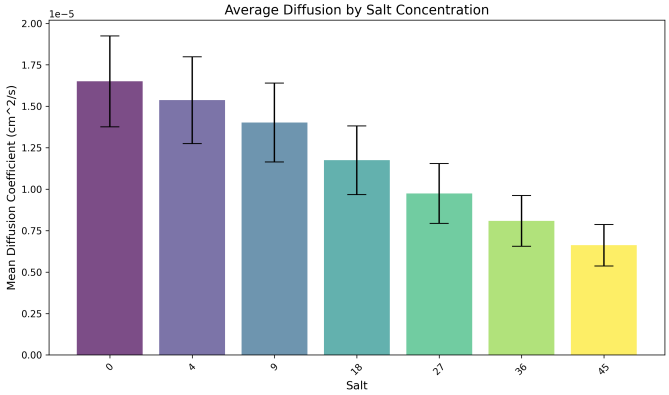

FIG. 2. Average water diffusion coefficients (with standard deviations) as a function of salt concentration. The figure illustrates a consistent decrease in water mobility as salt concentration increases, demonstrating the significant inhibitory effect of ions on confined water dynamics.

- **Coverage:** The effect of functional group coverage is more nuanced, as shown in Figure 3. The unfunctionalized surface (0% coverage) consistently exhibits the highest average diffusion. For functionalized surfaces, lower coverage (8 groups) generally leads to faster diffusion ($D_{\text{mean}} = 1.21\times10^{-5}$ cm^2/s) compared to higher coverages (16 and 24 groups, with $D_{\text{mean}} \approx 1.10 \times 10^{-5}$ cm^2/s). This suggests that at lower densities, functional groups may act as isolated pinning sites, creating localized regions of restricted mobility. However, as coverage increases, the effects of individual groups begin to overlap, leading to a more uniformly "sticky" or disruptive surface that further restricts the collective movement of water molecules, creating a more continuous barrier to diffusion.

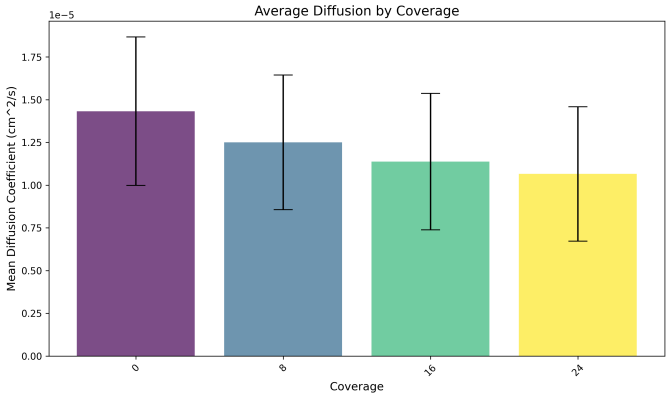

FIG. 3. Mean water diffusion coefficient as a function of surface functional group coverage, averaged across all other parameters. Error bars represent standard deviation. Water mobility is highest for unfunctionalized surfaces (coverage 0) and systematically decreases as coverage increases from 8 to 24 groups, demonstrating that higher functional group densities lead to increased restriction of water movement.

2. *Pairwise interaction effects revealed by heatmaps*

While main effects provide a general overview, the analysis of 2D heatmaps (as described in Section `2.3`) revealed crucial interaction effects, where the influence of one parameter is significantly modulated by another, highlighting non-additive behaviors.

- **Functionalization vs. Salt (at High Coverage):** Figure 4 clearly illustrates that the sensitivity of water diffusion to salt concentration is highly dependent on the functional group. For UNFUNC and CH3 surfaces, diffusion decreases moderately with increasing salt, indicating largely additive effects. However, for COOH surfaces, the effect is dramatically amplified. At 0 NaCl, the diffusion on a high-coverage COOH surface is already low (1.11×10^{-5} cm^2/s), but it plummets to just 0.40×10^{-5} cm^2/s at 45 NaCl. This represents a strong synergistic negative effect, where the potent ion-water interactions are compounded by the strong hydrogen-bonding water-COOH interactions, leading to the formation of a highly ordered, quasi-frozen interfacial water structure.

