

# MACHINE LEARNING BASED *AB-INITIO* MOLECULAR DYNAMICS

Roberto Car

Princeton University

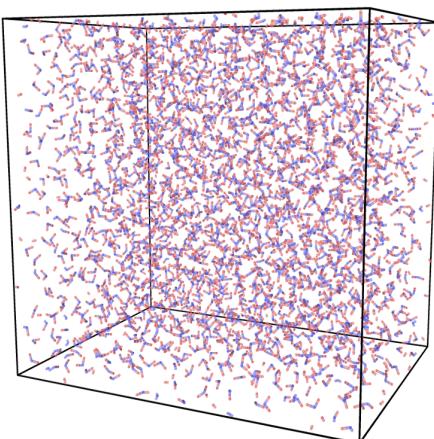


*MaX School on Advanced Materials and Molecular Modelling with Quantum Espresso*  
**ICTP/SISSA**  
May 17<sup>th</sup>, 2021



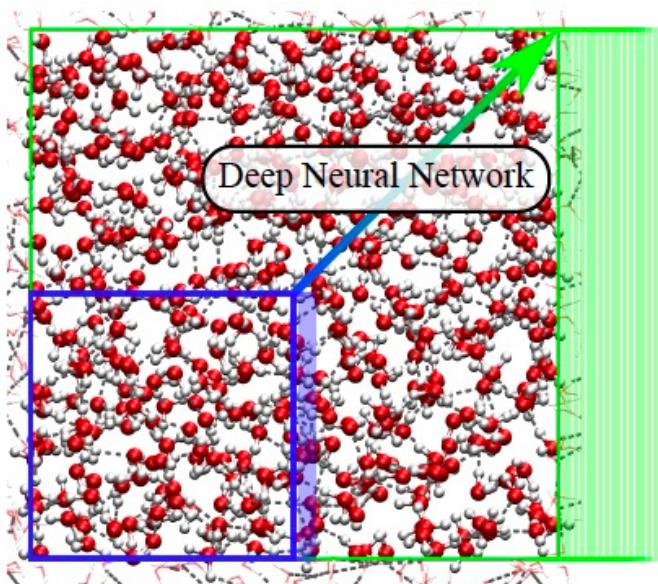
# Molecular Dynamics, *Ab-initio* Molecular Dynamics, and Machine Learning

$$m_k \ddot{\mathbf{r}}_k = -\frac{\partial E(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N)}{\partial \mathbf{r}_k}$$



- *Ab-initio* Molecular Dynamics (AIMD) is MD with forces derived on the fly from the quantum mechanical ground-state of the electrons.
- AIMD is more accurate than MD with empirical force fields (FF), but also more costly.
- By learning the potential energy surface (PES) from quantum mechanics, Machine Learning (ML) methods make possible simulations of AIMD quality at FF cost.
- One such method, the deep potential (DP) method, uses deep neural networks to model the PES and other ground state properties accessible to AIMD. It was developed in the Ph D thesis of **Linfeng Zhang**, directed by **Weinan E** (Math Dept) and myself. Several other people contributed to the effort.

# From AIMD to DPMD: overcoming size and time bottlenecks with Deep Neural Network (DNN) representations



Potential energy, polarization, and polarizability surfaces

$$\mathbf{o}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N) = \sum_i \mathbf{o}_i \quad \mathbf{o}_i = f(\{\mathbf{r}_k \in \mathcal{N}_i\})$$

$f$  is a **symmetry preserving continuous and differentiable** function of the atomic coordinates in environments with variable numbers of atoms.

Distinctive features:

- **Flexibility of DNN**
- **Learning on the fly (DP-GEN)**
- **Computational efficiency and HPC performance**

# Potential energy, polarization and polarizability surfaces

$$E(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N) = \sum_i E_i(\{\mathbf{r}_k \in \mathcal{N}_i\})$$

$$\mathbf{M}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N) = 6e \sum_l \mathbf{r}_{O_l} + e \sum_m \mathbf{r}_{H_m} - 8e \sum_i \mathbf{w}_i (\{\mathbf{r}_k \in \mathcal{N}_i\}) \quad \mathbf{w}_i = \frac{1}{4} \sum_{l_i}^4 \mathbf{w}_{l_i}$$

$$\boldsymbol{\alpha}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N) = \left( \frac{\partial \mathbf{M}}{\partial \boldsymbol{\varepsilon}} \right)_{\mathbf{r}^N} = -8e \sum_i \frac{\partial \mathbf{w}_i}{\partial \boldsymbol{\varepsilon}} = \sum_i \boldsymbol{\alpha}_i (\{\mathbf{r}_k \in \mathcal{N}_i\})$$

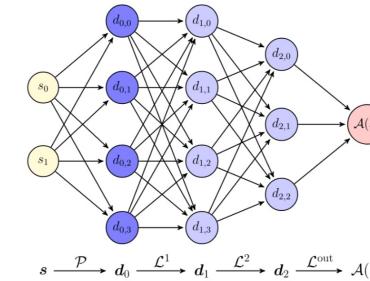
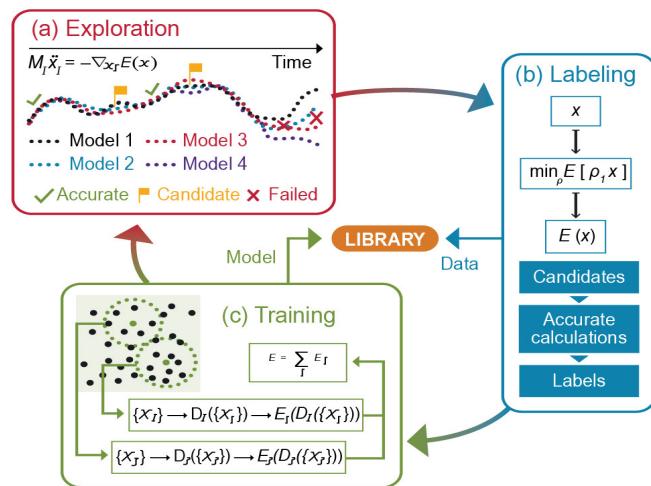


