Deep Neural Networks and Molecular Dynamics

Roberto Car

Princeton University

MAX Conference – ICTP, Trieste, Jan 29-31 2018

Deep Neural Networks

- **When properly trained, deep NN are capable of learning the complex functional dependence of physical properties from suitable descriptors**

- **Deep NN are extremely more powerful than standard interpolation approaches (e.g. spline, Fourier, etc.): seemingly they do not suffer the curse of dimensionality**

- **Why ?**

Deep NN and MD

- By learning the many-body potential energy surface (PES) from AIMD, Deep NN make simulations of AIMD quality possible at the cost of empirical FF, boosting the accessible size and time scales

- Deep NN open new perspectives for Coarse Graining (CG). In this case the free energy landscape replaces the PES

- Two schemes recently developed at Princeton: DeePMD and DeePCG

The DeePMD approach

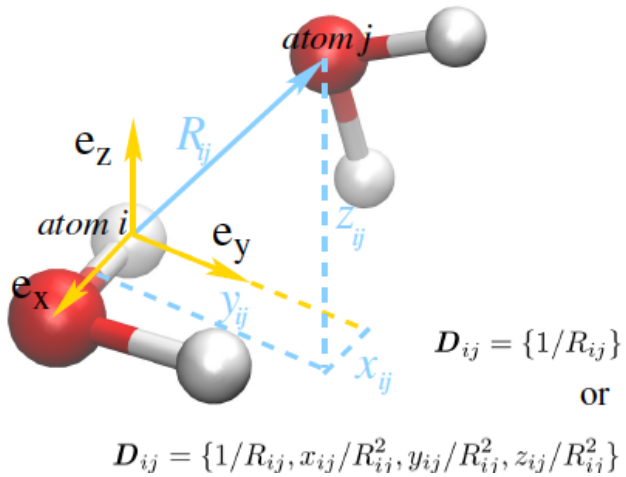
- **The atomic coordinates (*R*) of a multi-atom system are transformed into descriptors (*D*)**

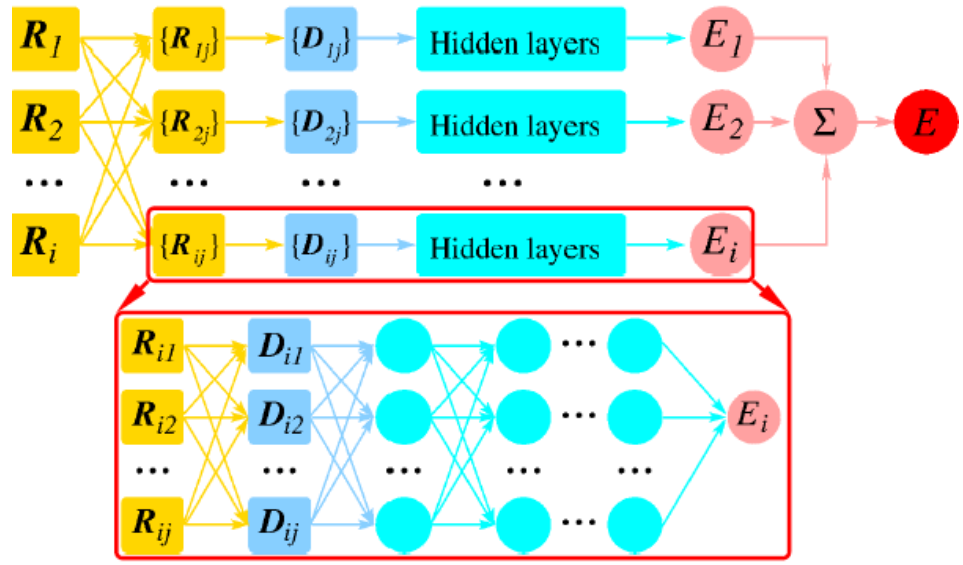
- **The descriptors are given in input to a deep NN in a way that preserves the translational, rotational, and permutational symmetry of the system**

- **In output the deep NN gives the PES of the system as a sum of atomic contributions; each one of these is an analytical function of the coordinates of the atoms within a cutoff radius R_c**

L. Zhang et al., arXiv: 1707.09571; see also Comm. Comp. Phys. **23**, 629 (2018)

DeePMD: schematics





To determine E_i the loss function L is minimized with the stochastic ADAM method on the configurations in an AIMD trajectory

