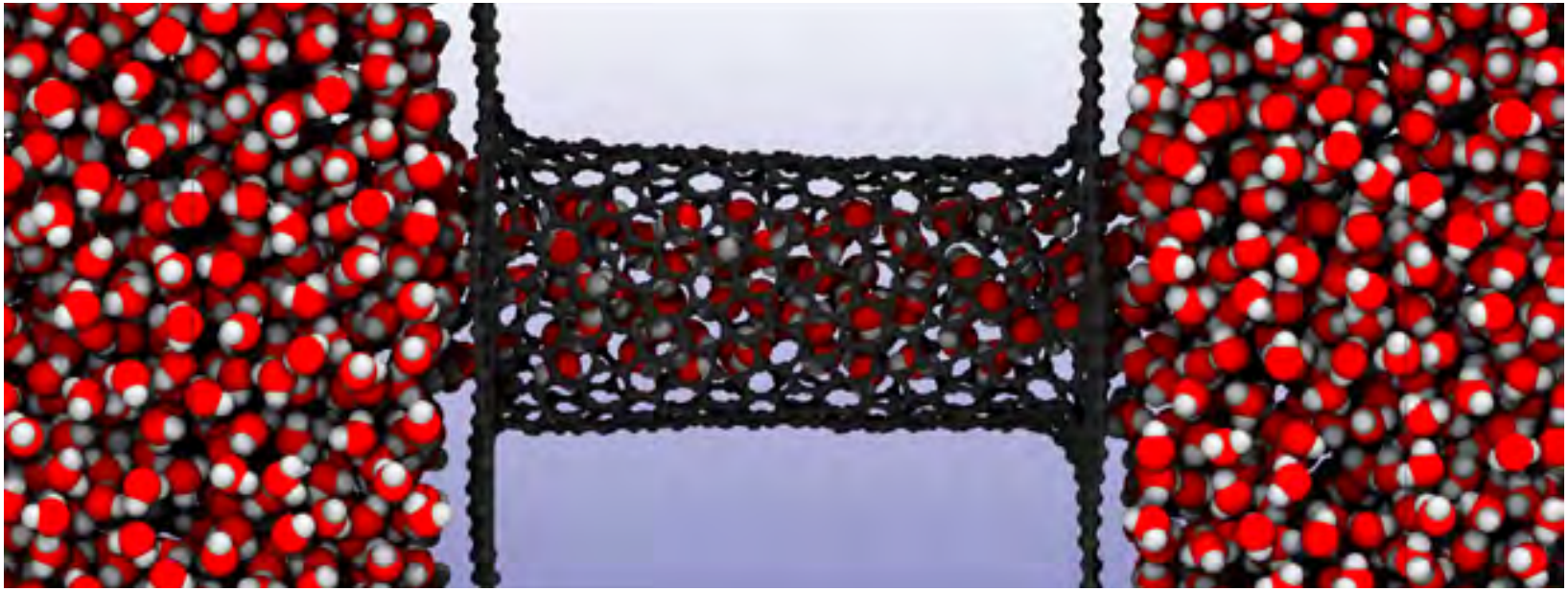
UNIVERSITY OF CAMBRIDGE

# Committee neural network potentials control generalization errors and enable active learning

Christoph Schran | University of Cambridge | Jun 23, 2022

# Water in Complex Environments

## Interfacial Water and Water under Confinement



## Relevance

- Nano filtration[1]
- Osmotic energy conversion[2]
- Desalination[3]

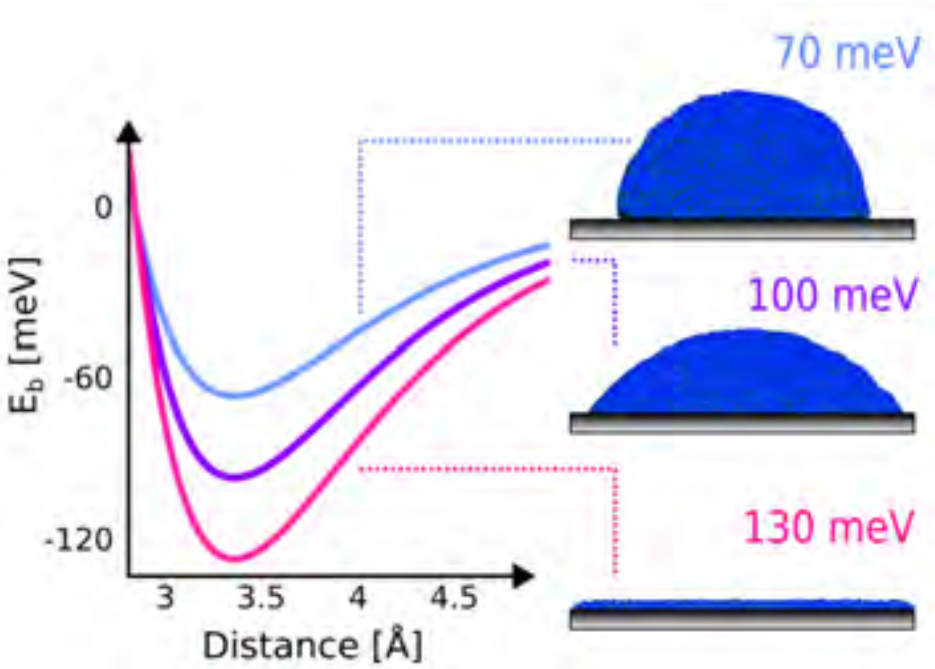
## Open Questions

- Water mobility and reactivity
- Material dependence
- Confinement effect


[1] Holt, J. K. et al. Science **2006**, *312*, 1034–1037.
[2] Siria, A. et al. Nature **2013**, *494*, 455–458.
[3] Shannon, M. A. et al. Nature **2008**, *452*, 301–310.

# Simulation of Water in Complex Environments

## We can't afford to make errors...




Complex materials need...

- high accuracy (sensitive properties)
- large system sizes and long simulation time