Maximally Localized Wannier functions for the valence electrons in water

L. Zhang *et al*, *Phys Rev Lett* **120**, 143001 (2018); L. Zhang et al., in *Advances in Neural Information Processing Systems* **31**, 4441 (2018); L. Zhang et al., *Phys Rev B* **102**, 041121(R) (2020); G. Sommers, M. Calegari Andrade et al., *PCCP* **22**, 10592 (2020)

# Deep Potential Generator (DP-GEN): an effective scheme for learning

The highly non-linear dependence of a DNN on the network parameters is exploited by the **active learning** procedure DP-GEN



$$\hat{A}_w(s) = \tanh(b_1 + W_1 \tanh(b_2 + W_2 \tanh(b_3 + W_3 \tanh(\dots \tanh(b_D + W_D s))))))$$

**DP-GEN: Labeling when model deviation is maximal (maximal disagreement)**

Typically less than 1/1000 of the visited configurations are labelled

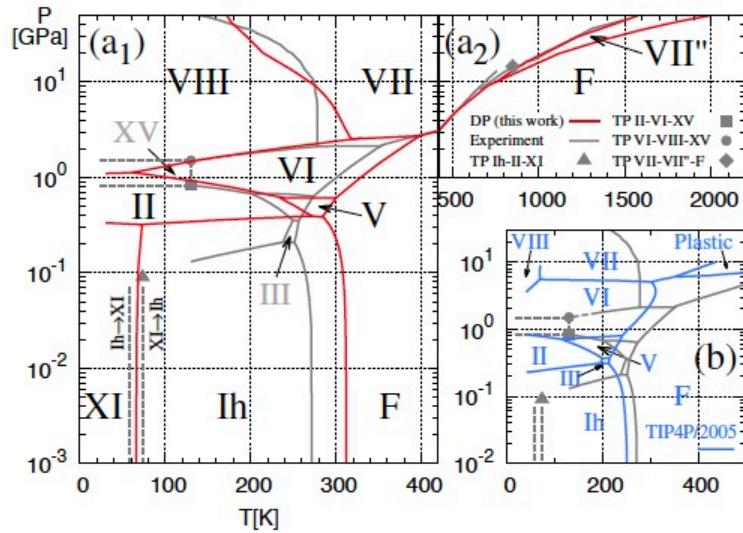
**Single shot** quantum mechanical (density functional theory) calculations are performed for the labelled configurations

**DPMD is a powerful proxy for AIMD: studies beyond the reach of direct AIMD simulations become possible**

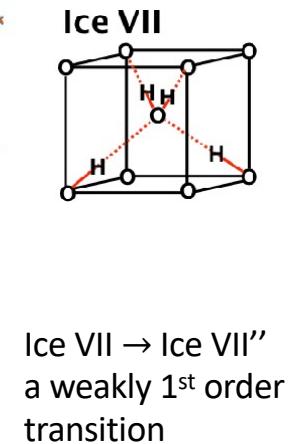
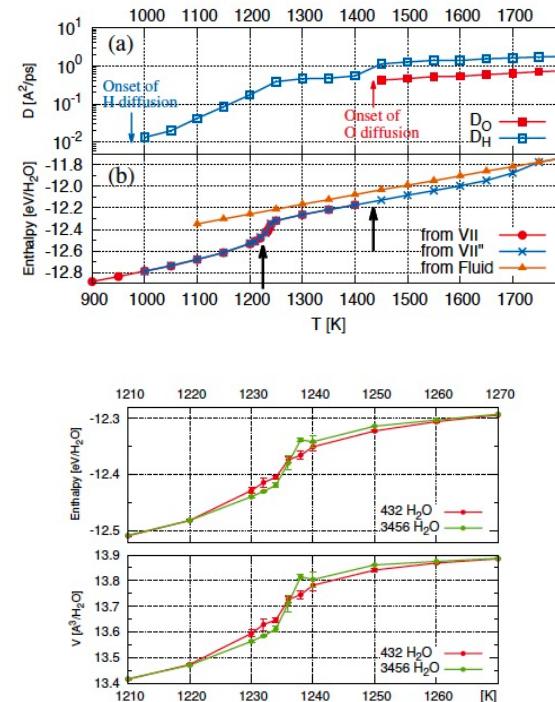
*Examples*

- Phase diagram of a SCAN-based DP model
- Auto-ionization of water
- Homogeneous nucleation of ice
- Vibrational spectra

