Deep Learning for Multiscale Molecular Modeling

Linfeng Zhang

Princeton University

June 19 2019, MoD-PMI2019, NIFS

Joint work with **Han Wang**, Roberto Car, Weinan E

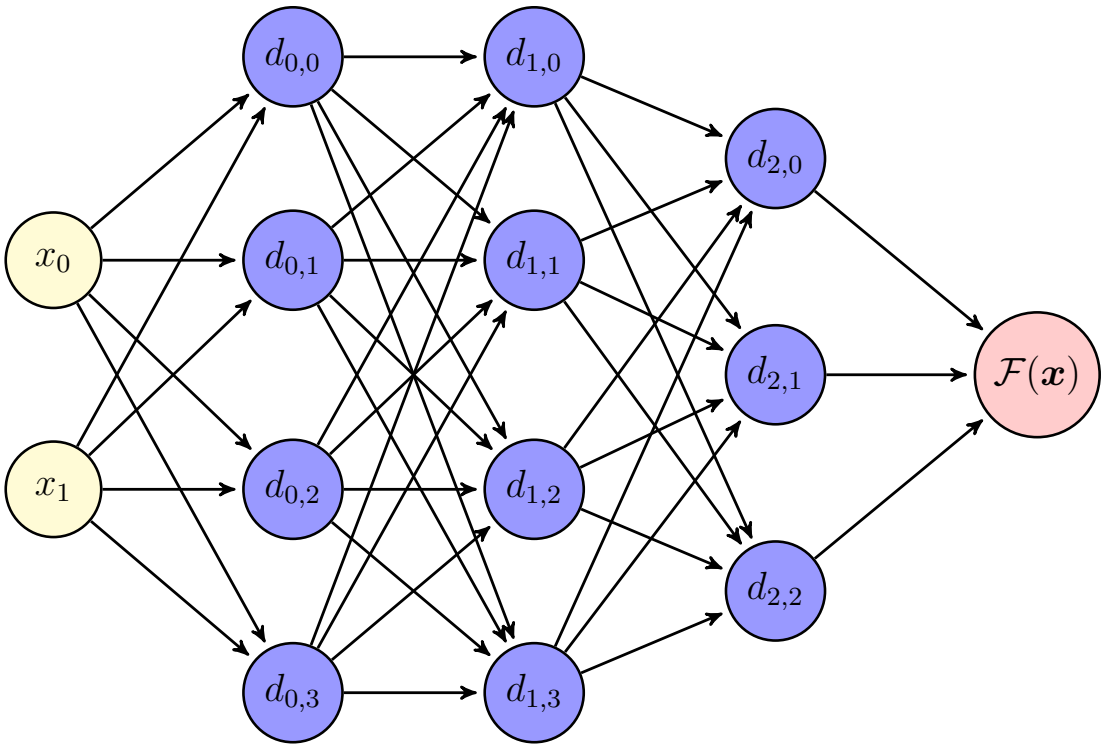
Outline

1. Introduction
2. Deep Potential
3. Deep Potential Generator (DP-GEN)
4. Free energy and Reinforced Dynamics
5. Conclusions

Outline

1. Introduction
2. Deep Potential
3. Deep Potential Generator (DP-GEN)
4. Free energy and Reinforced Dynamics
5. Conclusions

Where deep learning could help?



$$\boldsymbol{d}^p = \mathcal{L}^p(\boldsymbol{d}^{p-1}) = \phi\Big(\boldsymbol{W}^p \cdot \boldsymbol{d}^{p-1} + \boldsymbol{b}^p\Big)$$

Composition of analytical and nonlinear functions; Approximator for High-D functions.

Multi-scale Molecular Modeling

A few examples:

- *ab initio* molecular dynamics (MD):
  quantum mechanics (QM) to MD, potential energy surface (PES);
- Coarse-grained (CG) MD:
  atoms to CG "particles", free energy surface (FES)/CG potential;
- enhanced sampling/phase transition:
  atoms to fewer collective variables (CVs), FES.

Accuracy v.s. efficiency dilemma

PES as an example:

$$E = E(\boldsymbol{r}_1, ..., \boldsymbol{r}_i, ..., \boldsymbol{r}_N).$$

- First principle: accurate but very expensive.
  For example KS-DFT, $\sim 10^2$ atoms:
  $$E = \langle \Psi_0 | H_e^{KS} | \Psi_0 \rangle,$$

- Empirical potentials: fast but limited accuracy.
  For example Lennard-Jones potential
  $$E = \frac{1}{2} \sum_{i \neq j} V_{ij}, \quad V_{ij} = 4\epsilon[(\frac{\sigma}{r_{ij}})^{12} - (\frac{\sigma}{r_{ij}})^{6}].$$

Lennard-Jones, J. E. (1924), Proc. R. Soc. Lond. A, 106 (738): 463477

Two important aspects

Deep learning could help for a classical of problems in multi-scale molecular modeling.

$$\min_{\boldsymbol{w}} \frac{1}{\|\mathcal{D}\|} \sum_{i \in \mathcal{D}} l(f^{\boldsymbol{w}}, f)$$

- deep learning model $f^{\boldsymbol{w}}$;
- dataset $\mathcal{D}$;
- definition of l and optimization algorithm.

Outline

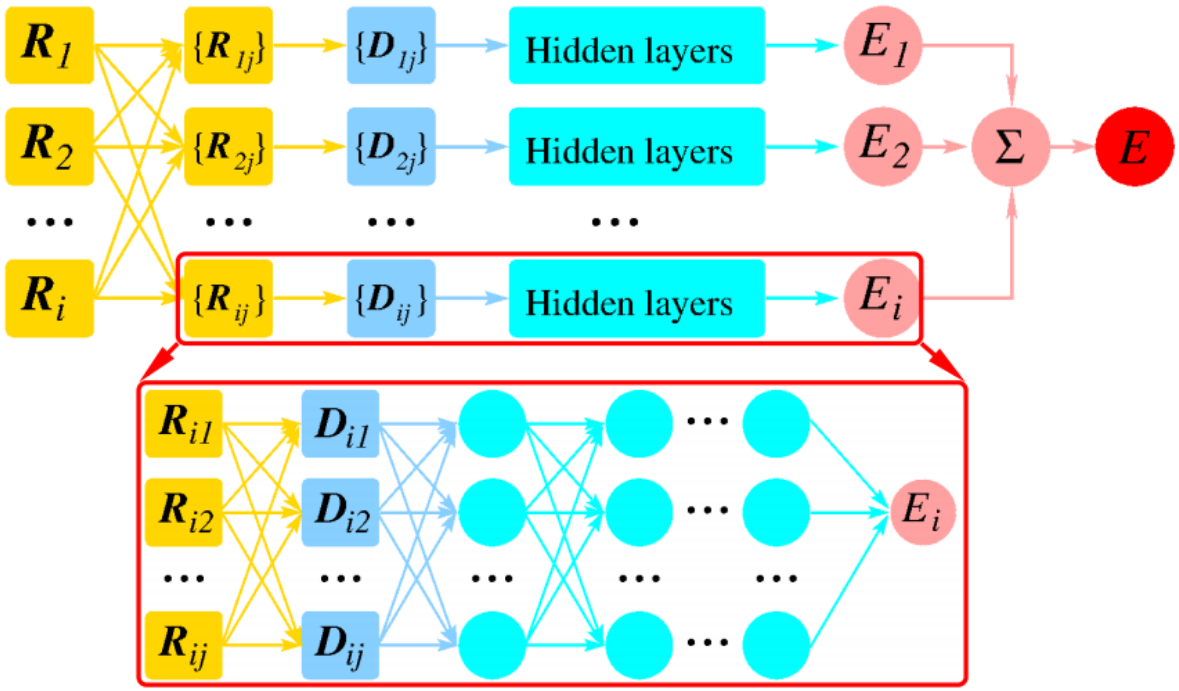
1. Introduction
2. **Deep Potential**
3. Deep Potential Generator (DP-GEN)
4. Free energy and Reinforced Dynamics
5. Conclusions

Requirement for a reliable PES model

- accuracy (e.g. uniform);
- efficiency (e.g. linear scaling);
- physical constraint (e.g. extensivity, symmetry);
- no human intervention/ end-to-end.

Typical construction

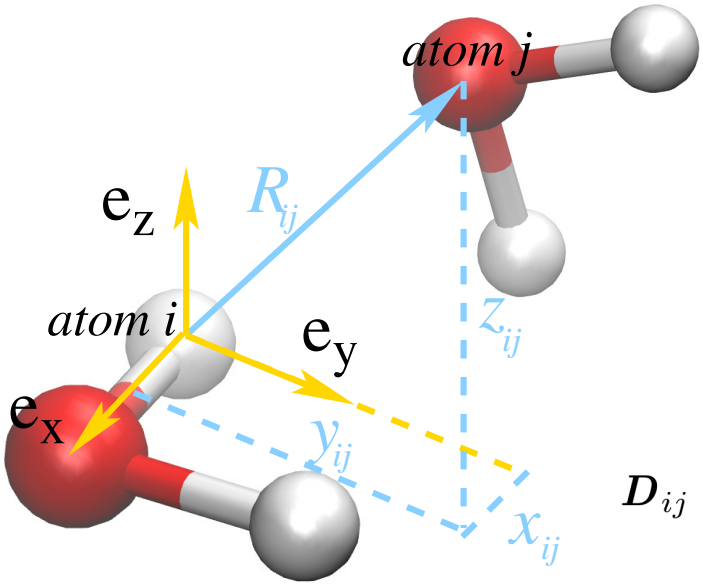
$$E = \sum_i E_i, \;\; E_i = E_{s(i)}(\boldsymbol{r}_i, \{\boldsymbol{r}_j\}_{j\in\mathcal{N}(i)}), \;\; \mathcal{N}(i) = \{j : r_{ij} = |\boldsymbol{r}_{ij}| \leq r_c\}$$



$E_i(\boldsymbol{r}_i, \{\boldsymbol{r}_j\}_{j\in\mathcal{N}(i)})$ represented by fully connected NNs with *symmetrized* inputs.

Behler, J., Parrinello, M. (2007). Phys. Rev. Lett., 98(14), 146401.

Descriptors: Local coordinates



$$\boldsymbol{D}_{ij} = \{1/R_{ij}\}$$

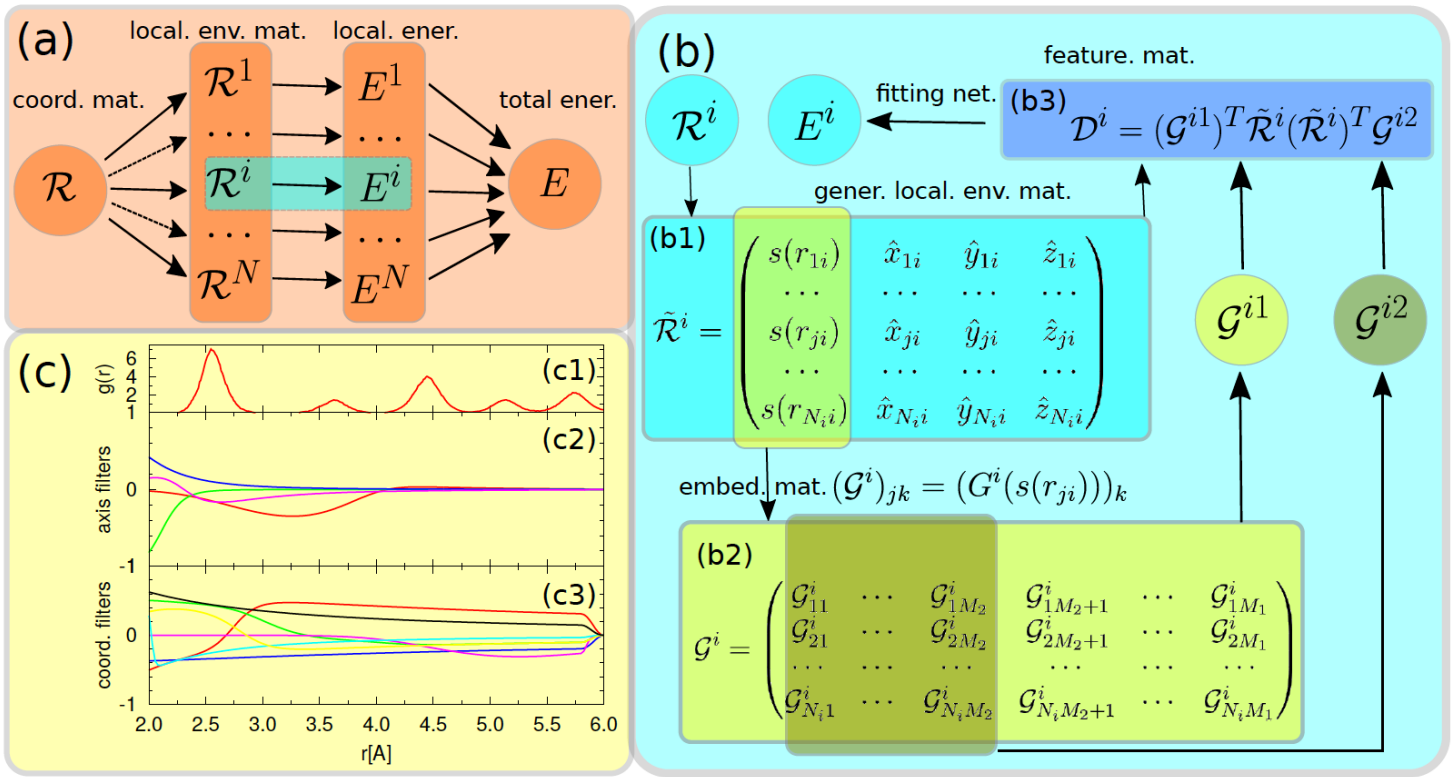
or

$$\boldsymbol{D}_{ij} = \{1/R_{ij}, x_{ij}/R_{ij}^2, y_{ij}/R_{ij}^2, z_{ij}/R_{ij}^2\}$$