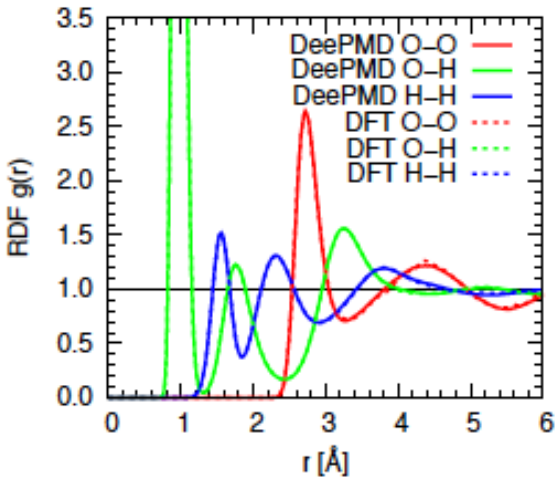
$$L(p_\epsilon, p_f, p_\xi) = p_\epsilon \Delta\epsilon^2 + \frac{p_f}{3N}\sum_i |\Delta \boldsymbol{F}_i|^2 + \frac{p_\xi}{9}||\Delta\xi||^2$$

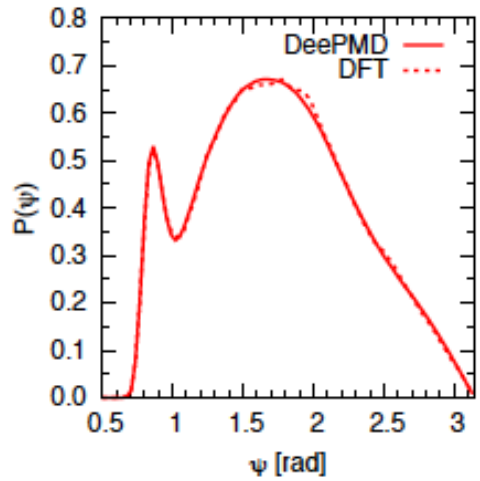
$$\text{PES:}\quad V(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N) = E = \sum_{i=1,N} E_i(\{\mathbf{R}\}_i)$$

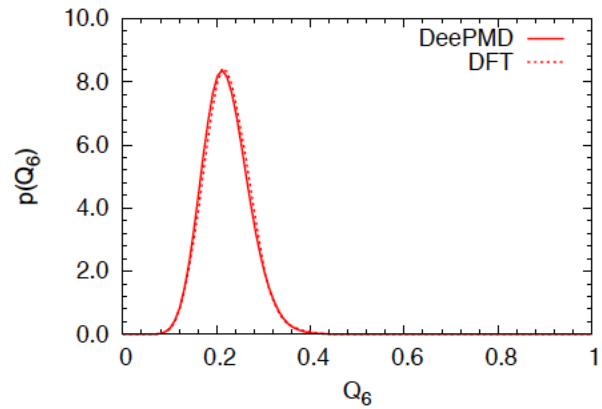
Code available at https://github.com/deepmodeling

How well does it work?