### *Ab initio* methods

\+ High accuracy

– Expensive / slow

### Force field methods

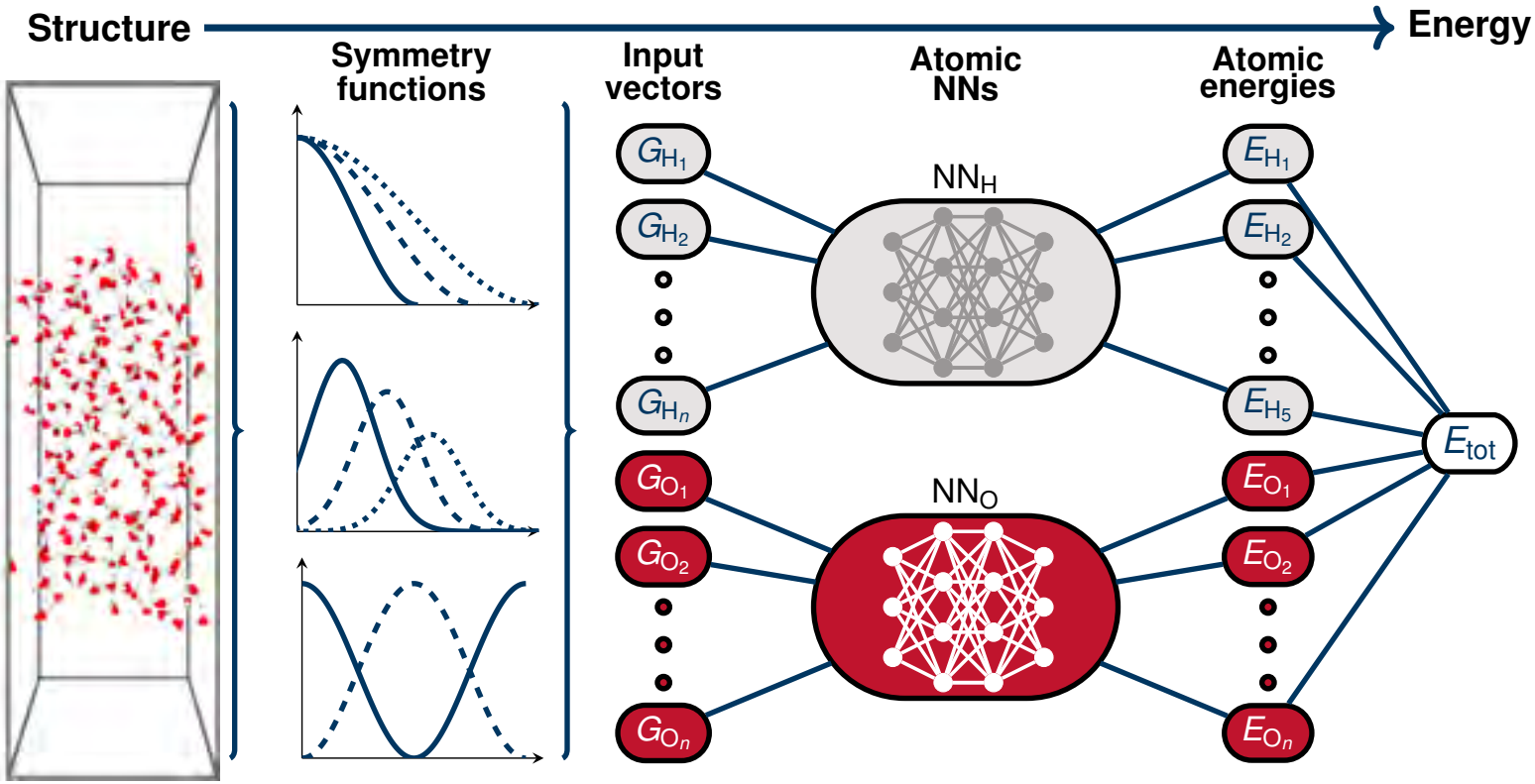
\+ Large system sizes

– Not accurate enough

➡ Use machine learning to get the best of both worlds:
High accuracy + fast evaluation

[1] Brandenburg, J. G. et al. J. Phys. Chem. Lett. **2019**, *10*, 358–368.

# High–dimensional Neural Network Potentials




- Analytical structure–energy relation[1][2]
- Transform structure via atom centered symmetry functions[3]
- Use in simulations: CP2K and Lammps


[1] Behler, J.; Parrinello, M. Phys. Rev. Lett. **2007**, *98*, 146401.
[2] Behler, J. Angew. Chemie - Int. Ed. **2017**, *56*, 12828–12840.
[3] Behler, J. J. Chem. Phys. **2011**, *134*, 074106.

# Committee Neural Network Potentials

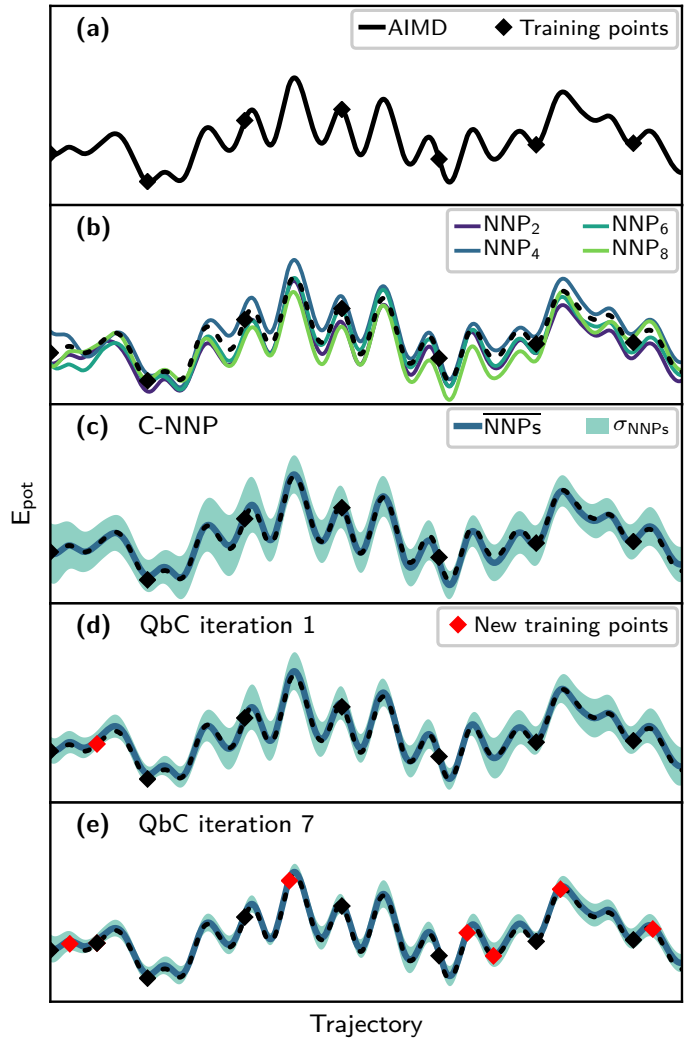
## Simple but Powerful Extension

$$E^{\text{C-NNP}} = \frac{1}{n_{\text{NNP}}} \sum_{i=1}^{n_{\text{NNP}}} E_i^{\text{NNP}}$$

$$\vec{F}^{\text{C-NNP}} = \frac{1}{n_{\text{NNP}}} \sum_{i=1}^{n_{\text{NNP}}} \vec{F}_i^{\text{NNP}}$$

Advantages[1][2][3][4]:

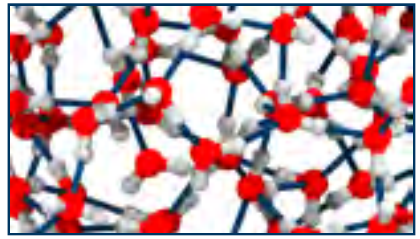
- More accurate prediction
- Estimation of uncertainty
- Reduced overfitting
- Active learning via QbC
- Efficient with shared descriptors





[1] Schran, C.; Brezina, K.; Marsalek, O. J. Chem. Phys. **2020**, *153*, 104105.
[2] Gastegger, M.; Behler, J.; Marquetand, P. Chem. Sci. **2017**, *8*, 6924–6935.
[3] Musil, F.; Willatt, M. J.; Langovoy, M. A.; Ceriotti, M. J. Chem. Theory Comput. **2019**, *15*, 906–915.
[4] Smith, J. S. et al. J. Chem. Phys. **2018**, *148*, 241733.