Han, et.al., CiCP, 23, 629 (2018). Zhang, et.al., PRL, 120, 143001 (2018)

Descriptors: a smooth descriptor by DNN

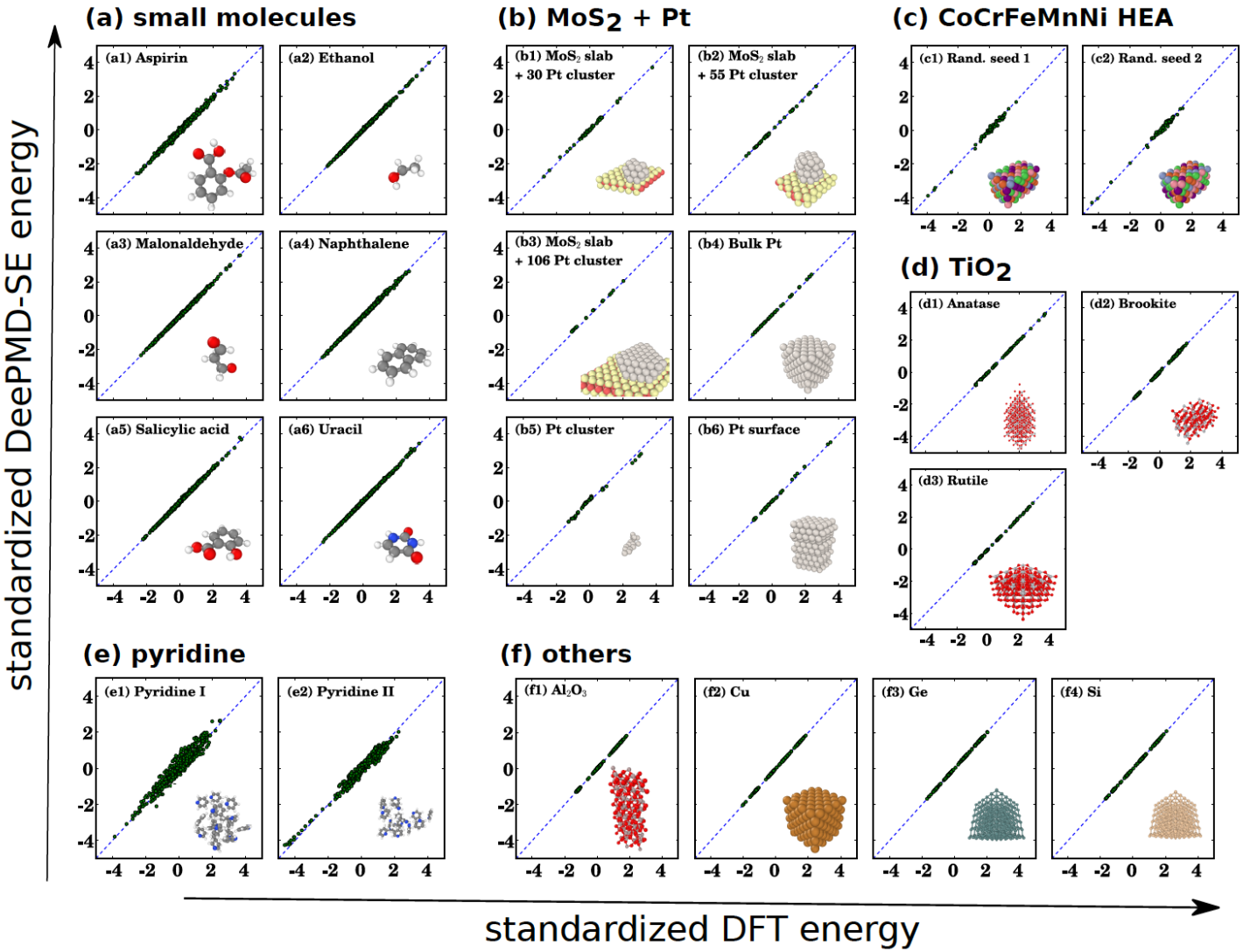
Key: complete and adaptive.



- Translation and Rotation: $(\mathcal{R}^i(\mathcal{R}^i)^T)$: $\Omega^i_{jk} = \boldsymbol{r}_{ji} \cdot \boldsymbol{r}_{ki}$,
- Permutation: $((\mathcal{G}^{i1})^T\mathcal{R}^i)$: $\sum_{j\in\mathcal{N}(i)} g(\boldsymbol{r}_{ji})\boldsymbol{r}_{ji}$,
- Finally, we propose: $\mathcal{D}^i = (\mathcal{G}^{i1})^T\mathcal{R}^i(\mathcal{R}^i)^T\mathcal{G}^{i2}$.

Zhang, et.al., NeurIPS 2018

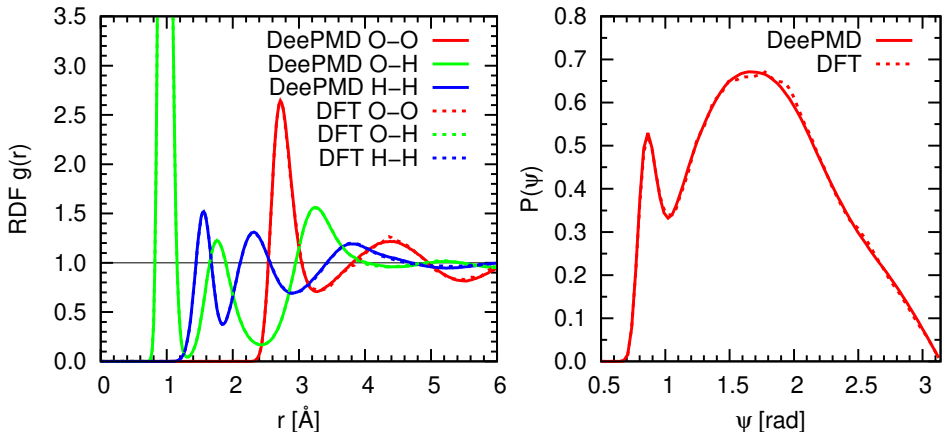
Various systems with the same principle



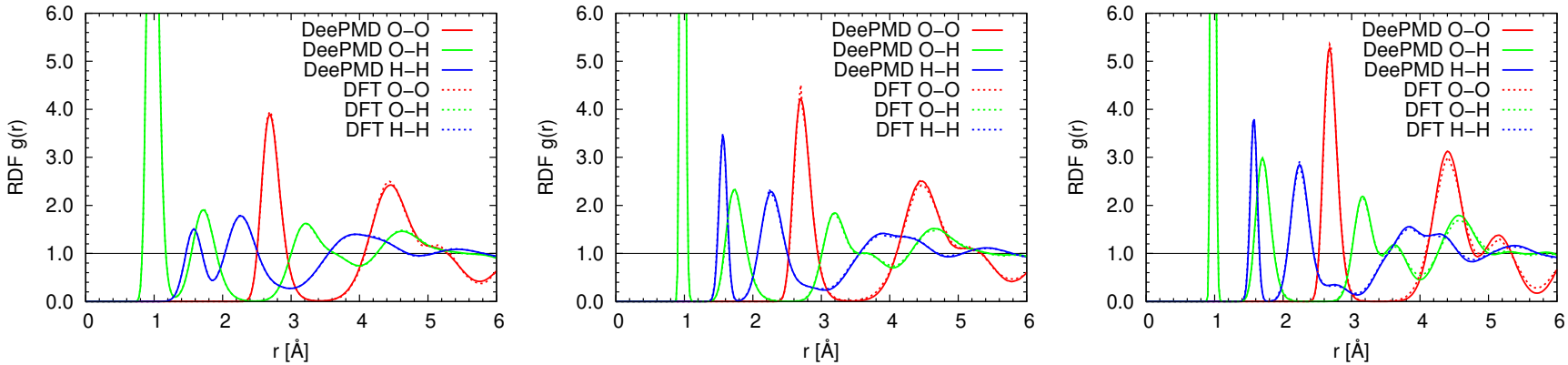
Zhang, et.al., NeurIPS 2018

Different thermodynamic conditions

- The path integral water structures (ambient cond.)