PI-AIMD (*N*,*p*,T) water with PBE0-TS$_{scf}$

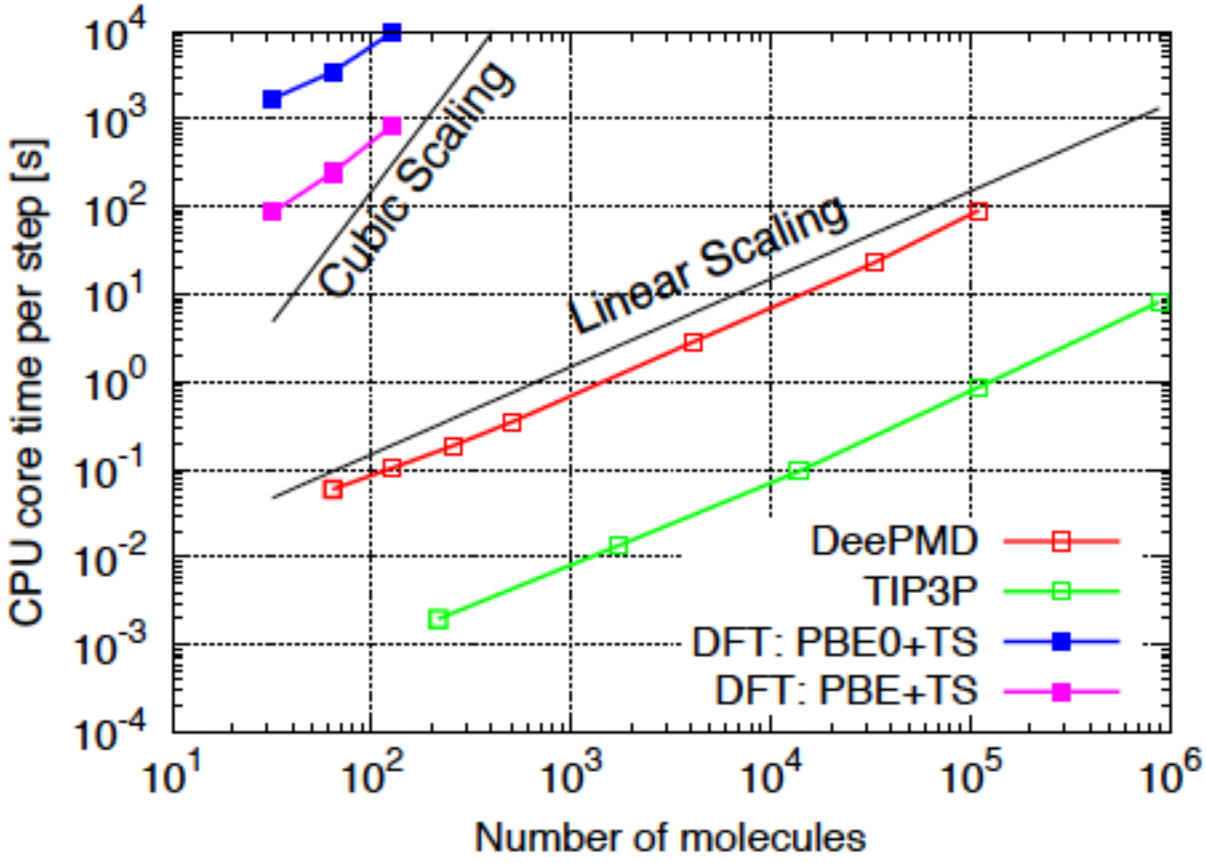






System	$\bar{E}$[eV/H_2O]		$\bar{\rho}$[g/m^3]	
liquid water	-467.678(2)	[-467.679(6)]	1.013(5)	[1.013(20)]
ice Ih (b)	-467.750(1)	[-467.747(4)]	0.967(1)	[0.966(6)]
ice Ih (c)	-468.0478(3)	[-468.0557(16)]	0.950(1)	[0.949(2)]
ice Ih (d)	-468.0942(2)	[-468.1026(9)]	0.986(1)	[0.985(2)]

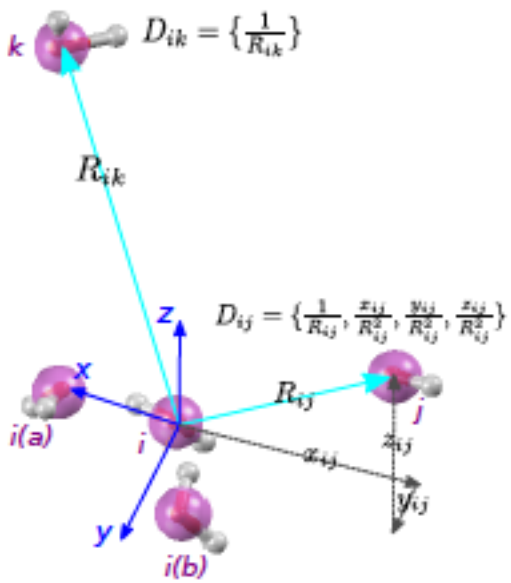
Computational cost



Note: TIP3P is a rigid molecule empirical force field, DeePMD uses flexible molecules and reproduces PI-AIMD results