# Active Learning for Water
## Query by Committee

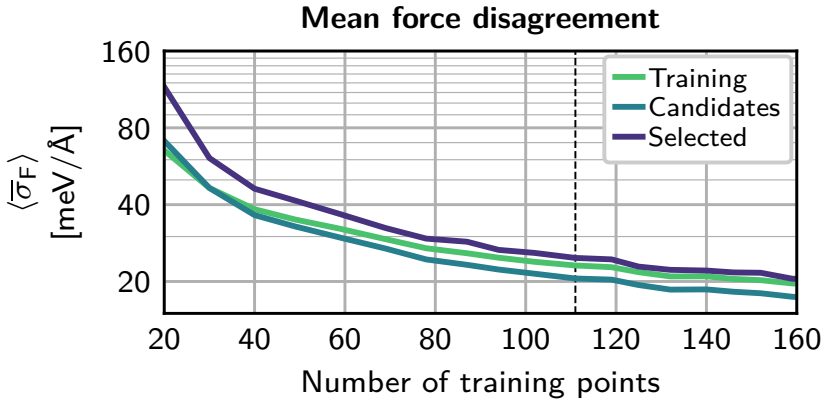


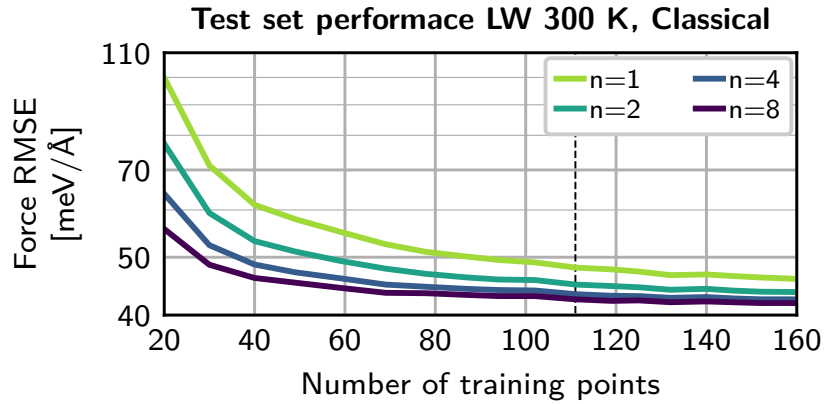
Starting point: AIMD water (revPBE0-D3) at 300K

- Start with 20 random structures
- Use force disagreement to select new training points
- 10 new points per iteration

➡ Committee average is much better predictor

➡ Convergence is reached after roughly 100 points



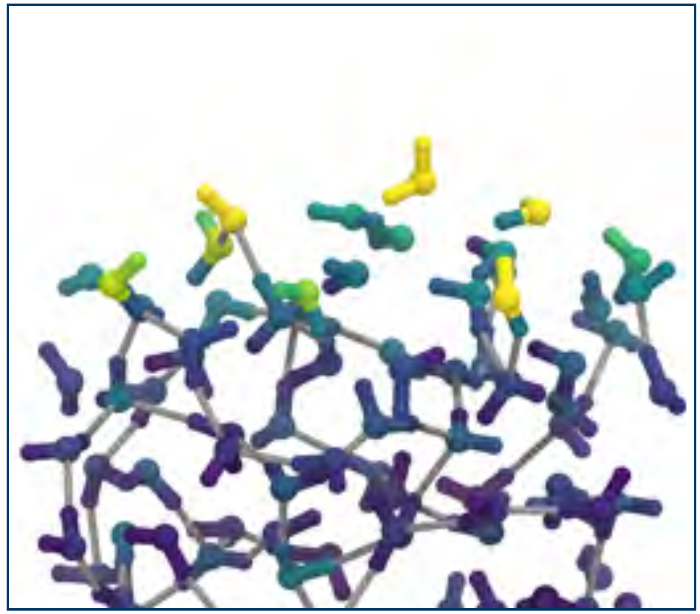

[1] Schran, C.; Brezina, K.; Marsalek, O. J. Chem. Phys. **2020**, *153*, 104105.

# C-NNPs: Robust and Accurate

## Tests on First Generation

Bulk water model trained to 100 points can:

- Treat higher T up to 500 K
- Simulate ice, gas phase clusters and interfaces
- ➡ Committee disagreement shows weakness of model
- Disagreement is directly related to error[1]



Force disagreement at the water-air interface

[1] Musil, F.; Willatt, M. J.; Langovoy, M. A.; Ceriotti, M. J. Chem. Theory Comput. **2019**, *15*, 906–915.

# C-NNPs: Controlling the Stability
## Biasing the Disagreement

Small disagreement = stable and accurate model

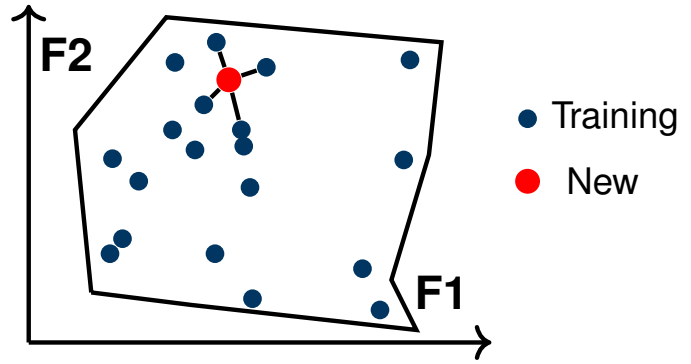
- Define simple biasing potential acting on energy disagreement

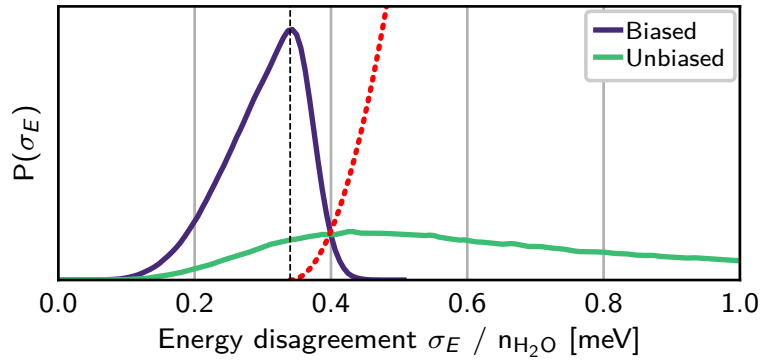
➡ Erect barrier around area with small error

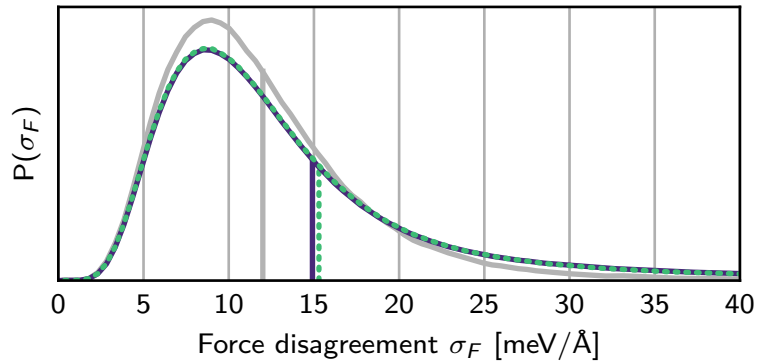
- Force disagreement not perturbed

➡ Mild influence on dynamics

$$E^{(\mathrm{b})} = \theta(\sigma_E - \sigma_0)\frac{1}{2}k^{(\mathrm{b})}\left(\sigma_E - \sigma_0\right)^2$$

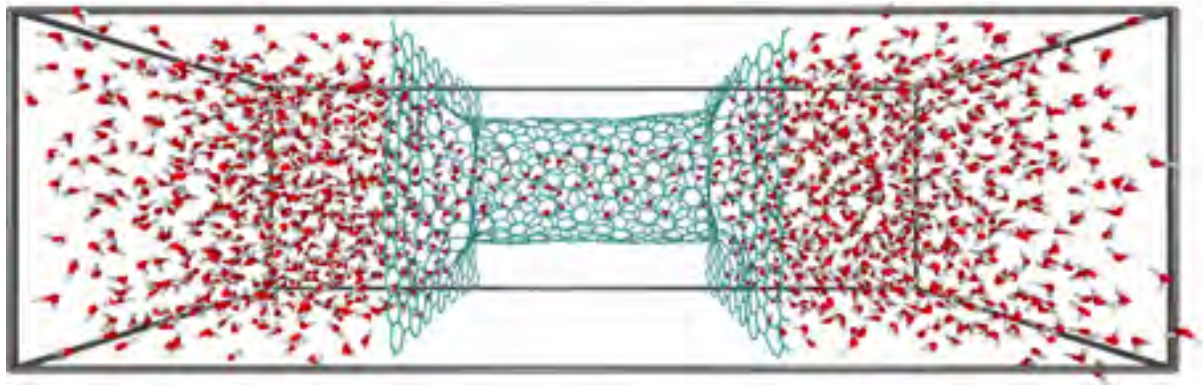



[1] Schran, C.; Brezina, K.; Marsalek, O. J. Chem. Phys. **2020**, *153*, 104105.