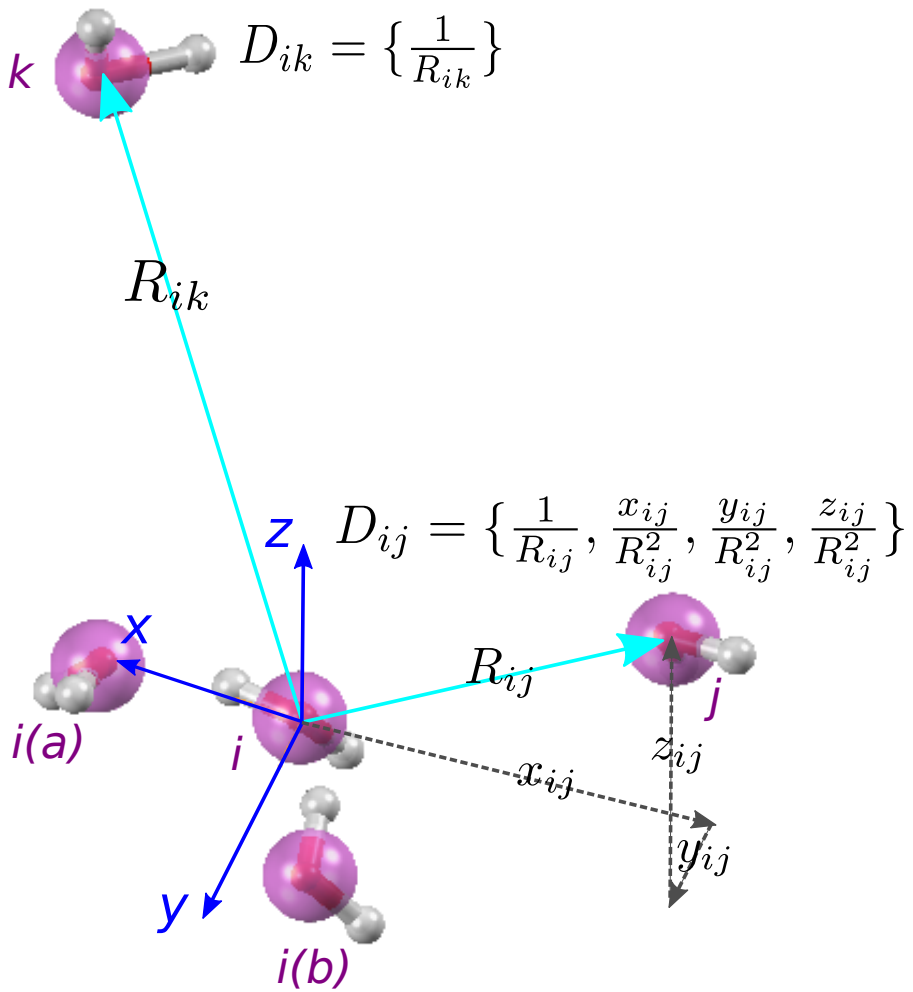


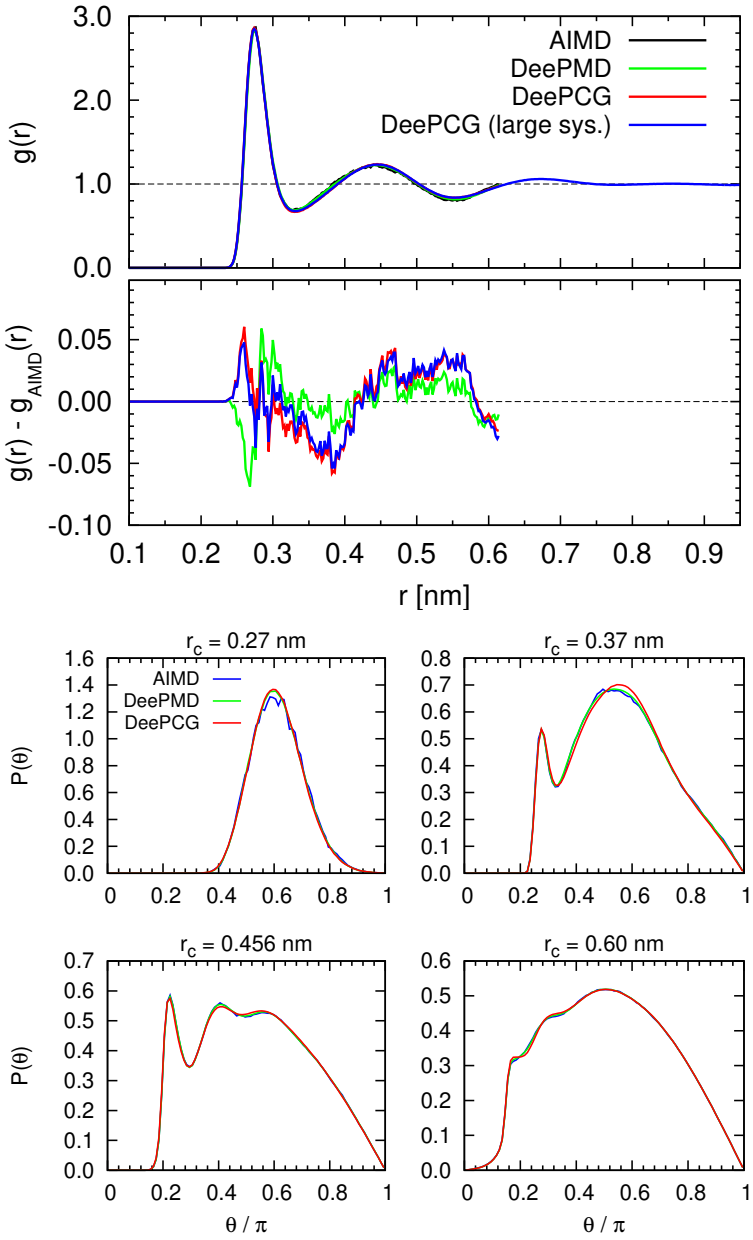
- Ice in different thermodynamic states



PI-ice, P=1.0 bar, T=273 K; ice P=1.0 bar,T=330 K; ice P=2.13 bar,T=238 K; Zhang et.al. Phys.Rev.Lett 120 143001 (2018)

Extension to coarse-graining

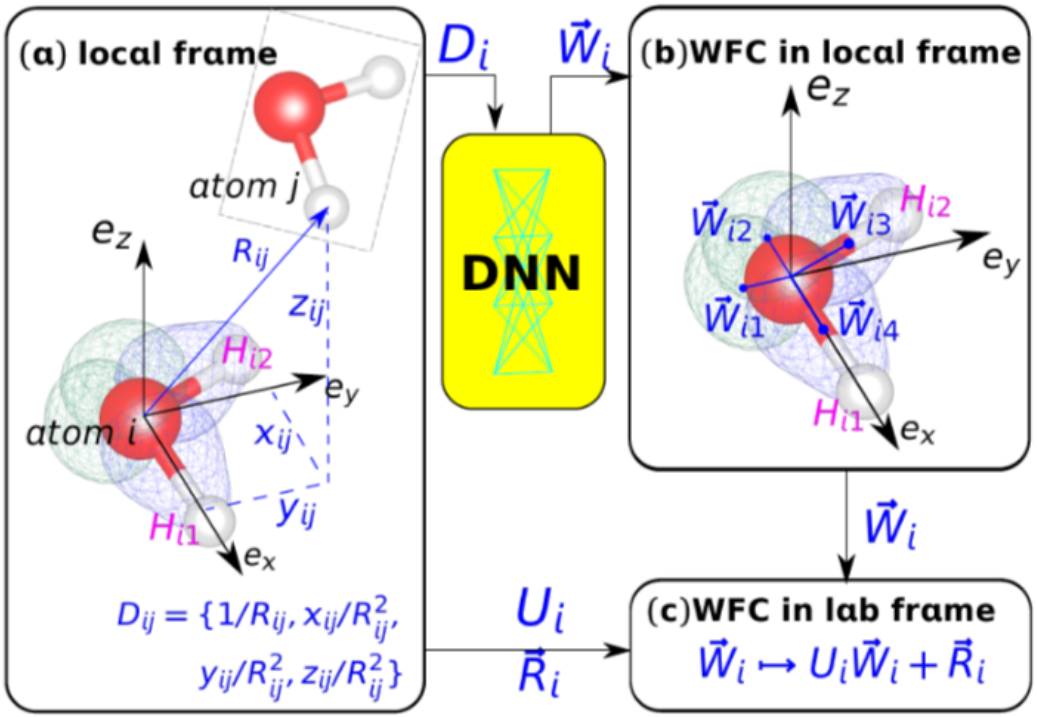




Zhang et.al. J. Chem. Phys., 149, 034101 (2018)

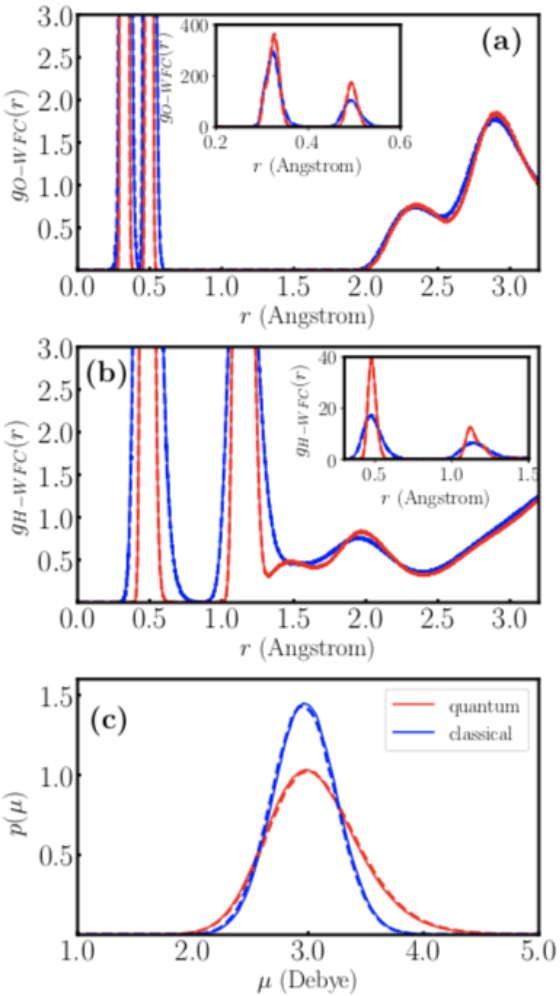
Extension to electronic information

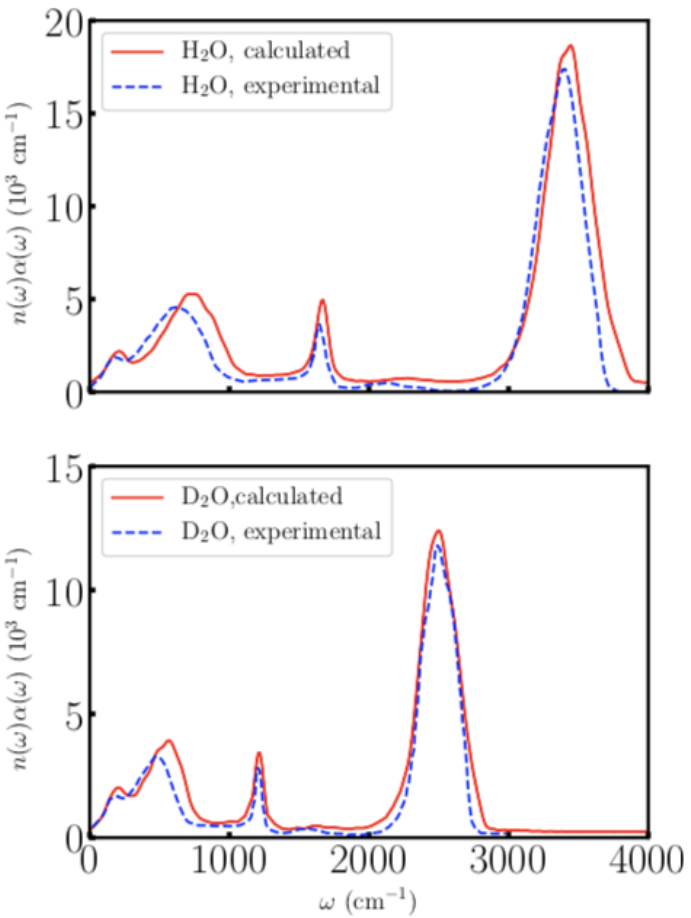
$$\{\vec{W}_n(\boldsymbol{r})\} = \min_{U_{\boldsymbol{k}}} \Omega[U_{\boldsymbol{k}}] = \min_{U_{\boldsymbol{k}}} \sum_n (\langle r^2 \rangle_n - \langle \bar{\boldsymbol{r}}^2 \rangle_n)$$



Nicola Marzari and David Vanderbilt. Phys. Rev. B 56.20 (1997): 12847.

Extension to electronic information





Left: Pair coorelation functions of ionic and MWLF center positions and disctribution of magnitude of molecular dipole.

Right: Infra-red spectrum calculated from 500 ps microcanonical DPMD simulation of 512 water molecules at $T = 300K$.

L. Zhang, et al., in preparation.

Extension to nonadiabatic excited state dynamics

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

Cite This: J. Phys. Chem. Lett. 2018, 9, 6702−6708

pubs.acs.org/JPCL

Deep Learning for Nonadiabatic Excited-State Dynamics

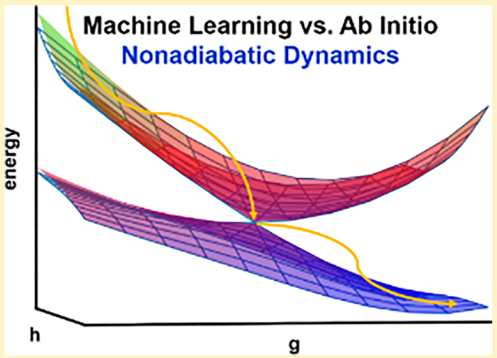
Wen-Kai Chen,[†] Xiang-Yang Liu,[†] Wei-Hai Fang,[†] Pavlo O. Dral,[‡] and Ganglong Cui*[,†]

[†]Key Laboratory of Theoretical and Computational Photochemistry, Ministry of Education, College of Chemistry, Beijing Normal University, Beijing 100875, China

[‡]Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany

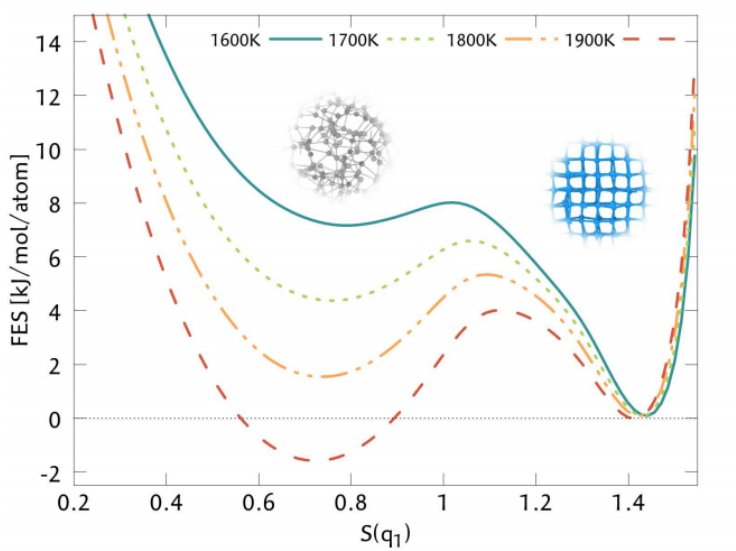
Supporting Information

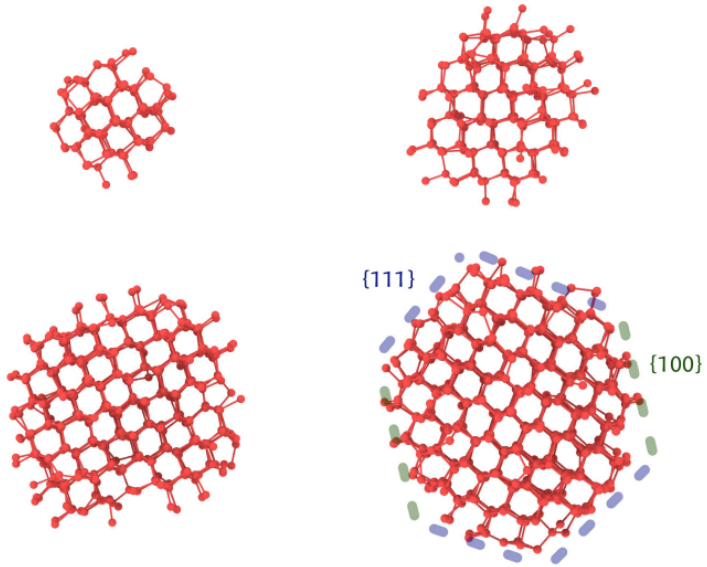
ABSTRACT: In this work we show that deep learning (DL) can be used for exploring complex and highly nonlinear multistate potential energy surfaces of polyatomic molecules and related nonadiabatic dynamics. Our DL is based on deep neural networks (DNNs), which are used as accurate representations of the CASSCF ground- and excited-state potential energy surfaces (PESs) of CH_2NH. After geometries near conical intersection are included in the training set, the DNN models accurately reproduce excited-state topological structures; photoisomerization paths; and, importantly, conical intersections. We have also demonstrated that the results from nonadiabatic dynamics run with the DNN models are very close to those from the dynamics run with the pure ab initio method. The present work should encourage further studies of using machine learning methods to explore excited-state potential energy surfaces and nonadiabatic dynamics of polyatomic molecules.



