# Deep Potential Molecular Dynamics on SUMMIT

**Roberto Car**

**Princeton University**

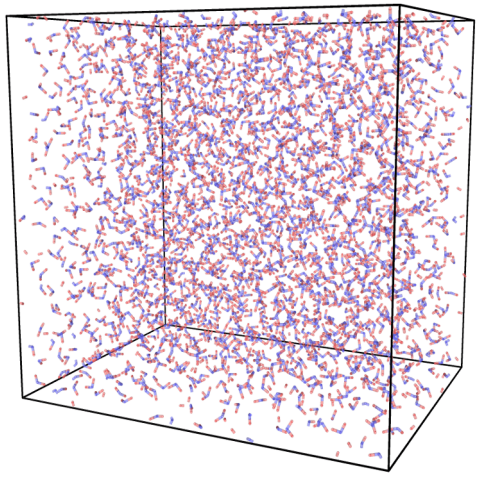

***OLCF User Meeting*** (June 22-24) Panel Discussion (June 22)

U.S. DEPARTMENT OF ENERGY | Office of Science

**Chemistry in Solution and at Interfaces**
*Computational Chemical Science Center*

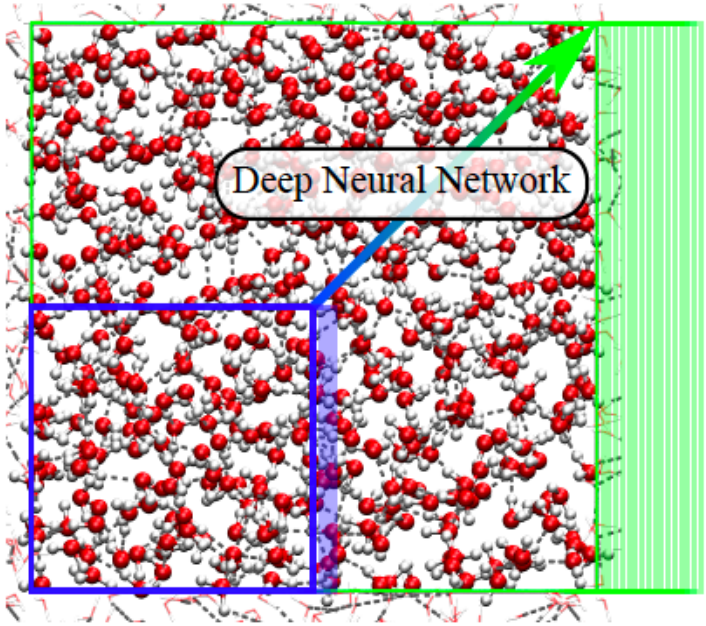
# Molecular Dynamics, *Ab-initio* Molecular Dynamics, and Machine Learning

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



- *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), it can describe chemical bond breaking/forming, but it is also much 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

$$\boldsymbol{O}(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3, \ldots, \boldsymbol{r}_N) = \sum_i \boldsymbol{o}_i \qquad \boldsymbol{o}_i = \boldsymbol{f}(\{\boldsymbol{r}_k \in \mathcal{N}_i\})$$

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

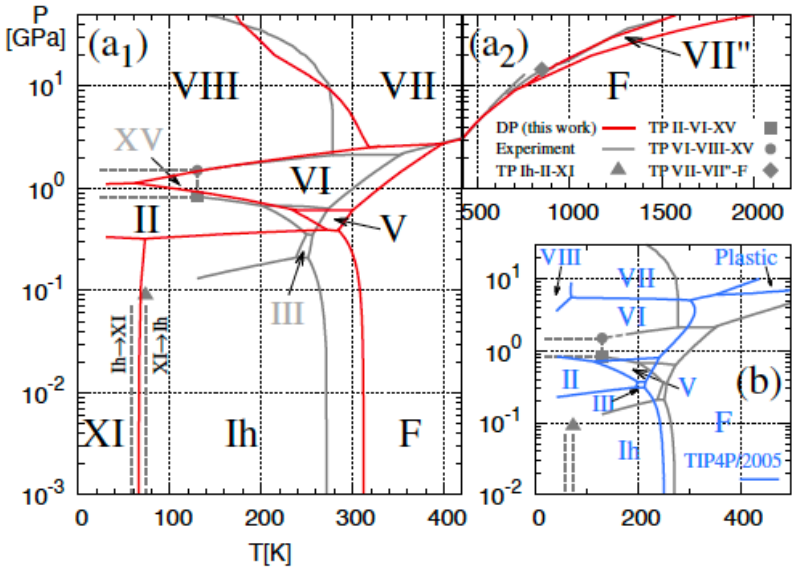
Distinctive features:

