

# Stochastic Multiscale Analysis



Continuum Mechanics Group  
Faculty of Mechanical Engineering  
Technical University Munich

September 4 2014  
Modeling and Numerical Methods for Uncertainty Quantification  
MNMUQ 2014 - Porquerolles Island, France

## Motivation

- Why *Multiscale* ?
  - all physical processes involve a *synergy* of multiple scales
  - often we can/choose to ignore finer-scales and employ coarsened descriptions (e.g. continuum models)
  - in a lot of problems we cannot ignore the fine-scale and we have to work with fine-scale or multi-scale models
- Why *Stochastic*?
  - the finer the scale we consider the less we know
  - when we coarse-grain, we loose information and information loss leads to uncertainty

Coarse-grained models of complex, many-body systems are useful in at least two ways:

- They enhance our understanding of the salient physical/chemical/bio mechanisms
- They enable efficient computations in multiscale/multiphysics regimes

Molecular Dynamics aim to *estimate* numerically *macroscopic* quantities

- equilibrium
- non-equilibrium

Atomistic Simulation is a commonly-used tool in many fields

- biology
- material science
- chemistry <sup>a</sup>
- engineering

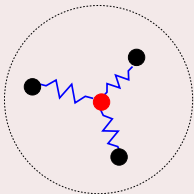
---

<sup>a</sup>2013 Nobel Prize in Chemistry to Karplus & Levitt & Warshel for the “development of multiscale models for complex chemical systems”

Mature software exists

- LAMMPS, GROMACS (free)
- AMBER, CHARMM

If  $\mathbf{q}_i$  is the position of atom  $i$ :



Newton's law:

$$m_i \ddot{\mathbf{q}}_i = \mathbf{f}_i = - \frac{\partial}{\partial \mathbf{q}_i} \sum_{j \neq i} V(\mathbf{q}_i, \mathbf{q}_j)$$

For an isolated system with  $N$  atoms, the total energy is the Hamiltonian  $H$ :

$$H(\mathbf{q}, \mathbf{p}) = \underbrace{\sum_i \frac{1}{2} \frac{\mathbf{p}_i^2}{m_i}}_{\text{kinetic energy}} + \underbrace{V(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N)}_{\text{potential energy}}$$

Hamiltonian equations:

$$\dot{\mathbf{q}}_i = \frac{\partial H}{\partial \mathbf{p}_i}, \quad \dot{\mathbf{p}}_i = - \frac{\partial H}{\partial \mathbf{q}_i}$$

where  $\mathbf{p}_i = m_i \dot{\mathbf{q}}_i$  are the *momenta*

One would need only to integrate in time (chaotic, non-linear equations):

Newtonian equations:

$$m_i \ddot{\mathbf{q}}_i = \mathbf{f}_i = -\frac{\partial V(\mathbf{q})}{\partial \mathbf{q}_i}$$

Hamiltonian equations:

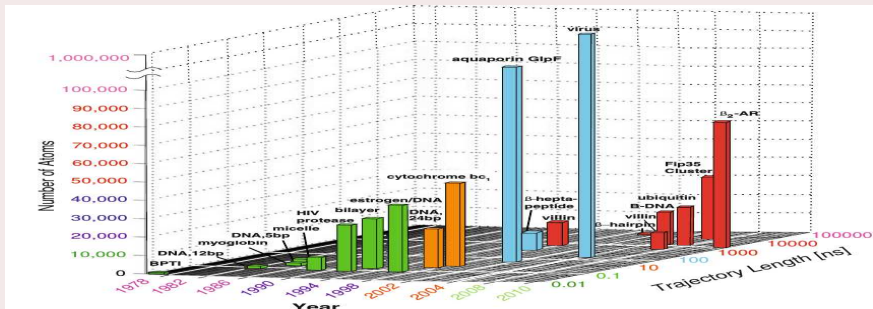
$$\dot{\mathbf{q}}_i = \frac{\partial H}{\partial \mathbf{p}_i}, \quad \dot{\mathbf{p}}_i = -\frac{\partial H}{\partial \mathbf{q}_i}$$

There are two + one problems with that:

- The number of atoms is very large,  $N \rightarrow 10^{23}$
- The dynamics is too fast,  $\Delta t \rightarrow 10^{-12} - 10^{-15}$

There are two + one problems with that:

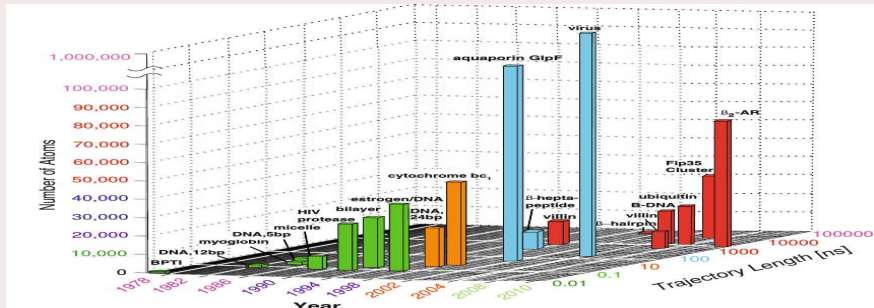
- The number of atoms is very large,  $N \rightarrow 10^{23}$
- The dynamics is too fast,  $\Delta t \rightarrow 10^{-12} - 10^{-15}$



The evolution of MD simulations (from *T. Schlick. Molecular Modeling: An Interdisciplinary Guide, 2010*)

There are two + one problems with that:

- The number of atoms is very large,  $N \rightarrow 10^{23}$
- The dynamics is too fast,  $\Delta t \rightarrow 10^{-12} - 10^{-15}$



The evolution of MD simulations (from *T. Schlick. Molecular Modeling: An Interdisciplinary Guide, 2010*)

- We do not know (exactly) the initial conditions i.e. there is *uncertainty*



- We do not know (exactly) the initial conditions for all DoFs  $\mathbf{q}_i, \mathbf{p}_i, i = 1, 2 \dots N$ .
- We prescribe an initial probability density of microstates  $\rho_0(\mathbf{q}, \mathbf{p})$  which is consistent with the initial *macroscopic* conditions
- This density will evolve in time,  $\rho(\mathbf{q}, \mathbf{p}; t)$ <sup>1</sup> :

$$\frac{\partial \rho}{\partial t} = -\frac{\partial H}{\partial \mathbf{p}} \cdot \frac{\partial \rho}{\partial \mathbf{q}} + \frac{\partial H}{\partial \mathbf{q}} \cdot \frac{\partial \rho}{\partial \mathbf{p}} \quad (\text{Liouville's equation})$$

- Is there an equilibrium/stationary distribution  $\rho_{\text{eq}}(\mathbf{q}, \mathbf{p}) = \lim_{t \rightarrow \infty} \rho(\mathbf{q}, \mathbf{p}; t)$ ?

---

<sup>1</sup>R. Zwanzig, Nonequilibrium Statistical Mechanics

- We do not know (exactly) the initial conditions for all DoFs  $\mathbf{q}_i, \mathbf{p}_i, i = 1, 2 \dots N$ .
- We prescribe an initial probability density of microstates  $\rho_0(\mathbf{q}, \mathbf{p})$  which is consistent with the initial *macroscopic* conditions
- This density will evolve in time,  $\rho(\mathbf{q}, \mathbf{p}; t)$ <sup>1</sup> :

$$\frac{\partial \rho}{\partial t} = -\frac{\partial H}{\partial \mathbf{p}} \cdot \frac{\partial \rho}{\partial \mathbf{q}} + \frac{\partial H}{\partial \mathbf{q}} \cdot \frac{\partial \rho}{\partial \mathbf{p}} \quad (\text{Liouville's equation})$$

- Is there an equilibrium/stationary distribution  $\rho_{\text{eq}}(\mathbf{q}, \mathbf{p}) = \lim_{t \rightarrow \infty} \rho(\mathbf{q}, \mathbf{p}; t)$ ?

---

<sup>1</sup>R. Zwanzig, Nonequilibrium Statistical Mechanics

## Equilibrium Statistical Mechanics

- Microcanonical ensemble (isolated system):

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) \propto \delta(H(\mathbf{q}, \mathbf{p}) - H_0)$$

- Canonical ensemble (system in a heat bath with temperature  $T$ ):

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) = \frac{\exp\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\}}{Z}$$

- Grand-canonical ensemble ...

## Equilibrium Statistical Mechanics

- Microcanonical ensemble (isolated system):

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) \propto \delta(H(\mathbf{q}, \mathbf{p}) - H_0)$$

- Canonical ensemble (system in a heat bath with temperature  $T$ ):

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) = \frac{\exp\left\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\right\}}{Z}$$

- Grand-canonical ensemble ...

## Equilibrium Statistical Mechanics

- Microcanonical ensemble (isolated system):

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) \propto \delta(H(\mathbf{q}, \mathbf{p}) - H_0)$$

- Canonical ensemble (system in a heat bath with temperature  $T$ ):

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) = \frac{\exp\left\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\right\}}{Z}$$

- Grand-canonical ensemble . . .