Chen, Wen-Kai, et al. J. P. C. Lett. 9.23 (2018): 6702-6708.

Combined with metadynamics



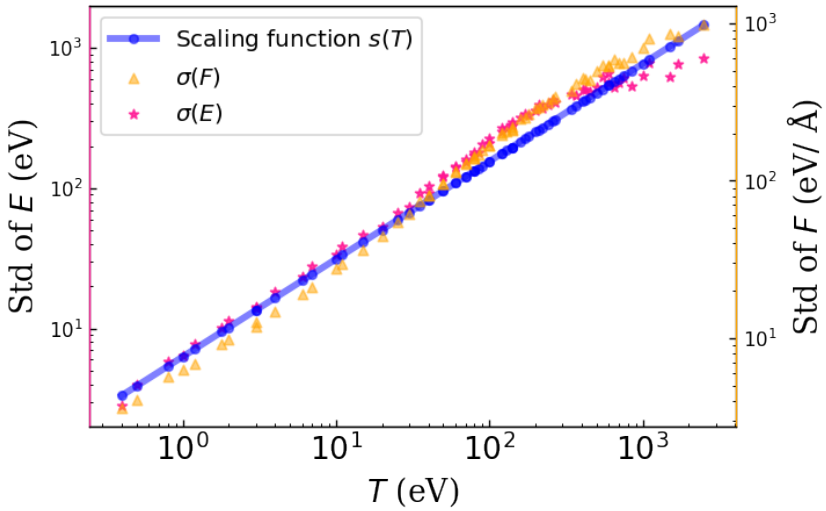


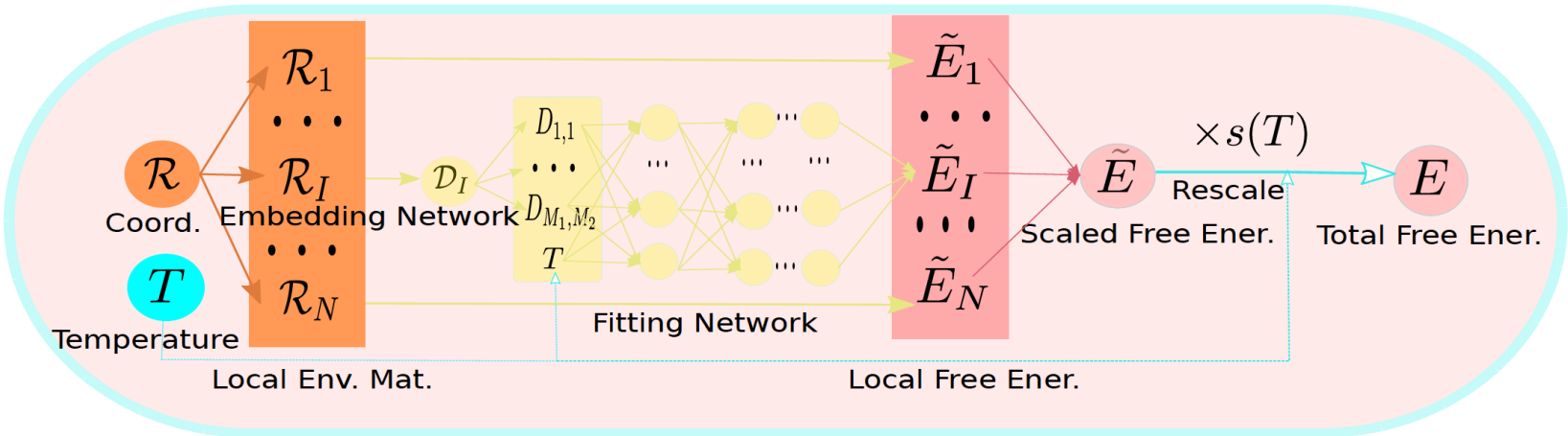
L. Bonati and M. Parrinello, Phys. Rev. Lett. 121, 265701

Extension to T-dependent free energy

$$E(\boldsymbol{R},T) = \min_{\{\psi_i\},\{f_i\}} E[\{\psi_i\},\{f_i\}]$$
$$= \min_{\{\psi_i\},\{f_i\}} U[\{\psi_i\},\{f_i\}] - TS[\{f_i\}].$$

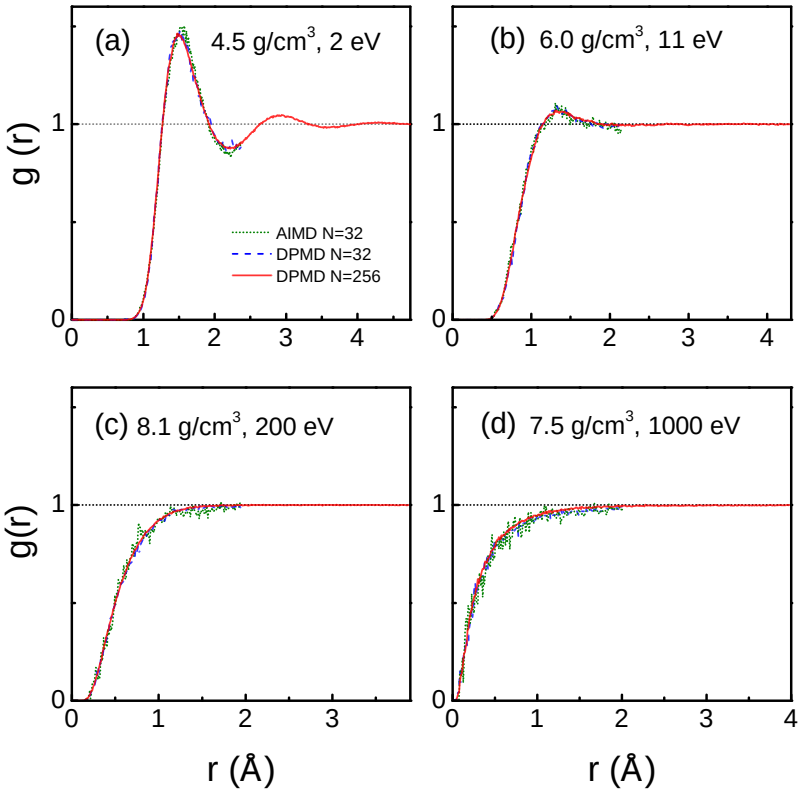
$$E^{\mathbf{w}}(\boldsymbol{R},T) = s(T)\sum_I E_I^{\mathbf{w}}(\mathcal{R}_I,T)$$

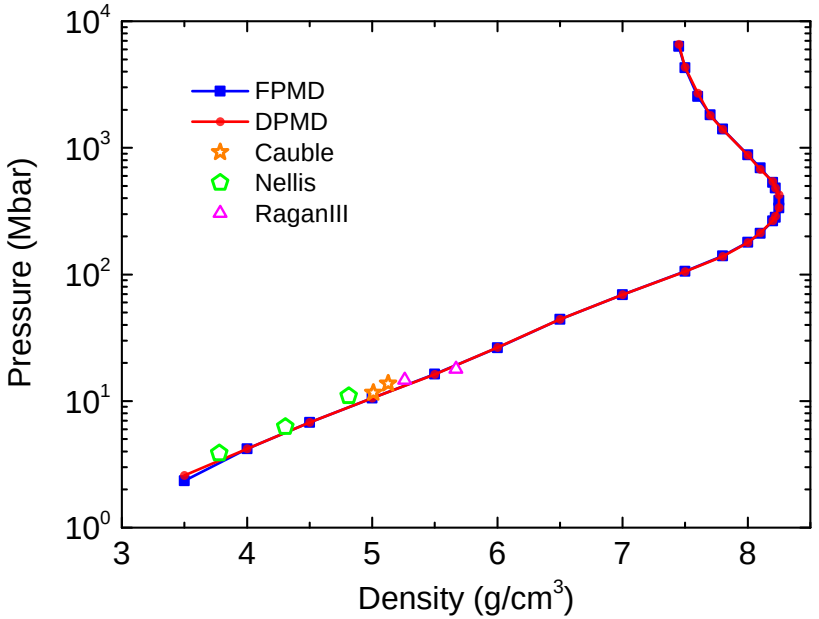




Extension to T-dependent free energy

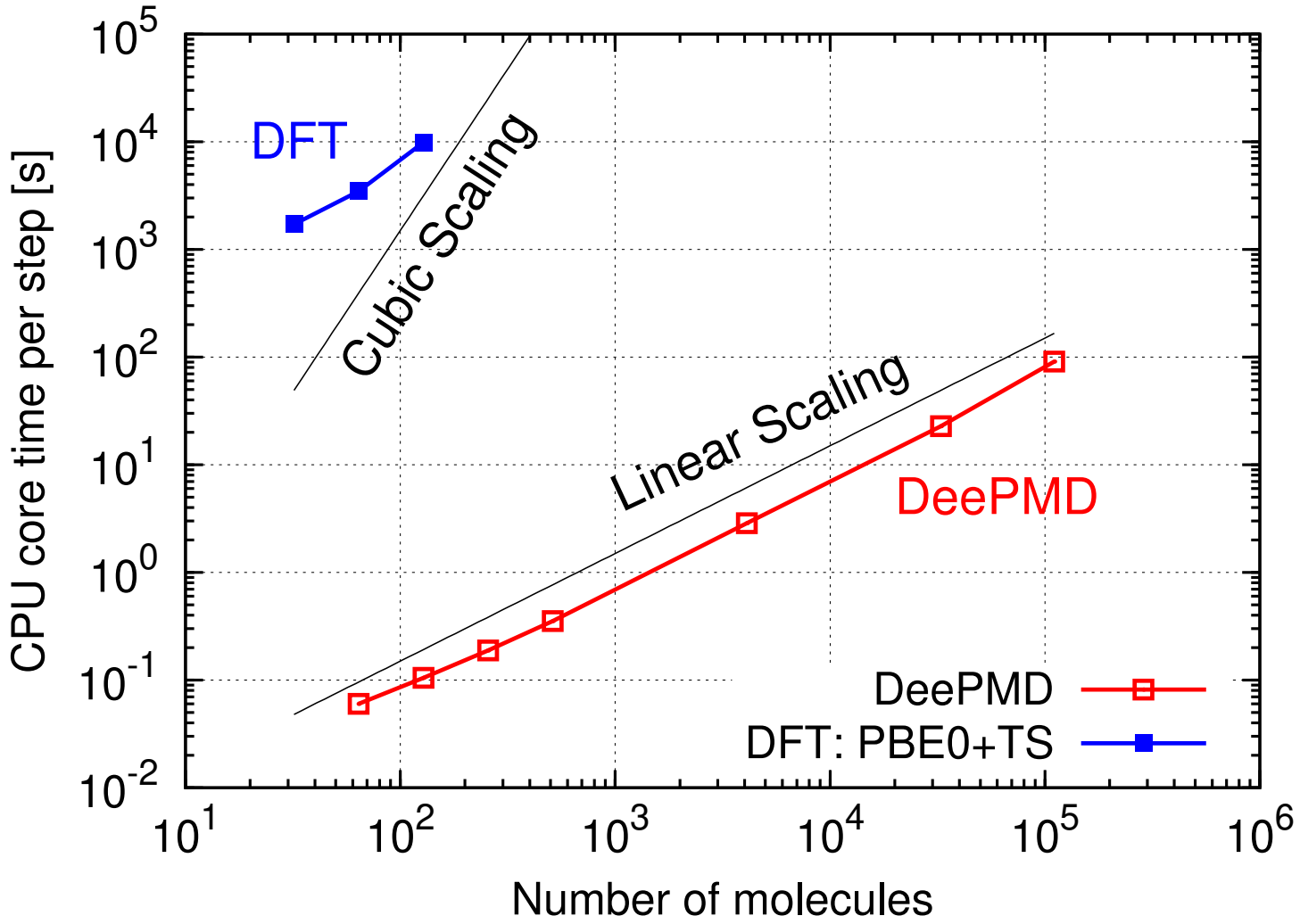
Left: Radial distribution functions (RDFs); Right: Rankine-Hugoniot curve.



