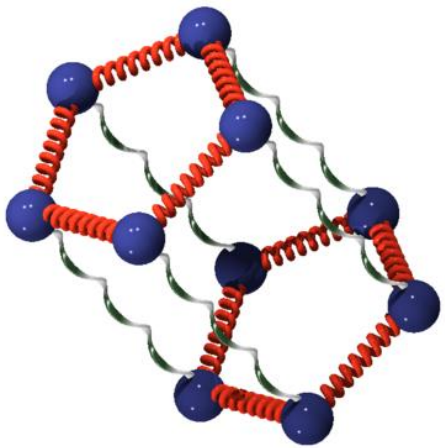
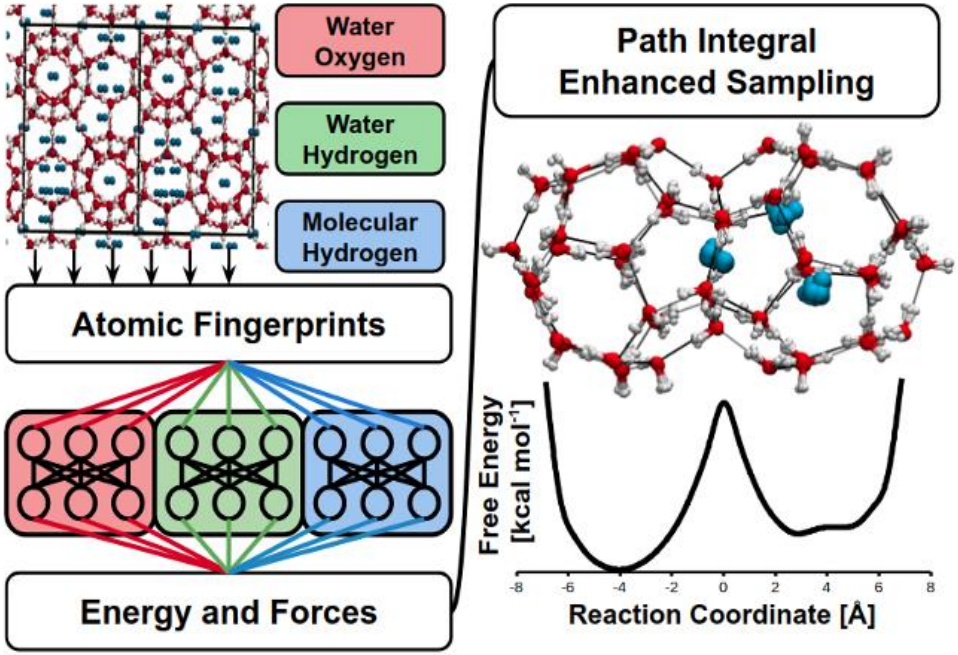
Combining path integral molecular dynamics with machine learning potentials for the study of complex quantum phenomena in condensed phases





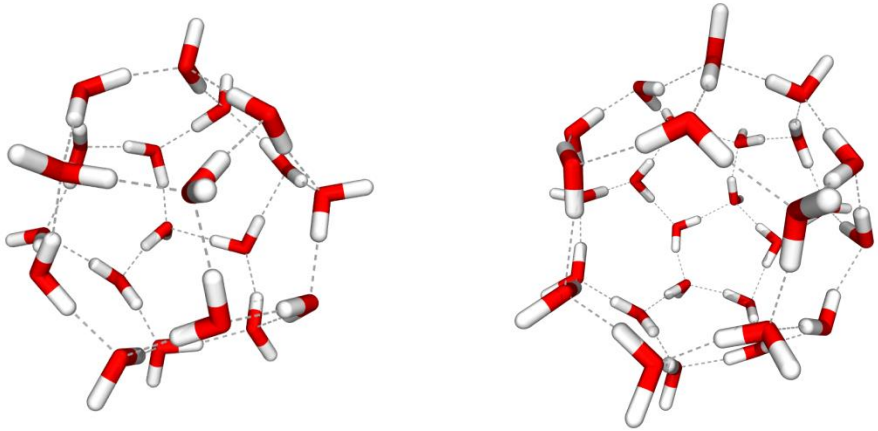
Mark E. Tuckerman

Dept. of Chemistry and Courant Institute of Mathematical Sciences
New York University, 100 Washington Square East, NY 10003
NYU-ECNU Center for Computational Chemistry at NYU Shanghai
3663 Zhongshan Road North, Shanghai 200062, China
华 东 师 范 大 学-纽约大学 计算化学联合研究中心