## Equilibrium Statistical Mechanics

- Microcanonical ensemble (isolated system):

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) \propto \delta(H(\mathbf{q}, \mathbf{p}) - H_0)$$

- Canonical ensemble (system in a heat bath with temperature  $T$ ):

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) = \frac{\exp\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\}}{Z}$$

- Grand-canonical ensemble ...

## Maximum Entropy connection <sup>a</sup>

<sup>a</sup>E.T. Jaynes, Information Theory and Statistical Mechanics, 1957

- Entropy increases with time  $t$
- The equilibrium density is the one that maximizes the entropy:

$$S = - \int \rho_{eq}(\mathbf{q}, \mathbf{p}) \log \rho_{eq}(\mathbf{q}, \mathbf{p}) d\mathbf{q}d\mathbf{p}$$

The objective of equilibrium statistical mechanics is to compute *averages* :

- E.g.:

$$E[f(\mathbf{q}, \mathbf{p})] = \int f(\mathbf{q}, \mathbf{p}) \rho_{eq}(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p}$$

- Other notations:  $\langle f(\mathbf{q}, \mathbf{p}) \rangle$ ,  $E_{eq}[f]$  ....
- These averages correspond to *macroscopic* thermodynamic quantities e.g.:
  - stress,
  - pressure
  - diffusivity,
  - viscosity,
  - specific heat etc.

The objective of equilibrium statistical mechanics is to compute *averages* :

$$E[f(\mathbf{q}, \mathbf{p})] = \int f(\mathbf{q}, \mathbf{p}) \rho_{eq}(\mathbf{q}, \mathbf{p}) d\mathbf{q}d\mathbf{p}$$

- Deterministic methods

- Generate a trajectory  $(\mathbf{q}(t), \mathbf{p}(t), \mathbf{x}(t))$
- The auxiliary variables  $\mathbf{x}(t)$  are appropriately selected to ensure states consistent with the canonical ensemble
- Typical examples: Nosé-Hoover, Nosé-Poincaré

- Stochastic Methods (Monte Carlo)

- Generate *random samples*  $\mathbf{q}, \mathbf{p}$ , or *sample trajectories*  $(\mathbf{q}(t), \mathbf{p}(t))$  from appropriately selected Stochastic PDEs.
- Typical examples: Importance Sampling, Markov Chain Monte Carlo, Sequential Monte Carlo
- Strong convergence results, increased flexibility that can incorporate deterministic methods.



The objective of equilibrium statistical mechanics is to compute *averages* :

$$E[f(\mathbf{q}, \mathbf{p})] = \int f(\mathbf{q}, \mathbf{p}) \rho_{eq}(\mathbf{q}, \mathbf{p}) d\mathbf{q}d\mathbf{p}$$

- Deterministic methods

- Generate a trajectory  $(\mathbf{q}(t), \mathbf{p}(t), \mathbf{x}(t))$
- The auxiliary variables  $\mathbf{x}(t)$  are appropriately selected to ensure states consistent with the canonical ensemble
- Typical examples: Nosé-Hoover, Nosé-Poincaré

- Stochastic Methods (Monte Carlo)

- Generate *random samples*  $\mathbf{q}_i, \mathbf{p}_i$  or *sample trajectories*  $(\mathbf{q}(t), \mathbf{p}(t))$  from appropriately selected Stochastic PDEs.
- Typical examples: Importance Sampling, Markov Chain Monte Carlo, Sequential Monte Carlo
- Strong convergence results, increased flexibility that can incorporate deterministic methods.

## Some considerations for Monte Carlo samplers:

- Suppose we wish to sample from:

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) = \frac{\exp\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\}}{Z}$$

where  $H(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^N \frac{1}{2} \frac{\mathbf{p}_i^2}{m_i} + V(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N)$

- $N$  is very large i.e. we can afford algorithms that scale  $O(N)$  but probably not  $O(N^2)$
- Correlations between  $\mathbf{q}$  and  $\mathbf{p}$  are strong (i.e. no random-walk scheme would work)
- Sampling  $\mathbf{p}$  is easy (i.e. multivariate Gaussian)

## Some considerations for Monte Carlo samplers:

- Suppose we wish to sample from:

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) = \frac{\exp\left\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\right\}}{Z}$$

where  $H(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^N \frac{1}{2} \frac{\mathbf{p}_i^2}{m_i} + V(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N)$

- $N$  is very large i.e. we can afford algorithms that scale  $O(N)$  but probably not  $O(N^2)$
- Correlations between  $\mathbf{q}$  and  $\mathbf{p}$  are strong (i.e. no random-walk scheme would work)
- Sampling  $\mathbf{p}$  is easy (i.e. multivariate Gaussian)

## Some considerations for Monte Carlo samplers:

- Suppose we wish to sample from:

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) = \frac{\exp\left\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\right\}}{Z}$$

where  $H(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^N \frac{1}{2} \frac{\mathbf{p}_i^2}{m_i} + V(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N)$

- $N$  is very large i.e. we can afford algorithms that scale  $O(N)$  but probably not  $O(N^2)$
- Correlations between  $\mathbf{q}$  and  $\mathbf{p}$  are strong (i.e. no random-walk scheme would work)
- Sampling  $\mathbf{p}$  is easy (i.e. multivariate Gaussian)

## Some considerations for Monte Carlo samplers:

- Suppose we wish to sample from:

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) = \frac{\exp\left\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\right\}}{Z}$$

where  $H(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^N \frac{1}{2} \frac{\mathbf{p}_i^2}{m_i} + V(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N)$

- $N$  is very large i.e. we can afford algorithms that scale  $O(N)$  but probably not  $O(N^2)$
- Correlations between  $\mathbf{q}$  and  $\mathbf{p}$  are strong (i.e. no random-walk scheme would work)
- Sampling  $\mathbf{p}$  is easy (i.e. multivariate Gaussian)

## Some considerations for Monte Carlo samplers:

- Suppose we wish to sample from:

$$\rho_{\text{eq}}(\mathbf{q}, \mathbf{p}) = \frac{\exp\left\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\right\}}{Z}$$

where  $H(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^N \frac{1}{2} \frac{\mathbf{p}_i^2}{m_i} + V(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N)$

- Techniques that employ first-order derivatives of  $H$  are generally employed. Their cost is generally determined by the number of times force evaluations have to be performed i.e.  $\frac{\partial V}{\partial \mathbf{q}}$ .
- Increasing the temperature  $T$  "flattens" the target density
- $\rho_{\text{eq}}$  is generally *multi-modal*. It is very difficult for usual MCMC schemes to escape these modes.

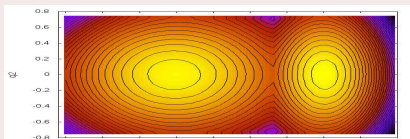
## Some considerations for Monte Carlo samplers:

- Suppose we wish to sample from:

$$\rho_{\text{eq}}(\mathbf{q}, \mathbf{p}) = \frac{\exp\left\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\right\}}{Z}$$

where  $H(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^N \frac{1}{2} \frac{\mathbf{p}_i^2}{m_i} + V(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N)$

- Techniques that employ first-order derivatives of  $H$  are generally employed. Their cost is generally determined by the number of times force evaluations have to be performed i.e.  $\frac{\partial V}{\partial \mathbf{q}}$ .
- Increasing the temperature  $T$  “flattens” the target density
- $\rho_{\text{eq}}$  is generally *multi-modal*. It is very difficult for usual MCMC schemes to escape these modes.



## Some considerations for Monte Carlo samplers:

- Suppose we wish to sample from:

$$\rho_{eq}(\mathbf{q}, \mathbf{p}) = \frac{\exp\left\{-\frac{1}{k_B T} H(\mathbf{q}, \mathbf{p})\right\}}{Z}, \quad H(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^N \frac{1}{2} \frac{\mathbf{p}_i^2}{m_i} + V(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N)$$

## Hybrid or Hamiltonian Monte Carlo (R. Neal, 2012)

Given the current state  $(\mathbf{q}_{old}, \mathbf{p}_{old})$

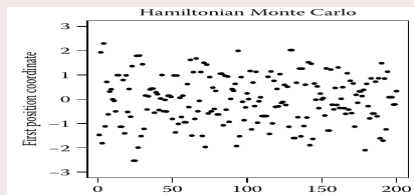
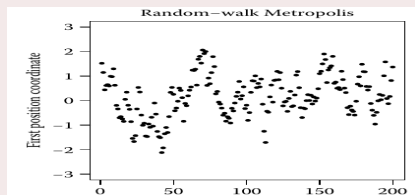
- Draw new  $\hat{\mathbf{p}}$  from multivariate Gaussian
- Perform  $L$  time-integration steps with your favorite, explicit (volume-perserving, reversible) integrator e.g. velocity Verlet:

$$(\mathbf{q}_{old}, \hat{\mathbf{p}}) \rightarrow \dots \underbrace{\left[ \begin{array}{l} \dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}} \\ \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}} \end{array} \right]}_{L\text{-steps}} \dots \rightarrow (\mathbf{q}_{new}, \mathbf{p}_{new})$$