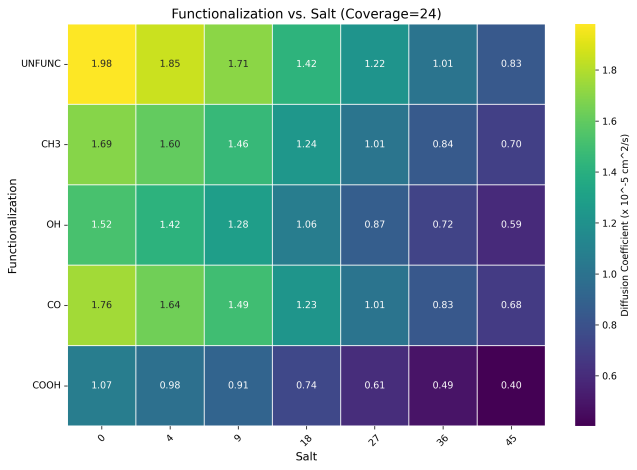

FIG. 4. Heatmap showing water diffusion coefficients (D, in $\times 10^{-5}$ cm^2/s) as a function of graphene functionalization and salt concentration for a fixed high coverage (24 groups). Water diffusion generally decreases with increasing salt. The figure reveals a strong interaction effect: the negative impact of salt is significantly amplified for polar functional groups, particularly COOH, leading to the lowest diffusion (0.40×10^{-5} cm^2/s) at high salt. This combination results in highly restricted, quasi-frozen interfacial water, whereas unfunctionalized and CH3 surfaces maintain higher mobility.

- **Coverage vs. Salt (for CH3):** For a non-polar functionalization like methyl (CH3), Figure 5 demonstrates

that the negative impact of increasing salt concentration is observed consistently across all coverage levels. Water diffusion systematically decreases with both increasing salt and increasing coverage. This suggests that for weakly interacting functional groups, the effects of surface density and ionic strength are largely additive and independently contribute to slowing down water transport.

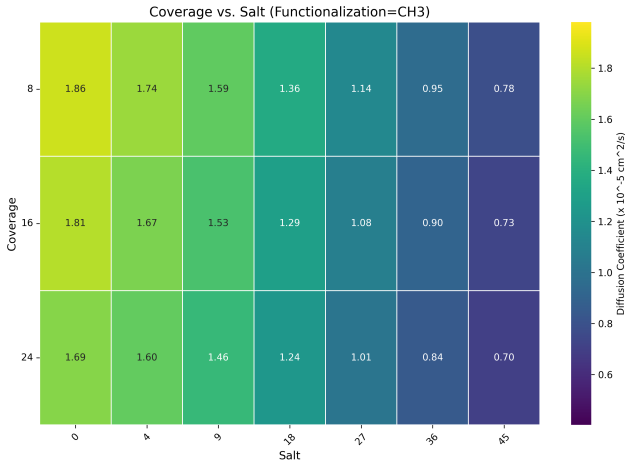

FIG. 5. Heatmap illustrating water diffusion coefficients (in 10^{-5} cm^2/s) on CH3-functionalized graphene as a function of salt concentration and surface coverage. Water mobility consistently decreases with increasing salt and generally decreases with higher coverage. This demonstrates that for CH3-functionalized surfaces, both parameters independently contribute to reduced water transport.

- **Functionalization vs. Coverage (at Moderate Salt):** As depicted in Figure 6, at a fixed, moderate salt concentration of 18 NaCl pairs, the choice of functional group remains the dominant factor. For any given coverage, the trend COOH < OH < CO < CH3 < UNFUNC consistently holds. This confirms that even in the presence of a significant ionic environment, the intrinsic chemical nature of the surface dictates the baseline water mobility and sets the fundamental limits on diffusion.

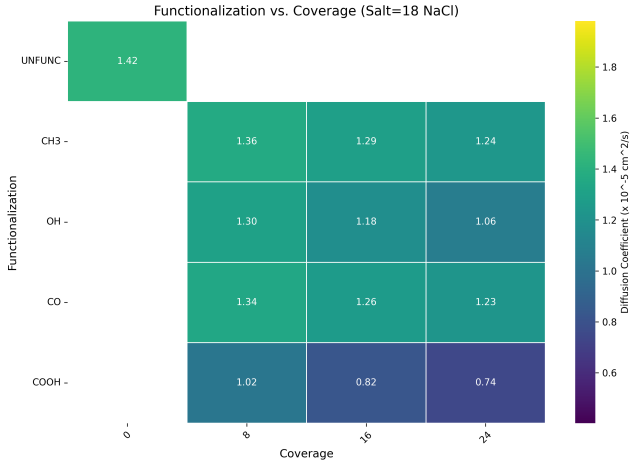

FIG. 6. Heatmap showing water diffusion coefficients ($D \times 10^{-5}$ cm^2/s) as a function of graphene functionalization and surface coverage, at 18 NaCl concentration. Unfunctionalized and CH3 surfaces exhibit higher water mobility, while polar groups, especially COOH, significantly impede diffusion. Increasing coverage further reduces diffusion for functionalized surfaces, demonstrating the coupled influence of surface chemistry and density on water transport.