# Rapid development of MLPs for complex aqueous systems

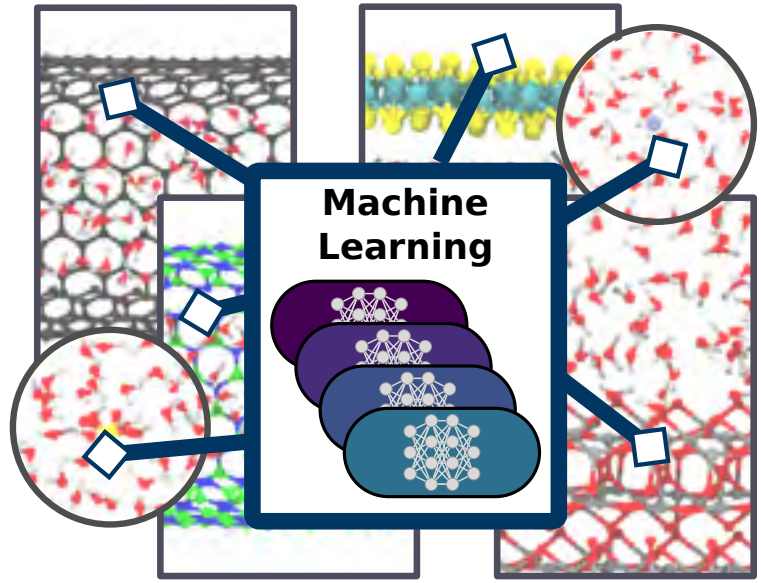


# Machine Learning Potentials Made Simple

Build end-to-end framework for complex systems in simple manner.

## Conceptual Idea

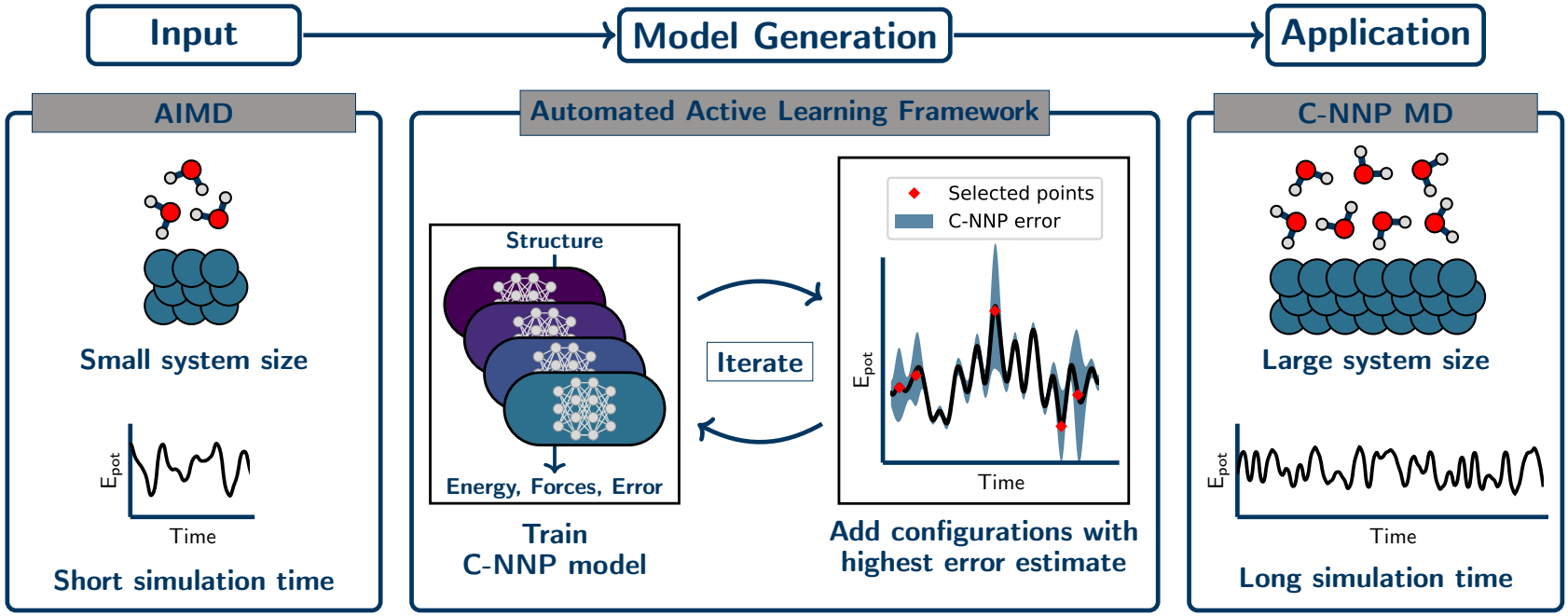
- Concentrate on specific thermodynamic condition
- Remove as much user input as possible
- Automate most of the development steps
- Establish automated validation scheme




➡ Training and simulation codes are open source + machine learning framework is made available

[1] Schran, C. et al. Proc. Natl. Acad. Sci. **2021**, *118*.

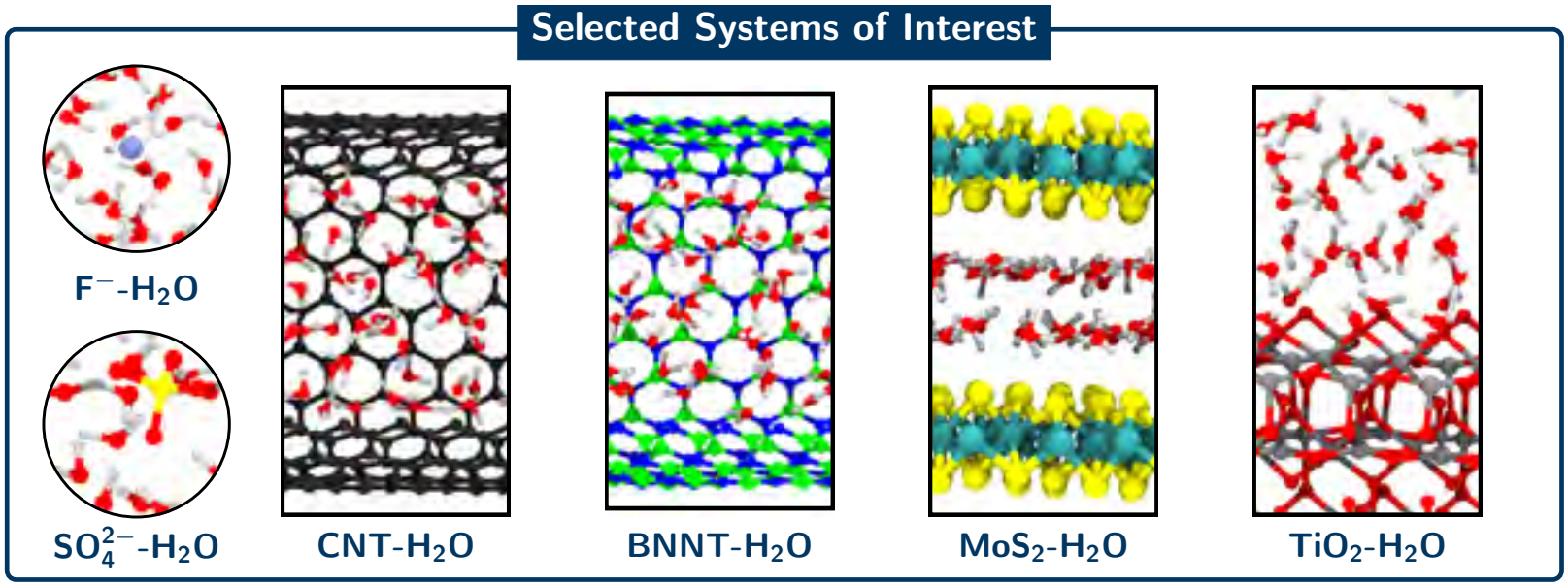
# Machine Learning Potentials made simple