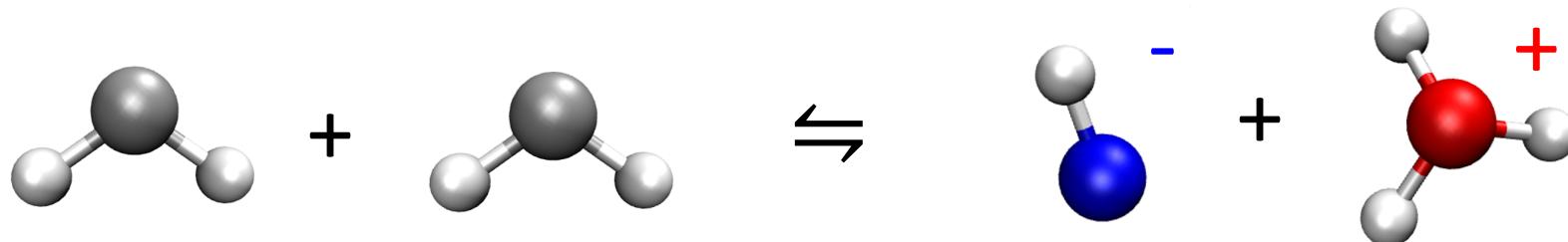
# A DP model for water based on SCAN-DFT



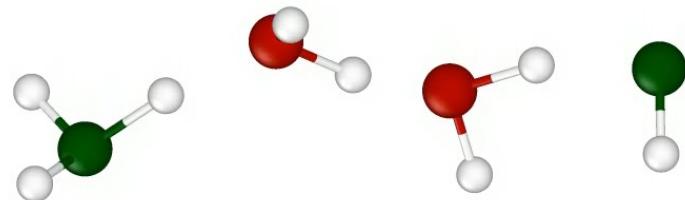
< 35,000 DFT minimizations to construct the PES, i.e. 0.05 %  
of the total configurations visited with DP-GEN  
Free energy error is approx. 1 meV/mol



# Water Self-Ionization (autoprotolysis)

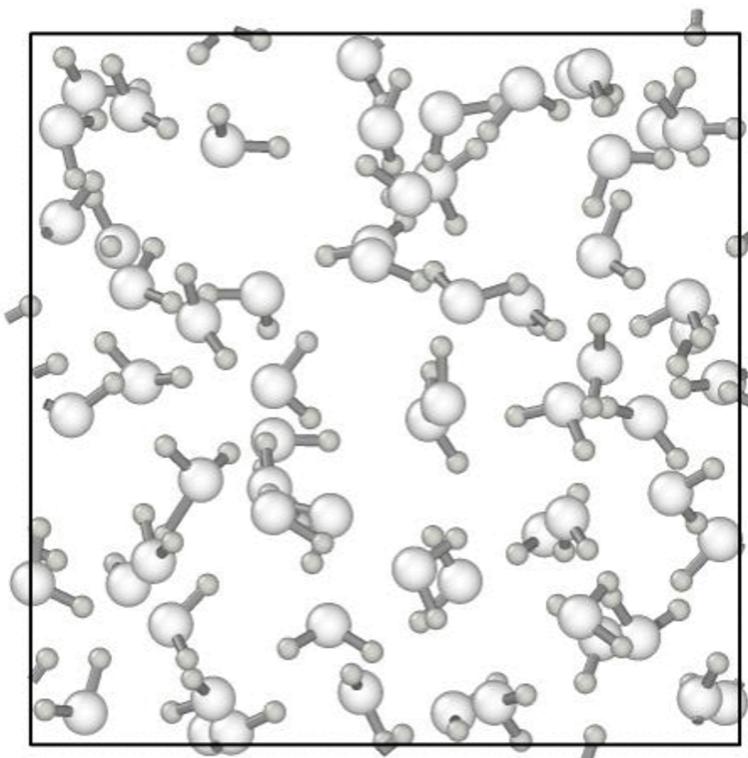


**Two in 1 billion** water molecules appear as ions (hydroxide and hydronium) at ambient conditions.  $K_W = [\text{H}_3\text{O}^+][\text{OH}^-]$



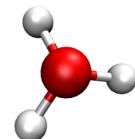
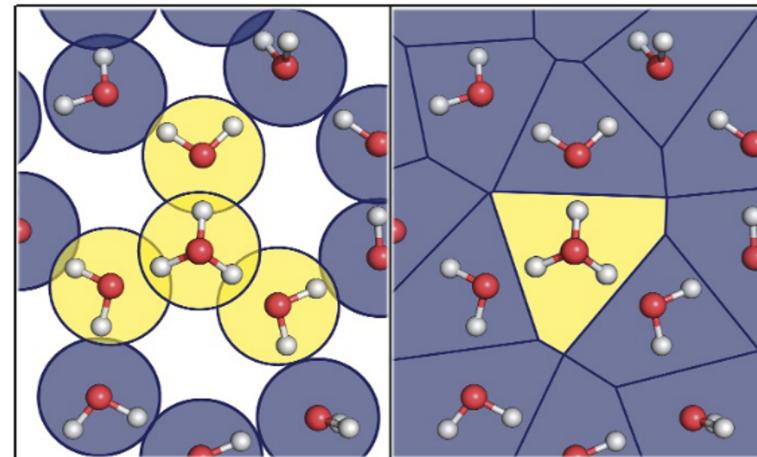
Recombination can be observed in AIMD simulations when the ions are close enough, but the spontaneous ionization requires enhanced sampling methods, such as metadynamics.