NEW YORK UNIVERSITY MRSEC

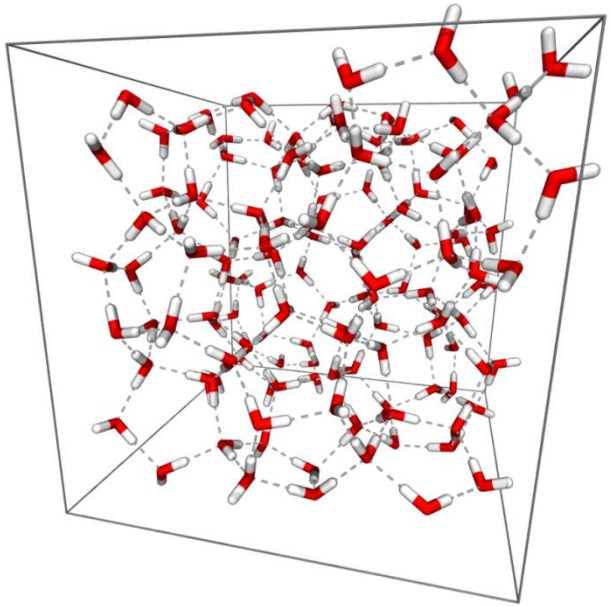
Hydrogen in structure-II clathrate

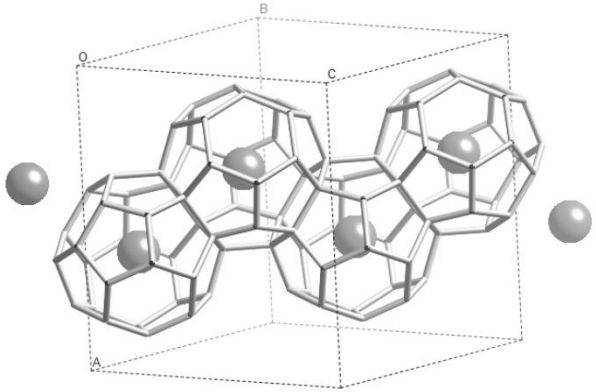
A. Witt, *et al. JPCC* (2010); J. R. Cendagorta, *et al. PCCP* (2016); J. R. Cendagorata *et al*. *Adv. Theor. Simulat*. (2021)



5^{12} cage $5^{12}6^{4}$ cage

Unit cell: 16 **:** 8





Strobel et al. (2007)

Site-site hopping model:

$$D(T) \approx \alpha k(T) l^2$$

S. Chandrasekhar (1943)

Occupancy: 4 H_2 in large cages, 1 in small cages ~ 3.8 mass %, 2 in small cages ~ 5.3 mass %
DOE target value ~ 5.5 mass %

Equilibrium nuclear quantum effects via Feynman path integrals

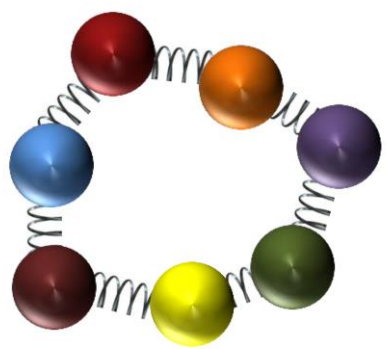
Quantum equilibrium properties derived from the canonical partition function:

$$Z(\beta) = \mathrm{Tr}\left[e^{-\beta\hat{H}}\right] = \int d\mathbf{r}\ \langle \mathbf{r} | e^{-\beta(\hat{T}+\hat{V})} | \mathbf{r} \rangle$$