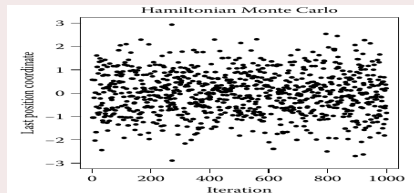
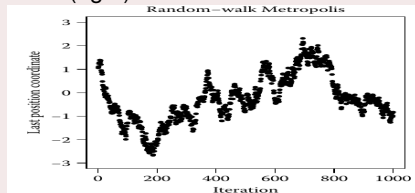
- M-H accept/reject based on:  $\exp\left\{-\frac{1}{k_B T} (H(\mathbf{q}_{new}, \mathbf{p}_{new}) - H(\mathbf{q}_{old}, \hat{\mathbf{p}}))\right\}$



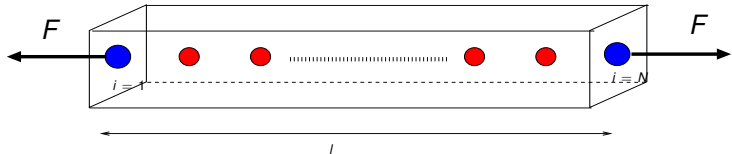
## Hybrid/Hamiltonian Monte Carlo (R. Neal, 2012)



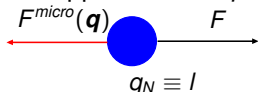
Samples of  $q_1$  for a **2-dimensional** problem using Random Walk Metropolis (left) and HMC (right)



Samples of  $q_1$  for a **100-dimensional** problem using Random Walk Metropolis (left) and HMC (right)



- Consider a very long ( $N \gg 1$ ) one-dimensional chain of atoms at equilibrium temperature  $T$
- We want to predict the *macroscopic* behavior of the chain i.e. the (force)  $F$  - (length)  $l$  relation.
- Suppose we fix  $q_1 \equiv 0$  and  $l = l(\mathbf{q}) = q_N$ .



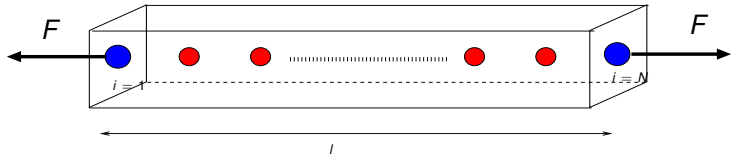
$$\begin{aligned}
 F &= - \langle \mathbf{F}^{micro} \rangle_{q_N=l} = \left\langle \frac{\partial V(\mathbf{q})}{\partial q_N} \right\rangle_{q_N=l} \\
 &= \int \frac{\partial V(\mathbf{q})}{\partial q_N} \rho_{eq}(\mathbf{q} | q_N = l) d\mathbf{q}
 \end{aligned}$$

- What is  $\rho_{eq}(\mathbf{q} | q_N = l)$ ?

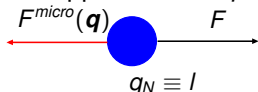
● Let:

$$Z(l) = \int \rho_{eq}(\mathbf{q}) \delta(l - q_N) d\mathbf{q}$$

i.e. the probability that the length of chain is  $l$ .



- Consider a very long ( $N \gg 1$ ) one-dimensional chain of atoms at equilibrium temperature  $T$
- We want to predict the *macroscopic* behavior of the chain i.e. the (force)  $F$  - (length)  $l$  relation.
- Suppose we fix  $q_1 \equiv 0$  and  $l = l(\mathbf{q}) = q_N$ .



$$F = - \langle \mathbf{F}^{micro} \rangle_{q_N=l} = \left\langle \frac{\partial V(\mathbf{q})}{\partial q_N} \right\rangle_{q_N=l}$$

$$= \int \frac{\partial V(\mathbf{q})}{\partial q_N} \rho_{eq}(\mathbf{q} | q_N = l) d\mathbf{q}$$

- What is  $\rho_{eq}(\mathbf{q} | q_N = l)$ ?
- Let:

$$Z(l) = \int \rho_{eq}(\mathbf{q}) \delta(l - q_N) d\mathbf{q}$$

i.e. the *probability that the length of chain is  $l$ .*

- Then:

$$\rho_{eq}(\mathbf{q}|q_N = l) = \frac{1}{Z(l)} \rho_{eq}(\mathbf{q}) \delta(l - q_N) = \frac{1}{Z(l)} e^{-\beta V(\mathbf{q})} \delta(l - q_N)$$

- Note that:

$$\begin{aligned} F &= \int \frac{\partial V(\mathbf{q})}{\partial q_N} \rho_{eq}(\mathbf{q}|q_N = l) d\mathbf{q} \\ &= \int \frac{\partial V(\mathbf{q})}{\partial q_N} \frac{e^{-\beta V(\mathbf{q})} \delta(l - q_N)}{Z(l)} d\mathbf{q} \\ &= \frac{\partial}{\partial q_N} \left( -\frac{1}{\beta} \log Z(l) \right) \quad (\beta = \frac{1}{k_B T}) \end{aligned}$$

- The quantity:

$$A(l) = -\frac{1}{\beta} \log Z(l) = -\frac{1}{\beta} \log \int e^{-\beta V(\mathbf{q})} \delta(l - q_N) d\mathbf{q}$$

is called **Helmholtz Free Energy**.

## Helmholtz Free Energy

$$A(l) = -\frac{1}{\beta} \log Z(l) = -\frac{1}{\beta} \log \int e^{-\beta V(\mathbf{q})} \delta(l - q_N) d\mathbf{q}$$

- Note that:

$$e^{-\beta A(l)} = Z(l) \equiv \text{probability that the length of chain is } l$$

- Hence  $A(l)$  is the *effective potential* for the *marginal distribution* of the *macroscopic* variable  $l$ .
- Since  $A(l)$  gives the *exact* marginal of  $l$ , there is no coarse-graining error i.e. we do not lose any information by “integrating out” the atomistic degrees of freedom.
- One can, in principle, compute free energies w.r.t to any macroscopic variables  $Q = \xi(\mathbf{q})$  (e.g. strain) i.e.:

$$A(Q) = -\frac{1}{\beta} \log \int \delta(Q - \xi(\mathbf{q})) e^{-\beta V(\mathbf{q})} d\mathbf{q}$$

## Helmholtz Free Energy

$$A(l) = -\frac{1}{\beta} \log Z(l) = -\frac{1}{\beta} \log \int e^{-\beta V(\mathbf{q})} \delta(l - q_N) d\mathbf{q}$$

- Note that:

$$e^{-\beta A(l)} = Z(l) \equiv \text{probability that the length of chain is } l$$

- Hence  $A(l)$  is the *effective potential* for the *marginal distribution* of the *macroscopic* variable  $l$ .
- Since  $A(l)$  gives the **exact** marginal of  $l$ , there is no coarse-graining error i.e. we do not lose any information by “integrating out” the atomistic degrees of freedom.
- One can, in principle, compute free energies w.r.t to any macroscopic variables  $Q = \xi(\mathbf{q})$  (e.g. strain) i.e.:

$$A(Q) = -\frac{1}{\beta} \log \int \delta(Q - \xi(\mathbf{q})) e^{-\beta V(\mathbf{q})} d\mathbf{q}$$

## Helmholtz Free Energy

$$A(l) = -\frac{1}{\beta} \log Z(l) = -\frac{1}{\beta} \log \int e^{-\beta V(\mathbf{q})} \delta(l - q_N) d\mathbf{q}$$

- Note that:

$$e^{-\beta A(l)} = Z(l) \equiv \text{probability that the length of chain is } l$$

- Hence  $A(l)$  is the *effective potential* for the *marginal distribution* of the *macroscopic* variable  $l$ .
- Since  $A(l)$  gives the **exact** marginal of  $l$ , there is no coarse-graining error i.e. we do not lose any information by “integrating out” the atomistic degrees of freedom.
- One can, in principle, compute free energies w.r.t to any macroscopic variables  $\mathbf{Q} = \xi(\mathbf{q})$  (e.g. strain) i.e.:

$$A(\mathbf{Q}) = -\frac{1}{\beta} \log \int \delta(\mathbf{Q} - \xi(\mathbf{q})) e^{-\beta V(\mathbf{q})} d\mathbf{q}$$

## Free Energy comprises contributions from:

- Energy/Potential
- Entropy

## Energetic contributions

- Suppose:

$$V(\mathbf{q}) = W(q_N) + \hat{V}(q_1, \dots, q_{N-1})$$

- Then for  $Q = q_N$ :

$$\begin{aligned} A(Q) &= -\frac{1}{\beta} \log \int e^{-\beta V(\mathbf{q})} \delta(Q - q_N) d\mathbf{q} \\ &= -\frac{1}{\beta} \log e^{-\beta W(Q)} \int e^{-\beta \hat{V}(q_1, \dots, q_{N-1})} dq_1 \dots dq_{N-1} \\ &= W(Q) \end{aligned}$$