## Rapid Development

- Use small scale AIMD trajectory (30 ps)
- Perform active learning
- Apply model for large scale C-NNP simulations

# C-NNPs for Water in Complex Environments




➡ C-NNP models for all 6 systems by identifying ≈300 structures from AIMD simulations

# Validation of the Models

## How to validate >6 models?

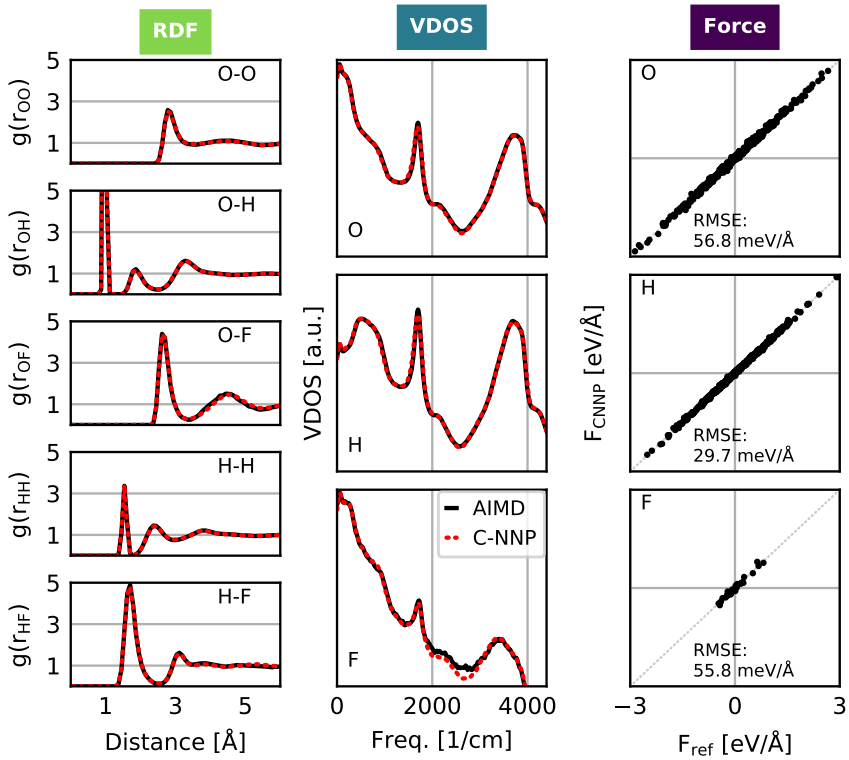
Automated testing:

1. Structural properties: RDFs
2. Dynamical properties: VDOS
3. ML properties: Force RMSE

Reduce to score (0-100%)

Scoring for RDFs and VDOS compared to AIMD:

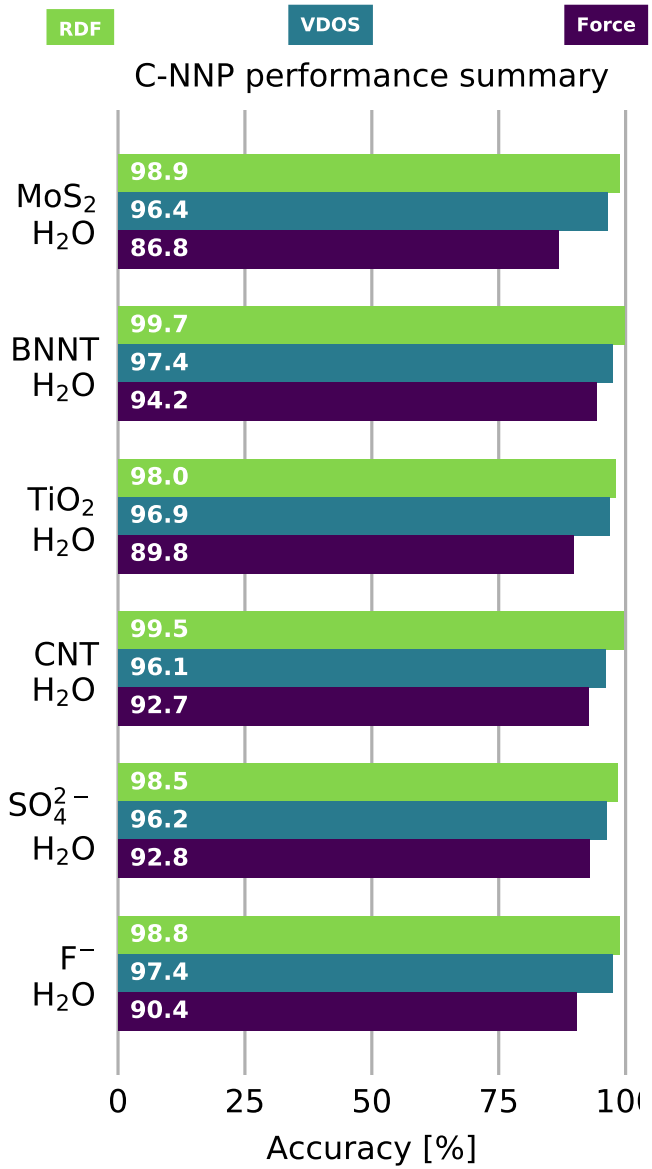
$$d = \frac{\int_{-\infty}^{+\infty} |\rho_P(x) - \rho(x)| \, \mathrm{d}x}{\int_{-\infty}^{+\infty} \rho_P(x) \mathrm{d}x + \int_{-\infty}^{+\infty} \rho(x) \mathrm{d}x}$$