Coarse graining: DeePCG

Use (again) water as an example and project out the H data. The coarse grained potential is a free energy



$$H(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}\mathcal{M}^{-1}\boldsymbol{p}^T + V(\boldsymbol{q}) \qquad p(\boldsymbol{q}) = \frac{1}{Z_{\boldsymbol{q}}} e^{-\beta V(\boldsymbol{q})}$$

$$p(\boldsymbol{\xi}^{(\mathrm{CG})}) = \frac{1}{Z_{\boldsymbol{q}}} \int e^{-\beta V(\boldsymbol{q})} \delta(\boldsymbol{\xi}^{(\mathrm{CG})}(\boldsymbol{q}) - \boldsymbol{\xi}^{(\mathrm{CG})})\, d\boldsymbol{q}$$

$$U(\boldsymbol{\xi}^{(\mathrm{CG})}) = -\frac{1}{\beta} \ln p(\boldsymbol{\xi}^{(\mathrm{CG})}) \qquad \boldsymbol{F}(\boldsymbol{\xi}^{(\mathrm{CG})}) = -\nabla_{\boldsymbol{\xi}^{(\mathrm{CG})}} U(\boldsymbol{\xi}^{(\mathrm{CG})})$$

While the microscopic potential *V* is available from a AIMD simulation, the coarse-grained potential *U* is not directly available

L. Zhang, et al., in preparation

$$F_i(\boldsymbol{\xi}) = -\partial_i U(\boldsymbol{\xi}) = \langle \mathcal{F}_i(\boldsymbol{q}) \rangle_{\boldsymbol{\xi}=\boldsymbol{\xi}(\boldsymbol{q})}$$

For the instantaneous force estimator we use the formula given by Ciccotti, Kapral, VandenEijnden, *ChemPhysChem* **6**, 1809 (2005)

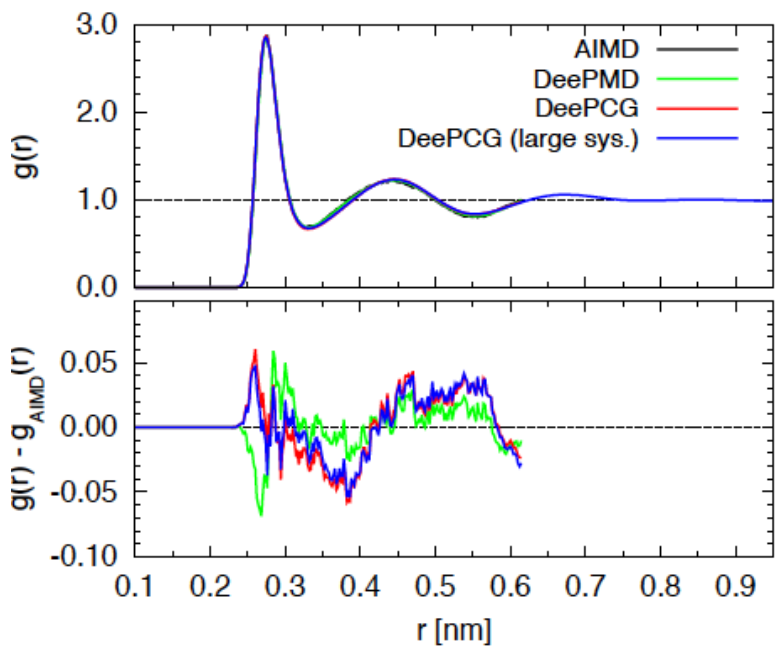
$$\hat{L}(\boldsymbol{w}) = \frac{1}{dDM} \sum_{n=1}^{D} \sum_{i=1}^{dM} \left| F_i(\boldsymbol{\xi}_n) + \partial_i U^{\boldsymbol{w}}(\boldsymbol{\xi}_n) \right|^2$$