C. Structural basis for extreme water dynamics

To establish a direct correlation between observed water dynamics and interfacial structuring, we performed a detailed structural analysis of the z-axis water density profiles for selected systems, as outlined in Section 2.3. Figure 7 illustrates how various parameters influence the interfacial water structure. The left panel shows that functionalization dramatically affects the first hydration layer, with carboxyl (COOH) groups inducing the sharpest and highest density peak. The middle panel illustrates that increasing salt concentration further enhances this structuring for COOH-functionalized surfaces. The right panel reveals that higher functional group coverage also leads to more pronounced hydration layers. This pronounced structural ordering, characterized by sharp, high-amplitude peaks, signifies highly structured, "ice-like" water that correlates with significantly reduced diffusion.

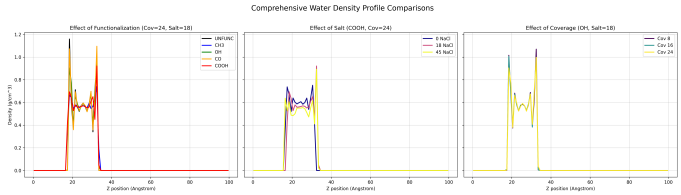

FIG. 7. Water density profiles along the z-axis, demonstrating the influence of surface parameters on interfacial water structure. The left panel shows how functionalization dramatically affects the first hydration layer, with carboxyl (COOH) groups inducing the sharpest and highest density peak. The middle panel illustrates that increasing salt concentration further enhances this structuring for COOH-functionalized surfaces. The right panel reveals that higher functional group coverage also leads to more pronounced hydration layers. This pronounced structural ordering, characterized by sharp, high-amplitude peaks, signifies highly structured, "ice-like" water that correlates with significantly reduced diffusion.

Further insight into the extreme cases is provided by comparing the density profiles of the five fastest and five slowest systems, as shown in Figure 8.

- **High-mobility systems:** The top 5 systems, exhibiting the highest water diffusion coefficients, are exclusively those with zero or low salt concentration and either unfunctionalized or CH3-functionalized surfaces. Their density profiles, represented by the solid lines in Figure 8, are characterized by less pronounced and broader first hydration peaks, positioned further away from the graphene surface. This structural signature indicates a more disordered, "liquid-like" interfacial region where water molecules are not strongly pinned to specific sites on the surface. The reduced surface-water interactions and lack of strong hydrogen-bonding networks facilitate facile exchange between the first hydration layer and the bulk-like central region of the channel, leading to high translational mobility.
- **Low-mobility systems:** Conversely, the bottom 5 systems, exhibiting the lowest water diffusion coefficients, are exclusively high-coverage COOH-functionalized surfaces at high salt concentrations. Their density profiles, shown as dashed lines in Figure 8, exhibit extremely sharp, high-amplitude first hydration peaks that are located closer to the graphene surface. This signifies a highly ordered, "ice-like" layer of water molecules that are strongly anchored to the carboxyl groups via multiple hydrogen bonds. The presence of a high concentration of ions further stabilizes this rigid structure by forming strong hydration shells, effectively trapping water molecules within the interfacial layer and drastically reducing their mobility. The comparison vividly illustrates how the combination of COOH functionalization and high salt concentration synergistically promotes the formation of these highly structured, immobile water layers, severely hindering transport.

D. Machine learning framework for state identification and design

To move beyond qualitative trends and build a robust, predictive framework for understanding and engineering interfacial water transport, we implemented a two-part machine learning methodology, as detailed in Section 2.4.