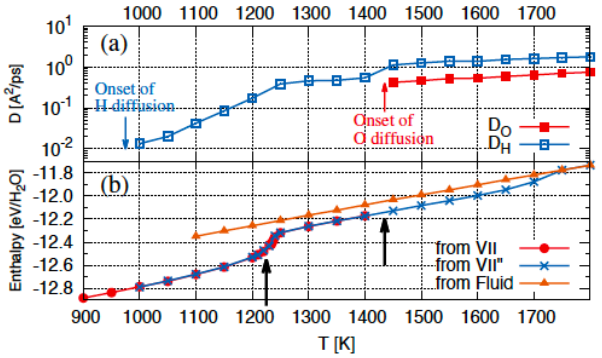
- **Flexibility of DNN**
- **Learning on the fly (DP-GEN)**
- **Computational efficiency and HPC performance**

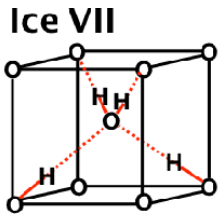
The DP representation uses 2 DNN: (1) the **embedding network** needed to construct the **feature matrix**, which is a general descriptor of the environment, and (2) the actual **fitting network** for the physical property of interest

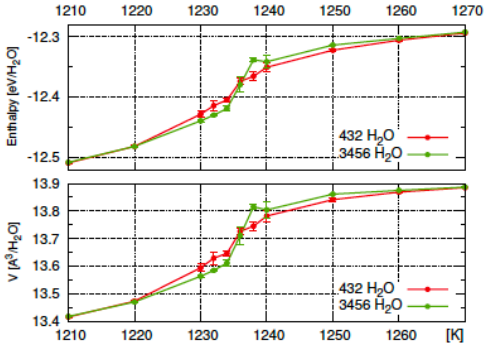
L. Zhang *et al*, *Phys Rev Lett* **120**, 143001 (2018); L. Zhang et al., in *Advances in Neural Information Processing Systems* **31**, 4441 (2018)