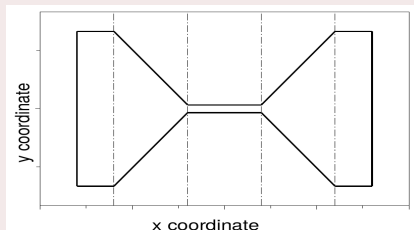
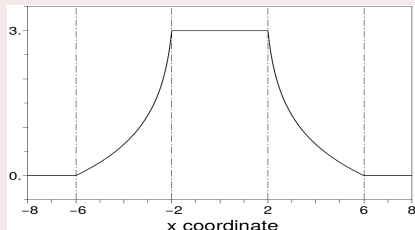


## Free Energy comprises contributions from:

- Energy/Potential
- Entropy

## Entropic contributions

- Suppose  $V(x, y)$  is 0 inside the box and  $+\infty$  outside. For  $Q = x$ :

(a)  $V(x, y)$ (b)  $A(x)$ 

From Lelievre et al. 2010. *Free Energy Computations: A mathematical perspective*

## Some considerations:

$$A(\mathbf{Q}) = -\frac{1}{\beta} \log \int \delta(\mathbf{Q} - \xi(\mathbf{q})) e^{-\beta V(\mathbf{q})} d\mathbf{q}$$

- The Helmholtz Free Energy  $A(\mathbf{Q})$  provides a **principled** and **exact** representation of the equilibrium atomistic ensemble w.r.t the *reduced coordinates*  $\mathbf{Q} = \xi(\mathbf{q})$ .
- If  $A(\mathbf{Q})$  is known we can reproduce **exactly** any macroscopic property depending on  $\mathbf{Q}$ .
- $A(\mathbf{Q})$  is really  $A(\mathbf{Q}, \beta)$  i.e. it **depends on the temperature**
- Free energy **differences** are of interest.

Helmholtz Free energy  $A(\mathbf{Q})$ :

$$A(\mathbf{Q}) = -\frac{1}{\beta} \log \int \delta(\mathbf{Q} - \xi(\mathbf{q})) e^{-\beta V(\mathbf{q})} d\mathbf{q}$$

- How can one estimate  $A(\mathbf{Q})$ ?

- Suppose  $\mathbf{Q} \equiv q_1$  and  $p_{\text{eq}}(q_1, q_2)$  is bimodal.

(a)

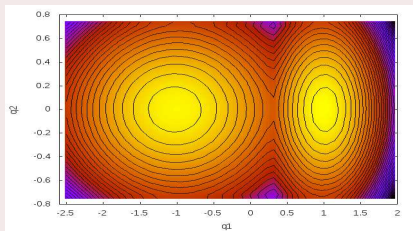
$p_{\text{eq}}(q_1, q_2)$

- Convergence is *very, very* slow -  $q_1$  is the *slow* variable

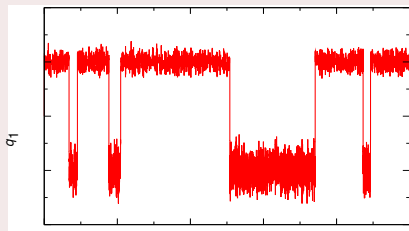
## Helmholtz Free energy $A(\mathbf{Q})$ :

$$A(\mathbf{Q}) = -\frac{1}{\beta} \log \int \delta(\mathbf{Q} - \xi(\mathbf{q})) e^{-\beta V(\mathbf{q})} d\mathbf{q}$$

- How can one estimate  $A(\mathbf{Q})$ ?
- Suppose  $\mathbf{Q} \equiv \mathbf{q}_1$  and  $\rho_{eq}(q_1, q_2)$  is bimodal.



(b)  $\rho_{eq}(q_1, q_2)$



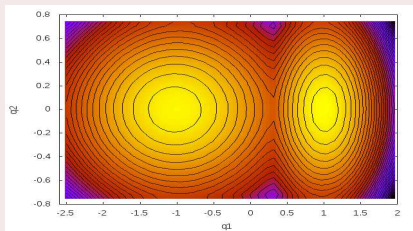
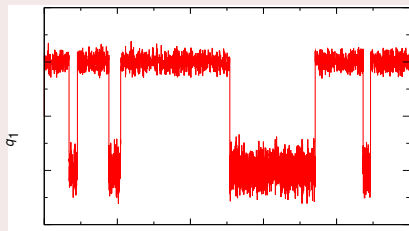
(c) Metastability

- Convergence is *very, very* slow -  $q_1$  is the *slow* variable

Helmholtz Free energy  $A(\mathbf{Q})$ :

$$A(\mathbf{Q}) = -\frac{1}{\beta} \log \int \delta(\mathbf{Q} - \xi(\mathbf{q})) e^{-\beta V(\mathbf{q})} d\mathbf{q}$$

- How can one estimate  $A(\mathbf{Q})$ ?
- Suppose  $\mathbf{Q} \equiv \mathbf{q}_1$  and  $\rho_{eq}(q_1, q_2)$  is bimodal.

(d)  $\rho_{eq}(q_1, q_2)$ 

(e) Metastability

- Convergence is **very, very** slow -  $q_1$  is the *slow* variable

## Coarse-grained variables $\mathbf{Q} = \xi(\mathbf{q})$ :

- Other names: relevant variables, macrostates, macroscopic variables, collective variables, *reaction coordinates*, internal variables, order parameters etc.
- How does one find the most appropriate coarse-grained variables  $\mathbf{Q}$  <sup>2</sup>?
  - based on the macroscopic prediction objectives.
  - based on physical insight
  - based on (principled) processing of computational data (*machine learning*)
- In the following we will assume that  $\mathbf{Q} = \xi(\mathbf{q})$  is given.

---

<sup>2</sup>Rohrdanz et al. 2013 *Discovering mountain passes via torchlight: methods for the definition of reaction coordinates and pathways in complex macromolecular reactions*

## Coarse-grained variables $\mathbf{Q} = \xi(\mathbf{q})$ :

- Other names: relevant variables, macrostates, macroscopic variables, collective variables, *reaction coordinates*, internal variables, order parameters etc.
- How does one find the most appropriate coarse-grained variables  $\mathbf{Q}$  <sup>a?</sup>
  - based on the macroscopic prediction objectives.
  - based on physical insight
  - based on (principled) processing of computational data (*machine learning*)
- In the following we will assume that  $\mathbf{Q} = \xi(\mathbf{q})$  is given.

---

<sup>a</sup>Rohrdanz et al. 2013 *Discovering mountain passes via torchlight: methods for the definition of reaction coordinates and pathways in complex macromolecular reactions*

Given *good*  $\mathbf{Q} = \xi(\mathbf{q})$  the computation of free energy (differences)  $A(\mathbf{Q})$  amounts to:

- sampling *efficiently* a multi-modal density
- computing the marginal w.r.t.  $\mathbf{Q}$ .

Some techniques:

- Thermodynamic integration (Kirkwood 1935)
  - Equilibrium methods
  - Nonequilibrium methods (Jarzynski 1997)
  - **Adaptive Biasing Potential (ABP) or Force (ABF)** methods (Wang-Landau 2001)
- 
- Keep in mind that a lot of these techniques can be used in completely different contexts in UQ e.g. Bayesian inference (Chopin et al. 2012)



# A unifying ABP framework for Free Energy Computations

- Recall that the *marginal*  $\rho_{\mathbf{Q}}(\mathbf{Q})$  w.r.t the coarse-grained variables  $\mathbf{Q} = \xi(\mathbf{q})$  is:

$$\rho_{\mathbf{Q}}(\mathbf{Q}) \propto e^{-\beta A(\mathbf{Q})} = \int e^{-\beta V(\mathbf{q})} \delta(\mathbf{Q} - \xi(\mathbf{q})) d\mathbf{q}$$

- Define an auxiliary density  $\hat{\rho}(\mathbf{q})$  as:

$$\hat{\rho}(\mathbf{q}) = \frac{e^{-\beta(V(\mathbf{q}) - \hat{A}(\xi(\mathbf{q}); \phi))}}{Z(\phi)}$$

where  $\hat{A}(\mathbf{Q}; \phi)$  is a function of the coarse-grained variables  $\mathbf{Q}$  that is parameterized by  $\phi$ .

- The corresponding marginal  $\hat{\rho}_{\mathbf{Q}}(\mathbf{Q})$  will be:

$$\begin{aligned} \hat{\rho}_{\mathbf{Q}}(\mathbf{Q}) &= \int \hat{\rho}(\mathbf{q}) \delta(\mathbf{Q} - \xi(\mathbf{q})) d\mathbf{q} \\ &= \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)} \end{aligned}$$

- Recall that the *marginal*  $\rho_{\mathbf{Q}}(\mathbf{Q})$  w.r.t the coarse-grained variables  $\mathbf{Q} = \xi(\mathbf{q})$  is:

$$\rho_{\mathbf{Q}}(\mathbf{Q}) \propto e^{-\beta A(\mathbf{Q})} = \int e^{-\beta V(\mathbf{q})} \delta(\mathbf{Q} - \xi(\mathbf{q})) d\mathbf{q}$$

- Define an auxiliary density  $\hat{\rho}(\mathbf{q})$  as:

$$\hat{\rho}(\mathbf{q}) = \frac{e^{-\beta(V(\mathbf{q}) - \hat{A}(\xi(\mathbf{q}); \phi))}}{Z(\phi)}$$

where  $\hat{A}(\mathbf{Q}; \phi)$  is a function of the coarse-grained variables  $\mathbf{Q}$  that is parameterized by  $\phi$ .