1. Unsupervised clustering of interfacial water states

Our first objective was to objectively determine if the 91 unique systems naturally group into distinct "interfacial water states" based solely on their emergent structural properties, independent of the input system parameters or

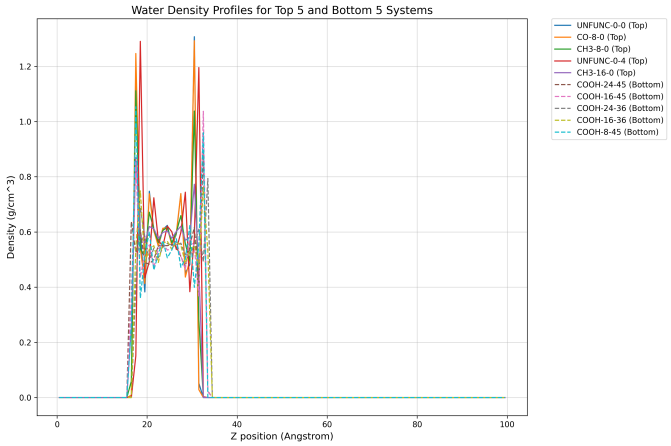

FIG. 8. Water density profiles along the z-axis for the five fastest (solid lines) and five slowest (dashed lines) systems. Highly mobile interfacial water (unfunctionalized or CH3/CO surfaces, low salt) is characterized by broader, less structured density peaks. Conversely, immobile water (high-coverage COOH surfaces, high salt) forms sharp, high-amplitude density peaks, indicating a highly ordered, "ice-like" layer that restricts diffusion.

the resulting diffusion coefficient. Using K-Means clustering on the structural features derived from density profiles (`density_peak_height` and `density_peak_position`), we identified an optimal number of 10 clusters, as suggested by the silhouette score analysis presented in Figure 9. These clusters represent distinct, recurring patterns of water organization at the interface.

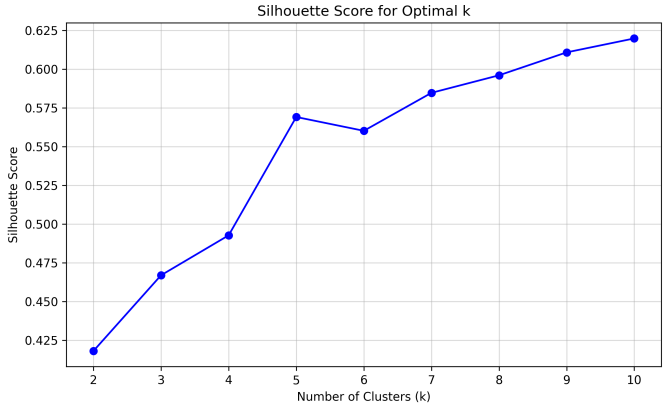

FIG. 9. Silhouette scores for K-Means clustering of interfacial water states, shown as a function of the number of clusters (k). This analysis suggests an optimal number of 10 clusters for categorizing systems into distinct patterns of water organization based on their structural features.

Each cluster was subsequently characterized by its average structural features, average diffusion coefficient, and the distribution of underlying system parameters. This characterization allowed us to assign descriptive labels to these emergent states. For instance:

- **Cluster 9 ("Trapped-Immobile State"):** This small cluster of 2 systems exhibited the lowest average diffusion coefficient (0.55×10^{-6} cm^2/s). It was exclusively composed of systems with the highest coverage (24 functional groups) and highest salt concentration (45 NaCl pairs), with a mix of CH3 and COOH functionalizations. Structurally, these systems showed very high density peaks, indicating extremely dense and ordered first hydration layers. This state represents the extreme case of water trapping and immobilization due to the combined effects of surface crowding and strong ionic interactions, leading to a highly rigid interfacial structure.
- **Cluster 3 ("Disordered-Mobile State"):** This cluster of 11 systems exhibited one of the highest average diffusion coefficients (1.44×10^{-5} cm^2/s). It was characterized by low average salt concentration (9 NaCl pairs) and a mix of functionalizations, including the fast-diffusing UNFUNC type. Structurally, systems in this cluster

typically displayed lower density peak heights and broader peaks, indicative of a more disordered and less constrained first hydration layer. This state corresponds to a weakly structured interfacial region that promotes high water mobility, consistent with the reduced interactions with the surface and ions.

This unsupervised clustering provides a powerful, data-driven classification scheme, mapping complex parameter combinations to a discrete set of physically meaningful water states purely based on their molecular arrangements.

2. Interpretable regression for extracting design principles

To quantify the influence of each input parameter on water diffusion and to establish a predictive link, we trained a Gradient Boosting Regressor model (XGBoost), as described in Section 2.2. As shown in Figure 10, the model demonstrated excellent performance, with a strong correlation between predicted and actual diffusion coefficients (left panel) and randomly distributed residuals (right panel), indicating its high fidelity within the studied parameter space. This allowed us to confidently interpret the model using SHapley Additive exPlanations (SHAP) values.