# A DP model for water based on SCAN-DFT




Ice VII → Ice VII''
a weakly 1st order transition

< 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

L. Zhang, H. Wang, R.C., W.E, *Phys Rev Lett* **126**, 236001 (2021)

# Pushing the limit of molecular dynamics with *ab initio* accuracy to 100 million atoms with machine learning

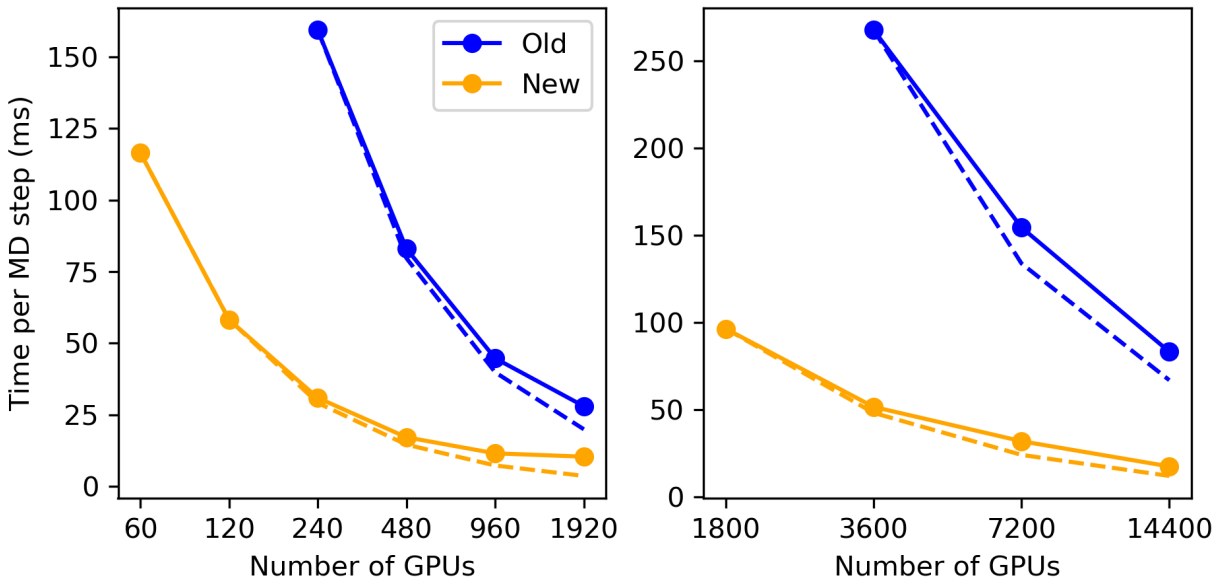
Weile Jia*, Han Wang†, Mohan Chen‡, Denghui Lu‡, Lin Lin*¶, Roberto Car‖, Weinan E‖, Linfeng Zhang‖ §
*University of California, Berkeley, Berkeley, USA
Email: jiaweile@berkeley.edu, linlin@math.berkeley.edu
† Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics, Beijing, China
Email: wang_han@iapcm.ac.cn
‡CAPT, HEDPS,College of Engineering, Peking University, Beijing, China
Email: mohanchen@pku.edu.cn, denghuilu@pku.edu.cn
¶Lawrence Berkeley National Laboratory, Berkeley, USA
‖ Princeton University, Princeton, USA
Email: rcar@princeton.edu, weinan@math.princeton.edu,linfengz@princeton.edu

**2020 Gordon Bell Prize** of the ACM at the SC20 Conference

Performance of Deep Potential Molecular Dynamics (DPMD) on Summit against other MD methods with *ab-initio* (DFT) accuracy

| Work | Year | Pot. | System | #atoms | #CPU cores | #GPUs | Machine | Peak[FLOPS] | TtS [s/step/atom] |
|---|---|---|---|---|---|---|---|---|---|
| Qbox [12] | 2006 | DFT | Mo | 1K | 262K | – | BlueGene/L | 207T | $2.8 \times 10^{-1}$ |
| LS3DF [14] | 2008 | LS-DFT | ZnTeO | 16K | 131K | – | BlueGene/P | 108T | $1.8 \times 10^{-2}$ |
| RSDFT [46] | 2011 | DFT | Si | 107K | 442K | – | K-computer | 3.1P | $2.6 \times 10^{0}$ |
| DFT-FE [13] | 2019 | DFT | Mg | 11K | 159K | 22.8K | Summit | 46P | $6.5 \times 10^{-2}$ |
| CONQUEST [17] | 2020 | LS-DFT | Si | 1M | 200K | – | K-computer | ? | $4.0 \times 10^{-3}$ |
| Simple-NN [47]* | 2019 | BP | $SiO_2$ | 14K | 80 | – | Unknown‡ | ? | $3.6 \times 10^{-5}$ |
| Singraber el.al. [48]* | 2019 | BP | $H_2O$ | 9K | 512 | – | VSC† | ? | $1.3 \times 10^{-6}$ |
| Baseline [45]** | 2018 | DP | $H_2O$ | 25K | 1 | 1 | Summit | – | $5.6 \times 10^{-5}$ |
| This work (double) | 2020 | DP | $H_2O$ | 679M | 27.3K | 27.3K | Summit | 80P | $3.0 \times 10^{-10}$ |
| This work (mixed-half) | 2020 | DP | $H_2O$ | 679M | 27.3K | 27.3K | Summit | 212P | $1.1 \times 10^{-10}$ |
| This work (double) | 2020 | DP | Cu | 127M | 27.3K | 27.3K | Summit | 91P | $8.1 \times 10^{-10}$ |
| This work (mixed-half) | 2020 | DP | Cu | 127M | 27.3K | 27.3K | Summit | 275P | $2.7 \times 10^{-10}$ |