Animation from Chunyi Zhang (Temple U.): it shows a recombination process as first described in Hassanali A, Prakash M K, Eshet H, et al. Proceedings of the National Academy of Sciences, 2011, 108(51): 20410-20415

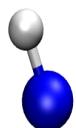


Self-ionization is facilitated by a bias potential depending on a reaction coordinate  $S$  in metadynamics simulations

Hydronium and Hydroxide sites are identified by Voronoi tessellation following Grifoni, Piccini, & Parrinello, *Proc. Natl. Acad. Sci.* **116**, 4054–4057 (2019)



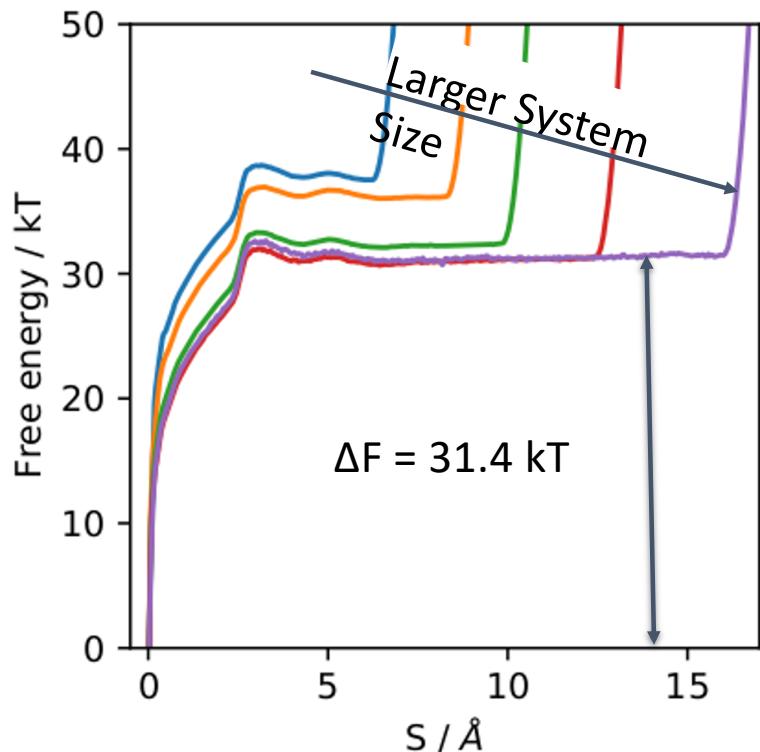
Hydronium



Hydroxide

## Water Auto-ionization

$$pK_W = -\log K_W = pH + pOH$$

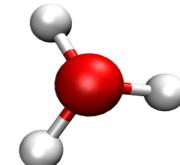
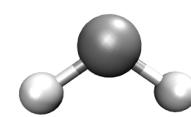


Heavy water at 330 K, ambient pressure

Simulation:  $pK_w = 13.6$  (classical DPMD)

Experiment:  $pK_w = 13.9$

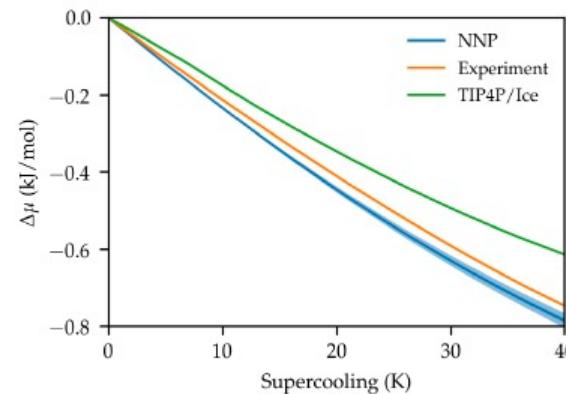
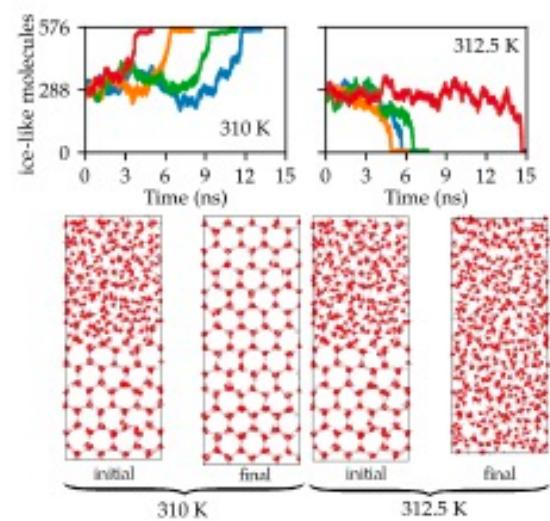
$$S = - \sum_{i>j} \delta_i \delta_j d_{ij}$$