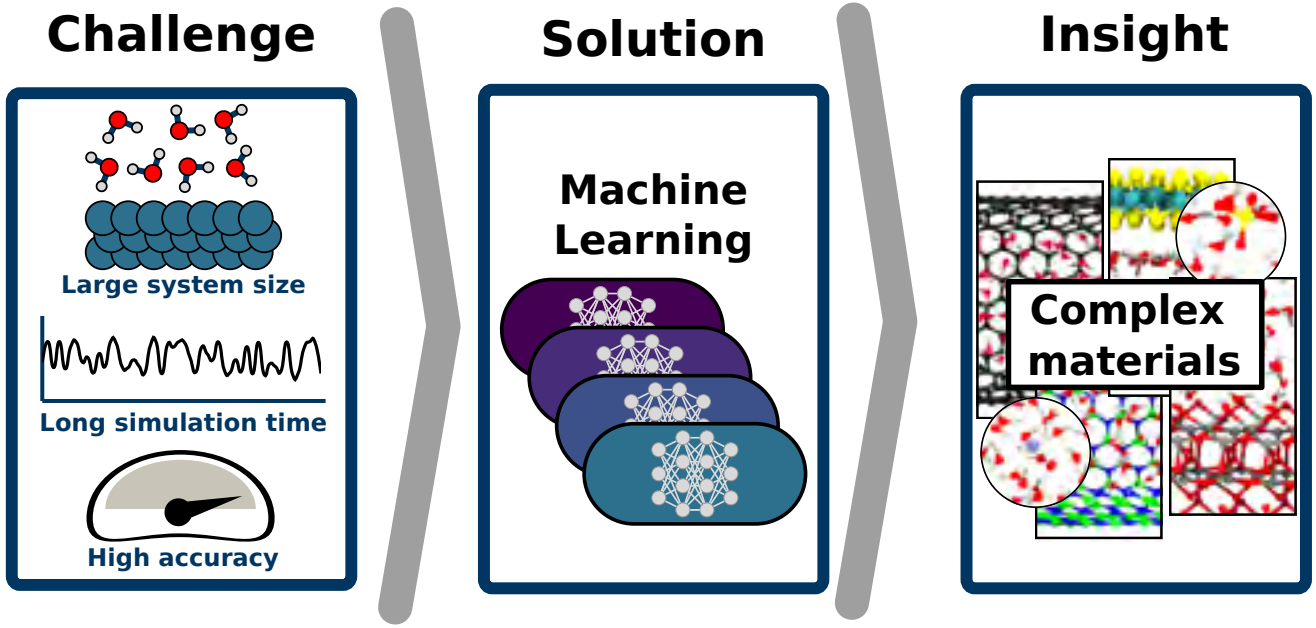
# Validation of the Models

## Accuracy overview

- All properties are reproduced very accurately
- Similar accuracy for all diverse systems
- Other properties (Hbonding, density profiles, water orientation) are also in very good agreement with reference



# Summary




## MLPs made simple

- Accurate and efficient description
- Data-driven and automated generation

➡ Open-source development including tutorials:

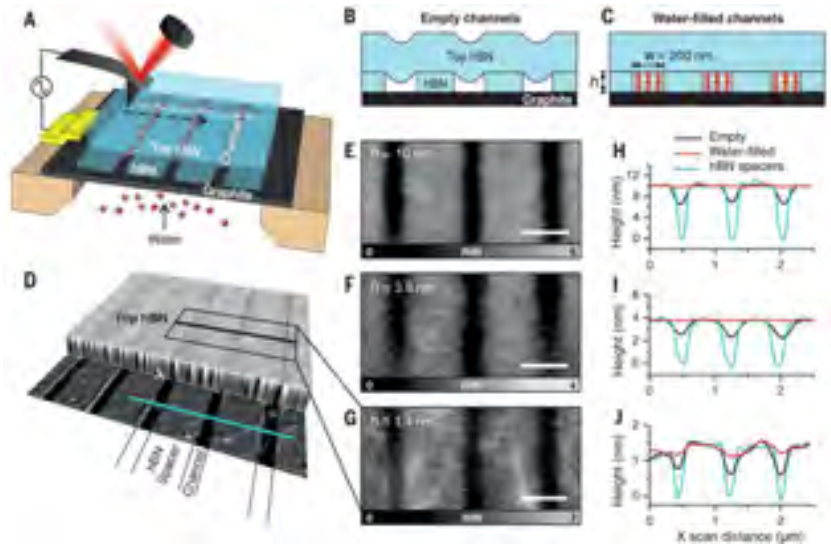
https://github.com/Marsalek-Group/aml

[1] Schran, C. et al. Proc. Natl. Acad. Sci. **2021**, *118*.

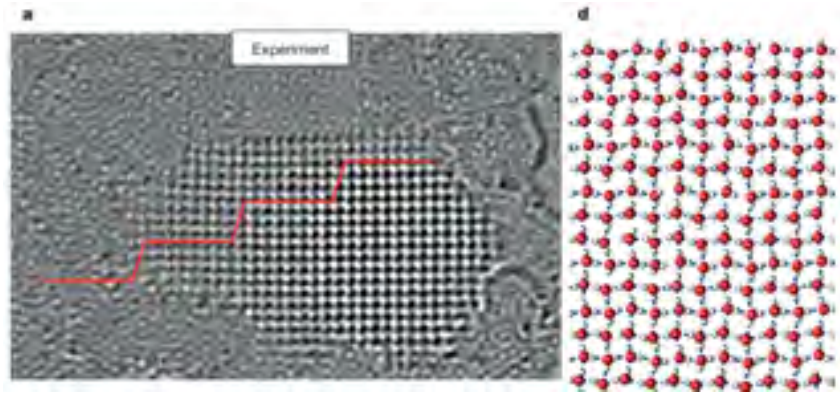
# Application: Phase behavior of monolayer confined water

# Motivation

## Experimental fabrication of hydrophobic capillaries with nanoscale dimensions




Measurement of anonymously low dielectric constant of nanoconfined water[1]



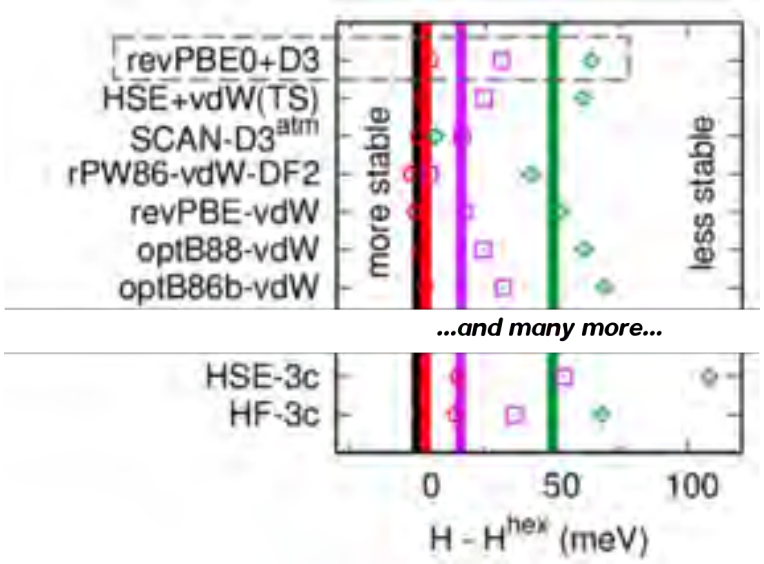

Possible existence of a square ice phase[2]

- ➡ Remarkably different properties of nanoscale water
- ➡ Provide molecular level understanding as foundation for technological application