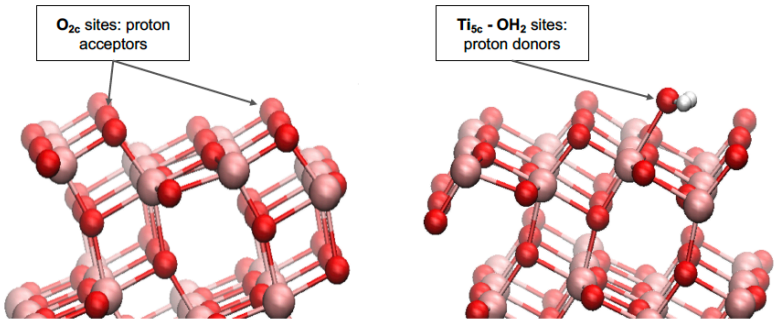
# A further optimization: compressed DPMD



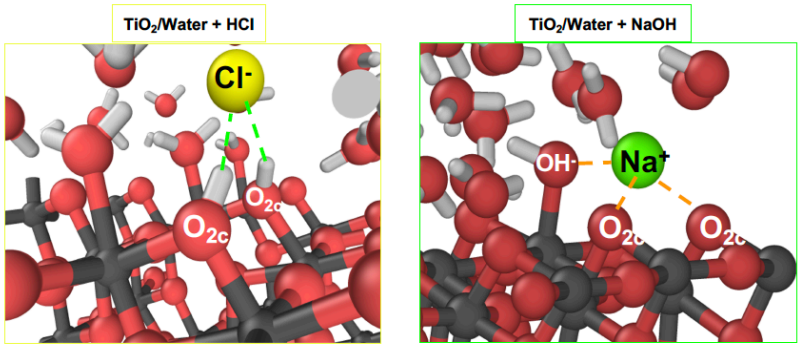

*Strong scaling on Summit using new (compressed) and old (non-compressed) DPMD code. Continuous and dashed lines represent the actual and ideal scaling, respectively. Left: 3 million atoms system (~900,000 water molecules and ~300,000 $TiO_2$ atoms). Right: 68 million atoms (~20.6 million water molecules and ~6 million TiO2 atoms).*

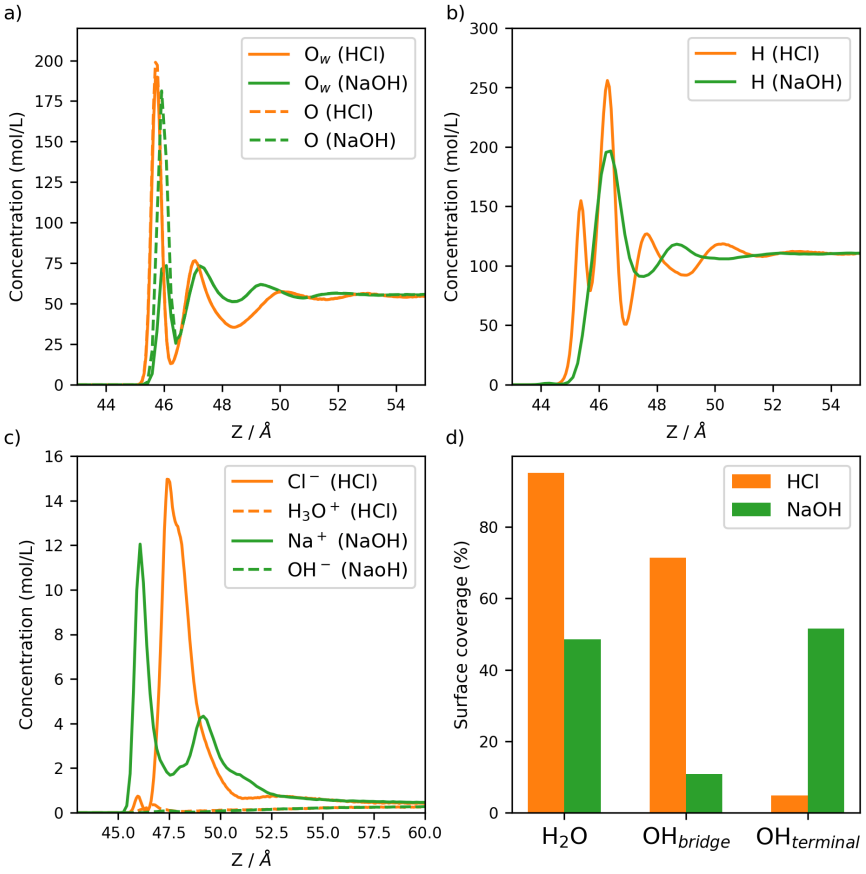
# Acidic and basic electrolytic solutions in contact with $TiO_2$