M. Callegari Andrade et al., unpublished (2021)

## Nucleation (CNT) from seeding simulations

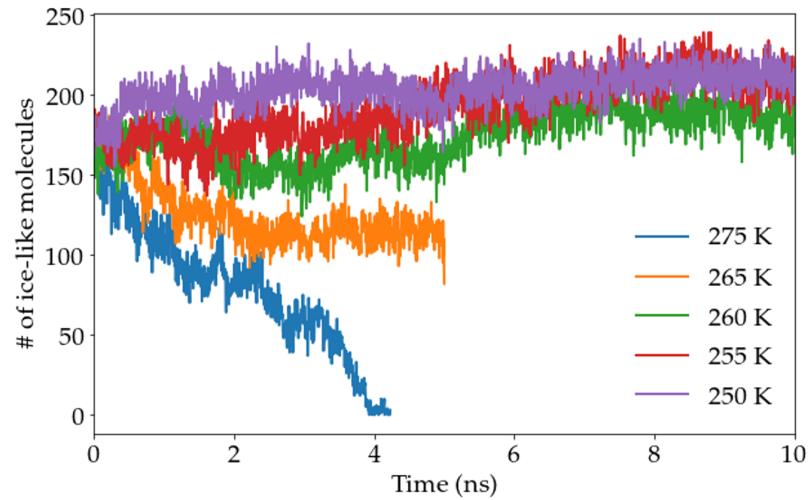
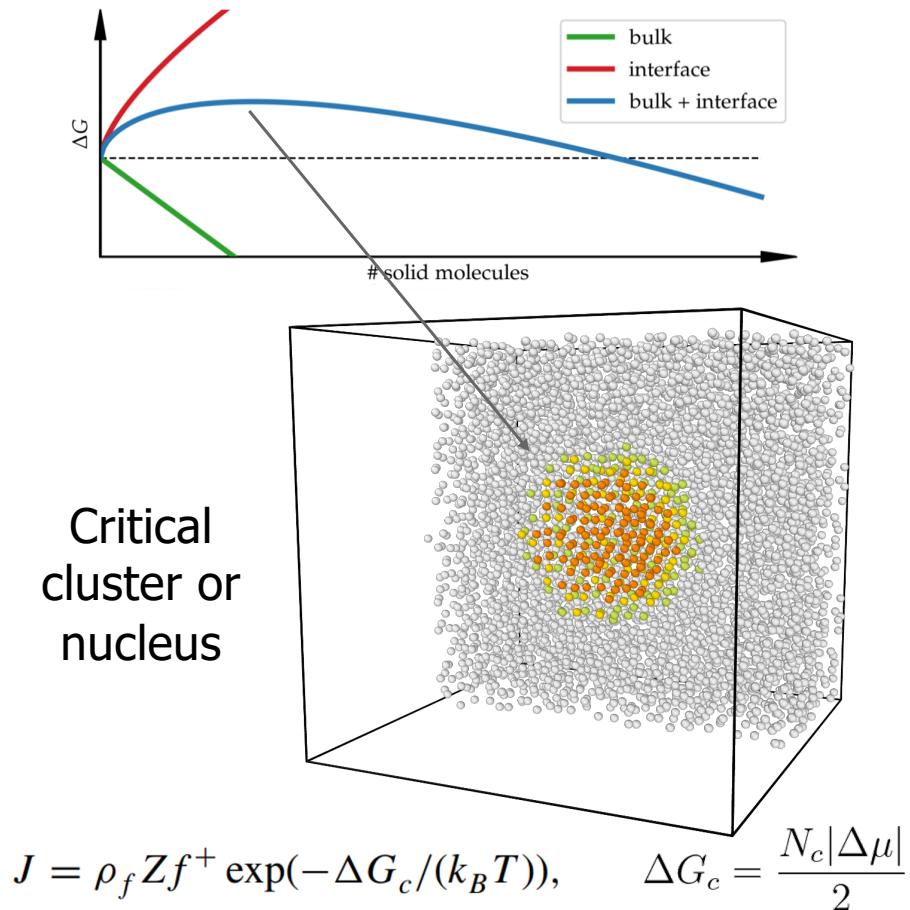
First, obtain the difference in chemical potential between solid and liquid from enhanced sampling simulations for a SCAN-based DP



Driving force for nucleation

P.M. Piaggi, A.Z. Panagiotopoulos, P.G. Debenedetti, R. Car, arXiv:2101.04806v1, and *J. Chem. Theory Comp.*, in press (2021), following techniques detailed in P. Piaggi and R.C., *J. Chem. Phys.* **152**, 204116 (2020)

# Seeding simulations



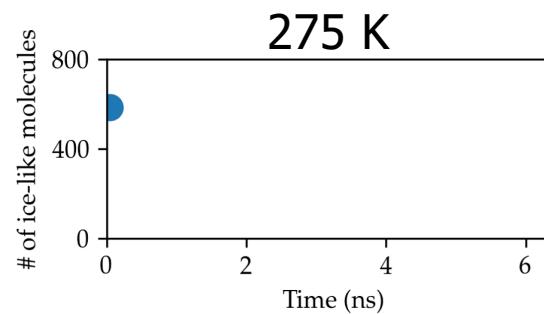
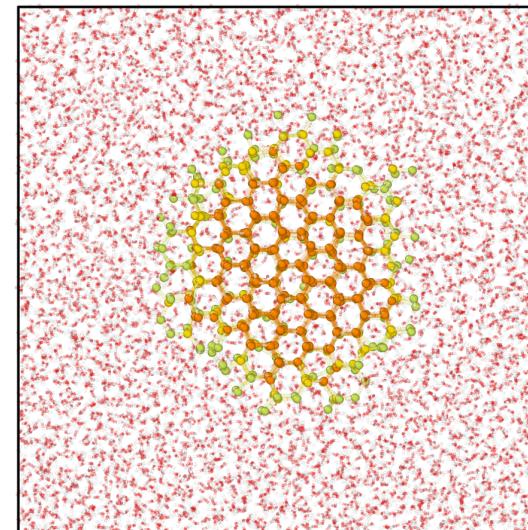
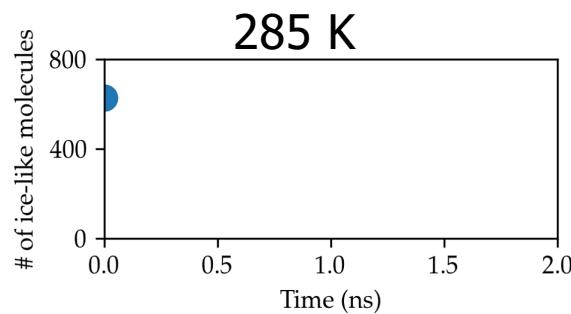
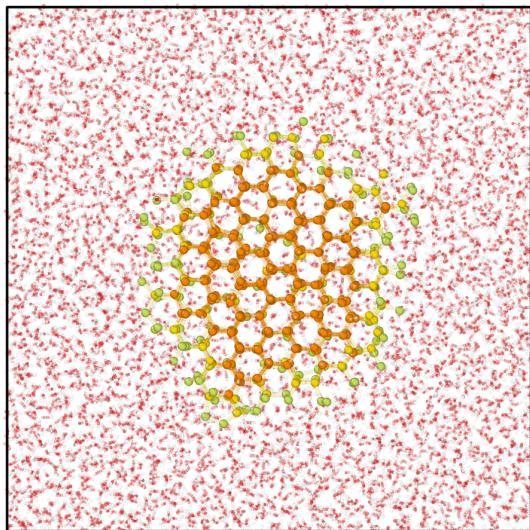
$$\Delta G = \frac{4}{3} \pi R^3 \rho_s \Delta \mu + 4 \pi R^2 \gamma$$