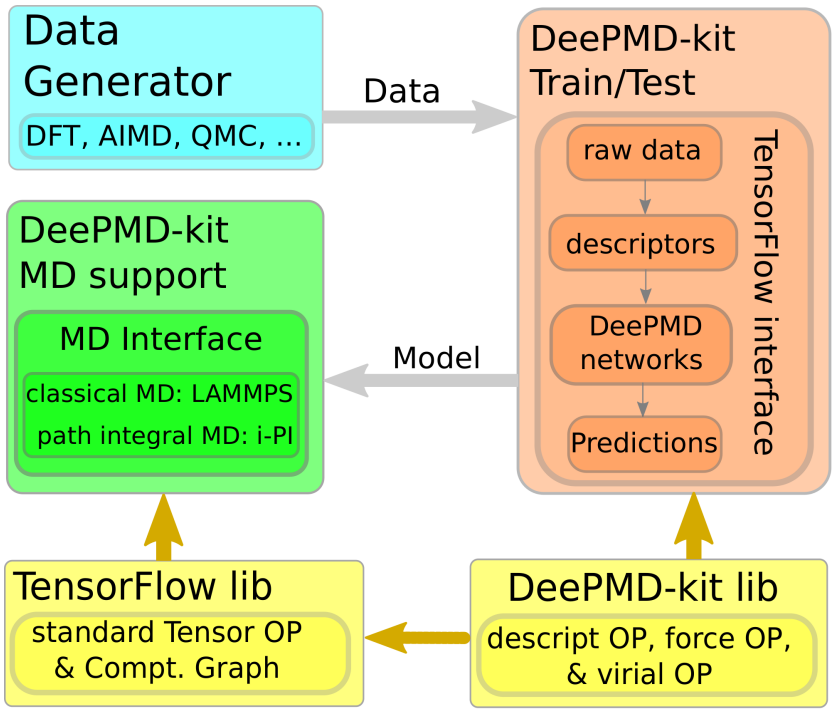


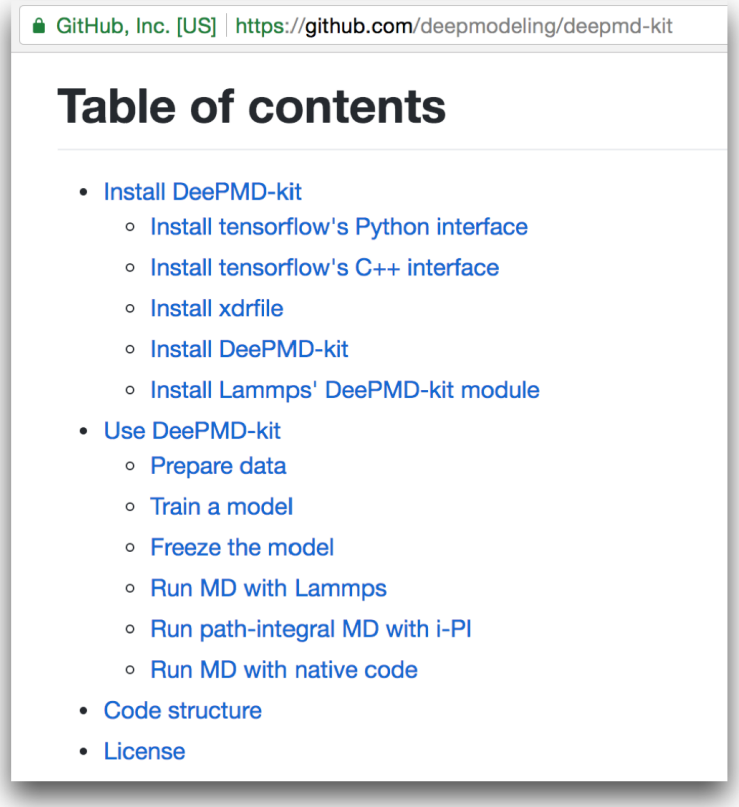
(in preparation)

Deep Potential: MD scalability



Open source software DeePMD-kit





- TensorFlow: efficient network operators
- LAMMPS, i-PI; MPI/GPU support.

Free download from https://github.com/deepmodeling/deepmd-kit

Comp.Phys.Comm., 0010-4655 (2018).

Outline

1. Introduction
2. Deep Potential
3. **Deep Potential Generator (DP-GEN)**
4. Free energy and Reinforced Dynamics
5. Conclusions

Two important aspects, revisited

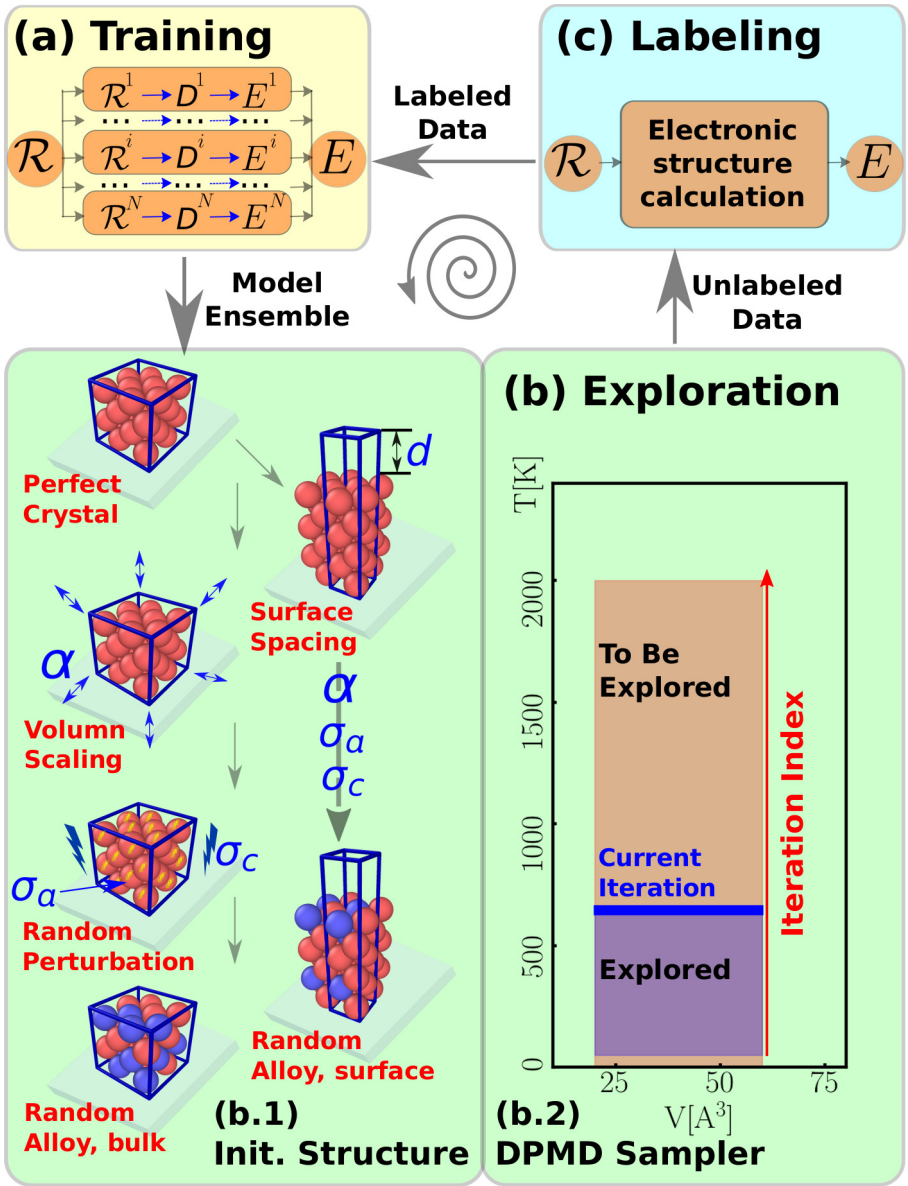
$$\min_{\boldsymbol{w}} \frac{1}{\|\mathcal{D}\|} \sum_{i \in \mathcal{D}} l(f^{\boldsymbol{w}}, f)$$

- deep learning model $f^{\boldsymbol{w}}$;
- dataset $\mathcal{D}$;
- definition of l and optimization algorithm.

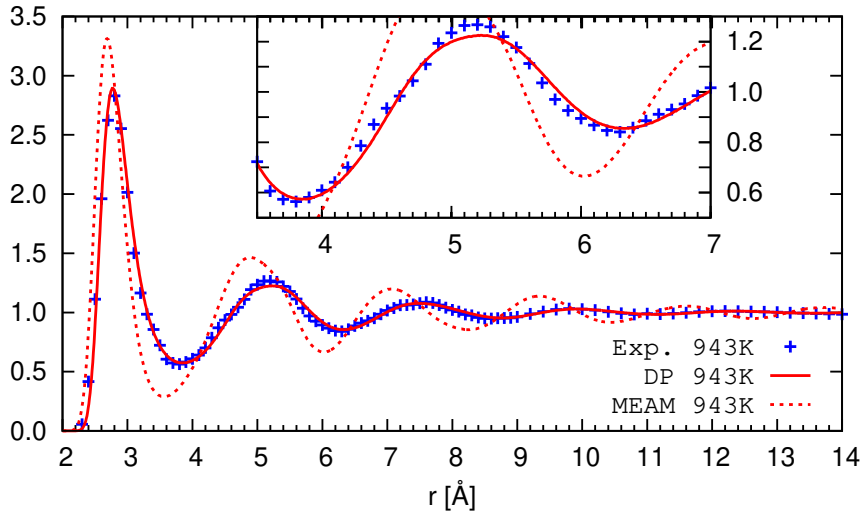
Active learning: the DP-GEN scheme

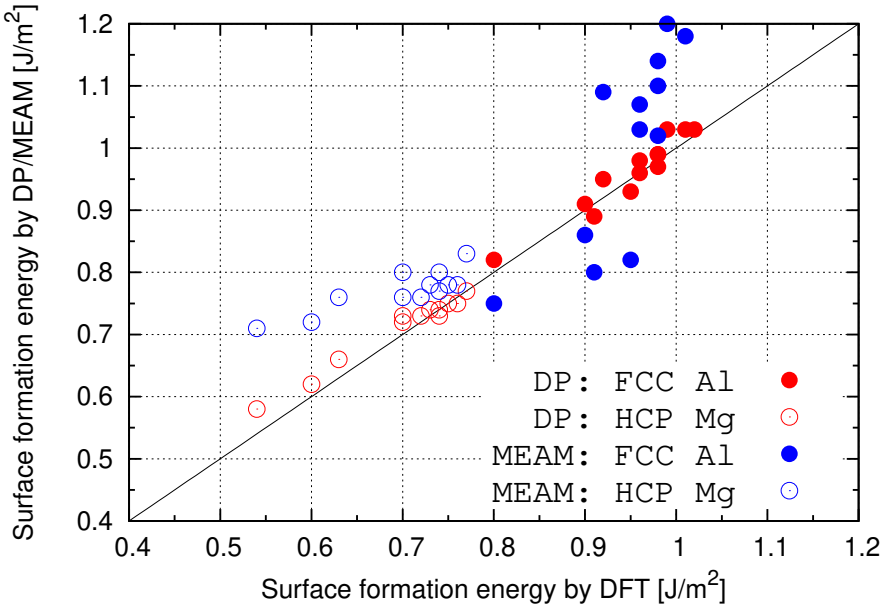
- **Training/Fitting**:
  model/representation.
- **Exploration**:
  sampler and error indicator;
  DPMD and model deviation
  $\epsilon = \max_i \sqrt{\langle \| \boldsymbol{f}_i - \langle \boldsymbol{f}_i \rangle \|^2 \rangle}$
- **Labeling**:
  ab initio calculator.
- Example: Al-Mg alloy
  0.0044 % explored confs. are labeled