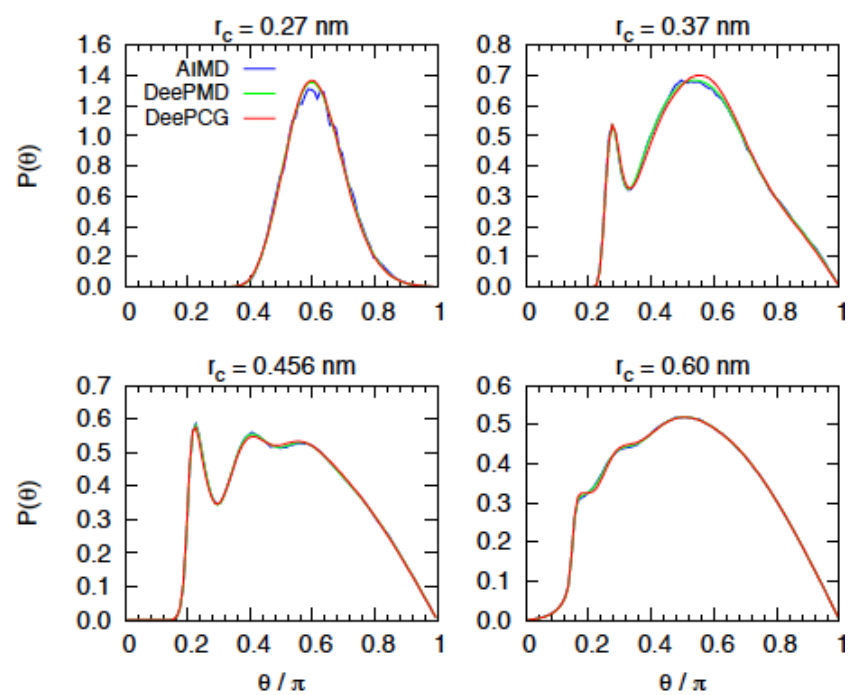
$$\hat{L}^{\mathrm{ins}}(\boldsymbol{w}) = \frac{1}{dDM} \sum_{n=1}^{D} \sum_{i=1}^{dM} \left| \mathcal{F}_i(\boldsymbol{\xi}_n) + \partial_i U^{\boldsymbol{w}}(\boldsymbol{\xi}_n) \right|^2$$

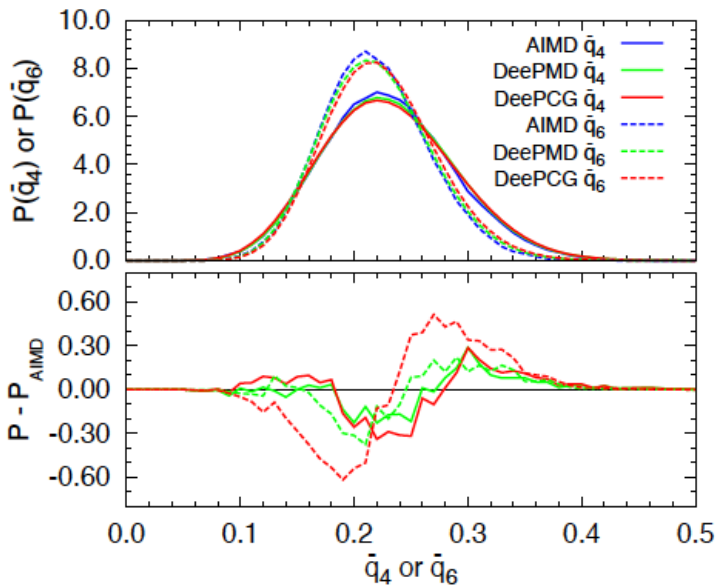
$$\begin{aligned} L^{\mathrm{ins}}(\boldsymbol{w}) &:= \frac{1}{dM} \sum_{i=1}^{dM} \left\langle \left| \mathcal{F}_i(\boldsymbol{\xi}(\boldsymbol{q})) + \partial_i U^{\boldsymbol{w}}(\boldsymbol{\xi}(\boldsymbol{q})) \right|^2 \right\rangle_{\boldsymbol{q}} \\ &\approx \hat{L}^{\mathrm{ins}}(\boldsymbol{w}). \end{aligned}$$

This amounts to an ergodicity requirement for the atomistic system and is always valid if the system samples an equilibrium thermodynamic state.

Very long trajectories are needed to learn the coarse-grained potential, due the HB (and Pauling ice rules) network of water. The solution to this problem is to use DeePMD water as the microscopic model.
15 ns long (*N,V,T*) DeePMD trajectories (PBE0-TS quality) have been used to extract *U*







Higher order correlations

- Equilibrium sampling with DeePCG is accelerated approximately ~8 times relative to DeePMD
- Significantly greater acceleration is expected in studies of crystal nucleation as the ice rules no longer apply (recall mW potential for water)
- It would be interesting to study if the freezing T of *U* is close to that of *V*