Extract  $T_m$   $\Delta \mu$   $N_c$   $f^+$  from simulations, as detailed in J.R. Espinosa et al., *J. Chem. Phys.* **141**, 18C529 (2014)

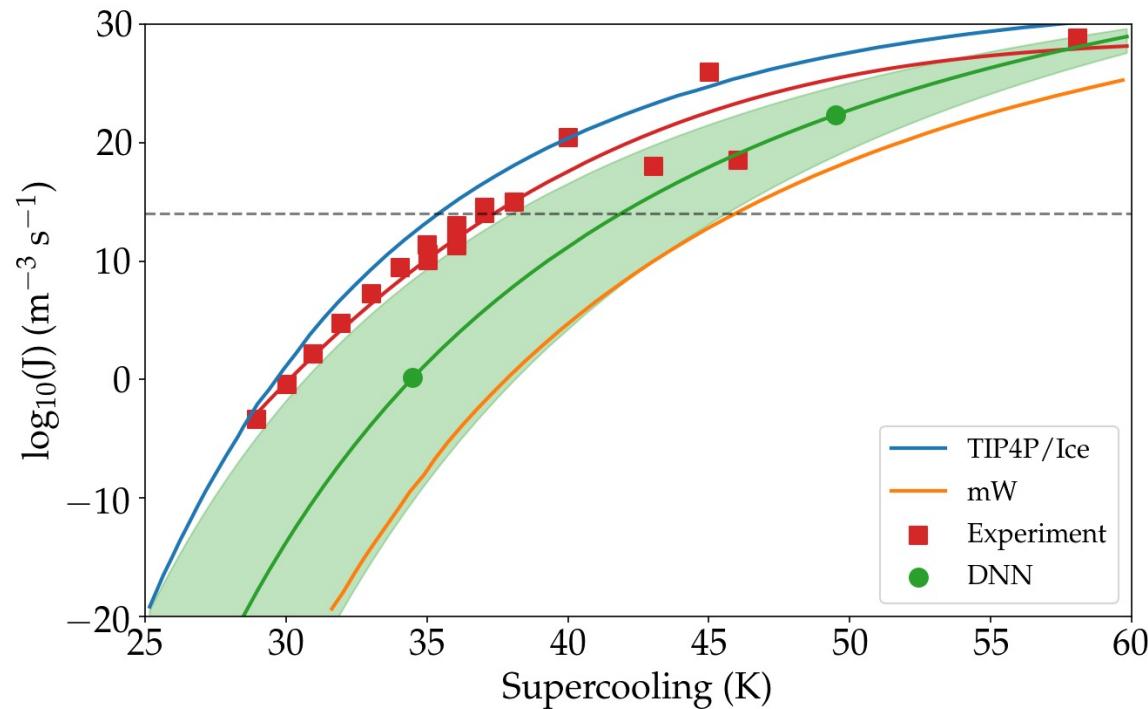
$\rho_f Z f^+$  is the kinetic prefactor,  $Z(\Delta \mu, T, N_c)$  is the Zeldovich factor

# Seeding simulations

~ 12'000 H<sub>2</sub>O molecules in total  
~ 600 H<sub>2</sub>O molecules in cluster



# Simulations Rates (preliminary results)



Ab-initio results for  
nucleation rate and surface  
free energy are in the right  
ballpark.

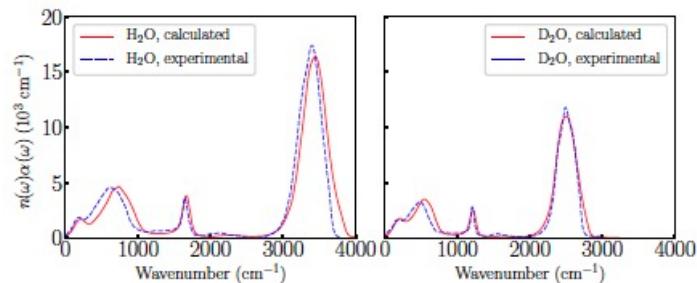
From P. Piaggi, et al. (2021)