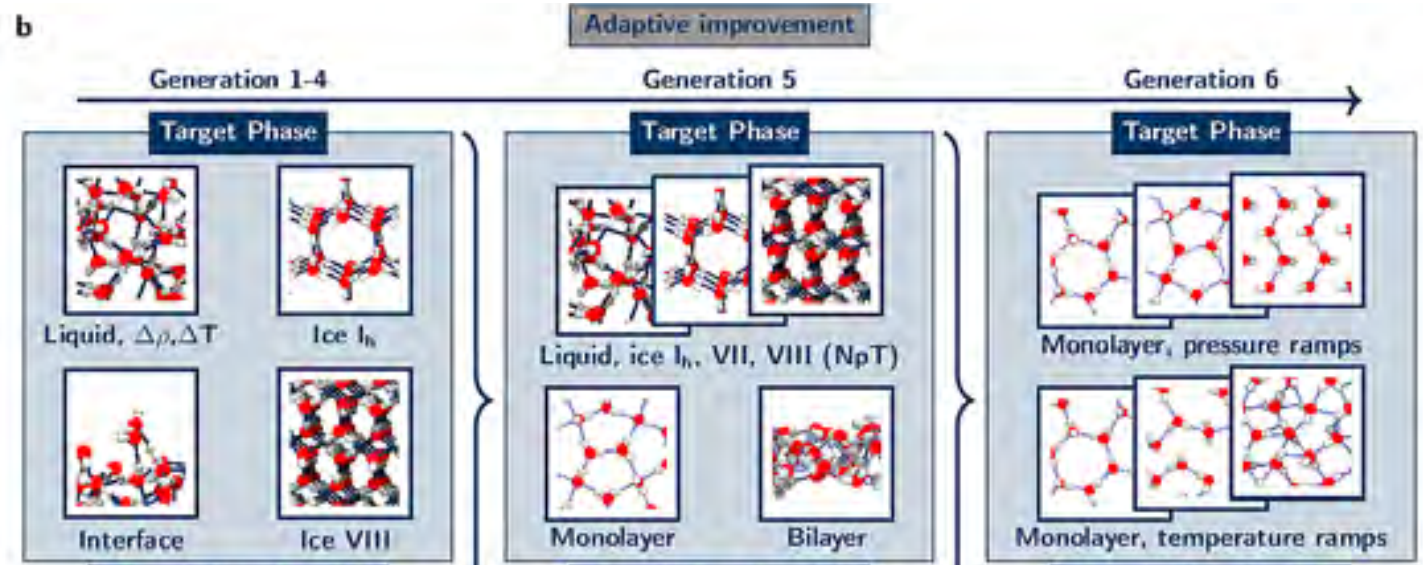

[1] Fumagalli, L. et al. Science **2018**, *360*, 1339–1342.
[2] Algara-Siller, G. et al. Nature **2015**, *519*, 443–445.

# Development of a Mono-layer Water Model

## Quantum Monte Carlo guided MLP




1. Use QMC to select DFT functional
2. Train MLP to DFT
3. RSS for all metastable phases
4. Free energy methods for phase diagram



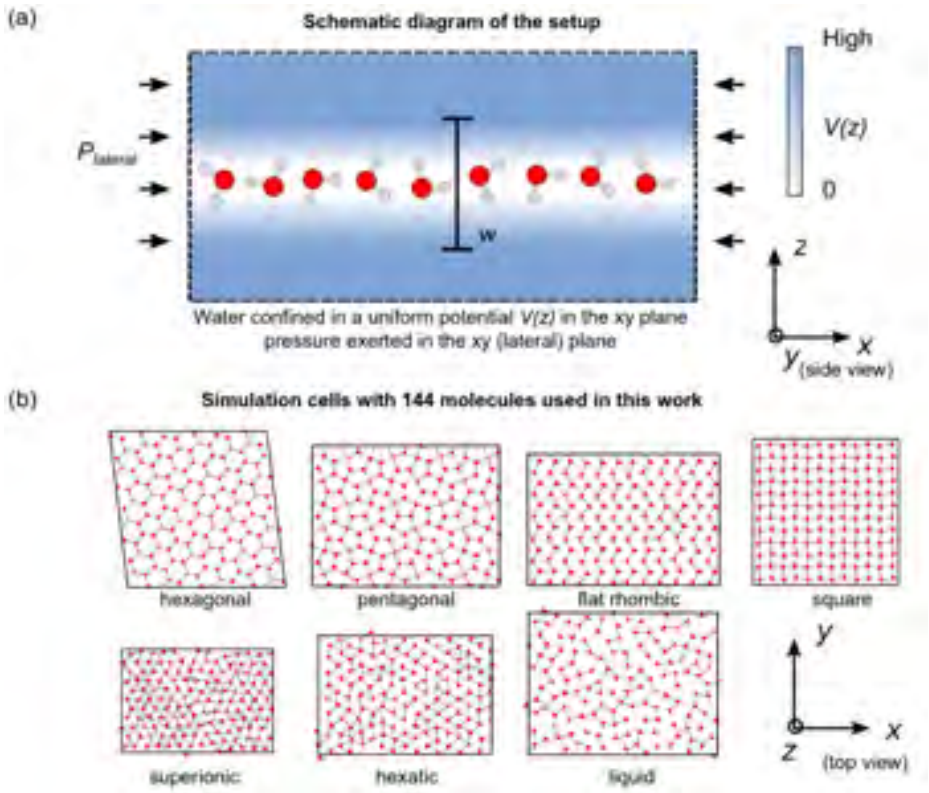
# Simulations of Monolayer confined water

## Setup

- Lennard-Jones confinement potential fitted to QMC data
- Lateral pressure
- 144 molecules per unit cell

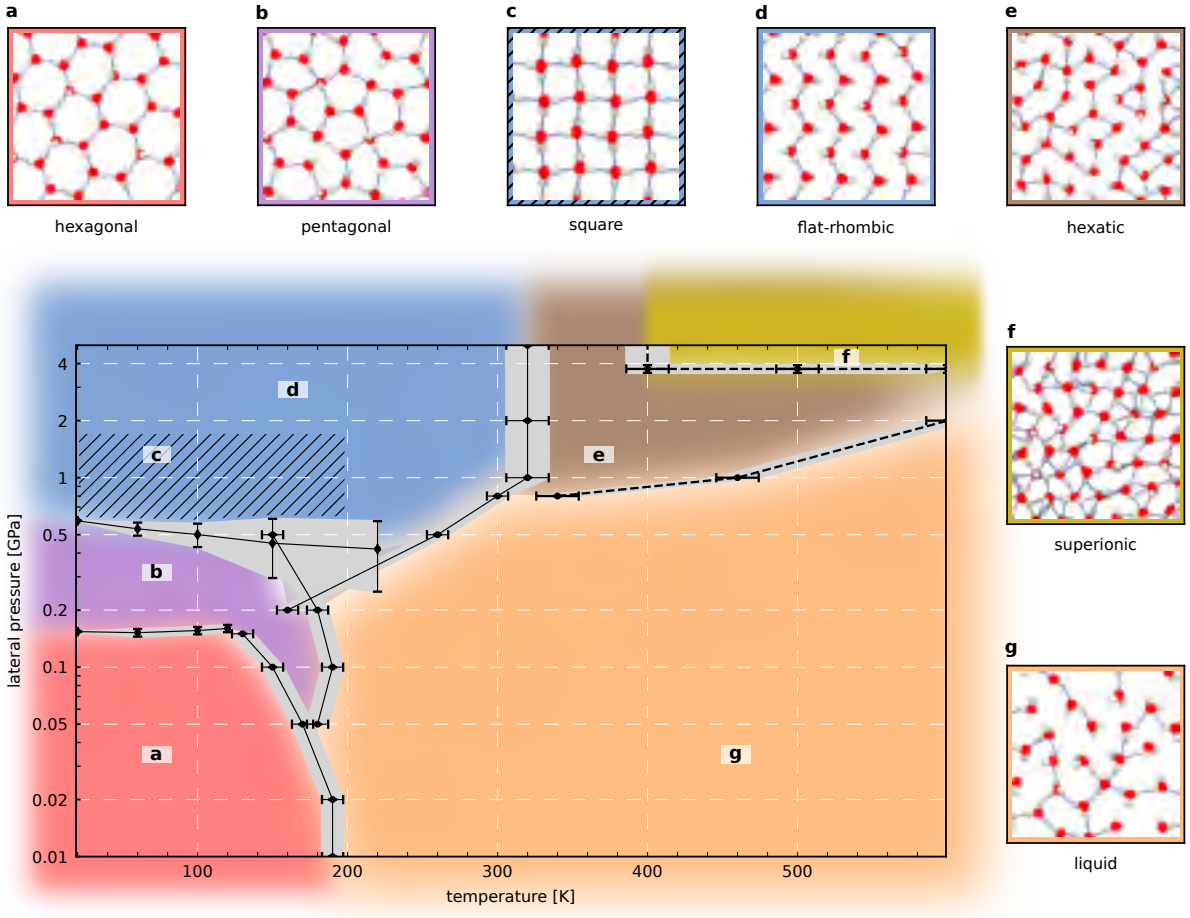
## Simulation techniques

- Random structure search to find stable ordered phases
- Thermodynamic integration to compute free energy
- Coexistence simulations to locate solid-liquid boundaries



# Phase Diagram of Monolayer Confined Water

## Rich phase behavior and unique properties




- Unique “hexatic” phase which is neither liquid nor solid
- Superionic behavior beyond 4 GPa and 400 K

[1] Kapil, V. et al. Nature **2022**, *just accepted*.