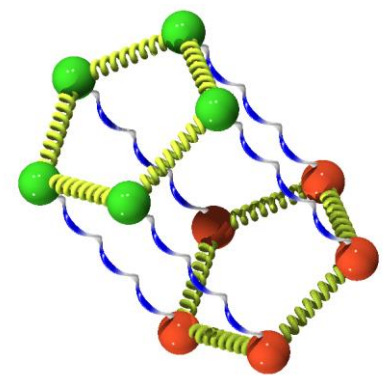
$$= \lim_{P\to\infty} \int d\mathbf{r}_1 \cdots d\mathbf{r}_P \prod_{i=1}^{P} \langle \mathbf{r}_i | \hat{\Omega} | \mathbf{r}_{i+1} \rangle_{\mathbf{r}_{P+1}=\mathbf{r}_1}, \qquad \hat{\Omega} = e^{-\beta\hat{T}/P} e^{-\beta\hat{V}/P}$$

$$Z_P(\beta) \equiv \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{3P/2} \int d\mathbf{r}_1 \cdots d\mathbf{r}_P \exp\left\{-\sum_{i=1}^{P}\left[\frac{mP}{2\beta\hbar^2}(\mathbf{r}_i - \mathbf{r}_{i+1})^2 + \frac{\beta}{P}V(\mathbf{r}_i)\right]\right\}\Bigg|_{\mathbf{r}_{P+1}=\mathbf{r}_1}$$

Classical particle



Quantum particle (thermodynamic view)



Interacting quantum particles

What are quantum free energy profiles?

Given system with classical configuration **r** and reaction coordinate $q(\mathbf{r})$, the classical free energy profile is given by

$$P(s) = \int \mathrm{d}\mathbf{r}\ \mathrm{d}\mathbf{p}\ e^{-\beta H(\mathbf{p},\mathbf{r})}\delta\big(q(\mathbf{r}) - s\big), \qquad F(s) = -\beta^{-1}\ln P(s)$$

Feynman-Kleinert centroid potential of mean force: $\mathbf{r}_c = \dfrac{1}{P}\sum_{i=1}^{P}\mathbf{r}_i$

$$e^{-\beta W_P(\mathbf{r}_c)} \propto \int d\mathbf{r}_1 \cdots d\mathbf{r}_P\ \delta\left(\frac{1}{P}\sum_{i=1}^{P}\mathbf{r}_i - \mathbf{r}_c\right)\exp\left\{-\sum_{i=1}^{P}\left[\frac{mP}{2\beta\hbar^2}\left(\mathbf{r}_i - \mathbf{r}_{i+1}\right)^2 + \frac{\beta}{P}V(\mathbf{r}_i)\right]\right\}\Bigg|_{\mathbf{r}_{P+1}=\mathbf{r}_1}$$