# Vibrational spectroscopies (Infrared and Raman)

Raman spectra probing H-bonding modes in liquid water

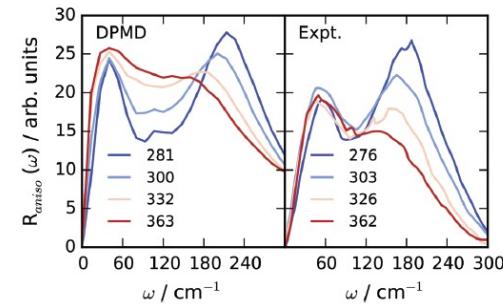
Infrared absorption of liquid water

$$\alpha(\omega)n(\omega) = \frac{2\pi\beta}{3cV} \int_{-\infty}^{+\infty} dt e^{-i\omega t} \langle \dot{\mathbf{M}}(0) \cdot \dot{\mathbf{M}}(t) \rangle$$



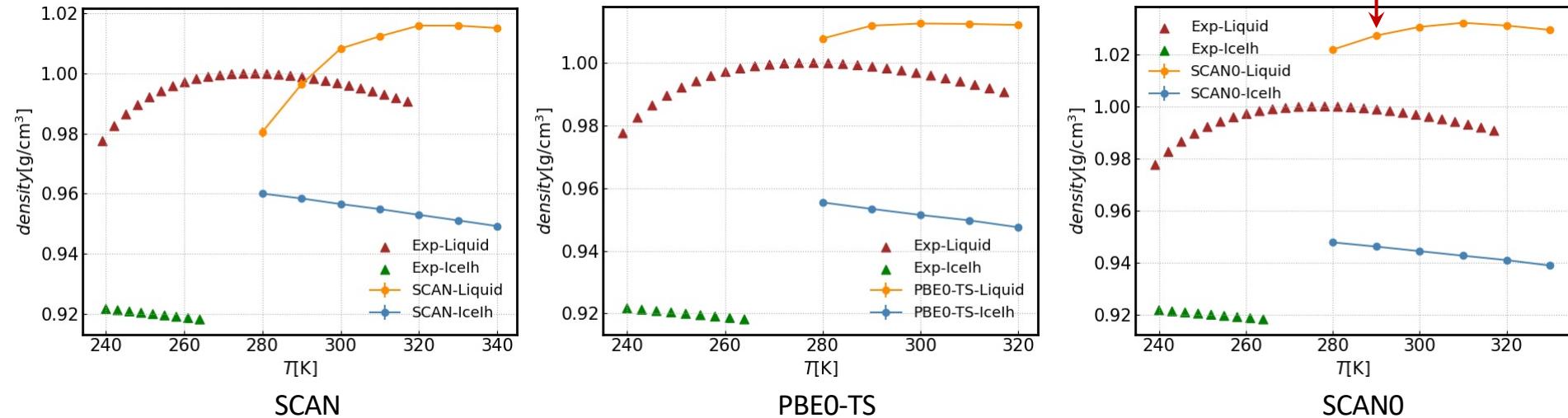
L. Zhang et al. (2020)

$$\beta = \alpha - 1\bar{\alpha} \quad R_{\text{aniso}}(\omega) \propto n_{\text{BE}}(\omega) \int_{-\infty}^{\infty} dt e^{-i\omega t} \frac{2}{15} \text{Tr} \langle \beta(0) \beta(t) \rangle$$



G. Sommers et al. (2020)

## Effects of the DFT approximation



**Fig. 1.** Density isobars for liquid water and ice Ih at  $P = 1$  bar computed via classical DPMD simulations at the indicated DFT levels. The experimental results were obtained from [1], [2] and [3].

- 
- [1] Eric W. Lemmon, Mark O. McLinden and Daniel G. Friend, "Thermophysical Properties of Fluid Systems" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds. P.J. Linstrom and W.G. Mallard, National Institute of Standards and Technology, Gaithersburg MD, 20899, <https://doi.org/10.18434/T4D303>, (retrieved May 14, 2021).
- [2] Hare, D. E., and C. M. Sorensen. "The density of supercooled water. II. Bulk samples cooled to the homogeneous nucleation limit." *The Journal of chemical physics* 87.8 (1987): 4840-4845.
- [3] Röttger, K., et al. "Lattice constants and thermal expansion of H<sub>2</sub>O and D<sub>2</sub>O ice Ih between 10 and 265 K." *Acta Crystallographica Section B: Structural Science* 50.6 (1994): 644-648.

## Large-Scale Hybrid DFT using $\alpha$ -exx: Current Capabilities

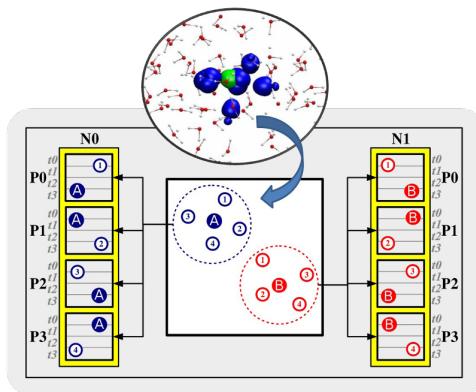
$O(N)$  algorithm via two levels of sparsity

$$E_{\text{xx}} = - \sum_{\langle ij \rangle} \int_{\Omega_{ij}} d\mathbf{r} \tilde{\rho}_{ij}(\mathbf{r}) \tilde{v}_{ij}(\mathbf{r})$$