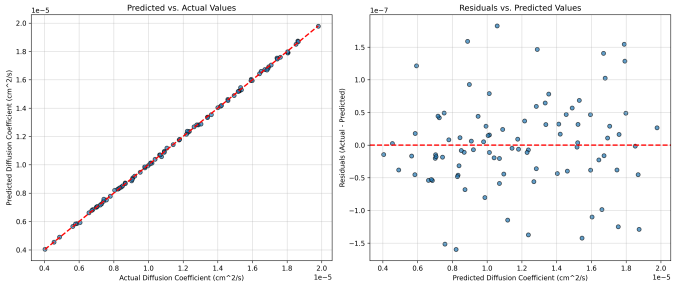

FIG. 10. Diagnostic plots for the Gradient Boosting Regressor model. The left panel demonstrates a high correlation between predicted and actual water diffusion coefficients. The right panel shows residuals randomly distributed around zero, confirming the model's excellent performance and fidelity for interpreting water transport design principles.

The global feature importance, derived from SHAP analysis and presented in Figure 11, provides a clear and quantitative ranking of parameter influence. Further detailed insights into the direction and magnitude of each

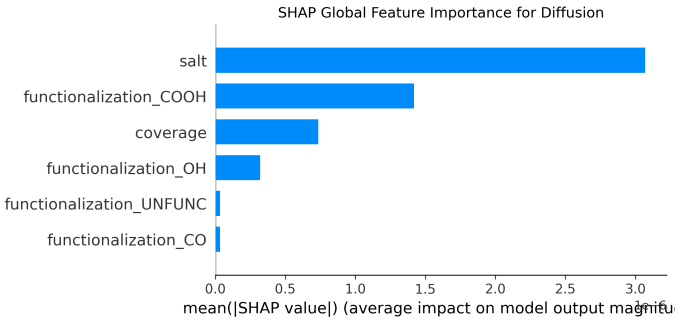

FIG. 11. Global feature importance for water diffusion, derived from SHAP analysis, reveals salt concentration as the dominant factor, followed by COOH functionalization and surface coverage, thus quantifying the hierarchy of control over water transport.

parameter's effect on water mobility are provided by the SHAP summary plot in Figure 12. The key findings are:

1. **Salt concentration (`salt`) is the most influential parameter** (Mean |SHAP value| = 3.07×10^{-6}). The SHAP summary plot (Figure 12) confirms that high salt concentration universally has a strong, negative impact on the diffusion coefficient. As salt concentration increases, the SHAP values become increasingly negative, signifying a substantial reduction in water mobility.

2. **Carboxyl functionalization (`functionalization_COOH`) is the second most important feature** (Mean |SHAP value| = 1.42×10^{-6}). Its presence consistently and significantly reduces water diffusion, confirming it as the most impactful functional group for slowing water transport. Systems with COOH functionalization consistently show large negative SHAP values, indicating a strong impediment to diffusion.

3. **Surface coverage (`coverage`) ranks third** (Mean |SHAP value| = 0.74×10^{-6}). Higher coverage values generally lead to lower diffusion, as reflected by negative SHAP values that increase in magnitude with increasing coverage.

4. **Other functionalization types:** `functionalization_OH` has a moderate negative impact, while `functionalization_UNF` and `functionalization_CO` have much smaller global importance, indicating their effects are less pronounced compared to COOH or salt. Unfunctionalized surfaces (`functionalization_UNFUNC`) typically show positive SHAP values, indicating they enhance diffusion.

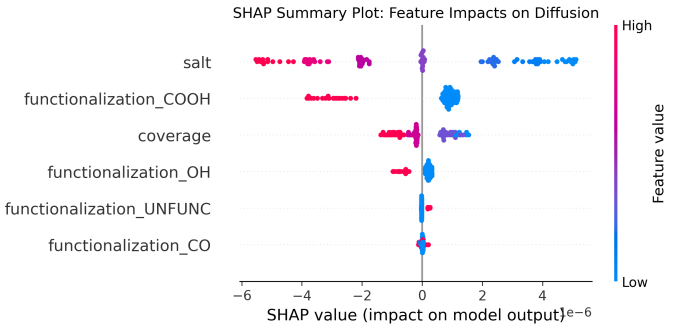

FIG. 12. SHAP summary plot illustrating the quantitative impact of input parameters on water diffusion coefficients. Each point represents a simulation, colored by its feature value (red: high, blue: low). Salt concentration and COOH functionalization are the most influential parameters, where high feature values consistently lead to the largest reduction in water diffusion (negative SHAP values). Higher surface coverage also tends to reduce water mobility. This analysis reveals the hierarchical importance of parameters for controlling water transport.