Acid/base sites on the $TiO_2$ anatase (101) surface



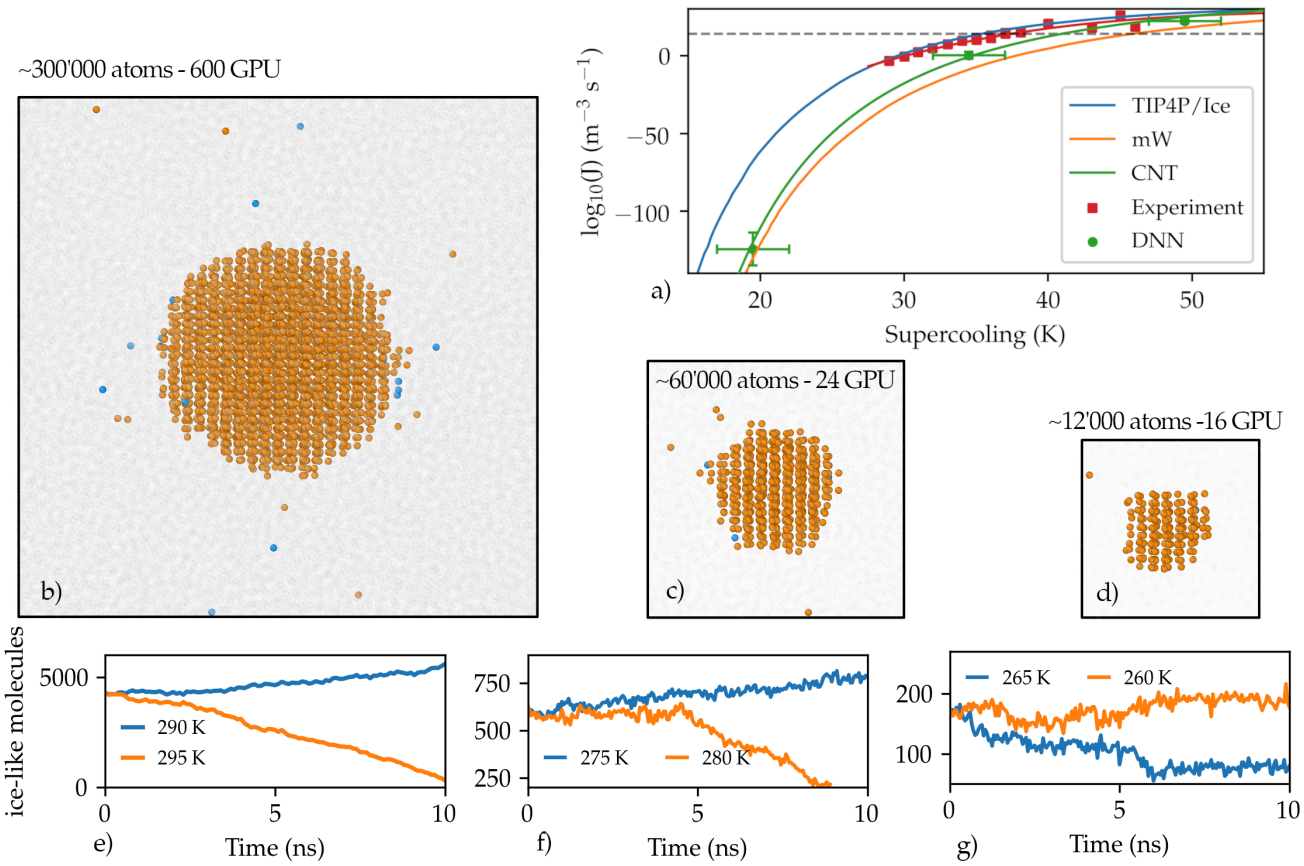

Ion solvation at the $TiO_2$/Water interface

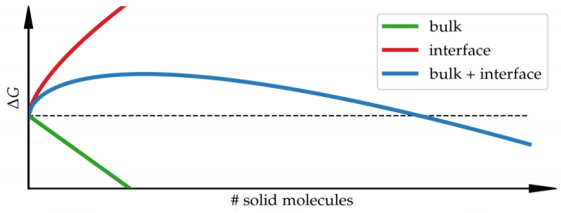



From M. Calegari Andrade et al., (2021)