Zhang et.al. Phys. Rev. Mat. 3, 023804

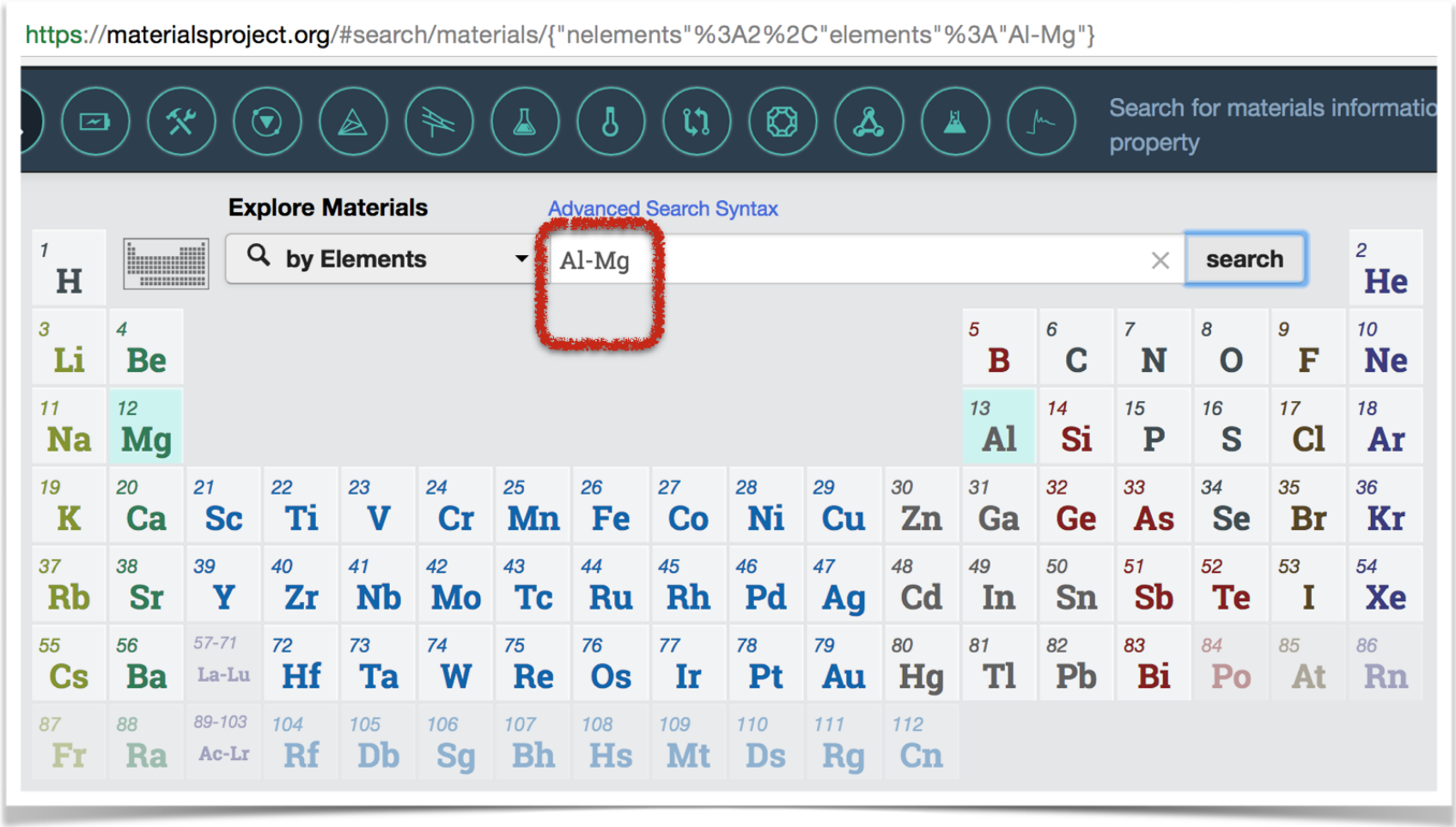


DP-GEN: test of Al

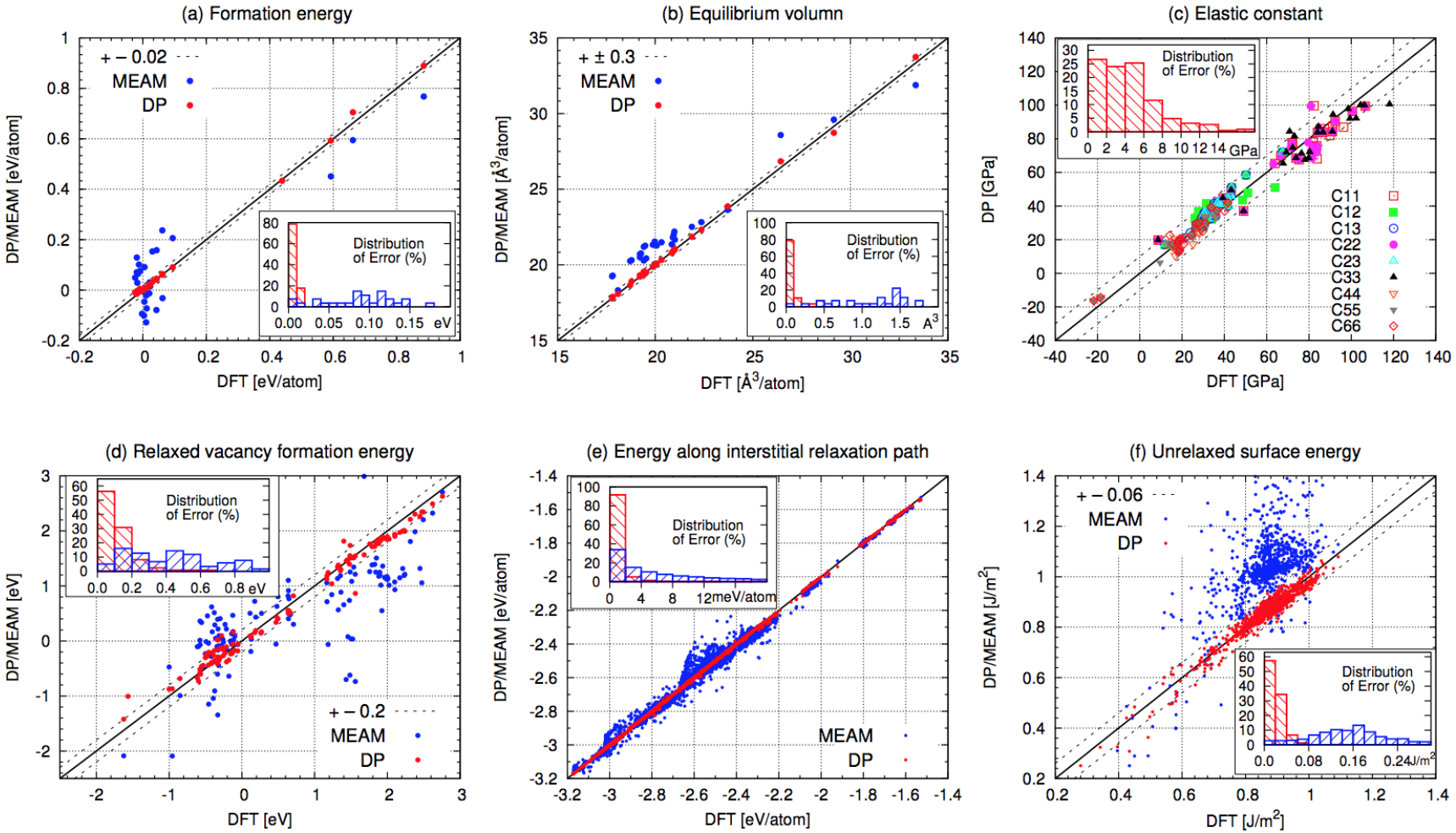




DP-GEN: tests based on Materials Project



DP-GEN: tests based on Materials Project



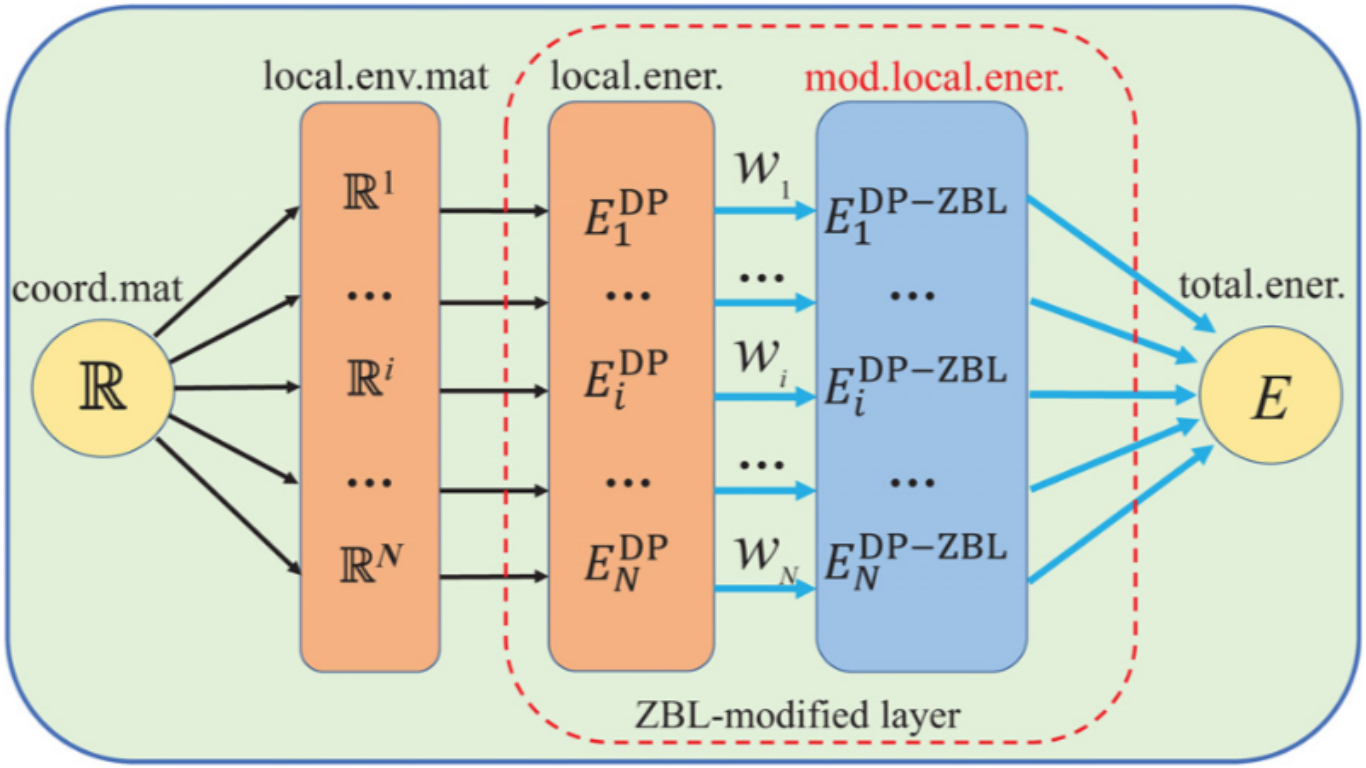
Irradiation damage simulation

Deep learning inter-atomic potential model for accurate irradiation damage simulations

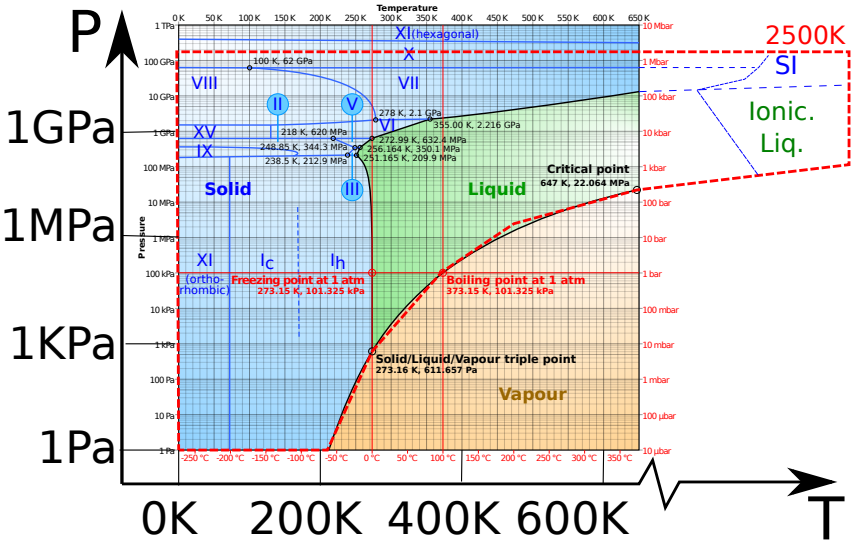
Cite as: Appl. Phys. Lett. **114**, 244101 (2019); doi: 10.1063/1.5098061
Submitted: 31 March 2019 · Accepted: 29 May 2019 ·
Published Online: 17 June 2019

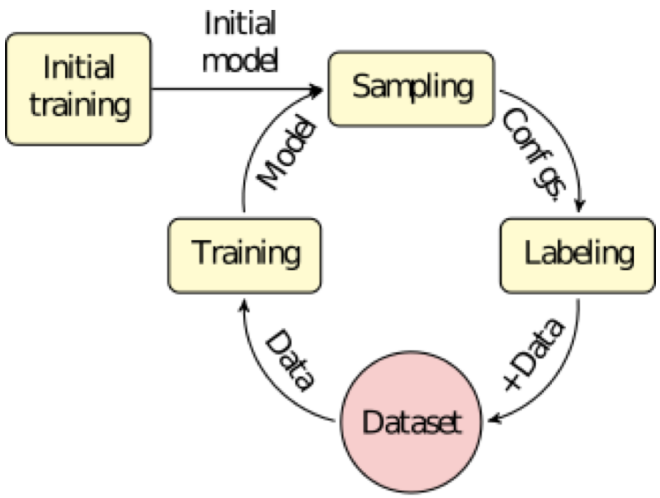
View Online Export Citation CrossMark

Hao Wang,[1,a)] Xun Guo,[1,a)] Linfeng Zhang,[2] Han Wang,[3,b)] and Jianming Xue[1,c)]



DP-GEN for water





- **Reference model:** DFT at the classical SCAN level;
- **Starting configurations:** relaxed Ice I-XV at $T = 0$ K and equilibrated liquid at $T = 330$ K;
- **Range of thermodynamic conditions:** red dashed box;
- **number of MD snapshots:** DPMD exploration: 1.4 billion, DFT calculation: 32 thousand ($\sim$0.002% of the former).