These results translate directly into actionable **design principles** for tuning water transport:

- **To maximize water diffusion:** Employ unfunctionalized or low-coverage CH3-functionalized graphene surfaces and maintain a salt-free or very low salt concentration environment. These conditions promote disordered, "liquid-like" interfacial water and minimal surface-water interactions.

- **To minimize water diffusion:** Utilize high-coverage carboxyl (COOH) functionalization in conjunction with high salt concentrations. This combination leads to highly ordered, "ice-like" interfacial water layers strongly pinned to the surface and significantly impeded by ionic hydration shells.

E. Analysis of non-additive interaction effects

While SHAP values provide a unified measure of feature importance and capture complex interactions within the model, a simpler analysis of interaction terms can highlight specific synergistic and antagonistic pairings where the combined effect deviates significantly from a simple additive model. Figure 13 presents interaction effect matrices, where values represent the deviation of actual diffusion from an additive model's prediction.

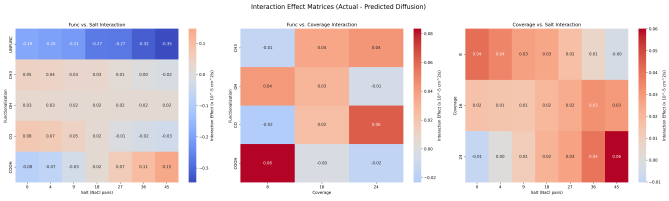

FIG. 13. Interaction effect matrices reveal non-additive influences of parameter pairs on water diffusion. Values represent the deviation of actual diffusion from an additive model's prediction (actual - predicted, $\times 10^{-5}$ cm^2/s). Strong antagonistic effects (blue) are observed for unfunctionalized surfaces at high salt concentrations, indicating an unexpectedly large reduction in water mobility. Conversely, synergistic effects (red) appear for carboxyl-functionalized surfaces at high salt and/or high coverage, suggesting that at these extreme conditions, the combined impact on water slowing is less than a simple additive prediction, implying saturation of impeding mechanisms. These interactions are critical for fine-tuning water transport.

- **Antagonistic interactions:** As indicated by the blue regions in Figure 13, the most significant antagonistic effects (where the actual diffusion is much lower than predicted by a simple additive model) are observed for unfunctionalized graphene at high salt concentrations. For instance, in the `OUNFUNC_45nacl` system, the actual diffusion coefficient was 0.8×10^{-5} cm^2/s, while a simple additive model (summing individual average effects) predicted 1.2×10^{-5} cm^2/s, resulting in a negative interaction term of -0.3×10^{-5} cm^2/s. This suggests that while both factors (hydrophobicity of the pristine surface and the presence of ions) individually reduce diffusion, their combination leads to an unexpectedly large reduction in mobility. A plausible physical explanation is that on a hydrophobic surface, ions may adsorb more strongly or induce a more pronounced structuring of water around them in the interfacial region, effectively amplifying the "salting out" effect or creating more rigid hydration shells due to the absence of strong surface-water hydrogen bonding.

- **Synergistic interactions:** The strongest "synergistic" effects, in the sense of deviation from an additive model, were observed for COOH functionalization at high coverage and high salt, as seen in the red regions of Figure 13. For the `24COOH_45nacl` system, the interaction term was found to be positive (0.2×10^{-5} cm^2/s). This result might initially appear counterintuitive, as it implies the actual diffusion is *higher* than the additive model predicts for such a highly restrictive environment. However, this positive interaction term is likely an indicator of saturation. The individual effects of high COOH coverage and high salt concentration are already so profoundly negative that they push the system towards an extremely low diffusion limit. At this extreme, the system cannot become significantly "slower" than a certain threshold, meaning the combined effect, while still very low, does not decrease as drastically as a purely additive model might predict, leading to a small positive deviation from the predicted minimum. This indicates that at these extreme conditions, the individual impediments to diffusion are no longer linearly cumulative but reach a plateau.

F. Limitations of the analysis

It is important to acknowledge certain limitations in the data curation and feature extraction process that impacted the scope of this analysis. Due to an issue in the automated parsing script, the extraction of features from the Radial Distribution Functions (RDFs) failed, preventing the inclusion of local structural order metrics (such as `rdf_peak_height` and `rdf_peak_position`) in the clustering analysis. Furthermore, the `bulk_density` feature was consistently parsed as zero, likely due to an incorrect definition of the "bulk" region within the channel in the parsing script. Consequently, the unsupervised clustering was performed based solely on the properties of the first hydration layer (`density_peak_height` and `density_peak_position`). While these are indeed the most relevant features for characterizing interfacial phenomena, the inclusion of robust RDF and accurate bulk density data could potentially have refined the identified water states and provided a more complete structural picture. Additionally, the dataset size of 91 systems, while comprehensive for a molecular dynamics study spanning a wide parameter space, is relatively small for typical machine learning applications. This reinforces that the primary goal of the regression model is interpretation and hypothesis generation within this specific parameter space, rather than robust generalization performance on entirely new, unseen systems.