- The corresponding marginal  $\hat{\rho}_{\mathbf{Q}}(\mathbf{Q})$  will be:

$$\begin{aligned} \hat{\rho}_{\mathbf{Q}}(\mathbf{Q}) &= \int \hat{\rho}(\mathbf{q}) \delta(\mathbf{Q} - \xi(\mathbf{q})) d\mathbf{q} \\ &= \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)} \end{aligned}$$

- Note that if  $A(\mathbf{Q}) = \hat{A}(\mathbf{Q}; \phi) + \text{constant}$  then:

$$\hat{\rho}_{\mathbf{Q}}(\mathbf{Q}) = \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)} \equiv (\text{uniform})$$

- In practice we define a (small or large) subset of  $\mathbf{Q}$ -values, say  $\mathcal{D}$ , that is of interest. Let  $\pi(\mathbf{Q}) \propto \mathbf{1}_{\mathcal{D}}(\mathbf{Q})$  be the uniform on  $\mathcal{D}$

- Hence the optimal free energy approximation  $\hat{A}(\mathbf{Q}; \phi)$  is found by:

$$\phi^{opt} = \arg \min_{\phi} f(\phi) = \text{DISTANCE}(\underbrace{\pi(\mathbf{Q})}_{\text{uniform on } \mathcal{D}}, \hat{\rho}_{\mathbf{Q}}(\mathbf{Q}))$$

- We propose using the Kullback-Leibler divergence to measure the distance between densities i.e.:

$$\begin{aligned} f(\phi) = \text{KL}(\pi(\mathbf{Q}) || \hat{\rho}_{\mathbf{Q}}(\mathbf{Q})) &= - \int \pi(\mathbf{Q}) \log \frac{\hat{\rho}_{\mathbf{Q}}(\mathbf{Q})}{\pi(\mathbf{Q})} d\mathbf{Q} \\ &= - \int \pi(\mathbf{Q}) \left( -\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi)) \right) d\mathbf{Q} + \log Z(\phi) + \dots \end{aligned}$$

- Note that if  $A(\mathbf{Q}) = \hat{A}(\mathbf{Q}; \phi) + \text{constant}$  then:

$$\hat{\rho}_{\mathbf{Q}}(\mathbf{Q}) = \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)} \equiv (\text{uniform})$$

- In practice we define a (small or large) subset of  $\mathbf{Q}$ -values, say  $\mathcal{D}$ , that is of interest. Let  $\pi(\mathbf{Q}) \propto \mathbf{1}_{\mathcal{D}}(\mathbf{Q})$  be the uniform on  $\mathcal{D}$
- Hence the *optimal* free energy approximation  $\hat{A}(\mathbf{Q}; \phi)$  is found by:

$$\phi^{opt} = \arg \min_{\phi} f(\phi) = \text{DISTANCE}(\underbrace{\pi(\mathbf{Q})}_{\text{uniform on } \mathcal{D}}, \hat{\rho}_{\mathbf{Q}}(\mathbf{Q}))$$

- We propose using the Kullback-Leibler divergence to measure the distance between densities i.e.:

$$\begin{aligned} f(\phi) = \text{KL}(\pi(\mathbf{Q}) \parallel \hat{\rho}_{\mathbf{Q}}(\mathbf{Q})) &= - \int \pi(\mathbf{Q}) \log \frac{\hat{\rho}_{\mathbf{Q}}(\mathbf{Q})}{\pi(\mathbf{Q})} d\mathbf{Q} \\ &= - \int \pi(\mathbf{Q}) \left( -\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi)) \right) d\mathbf{Q} + \log Z(\phi) + \dots \end{aligned}$$

# A unifying ABP framework for Free Energy Computations

- Note that if  $A(\mathbf{Q}) = \hat{A}(\mathbf{Q}; \phi) + \text{constant}$  then:

$$\hat{\rho}_{\mathbf{Q}}(\mathbf{Q}) = \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)} \equiv (\text{uniform})$$

- In practice we define a (small or large) subset of  $\mathbf{Q}$ -values, say  $\mathcal{D}$ , that is of interest. Let  $\pi(\mathbf{Q}) \propto \mathbf{1}_{\mathcal{D}}(\mathbf{Q})$  be the uniform on  $\mathcal{D}$
- Hence the *optimal* free energy approximation  $\hat{A}(\mathbf{Q}; \phi)$  is found by:

$$\phi^{opt} = \arg \min_{\phi} f(\phi) = \text{DISTANCE}(\underbrace{\pi(\mathbf{Q})}_{\text{uniform on } \mathcal{D}}, \hat{\rho}_{\mathbf{Q}}(\mathbf{Q}))$$

- We propose using the Kullback-Leibler divergence to measure the distance between densities i.e.:

$$\begin{aligned} f(\phi) = KL(\pi(\mathbf{Q}) || \hat{\rho}_{\mathbf{Q}}(\mathbf{Q})) &= - \int \pi(\mathbf{Q}) \log \frac{\hat{\rho}_{\mathbf{Q}}(\mathbf{Q})}{\pi(\mathbf{Q})} d\mathbf{Q} \\ &= - \int \pi(\mathbf{Q}) \left( -\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi)) \right) d\mathbf{Q} + \log Z(\phi) + \dots \\ &= -\beta \int \pi(\mathbf{Q}) \hat{A}(\mathbf{Q}; \phi) d\mathbf{Q} + \log Z(\phi) + \dots \end{aligned}$$

# A unifying ABP framework for Free Energy Computations

- Note that if  $A(\mathbf{Q}) = \hat{A}(\mathbf{Q}; \phi) + \text{constant}$  then:

$$\hat{\rho}_{\mathbf{Q}}(\mathbf{Q}) = \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)} \equiv (\text{uniform})$$

- In practice we define a (small or large) subset of  $\mathbf{Q}$ -values, say  $\mathcal{D}$ , that is of interest. Let  $\pi(\mathbf{Q}) \propto \mathbf{1}_{\mathcal{D}}(\mathbf{Q})$  be the uniform on  $\mathcal{D}$
- Hence the optimal free energy approximation  $\hat{A}(\mathbf{Q}; \phi)$  is found by:

$$\phi^{opt} = \arg \min_{\phi} f(\phi) = \underset{\substack{\text{uniform on } \mathcal{D}}}{\text{DISTANCE}(\underbrace{\pi(\mathbf{Q})}_{\text{uniform on } \mathcal{D}}, \hat{\rho}_{\mathbf{Q}}(\mathbf{Q}))}$$

- We propose using the Kullback-Leibler divergence to measure the distance between densities i.e.:

$$\begin{aligned} f(\phi) = KL(\pi(\mathbf{Q}) || \hat{\rho}_{\mathbf{Q}}(\mathbf{Q})) &= - \int \pi(\mathbf{Q}) \log \frac{\hat{\rho}_{\mathbf{Q}}(\mathbf{Q})}{\pi(\mathbf{Q})} d\mathbf{Q} \\ &= - \int \pi(\mathbf{Q}) \left( -\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi)) \right) d\mathbf{Q} + \log Z(\phi) + \dots \\ &= -\beta \underbrace{\int \pi(\mathbf{Q}) \hat{A}(\mathbf{Q}; \phi) d\mathbf{Q}}_{\text{easy}} + \underbrace{\log Z(\phi)}_{\text{difficult}} \end{aligned}$$

# A unifying ABP framework for Free Energy Computations

- Objective:

$$\min_{\phi} f(\phi) = -\beta \int \pi(\mathbf{Q}) \hat{A}(\mathbf{Q}; \phi) d\mathbf{Q} + \log Z(\phi)$$

- Suppose one expresses:

$$\hat{A}(\mathbf{Q}; \phi) = \sum_{i=1}^M \phi_i K_i(\mathbf{Q})$$

where  $K_i(\mathbf{Q})$  are some **given** basis/kernel functions <sup>2</sup>.

- Derivatives:

$$\begin{aligned} \frac{\partial f}{\partial \phi_i} &= -\beta E_{\pi} \left[ \frac{\partial \hat{A}(\mathbf{Q}; \phi)}{\partial \phi_i} \right] + \frac{d \log Z(\phi)}{d \phi_i} \\ &= -\beta (E_{\pi} [K_i(\mathbf{Q})] - E_{\rho_{\phi}} [K_i(\mathbf{Q})]), \quad \rho_{\phi}(\mathbf{Q}) = \frac{e^{-\beta(\hat{A}(\mathbf{Q}; \phi) - \lambda(\phi))}}{Z(\phi)} \end{aligned}$$

$$\frac{\partial^2 f}{\partial \phi_i \partial \phi_j} = \beta^2 \text{Cov}_{\rho_{\phi}} [K_i(\mathbf{Q}), K_j(\mathbf{Q})] > 0 \quad \rightarrow \text{CONVEX}$$

<sup>2</sup>You will ask how did you pick  $K_i$  and how do you know how many  $M$  you must use?