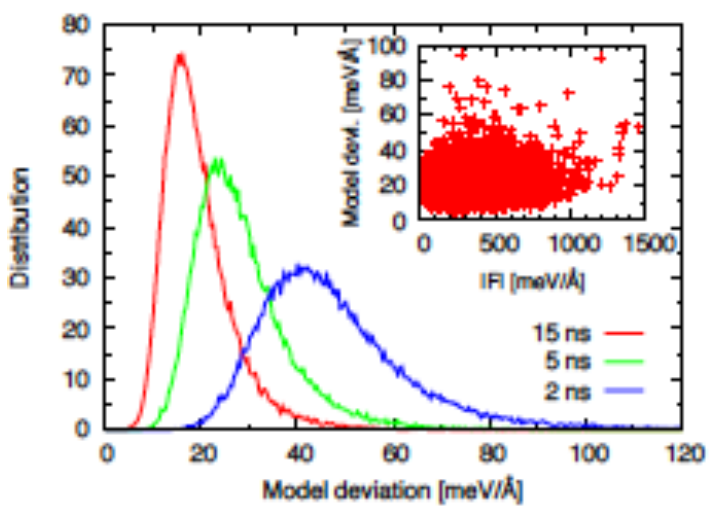
Random noise, multiple minima

The instantaneous force can be seen as the mean force plus noise with zero mean

$$\mathcal{F}_i(q) = F_i(\xi(q)) + R_i(q)$$

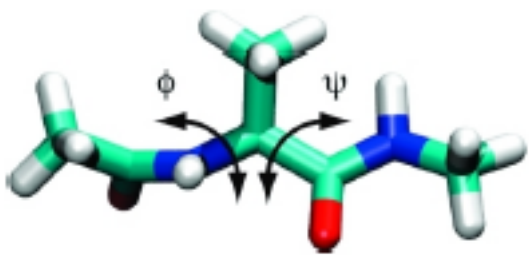
Yet the noise does not seem to affect the ability of DeePCG to learn *U*

Deep NN are highly nonlinear functions of the network parameters and the ADAM optimization converges to different local minima when different initializations are used



$$\Sigma_i = \sqrt{\left\langle \left\| \nabla_i U^{\boldsymbol{w}} - \left\langle \nabla_i U^{\boldsymbol{w}} \right\rangle_{\boldsymbol{w}} \right\|^2 \right\rangle_{\boldsymbol{w}}}$$

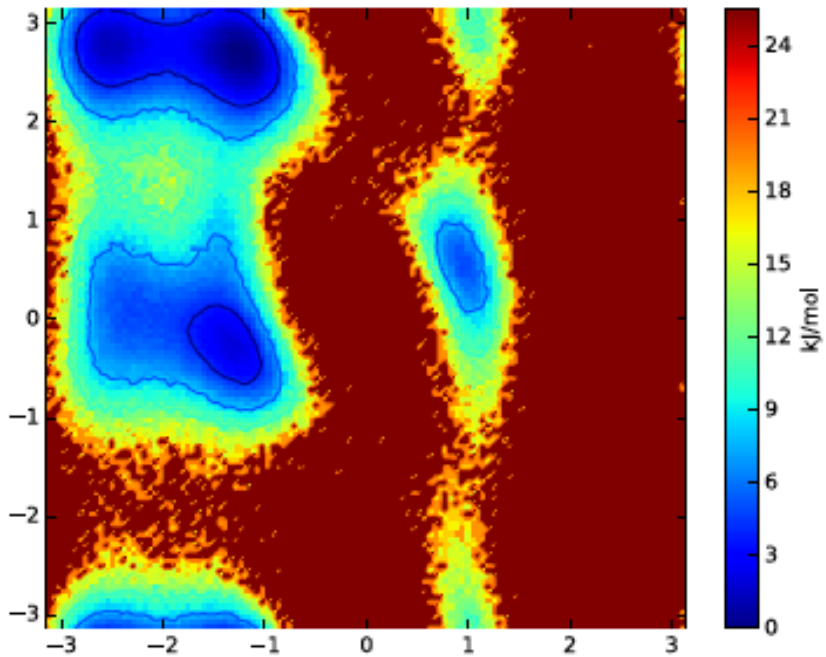
Using DeePCG to represent the Landau free energy surface