  Typical AIMD trajectory: 100 thousand snapshots (50-100 *ps*).
- **number of DP-GEN iterations:** 100.

Thermodynamic integration (TI) for the phase diagram

Hamiltonian TI

$(H^{(ref)}, p, T) \mapsto (H, p, T)$

↓

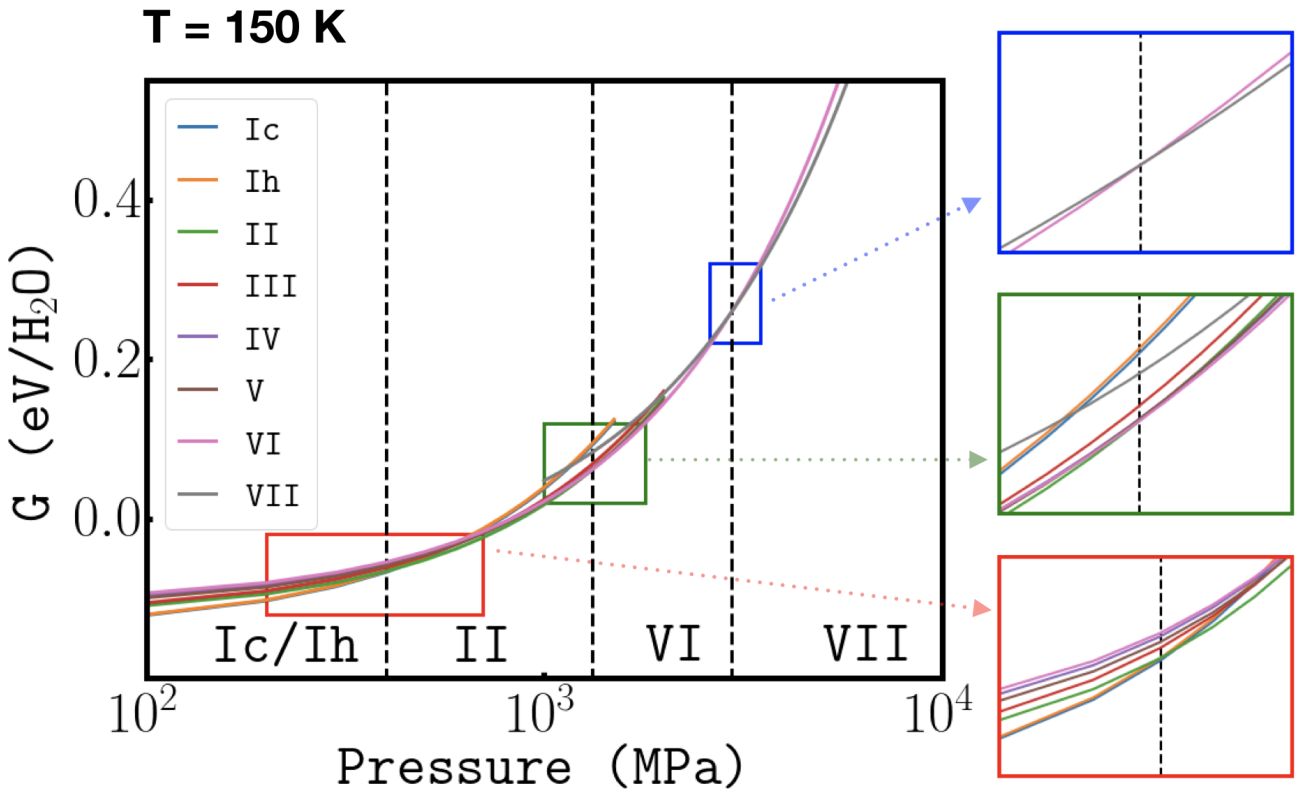
TI

$(H, p_1, T_1) \mapsto (H, p_2, T_2)$

↓

Clapeyron Eqn.

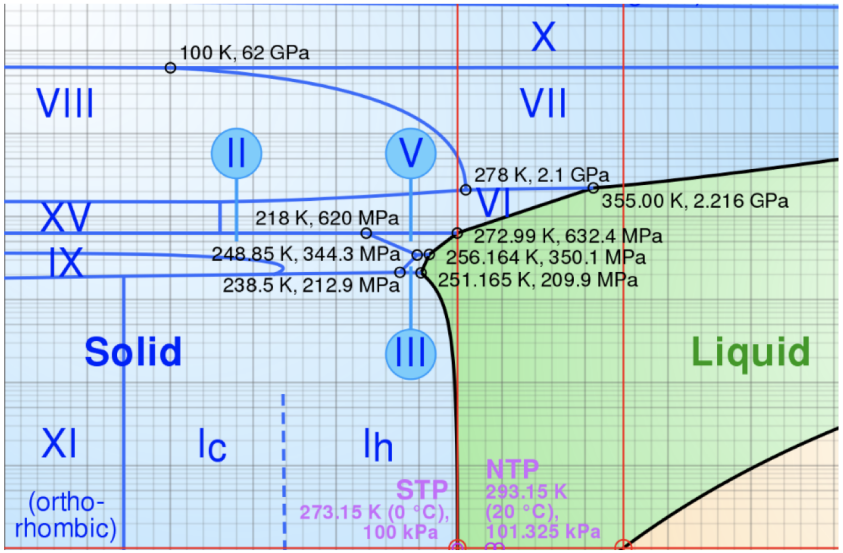
$$\frac{dp}{dT} = \frac{h_{II} - h_I}{T(v_{II} - v_I)}$$



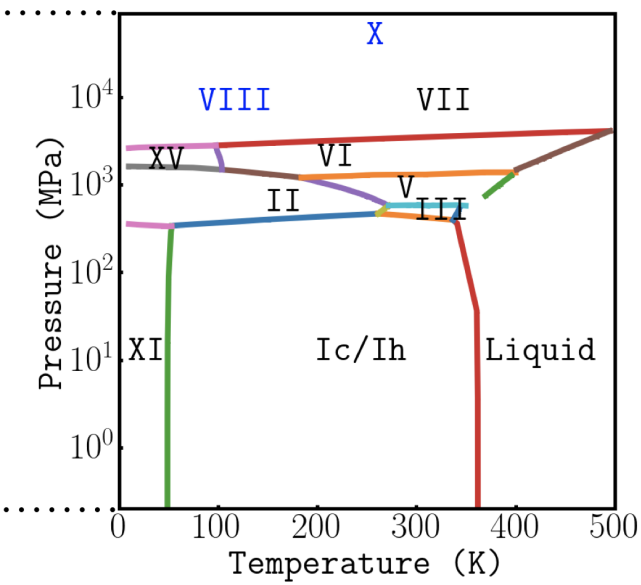
Special issues: size effect; proton disorder, etc.

Water phase diagram modeled by DP+SCAN

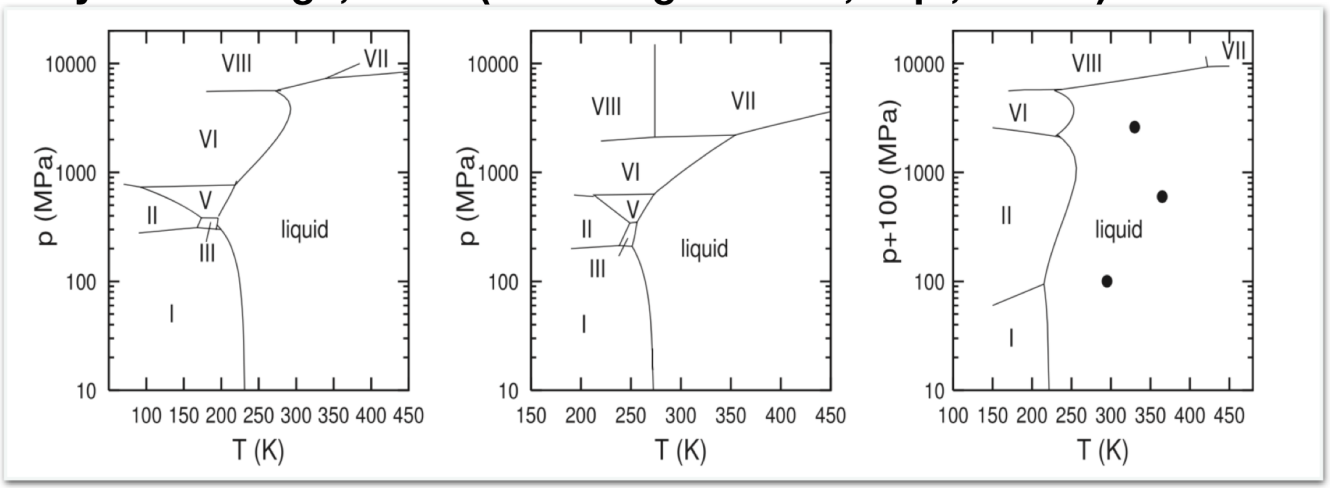
(a) Exp. collected by Martin Chaplin



(b) current result

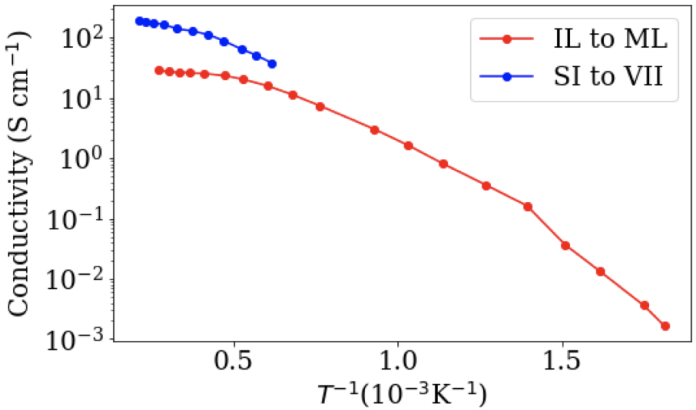


(c) Work by Carlos Vega, et al. (Left to right: TIP4P, Exp., SPC/E)