# A unifying ABP framework for Free Energy Computations

- Objective:

$$\min_{\phi} f(\phi) = -\beta \int \pi(\mathbf{Q}) \hat{A}(\mathbf{Q}; \phi) d\mathbf{Q} + \log Z(\phi)$$

- Suppose one expresses:

$$\hat{A}(\mathbf{Q}; \phi) = \sum_{i=1}^M \phi_i K_i(\mathbf{Q})$$

where  $K_i(\mathbf{Q})$  are some **given** basis/kernel functions <sup>2</sup>.

- Derivatives:

$$\begin{aligned} \frac{\partial f}{\partial \phi_i} &= -\beta E_{\pi} \left[ \frac{\partial \hat{A}(\mathbf{Q}; \phi)}{\partial \phi_i} \right] + \frac{d \log Z(\phi)}{d \phi_i} \\ &= -\beta (E_{\pi} [K_i(\mathbf{Q})] - E_{\hat{\rho}_{\mathbf{Q}}} [K_i(\mathbf{Q})]), \quad \hat{\rho}_{\mathbf{Q}}(\mathbf{Q}) = \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)} \end{aligned}$$

$$\frac{\partial^2 f}{\partial \phi_i \partial \phi_j} = \beta^2 \text{Cov}_{\hat{\rho}_{\mathbf{Q}}} [K_i(\mathbf{Q}), K_j(\mathbf{Q})] > 0 \quad \rightarrow \text{CONVEX}$$

<sup>2</sup>You will ask how did you pick  $K_i$  and how do you know how many  $M$  you must use?



# A unifying ABP framework for Free Energy Computations

$$\frac{\partial f}{\partial \phi_i} = -\beta (E_{\pi}[K_i(\mathbf{Q})] - E_{\hat{\rho}_{\mathbf{Q}}}[K_i(\mathbf{Q})]), \quad \hat{\rho}_{\mathbf{Q}}(\mathbf{Q}) = \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)}$$

- Q: But one does not know  $\hat{\rho}_{\mathbf{Q}}(\mathbf{Q})$ , after all, isn't  $A(\mathbf{Q})$  is what we are looking for?
- A: We will use the auxiliary (joint) density  $\tilde{\rho}(\mathbf{q})$  of which  $\hat{\rho}_{\mathbf{Q}}(\mathbf{Q})$  is the marginal:

$$\tilde{\rho}(\mathbf{q}) = \frac{e^{-\beta(V(\mathbf{q}) - \hat{A}(\boldsymbol{\epsilon}(\mathbf{q}); \phi))}}{Z(\phi)}$$

- Q: Even if you do Monte Carlo for  $\tilde{\rho}(\mathbf{q})$  you will get a very poor estimates (especially in the beginning i.e. when  $\hat{A} \equiv 0$ )?
- A: Stochastic Approximation (Robbins-Monro, 1951)

$$\phi^{(m+1)} = \phi^{(m)} - \eta_m \widetilde{\nabla f(\phi^{(m)})}$$

It is OK if  $\widetilde{\nabla f(\phi^{(m)})}$  is a *noisy* estimate of  $\nabla f(\phi^{(m)})$  as long as  $\sum_{m=0}^{\infty} \eta_m = +\infty$ ,  
 $\sum_{m=0}^{\infty} \eta_m^2 < +\infty$

# A unifying ABP framework for Free Energy Computations

$$\frac{\partial f}{\partial \phi_i} = -\beta (E_\pi[K_i(\mathbf{Q})] - E_{\hat{\rho}_\mathbf{Q}}[K_i(\mathbf{Q})]), \quad \hat{\rho}_\mathbf{Q}(\mathbf{Q}) = \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)}$$

- Q: But one does not know  $\hat{\rho}_\mathbf{Q}(\mathbf{Q})$ , after all, isn't  $A(\mathbf{Q})$  is what we are looking for?
- A: We will use the auxiliary (joint) density  $\hat{\rho}(\mathbf{q})$  of which  $\hat{\rho}_\mathbf{Q}(\mathbf{Q})$  is the marginal:

$$\hat{\rho}(\mathbf{q}) = \frac{e^{-\beta(V(\mathbf{q}) - \hat{A}(\xi(\mathbf{q}); \phi))}}{Z(\phi)}$$

- Q: Even if you do Monte Carlo for  $\hat{\rho}(\mathbf{q})$  you will get a very poor estimates (especially in the beginning i.e. when  $\hat{A} \equiv 0$ )?
- A: Stochastic Approximation (Robbins-Monro, 1951)

$$\phi^{(m+1)} = \phi^{(m)} - \eta_m \widetilde{\nabla f(\phi^{(m)})}$$

It is OK if  $\widetilde{\nabla f(\phi^{(m)})}$  is a *noisy* estimate of  $\nabla f(\phi^{(m)})$  as long as  $\sum_{m=0}^{\infty} \eta_m = +\infty$ ,  
 $\sum_{m=0}^{\infty} \eta_m^2 < +\infty$

$$\frac{\partial f}{\partial \phi_i} = -\beta (E_\pi[K_i(\mathbf{Q})] - E_{\hat{\rho}_\mathbf{Q}}[K_i(\mathbf{Q})]), \quad \hat{\rho}_\mathbf{Q}(\mathbf{Q}) = \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)}$$

- Q: But one does not know  $\hat{\rho}_\mathbf{Q}(\mathbf{Q})$ , after all, isn't  $A(\mathbf{Q})$  is what we are looking for?
- A: We will use the auxiliary (joint) density  $\hat{\rho}(\mathbf{q})$  of which  $\hat{\rho}_\mathbf{Q}(\mathbf{Q})$  is the marginal:

$$\hat{\rho}(\mathbf{q}) = \frac{e^{-\beta(V(\mathbf{q}) - \hat{A}(\xi(\mathbf{q}); \phi))}}{Z(\phi)}$$

- Q: Even if you do Monte Carlo for  $\hat{\rho}(\mathbf{q})$  you will get a very poor estimates (especially in the beginning i.e. when  $\hat{A} \equiv 0$ )?
- A: Stochastic Approximation (Robbins-Monro, 1951)

$$\phi^{(m+1)} = \phi^{(m)} - \eta_m \widetilde{\nabla f(\phi^{(m)})}$$

It is OK if  $\widetilde{\nabla f(\phi^{(m)})}$  is a *noisy* estimate of  $\nabla f(\phi^{(m)})$  as long as  $\sum_{m=0}^{\infty} \eta_m = +\infty$ ,  
 $\sum_{m=0}^{\infty} \eta_m^2 < +\infty$

$$\frac{\partial f}{\partial \phi_i} = -\beta (E_\pi[K_i(\mathbf{Q})] - E_{\hat{\rho}_\mathbf{Q}}[K_i(\mathbf{Q})]), \quad \hat{\rho}_\mathbf{Q}(\mathbf{Q}) = \frac{e^{-\beta(A(\mathbf{Q}) - \hat{A}(\mathbf{Q}; \phi))}}{Z(\phi)}$$

- Q: But one does not know  $\hat{\rho}_\mathbf{Q}(\mathbf{Q})$ , after all, isn't  $A(\mathbf{Q})$  is what we are looking for?
- A: We will use the auxiliary (joint) density  $\hat{\rho}(\mathbf{q})$  of which  $\hat{\rho}_\mathbf{Q}(\mathbf{Q})$  is the marginal:

$$\hat{\rho}(\mathbf{q}) = \frac{e^{-\beta(V(\mathbf{q}) - \hat{A}(\xi(\mathbf{q}); \phi))}}{Z(\phi)}$$

- Q: Even if you do Monte Carlo for  $\hat{\rho}(\mathbf{q})$  you will get a very poor estimates (especially in the beginning i.e. when  $\hat{A} \equiv 0$ )?
- A: Stochastic Approximation (Robbins-Monro, 1951)

$$\phi^{(m+1)} = \phi^{(m)} - \eta_m \widetilde{\nabla f(\phi^{(m)})}$$

It is OK if  $\widetilde{\nabla f(\phi^{(m)})}$  is a *noisy* estimate of  $\nabla f(\phi^{(m)})$  as long as  $\sum_{m=0}^{\infty} \eta_m = +\infty$ ,  
 $\sum_{m=0}^{\infty} \eta_m^2 < +\infty$