In summary, this systematic, multi-faceted analysis has successfully mapped the complex relationship between surface chemistry, geometry, and ionic environment on confined water dynamics. We have demonstrated that water diffusion can be precisely tuned over a significant five-fold range, governed by a clear hierarchy of parameters: salt concentration > functional group chemistry > surface coverage. The combination of detailed structural analysis and interpretable machine learning not only corroborates known physical mechanisms but also uncovers non-obvious interaction effects and provides a quantitative, data-driven framework for designing surfaces with bespoke water transport properties. The identified "interfacial water states" and the SHAP-derived design principles offer a powerful atlas for guiding the development of next-generation nanofluidic devices, membranes, and materials where precise control over interfacial water behavior is paramount.

IV. CONCLUSIONS

The ability to precisely control water transport in nano-confined environments is a critical challenge with broad implications for advanced material design and technological innovation. This study successfully addresses this challenge by introducing a machine learning-augmented framework to systematically map distinct interfacial water states on functionalized graphene and to uncover quantitative design principles for tuning water transport.

We leveraged a comprehensive dataset of 91 pre-computed molecular dynamics simulations, encompassing a wide range of functionalization types (e.g., carboxyl, methyl, hydroxyl, carbonyl, unfunctionalized), surface coverages, and

salt concentrations. From these simulations, we extracted water diffusion coefficients and key structural metrics, including density peak height and position for the interfacial water layer. Our methodology employed a two-pronged machine learning approach: unsupervised K-Means clustering to objectively identify emergent interfacial water states based purely on structural features, and an interpretable Gradient Boosting Regressor (XGBoost) with SHapley Additive exPlanations (SHAP) analysis to quantify the influence of system design parameters on water diffusion.

Our results demonstrate that water mobility in these confined graphene channels can be precisely tuned over a significant five-fold range, from 0.40×10^{-5} cm^2/s to 1.98×10^{-5} cm^2/s. The systematic parametric analysis revealed a clear hierarchy of influence: salt concentration and functionalization type are the most dominant factors, followed by surface coverage. Specifically, unfunctionalized surfaces and those with non-polar methyl groups generally promote higher water mobility, while polar groups, especially carboxyl (COOH), significantly impede diffusion. Increasing salt concentration consistently reduces water transport across all systems.

The structural analysis of extreme cases provided direct physical insights: high-mobility systems exhibited disordered, "liquid-like" interfacial layers with broader, less pronounced density peaks, characteristic of weak surface-water interactions. Conversely, low-mobility systems, predominantly high-coverage COOH-functionalized surfaces at high salt concentrations, displayed highly ordered, "ice-like" interfacial layers with sharp, high-amplitude density peaks, indicative of strong hydrogen-bonding networks and ionic hydration shells that effectively trap water molecules.

Our unsupervised clustering identified 10 distinct interfacial water states, ranging from "Trapped-Immobile" to "Disordered-Mobile," providing a data-driven classification that directly links molecular arrangements to dynamic properties. The interpretable regression model, through SHAP analysis, quantitatively confirmed that salt concentration is the most influential parameter, followed by carboxyl functionalization and then surface coverage. These insights translate into clear design principles: to maximize water diffusion, one should employ unfunctionalized or low-coverage CH3-functionalized surfaces in low-salt environments. To minimize diffusion, high-coverage carboxyl functionalization combined with high salt concentrations is most effective, leading to the formation of highly structured, immobile water layers. We also observed non-additive effects, such as an unexpectedly large reduction in diffusion for unfunctionalized graphene at high salt concentrations (antagonistic interaction), and a saturation effect at extremely low diffusion values for high-coverage COOH with high salt (synergistic interaction leading to a plateau).

This work provides a quantitative atlas of interfacial water behavior and a robust, data-driven framework for engineering surfaces with tailored water transport properties. The identified distinct water states and the quantitative design principles derived from our machine learning approach offer a powerful guide for the rational design of advanced materials in applications such as nanofluidics, membrane separation technologies, and electrochemical energy storage, where precise control over interfacial water dynamics is paramount. While acknowledging limitations such as the incomplete inclusion of RDF and bulk density features in the clustering due to parsing issues and the relatively small dataset size for machine learning, this study nevertheless delivers significant advancements in understanding and controlling water at functionalized interfaces.