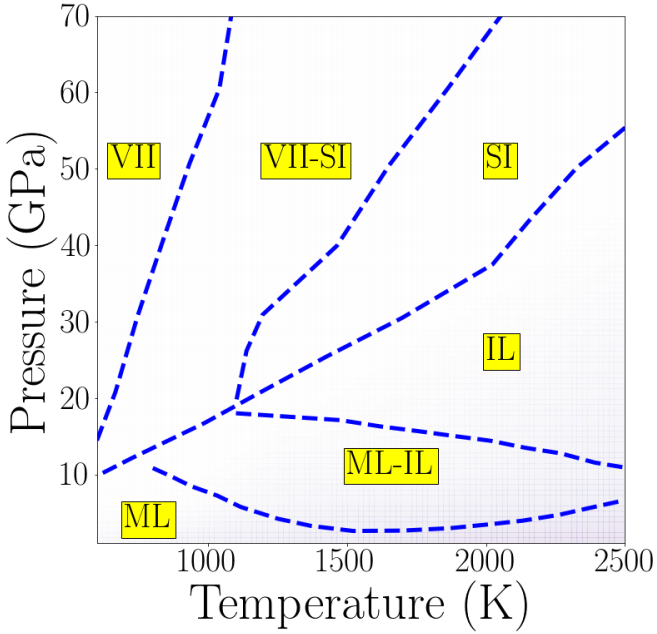


High-pressure phases modeled by DP+SCAN

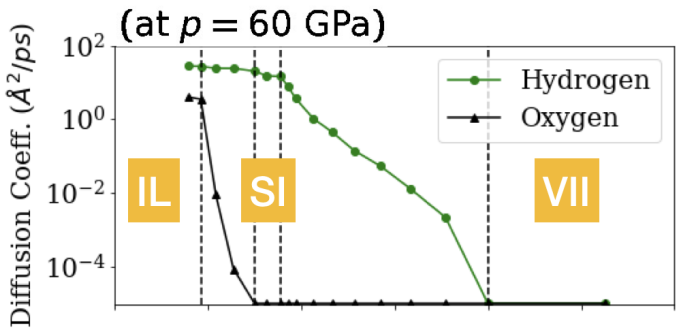
Exp. by Marius, et. al

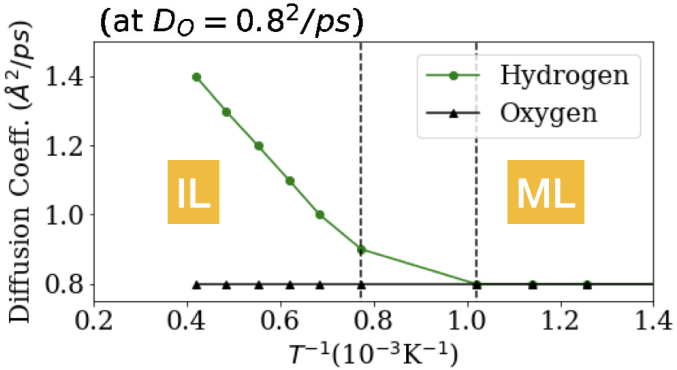


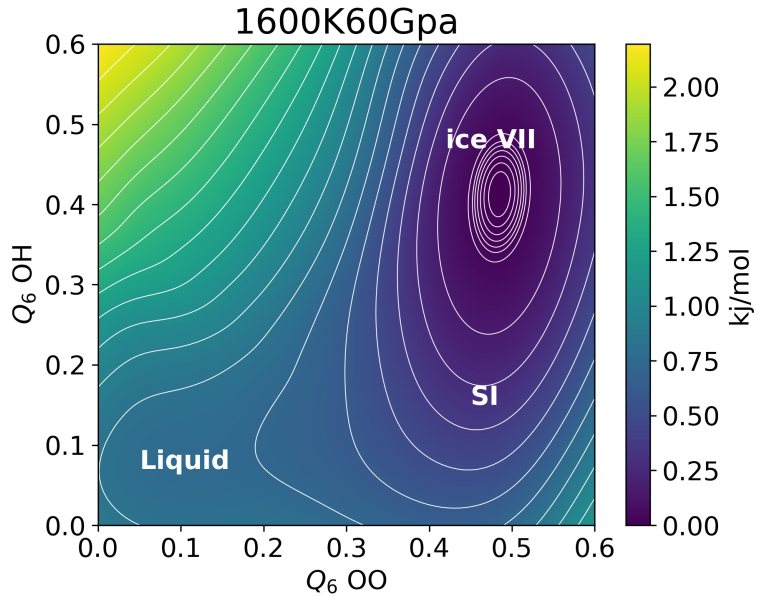
("Experimental evidence for superionic water ice using shock compression", Nature Physics 14, 297–302 (2018))



DPMD simulation:





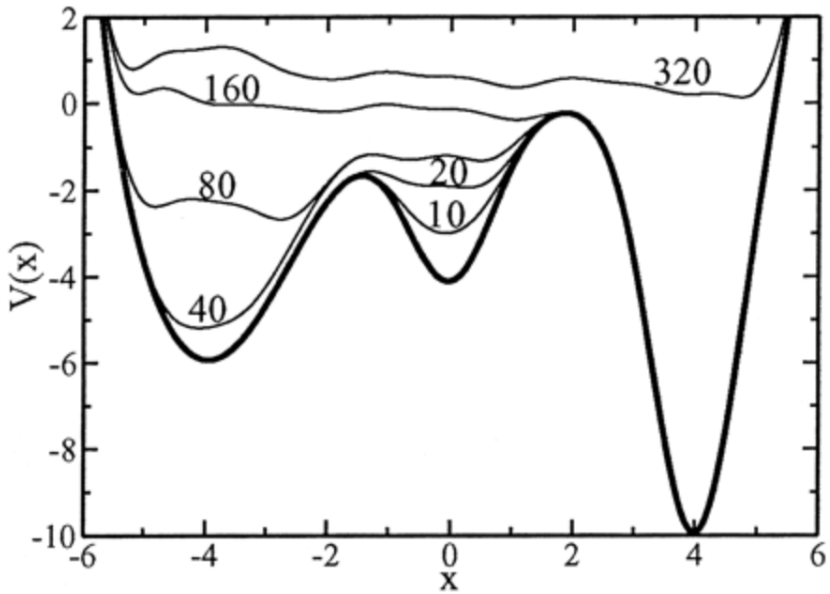


Outline

1. Introduction
2. Deep Potential
3. Deep Potential Generator (DP-GEN)
4. **Free energy and Reinforced Dynamics**
5. Conclusions

Free energy and deep neural networks

- Exploring configuration space, phase transition, ...
  - high dimensionality of the collective variable space;
  - high energy barriers and complex energy landscape.
- Metadynamics PNAS 99(20):1256212566, 2002):



- Temperature accelerated (Chem. Phys. Lett., 426(1):168175, 2006.)
- curse of dimensionality.

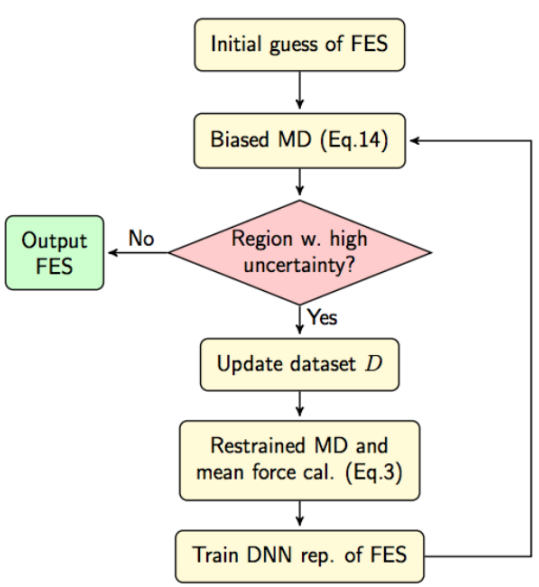
Reinforced dynamics

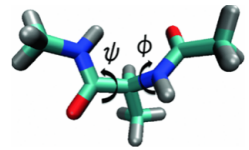
	Potential energy	Free energy
Method	DP-GEN	Reinforced dynamics
Model	Deep potential	ResNet
Sampler	Deep potential MD	Biased MD
Label	Electronic struct.	Restrained MD

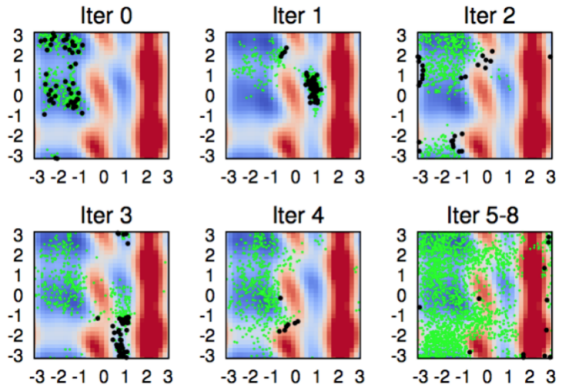
Reinforced dynamics

- reinforcement learning: state, $action$, best policy, $reward$;
- reinforced dynamics (RiD): atomic system, $biased\ potential$, FES, $model\ deviation$.

$$\epsilon^2(\boldsymbol{s}) = \left\langle \|\boldsymbol{\mathcal{F}}(\boldsymbol{s}) - \bar{\boldsymbol{\mathcal{F}}}(\boldsymbol{s})\|^2 \right\rangle, \tilde{\boldsymbol{f}}_i(\boldsymbol{r}) = -\nabla_{\boldsymbol{r}_i} U(\boldsymbol{r}) + \sigma(\epsilon(\boldsymbol{s}(\boldsymbol{r}))) \nabla_{\boldsymbol{r}_i} \boldsymbol{\mathcal{A}}(\boldsymbol{s}(\boldsymbol{r}))$$



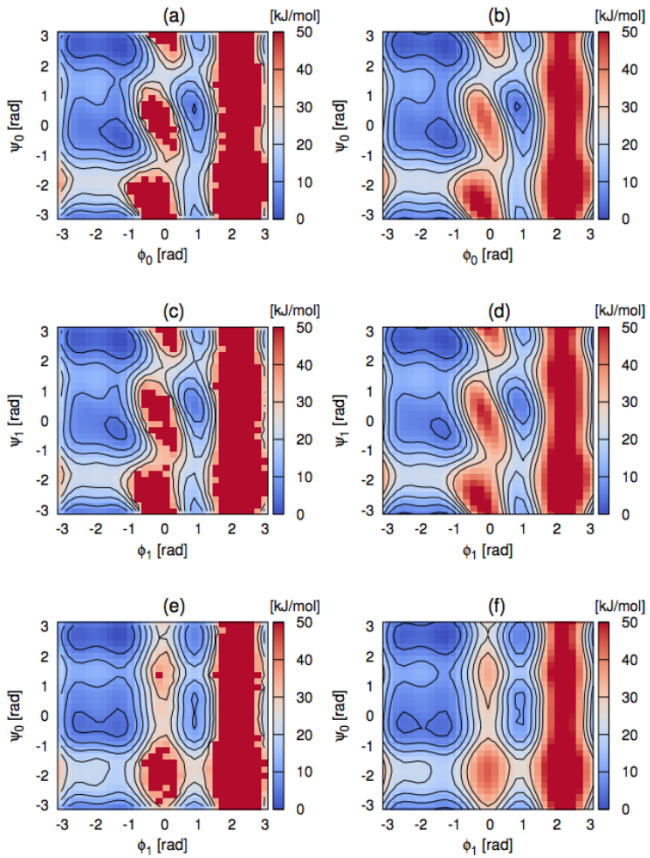


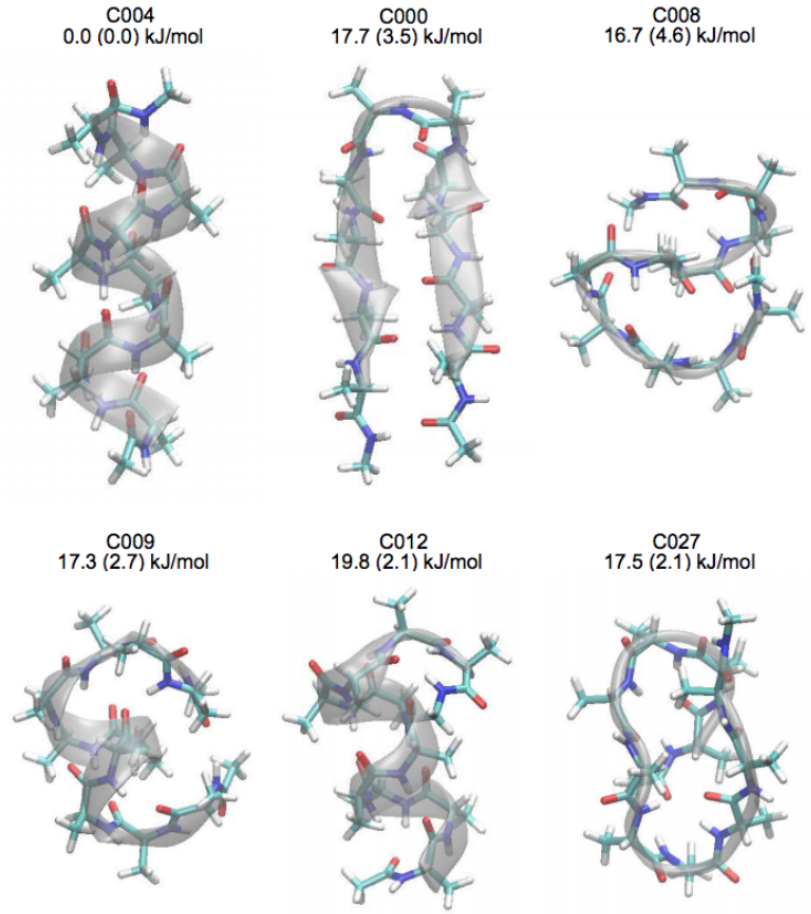


Zhang, et.al. J.Chem.Phys 148, 124113 (2018).

Reinforced dynamics

- Left: Tripeptide: brute-force simulation ($\sim$50 μs) $v.s.$ RiD (10 ns biased + 190 ns restrained):
- Right: higher dimensional FES: ala-10 and 20 CVs.





Outline

1. Introduction
2. Deep Potential
3. Deep Potential Generator (DP-GEN)
4. Free energy and Reinforced Dynamics
5. **Conclusions**

Conclusions

- Model construction and data exploration for PES and FES;
- Useful models: Deep Potential, DP-GEN, reinforced dynamics; check `https://github.com/deepmodeling/deepmd-kit`
- Fundamental problems: quantum many-body problem, DFT, dynamics.

Acknowledgements

Advisors

- Roberto Car, Weinan E

Collaborators

- Han Wang, De-Ye Lin (IAPCM)
- Jiequn Han, Yixiao Chen, Hsin-Yu Ko, Marcos Andrade (Princeton),
- Wissam A Saidi (Univ. of Pittsburgh), Xifan Wu (Temple)
- Mohan Chen, Yuzhi Zhang (Peking Univ.)

Fundings and computational resources

- Tiger@Princeton, BIBDR, & NERSC;
- ONR grant N00014-13-1-0338, DOE grants DE-SC0008626 and DE-SC0009248, and NSFC grants U1430237 and 91530322;
- Computational Chemical Science Center: Chemistry in Solution and at Interfaces (DE-SC0019394).