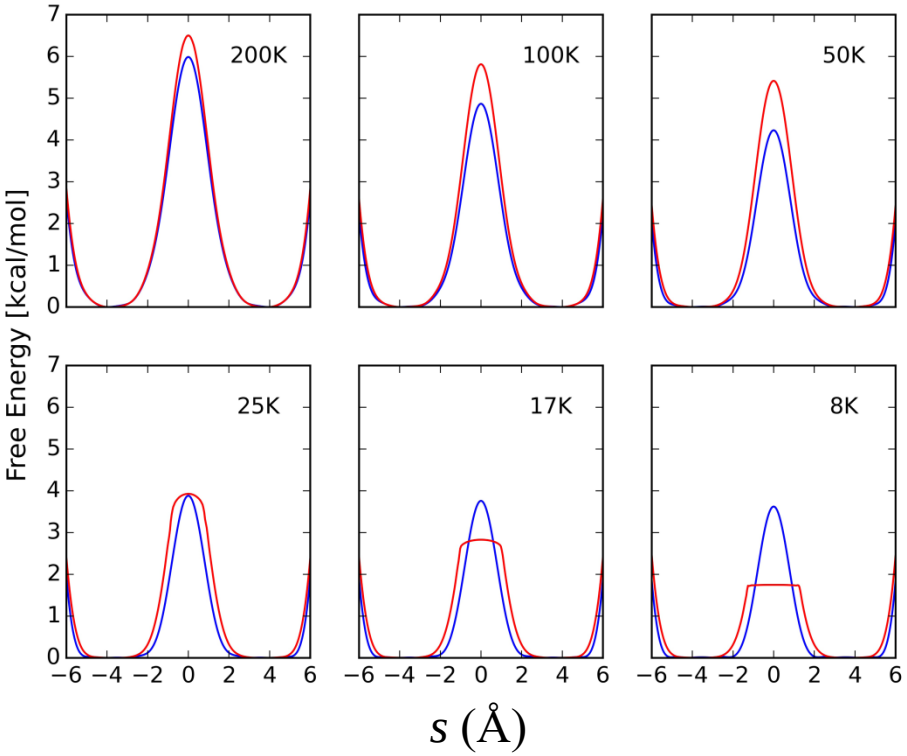
Recovering the partition function:

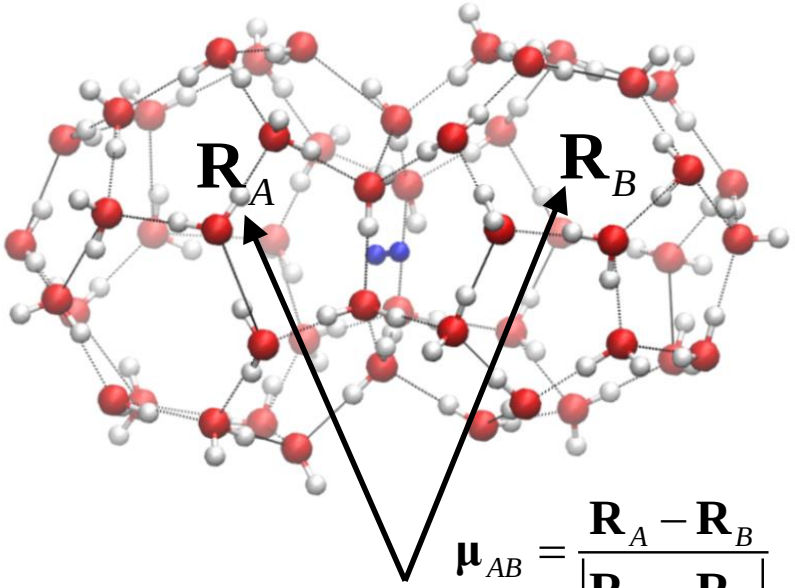
$$Z_P(\beta) = \int d\mathbf{r}_c\ e^{-\beta W_P(\mathbf{r}_c)}$$

Centroid based path integral free-energy estimator:

$$\pi_c(\mathbf{r}_1,\ldots,\mathbf{r}_P;s) = \delta\left(\frac{1}{P}\sum_{i=1}^{P}q(\mathbf{r}_i) - s\right), \qquad F_c(s) = -\beta^{-1}\ln\left\langle \pi(\mathbf{r}_1,\ldots,\mathbf{r}_P;s)\right\rangle_{\mathbf{r}}$$

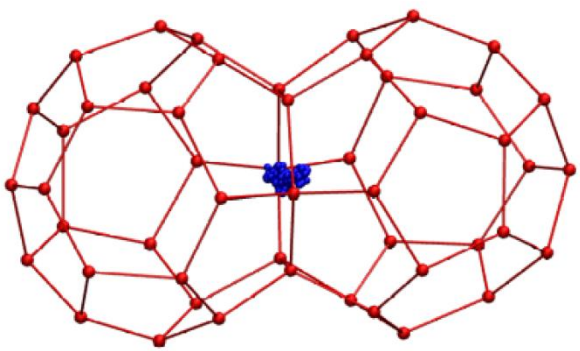
Unit cell of Mak *et al.*
JCP (1965)