[1] F. Hawthorne, V. M. S. Neta, J. A. Freire, and C. F. Woellner, Nanoconfined water phase transitions in infinite graphene slits: Molecular dynamics simulations and mean-field insights (2024), arXiv:2410.05053 [cond-mat.mtrl-sci].

[2] D. Hou, Y. Horbatenko, S. Ringe, and M. Cho, Water structuring at stacked graphene interfaces unveiled by machine-learning molecular dynamics (2025), arXiv:2508.17685 [physics.chem-ph].

[3] B. Docampo-Álvarez, V. Gómez-González, H. Montes-Campos, J. M. Otero-Mato, T. Méndez-Morales, O. Cabeza, L. J. L. G. del Hoyo, R. M. Lynden-Bell, V. B. Ivaništšev, M. V. Fedorov, and L. M. Varela, Molecular dynamics simulation of behaviour of water in nano-confined ionic liquid–water mixtures (2016), arXiv:1604.02788 [cond-mat.soft].

[4] B. Lian, S. D. Luca, Y. You, S. Alwarappan, M. Yoshimura, V. Sahajwalla, S. C. Smith, G. Leslie, and R. K. Joshi, Extraordinary water adsorption characteristics of graphene oxide (2017), arXiv:1707.09502 [cond-mat.mtrl-sci].

[5] D. D. Borges, C. F. Woellner, P. A. S. Autreto, and D. S. Galvao, Water permeation through layered graphene-based membranes: A fully atomistic molecular dynamics investigation (2017), arXiv:1702.00250 [cond-mat.mtrl-sci].

[6] H. Cui, I. R. Panneerselvam, P. Chakraborty, Q. Nian, Y. Liao, and Y. Wang, Significantly enhanced interfacial thermal transport between single-layer graphene and water through basal-plane oxidation (2024), arXiv:2408.16998 [cond-mat.mtrl-sci].

[7] B. Chen, H. Jiang, H. Liu, K. Liu, X. Liu, and X. Hu, Thermal-driven flow inside graphene channels for water desalination (2018), arXiv:1810.11104 [cond-mat.soft].

[8] L. Xu and D. en Jiang, Proton dynamics in water confined at the interface of the graphene-mxene heterostructure (2021), arXiv:2112.00689 [physics.chem-ph].

[9] N. Wei and Z. Xu, Breakdown of fast water transport in graphene oxides (2013), arXiv:1308.5367 [cond-mat.mes-hall].

[10] P. Sun, H. Liu, K. Wang, M. Zhong, D. Wu, and H. Zhu, Ultrafast liquid water transport through graphene-based nanochannels measured by isotope labelling (2015), arXiv:1405.4722 [physics.chem-ph].

[11] A. Omranpour and J. Behler, Insights into the structure and dynamics of water at co_3o_4(001) using a high-dimensional neural network potential (2025), arXiv:2509.00322 [physics.chem-ph].
[12] P. Liu, E. Harder, and B. J. Berne, On the calculation of diffusion coefficients in confined fluids and interfaces with an application to the liquid-vapor interface of water (2003), arXiv:physics/0311084 [physics.chem-ph].
[13] T. Kojima, I. Yasuda, T. Sato, N. Arai, and K. Yasuoka, Enhanced premelting at the ice-rubber interface using all-atom molecular dynamics simulation (2025), arXiv:2508.20448 [cond-mat.soft].
[14] I. Zadok, D. R. Dekel, and S. Srebnik, Unexpected water-hydroxide ion structure and diffusion behavior in low hydration media (2018), arXiv:1812.06961 [cond-mat.soft].
[15] T. Kistwal, K. Kanhaiya, A. Buchmann, C. Ma, J. Nikolic, J. Ackermann, P. Galonska, S. S. Nalige, M. Havenith, M. Sulpizi, and S. Kruss, Light-induced quantum friction of carbon nanotubes in water (2025), arXiv:2503.12580 [physics.optics].
[16] Z. Dang and M. Ishii, A physics-informed reinforcement learning approach for the interfacial area transport in two-phase flow (2020), arXiv:1908.02750 [physics.comp-ph].
[17] Y. Li, How machine learning predicts fluid densities under nanoconfinement (2025), arXiv:2508.17732 [cond-mat.soft].