Enforcing Sparsity - Greedy addition of kernels (Berger et al 1996. A *Maximum Entropy Approach to Natural Language*)

$$\hat{A}(\mathbf{Q}; \phi) = \sum_{i=1}^M \phi_i K_i(\mathbf{Q})$$

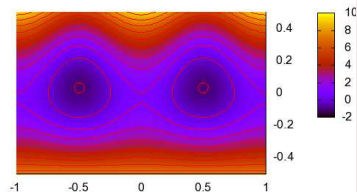
- Let  $\hat{A}(\mathbf{Q})^{(M)}$  the approximation with  $M$  basis/kernel functions
- If

$$\hat{\rho}(\mathbf{q}) = \frac{e^{-\beta(V(\mathbf{q}) - \hat{A}^{(M)}(\xi(\mathbf{q}); \phi))}}{Z(\phi)}$$

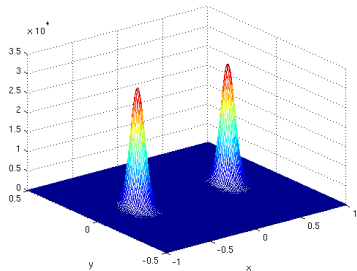
is the corresponding *auxiliary* density, then select  $K_{M+1}(\mathbf{Q})$  such as:

$$K_{M+1}(\mathbf{Q}) = \arg \max_{K(\mathbf{Q}) \in \mathcal{K}} (E_{\pi}[K(\mathbf{Q})] - E_{\hat{\rho}_{\mathbf{Q}}}[K(\mathbf{Q})])^2$$

## Example: Bimodality



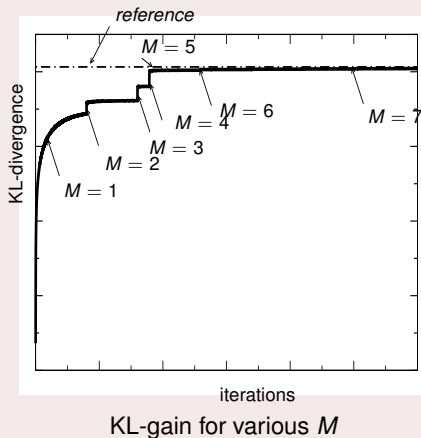
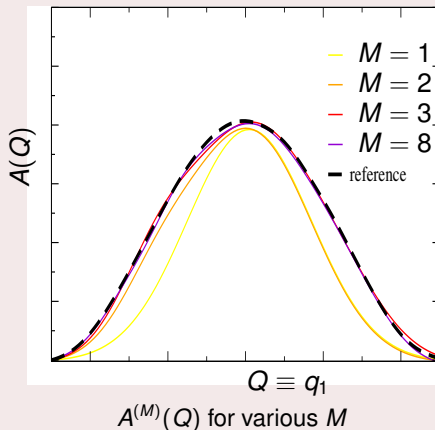
$$V(q_1, q_2)$$



$$\rho_{eq}(q_1, q_2) \propto e^{-\beta V(q_1, q_2)}$$

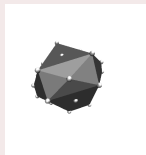
# A unifying ABP framework for Free Energy Computations

## Example: Bimodality

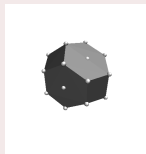


# A unifying ABP framework for Free Energy Computations

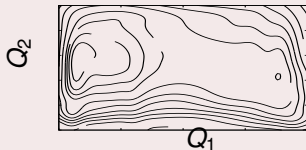
## $LJ_{38}$ : Two reaction coordinates



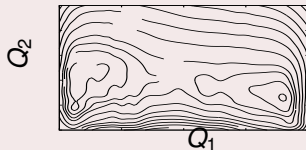
$Q_1 = 0.01$   
(icosahedron)



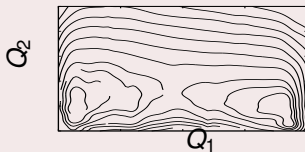
$Q_1 = 0.19$   
(octahedron)



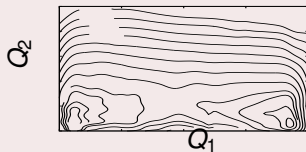
(a)  $T = 0.21$ .



(b)  $T = 0.17$ .



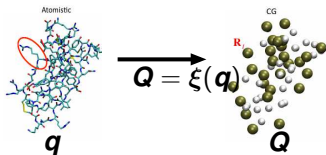
(c)  $T = 0.14$ .



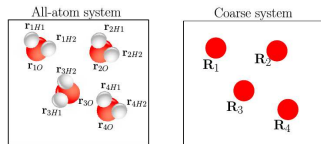
(d)  $T = 0.11$ .

Free energy contours  $A(Q_1, Q_2)$  at various temperatures.





From Noid et al. 2008



CG of water molecules. (from Billionis et al. 2013)

- e.g. define *pseudoatoms*  $I$  such that for all atoms  $i \in I$ :

$$\mathbf{Q}_I = \frac{\sum_{i \in I} m_i \mathbf{q}_i}{\sum_{i \in I} m_i}$$

- Is there an optimal CG-description? How should two alternative CG-coordinate systems be compared?

- If  $V(\mathbf{q})$  is the reference potential and  $\rho_{eq}(\mathbf{q})$ :

$$\rho_{eq}(\mathbf{q}) = \frac{e^{-\beta V(\mathbf{q})}}{Z}$$

the equilibrium distribution.

- The *exact* coarse-graining w.r.t. to  $\mathbf{Q} = \xi(\mathbf{q})$  requires the computation of the corresponding Helmholtz free energy  $A(\mathbf{Q})$ :

$$A(\mathbf{Q}) = -\frac{1}{\beta} \log \int \frac{e^{-\beta V(\mathbf{q})}}{Z} \delta(\mathbf{Q} - \xi(\mathbf{q})) d\mathbf{q}$$

- In principle any of the Free Energy computational techniques can be employed.
- The problem is that all of them are applicable to low-dimensional  $\mathbf{Q}$  (e.g.  $\dim(\mathbf{Q}) < 5$  or  $10$ )
- When  $\dim(\mathbf{Q}) \gg 1$ , physical insight is necessary.

- If  $V(\mathbf{q})$  is the reference potential and  $\rho_{eq}(\mathbf{q})$ :

$$\rho_{eq}(\mathbf{q}) = \frac{e^{-\beta V(\mathbf{q})}}{Z}$$

the equilibrium distribution.

- The *exact* coarse-graining w.r.t. to  $\mathbf{Q} = \xi(\mathbf{q})$  requires the computation of the corresponding Helmholtz free energy  $A(\mathbf{Q})$ :

$$A(\mathbf{Q}) = -\frac{1}{\beta} \log \int \frac{e^{-\beta V(\mathbf{q})}}{Z} \delta(\mathbf{Q} - \xi(\mathbf{q})) d\mathbf{q}$$

- In principle any of the Free Energy computational techniques can be employed.
- The problem is that all of them are applicable to low-dimensional  $\mathbf{Q}$  (e.g.  $dim(\mathbf{Q}) < 5$  or  $10$ )
- When  $dim(\mathbf{Q}) \gg 1$ , physical insight is necessary.

- Let an *approximation* to the true free energy  $A(\mathbf{Q})$  be:

$$\hat{A}(\mathbf{Q}; \phi) \rightarrow \hat{\rho}(\mathbf{Q}) = \frac{e^{-\beta \hat{A}(\mathbf{Q}; \phi)}}{Z(\phi)}$$

that depends on some parameters  $\phi$  .E.g.:

$$\left. \begin{aligned} \hat{A}(\mathbf{Q}; \phi) &= \sum_{I < J} \underbrace{V(|\mathbf{Q}_I - \mathbf{Q}_J|; \phi)}_{\text{pairwise-potential}} \\ V(r; \phi) &= \sum_{i=1}^M \phi_i \underbrace{\psi_i(r)}_{\text{known}} \end{aligned} \right\} \rightarrow \hat{A}(\mathbf{Q}; \phi) = \sum_{i=1}^M \phi_i (\sum_{I < J} \psi_i(|\mathbf{Q}_I - \mathbf{Q}_J|)) \\ = \sum_{i=1}^M \phi_i K_i(\mathbf{Q})$$

- What is the *optimal*  $\phi^{opt}$ ?