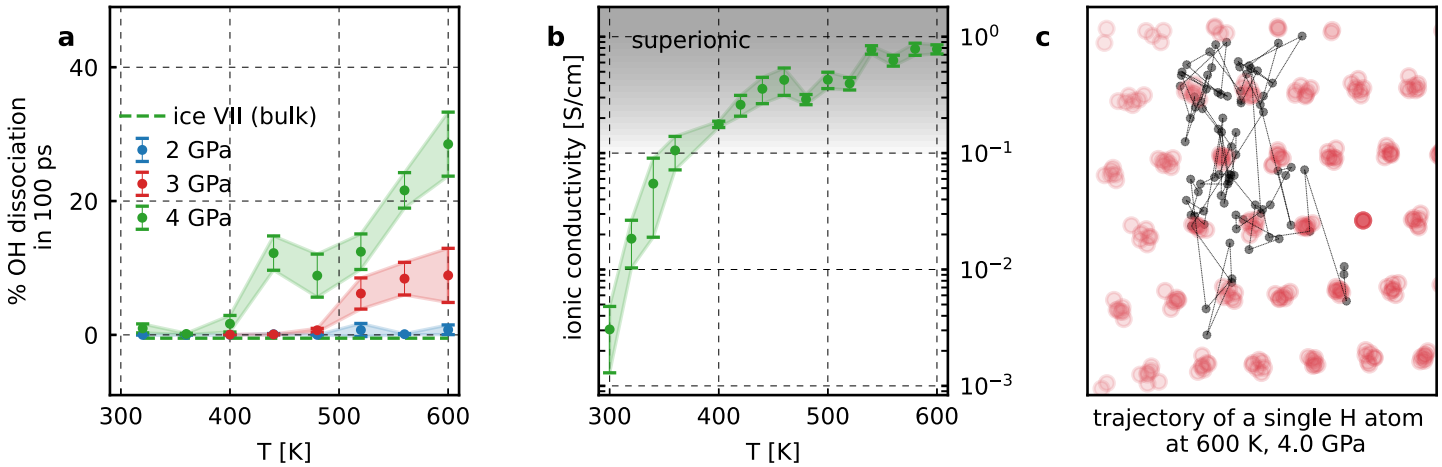
# Phase Diagram of Monolayer Confined Water

## Neither liquid, nor solid

- No translational long range order
- Hexagonal orientational short range order
- Resemblance with free rotor phase of H2

[1] Kapil, V. et al. Nature **2022**, *just accepted*.

# Phase Diagram of Monolayer Confined Water
## Superionic phase




- Pressure increases OH dissociation significantly
- Threshold for superionic conductivity reached at 400 K for 2 GPa
- Fixed lattice of oxygen atoms, diffusing hydrogen atoms

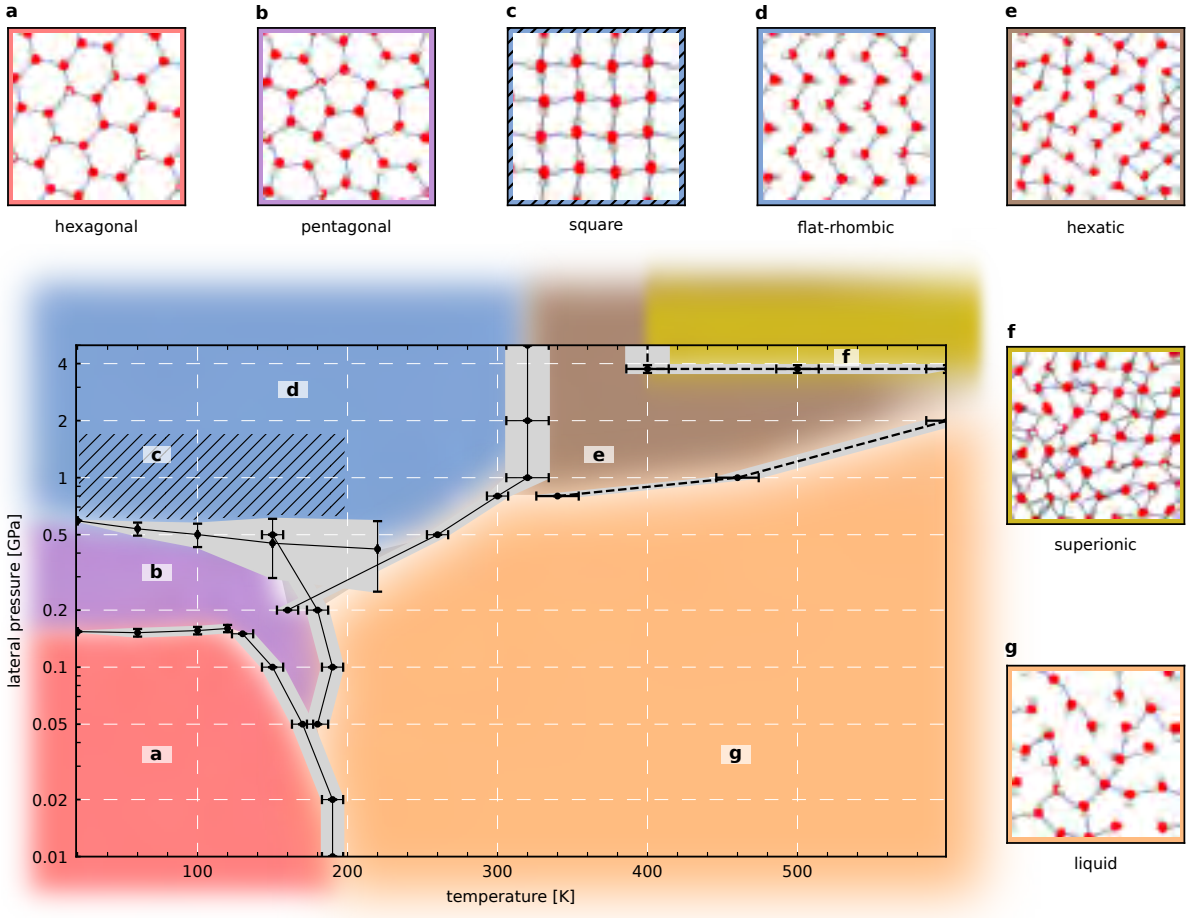
[1] Kapil, V. et al. Nature **2022**, *just accepted*.

# Phase Diagram of Monolayer Confined Water
## Superionic phase

- Facile proton diffusion through fixed oxygen lattice
- No adsorption of protons on explicit carbon validated via AIMD

[1] Kapil, V. et al. Nature **2022**, *just accepted*.

# Summary




- Unique “hexatic” phase which is neither liquid nor solid
- Superionic behavior beyond 4 GPa and 400 K

➡ **Enabled by error control via committee NNPs**

[1] Kapil, V. et al. Nature **2022**, *just accepted*.

# Acknowledgement

## Cambridge/UCL/Imperial

Angelos Michaelides

Erich A. Müller

Venkat Kapil

Patrick Rowe

Fabian Thiemann

Niamh O'Neill

## Bochum

Dominik Marx

Jörg Behler

Fabien Brieuc

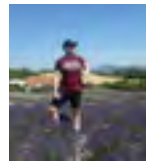
Harald Forbert

Laura Duran

Richard Beckmann

## Prague

Ondrej Marsalek



Krystof Brezina

# Funding


Unterstützt von / Supported by

Alexander von Humboldt Stiftung/Foundation

archer

archer2