# Homogeneous ice nucleation with the seeding technique




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

$$J = \rho_f Z f^+ \exp(-\Delta G_c/(k_B T)),$$

Homogeneous nucleation limit:
40-41 K (predicted from simulation)
38 K (experiment)

From P. Piaggi et al., (2021)

# Conclusions and Outlook

- Long term goals in the two studies that were mentioned: (a) model charge distribution in electrolyte in contact with reactive interface under applied bias; (b) model heterogeneous nucleation (e.g. ice nucleation in the atmosphere)
- New algorithmic implementation/optimization will be necessary
- Properties that were previously out of reach become accessible to first-principles calculations, providing a new test ground for quantum mechanical approximations (e.g. DFT)

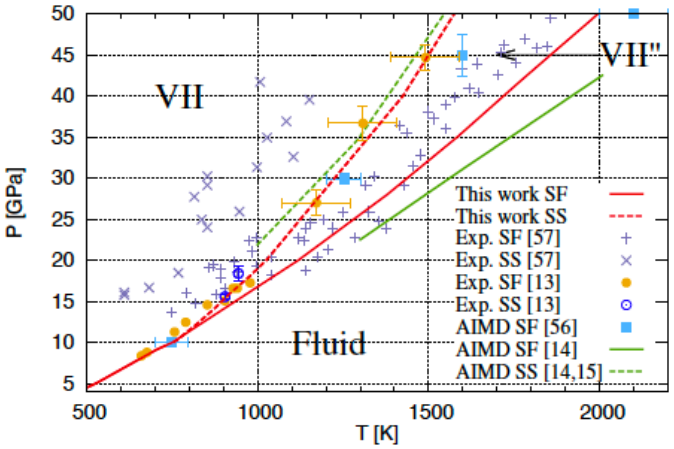
# Acknowledgements

U.S. DEPARTMENT OF ENERGY | Office of Science

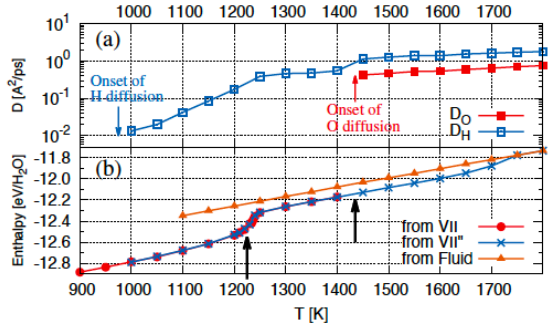
Chemistry in Solution and at Interfaces
Computational Chemical Science Center

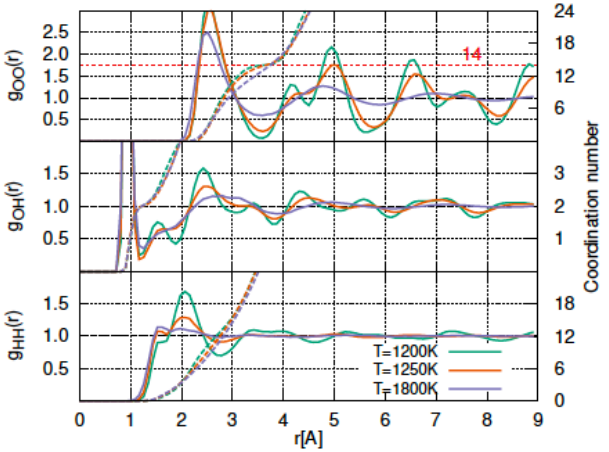
http://chemlabs.princeton.edu/ccsc/

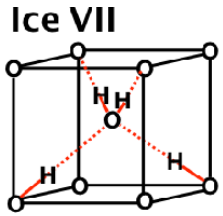
# Temperature induced molecular to ionic transition at high pressure




Melting of ice VII is a 2 step process




Strong covalent fluctuations persist in spite of ice rule breaking

Ionic defect (hydronium – hydroxide like) population is approx. 11 % in the fluid at 1450 K