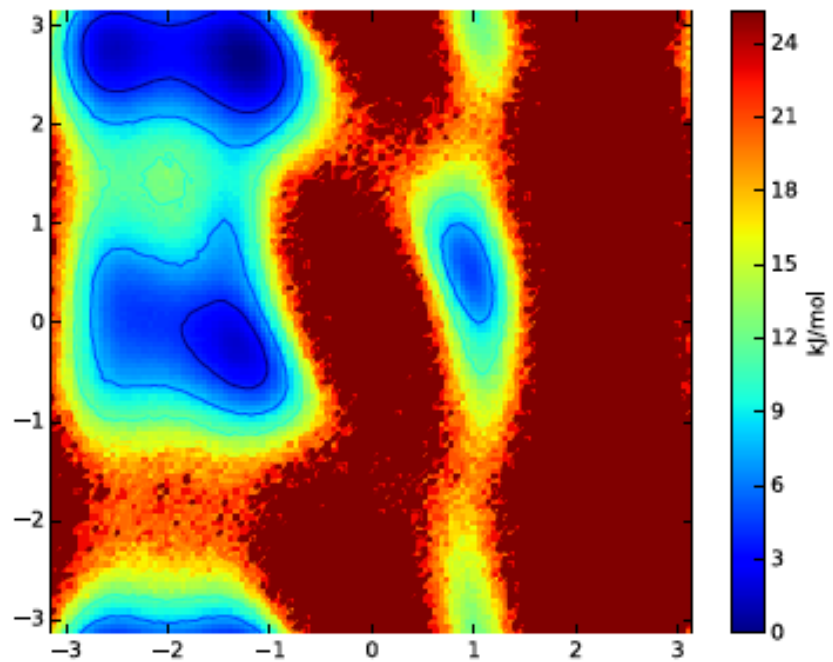
Alanine dipeptide is solvated in 342 TIP3P molecules. The AMBER FF is used for alanine dipeptide.

The microscopic model refers to the model with explicit water. The CG model does not include the solvating water

For illustration purposes the free energy surface is sampled by brute force: 5 μs are used for the microscopic model, 655 ns are used for the CG model. Enhanced sampling methods could be used without difficulty to accelerate convergence.

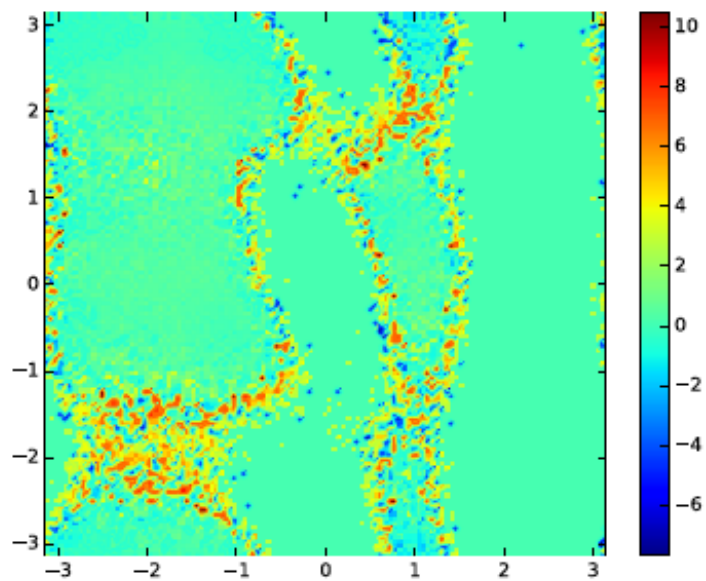


(A) With solvent



(B) Solvent free

Difference A - B



The difference is larger at the edges of a basin because the resolution of our brute force calculation at barriers and saddle points is bad. Much better agreement would be possible by improving the resolution with accurate enhanced sampling methods

Open issues

- **Transferability to different thermodynamic conditions**
- **Non uniform systems: defects and interfaces**
- **Combine consistently potentials obtained by learning in different regions of space for a non uniform system**

Sampling and dynamics

- **The dynamical laws of the microscopic system are known but equations of motion in close deterministic form do not exist for the coarse-grained system**
- **It is an issue of great interest to investigate if dynamical equations can be formulated for the coarse-grained system to approximate the real system dynamics of the coarse-grained variables: this usually involves approximation like time scale separation. Then techniques like the Mori-Zwanzig projection technique can be used**

Acknowledgements

Collaborators

Linfeng Zhang, Jiequn Han, *Program in Applied and Computational Mathematics, Princeton (USA)*

Han Wang, *Institute of Applied Physics and Computational Mathematics, Beijing (China)*

Weinan E, Department of Mathematics and Program in Applied and Computational Mathematics, Princeton (USA) and Center for Data Science, Beijing University (China)

Funding (RC and LZ): DOE-BES