Linear number of overlapping MLWF pairs

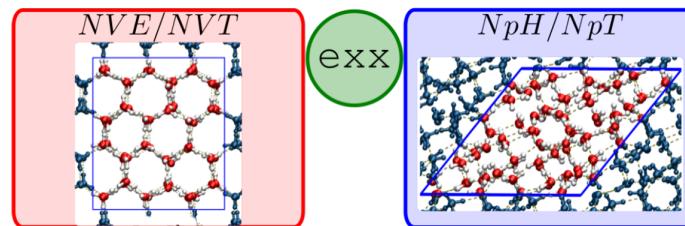
Compact system-size-independent Poisson domains

Massively parallel MPI/OpenMP implementation available in Quantum ESPRESSO

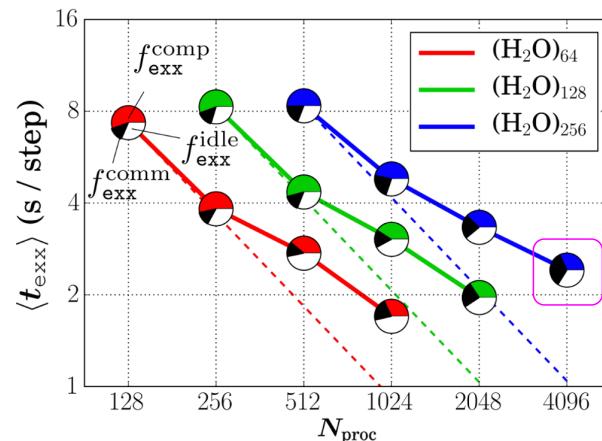


Ko, Jia, Santra, Wu, Car, DiStasio, *J. Chem. Theory Comput.* (2020).  
Ko, Santra, DiStasio, *J. Chem. Theory Comput.* (2021).

Support for general cells & analytical cell derivatives enable EXX-CPMD in *NVE/NVT* & *NpH/NpT* ensembles



EXX-CPMD for  $(\text{H}_2\text{O})_{256}$  in < 2.4 s/step—split between computation, communication, and idling



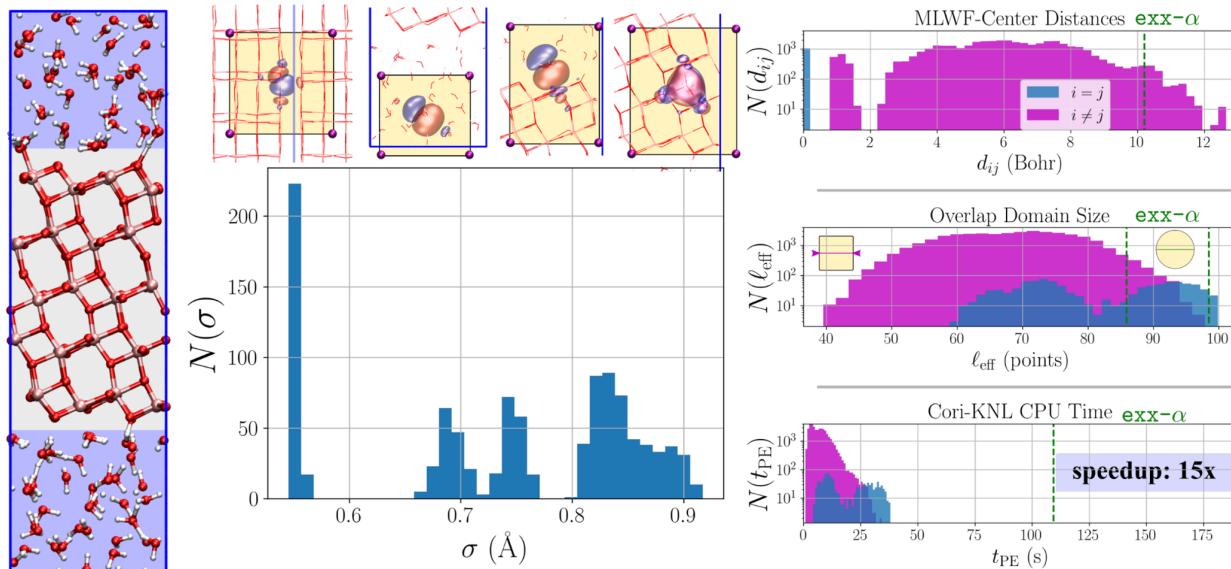
# Large-Scale Hybrid DFT using $\beta$ -exx: Near-Future Capabilities

Designed to treat anisotropic & heterogeneous systems (*e.g.*, interfaces, multi-component/multi-phase systems) without compromising accuracy for efficiency

Comprehensive reimplementation with improvements to computation, communication, & idling:

- **6x speedup** for liquid H<sub>2</sub>O
- **15x speedup** for TiO<sub>2</sub>–H<sub>2</sub>O interface
- **30x speedup** for metal-doped molten salt (bipolaron)

Includes advanced vectorization, GPU support, asynchronous MPI, ...



# Conclusions and Outlook

- DPMD is several order of magnitude more efficient than direct AIMD and has linear scaling with size, opening the way to studies well beyond the reach of AIMD (complex phase behavior, reactions in solution, nucleation, dynamic response...)
- DP is a proxy for DFT, quantification of the model deviation from the reference is important. Algorithms to achieve this goal are crucial (e.g., spread of a network ensemble, re-weighting of DP data, etc.)
- More properties are accessible to first-principles calculation: check DFT approximations
- Issues: long range electrostatics, quantum corrections to statics, quantum corrections to dynamics and spectroscopy, chemical accuracy (DFT and beyond)...

# Acknowledgements



<http://chemlabs.princeton.edu/ccsc/>