- The Relative Entropy method<sup>3</sup>

$$\phi^{opt} = \arg \min_{\phi} f(\phi) = KL(\rho_{eq}(\mathbf{Q}) || \hat{\rho}(\mathbf{Q}))$$

<sup>3</sup>Shell et al 2008, Billionis et al. 2013

- Let an *approximation* to the true free energy  $A(\mathbf{Q})$  be:

$$\hat{A}(\mathbf{Q}; \phi) \rightarrow \hat{\rho}(\mathbf{Q}) = \frac{e^{-\beta \hat{A}(\mathbf{Q}; \phi)}}{Z(\phi)}$$

that depends on some parameters  $\phi$  .E.g.:

$$\left. \begin{aligned} \hat{A}(\mathbf{Q}; \phi) &= \sum_{I < J} \underbrace{V(|\mathbf{Q}_I - \mathbf{Q}_J|; \phi)}_{\text{pairwise-potential}} \\ V(r; \phi) &= \sum_{i=1}^M \phi_i \underbrace{\psi_i(r)}_{\text{known}} \end{aligned} \right\} \rightarrow \begin{aligned} \hat{A}(\mathbf{Q}; \phi) &= \sum_{i=1}^M \phi_i (\sum_{I < J} \psi_i(|\mathbf{Q}_I - \mathbf{Q}_J|)) \\ &= \sum_{i=1}^M \phi_i K_i(\mathbf{Q}) \end{aligned}$$

- What is the *optimal*  $\phi^{opt}$ ?
- The Relative Entropy method <sup>3</sup>

$$\phi^{opt} = \arg \min_{\phi} f(\phi) = KL(\rho_{eq}(\mathbf{Q}) || \hat{\rho}(\mathbf{Q}))$$

<sup>3</sup>Shell et al 2008, Billionis et al. 2013

## Relative Entropy method

$$\phi^{opt} = \arg \min_{\phi} f(\phi) = KL(\rho_{eq}(\mathbf{Q}) || \hat{\rho}(\mathbf{Q}))$$

## Alternatives

- **Inversion based methods.** E.g. Boltzman Inversion (Tchöp et al. 1998), Iterative Boltzmann Inversion (Reith et al. (2003), Inverse Monte Carlo [Lyubartsev et al. 1995, Soper 1996).
- **Variational methods.** E.g. Force Matching (Ercolessi et al. 1994, Izvekov et al. 2005, Noid et al. 2007).

## Relative Entropy method

$$\begin{aligned} f(\phi) &= KL(\rho_{eq}(\mathbf{Q}) || \hat{\rho}(\mathbf{Q})) \\ &= - \int \rho_{eq}(\mathbf{Q}) \log \frac{e^{-\beta \hat{A}(\mathbf{Q}; \phi)} / Z(\phi)}{\rho_{eq}(\mathbf{Q})} \\ &= \beta \int \hat{A}(\mathbf{Q}; \phi) \rho_{eq}(\mathbf{Q}) d\mathbf{Q} + \log Z(\phi) + \dots \\ &= E_{\rho_{eq}}[\hat{A}(\mathbf{Q}; \phi)] + \log Z(\phi) \end{aligned}$$

- Derivatives (If  $\hat{A}(\mathbf{Q}; \phi) = \sum_{i=1}^M \phi_i K_i(\mathbf{Q})$ ):

$$\frac{\partial f}{\partial \phi_i} = \beta (E_{\rho_{eq}}[K_i(\mathbf{Q})] - E_{\hat{\rho}}[K_i(\mathbf{Q})])^a$$

$$\frac{\partial^2 f}{\partial \phi_i \partial \phi_j} = \beta^2 \text{Cov}_{\hat{\rho}}[K_i(\mathbf{Q}), K_j(\mathbf{Q})] > 0 \quad \rightarrow \text{CONVEX}$$

---

<sup>a</sup>This is exactly what one would get with MAXENT

## Relative Entropy method

$$\frac{\partial f}{\partial \phi_i} = \beta(E_{\rho_{\text{eq}}}[K_i(\mathbf{Q})] - E_{\hat{\rho}}[K_i(\mathbf{Q})])$$

- Q: How can one estimate  $E_{\rho_{\text{eq}}}[K_i(\mathbf{Q})]$ ?
- A: *All* existing methods assume that a very long MCMC run of  $\rho_{\text{eq}}(\mathbf{q})$  is available and:

$$E_{\rho_{\text{eq}}}[K_i(\mathbf{Q})] \approx \frac{1}{N} \sum_{j=1}^N K_i(\xi(\mathbf{q}^{(j)}))$$

- As we have seen this is not always straightforward.
- Q: How can one estimate  $E_{\hat{\rho}}[K_i(\mathbf{Q})]$  where  $\hat{\rho}(\mathbf{Q}) \propto e^{-\beta \lambda(\mathbf{Q}, \phi)}$
- A: With ones favorite MCMC sampler (sampling in the reduced coordinates)
- Both estimates will be *noisy*  $\rightarrow$  Stochastic Approximations are needed.



## Relative Entropy method

$$\frac{\partial f}{\partial \phi_i} = \beta(E_{\rho_{\text{eq}}} [K_i(\mathbf{Q})] - E_{\hat{\rho}} [K_i(\mathbf{Q})])$$

- Q: How can one estimate  $E_{\rho_{\text{eq}}} [K_i(\mathbf{Q})]$ ?
- A: *All* existing methods assume that a very long MCMC run of  $\rho_{\text{eq}}(\mathbf{q})$  is available and:

$$E_{\rho_{\text{eq}}} [K_i(\mathbf{Q})] \approx \frac{1}{N} \sum_{j=1}^N K_i(\xi(\mathbf{q}^{(j)}))$$

- As we have seen this is not always straightforward.
- Q: How can one estimate  $E_{\hat{\rho}} [K_i(\mathbf{Q})]$  where  $\hat{\rho}(\mathbf{Q}) \propto e^{-\beta \hat{A}(\mathbf{Q}; \phi)}$
- A: With ones favorite MCMC sampler (sampling in the reduced coordinates)

• Both estimates will be *noisy*  $\rightarrow$  Stochastic Approximations are needed.

## Relative Entropy method

$$\frac{\partial f}{\partial \phi_i} = \beta(E_{\rho_{\text{eq}}} [K_i(\mathbf{Q})] - E_{\hat{\rho}} [K_i(\mathbf{Q})])$$

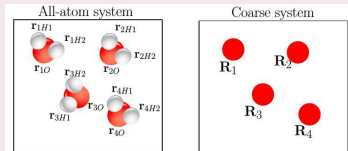
- Q: How can one estimate  $E_{\rho_{\text{eq}}} [K_i(\mathbf{Q})]$ ?
- A: *All* existing methods assume that a very long MCMC run of  $\rho_{\text{eq}}(\mathbf{q})$  is available and:

$$E_{\rho_{\text{eq}}} [K_i(\mathbf{Q})] \approx \frac{1}{N} \sum_{j=1}^N K_i(\xi(\mathbf{q}^{(j)}))$$

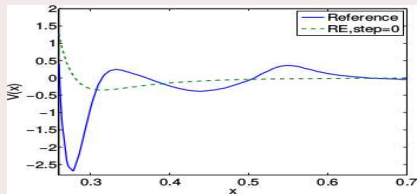
- As we have seen this is not always straightforward.
- Q: How can one estimate  $E_{\hat{\rho}} [K_i(\mathbf{Q})]$  where  $\hat{\rho}(\mathbf{Q}) \propto e^{-\beta \hat{A}(\mathbf{Q}; \phi)}$
- A: With ones favorite MCMC sampler (sampling in the reduced coordinates)
- Both estimates will be *noisy*  $\rightarrow$  **Stochastic Approximations** are needed.

## Example (Billionis et al. 2013):

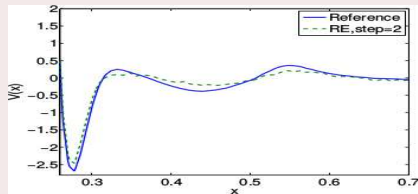
- Full-order system: 2180 *water* molecules (i.e.  $\dim(\mathbf{q}) > 19620$ )
- $T = 300K$ , SPC/E water model (Kusalik et al. 1994)
- 220ns in total are simulated with a time step of 0.002ps (Gromacs)
- 5hrs on 240 CPUs!
- $N = 10^5$  samples are used to estimate expectations w.r.t  $\rho_{eq}(\mathbf{q})$
- Coarse-grained model with  $\dim(\mathbf{Q}) = 6540$
- Cubic splines were used as the basis functions  $\phi_i(r)$  in the pairwise coarse-grained potential



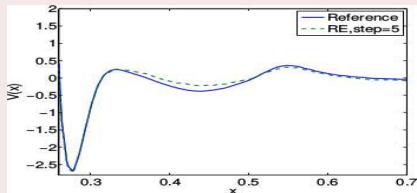
## Example (Bilionis et al. 2013):



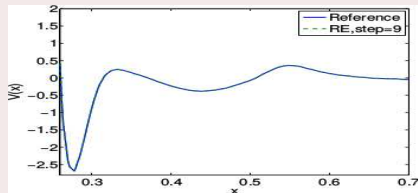
(a)



(b)



(c)



(d)

Learned pairwise coarse-grained potential (Bilionis et al. 2013)

- What is the best set of coarse-grained coordinates  $\mathbf{Q}$ ?
- How does this change depending on the prediction objectives?
- Is there a way to learn a hierarchy of coarse-grained coordinates  $\mathbf{Q}_1 = \xi_1(\mathbf{Q}_2) = \xi_1 \circ \xi_2(\mathbf{Q}_3) = \dots = \xi_1 \circ \xi_2 \dots \circ \xi_s(\mathbf{q})$ ?
- Is there a way to learn this exclusively from (short bursts of) simulation data?
- How can one learn free-energy surfaces in high-dimensions?
- Since we are learning from finite amounts of simulation data, what is the inferential uncertainty?
- Can we be *Bayesian* in our coarse-graining (e.g. Español et al 2011, Wright et al. 2013 ICES report 13-31)
- *Non-equilibrium* statistical mechanics?