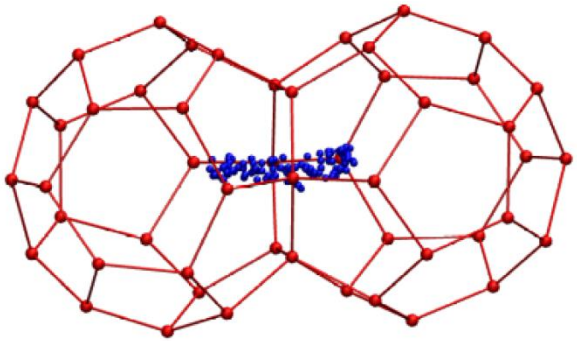


$$\boldsymbol{\mu}_{AB} = \frac{\mathbf{R}_A - \mathbf{R}_B}{\left|\mathbf{R}_A - \mathbf{R}_B\right|}$$

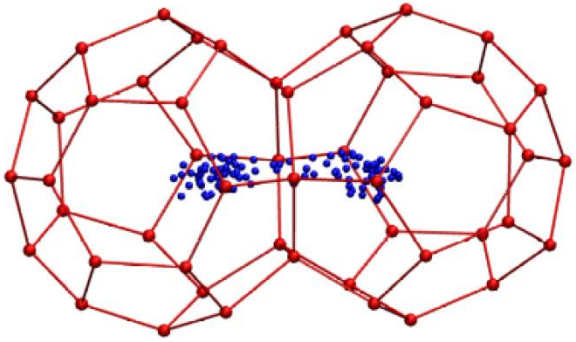
$$q(\mathbf{r}) = \left[\frac{1}{2}\left(\mathbf{r}_{\mathrm{H}_1} + \mathbf{r}_{\mathrm{H}_2}\right) - \mathbf{R}_A\right] \cdot \boldsymbol{\mu}_{AB} - \frac{\left|\mathbf{R}_B - \mathbf{R}_A\right|}{2}$$



50 K



25 K



8 K

What are quantum free energy profiles?

Quantum position-dependent observable

$$\left\langle \hat{O}(\mathbf{r}) \right\rangle = \frac{1}{Z(\beta)} \mathrm{Tr}\left[\hat{O}(\mathbf{r}) e^{-\beta \hat{H}} \right] = \left\langle \frac{1}{P} \sum_{i=1}^{P} O(\mathbf{r}_i) \right\rangle$$

Quantum marginal distribution:

$$P(s) = \frac{1}{Z} \mathrm{Tr}\left[e^{-\beta \hat{H}} \delta\left(\hat{q}(\hat{\mathbf{r}}) - s\mathrm{I} \right) \right]$$

Corresponding path integral bead estimator:

$$\pi_b(\mathbf{r}_1, \ldots, \mathbf{r}_P) = \frac{1}{P} \sum_{i=1}^{P} \delta\left(q(\mathbf{r}_i) - s \right), \qquad F_b(s) = -\beta^{-1} \ln \left\langle \pi_b(\mathbf{r}_1, \ldots, \mathbf{r}_P; s) \right\rangle_{\mathbf{r}}$$

What are quantum free energy profiles?

A centroid-based estimator is much easier to evaluate, so let's use it to bias the sampling:

$$\left\langle \pi_b(\mathbf{r}_1,\ldots,\mathbf{r}_P;s)\right\rangle = \frac{\int d\sigma\ \left\langle \pi_b(\mathbf{r}_1,\ldots,\mathbf{r}_P;s)\right\rangle_{\mathbf{r}}(\sigma)e^{-\beta F_c(\sigma)}}{\int d\sigma\ e^{-\beta F_c(\sigma)}} = e^{-\beta F_b(s)}$$

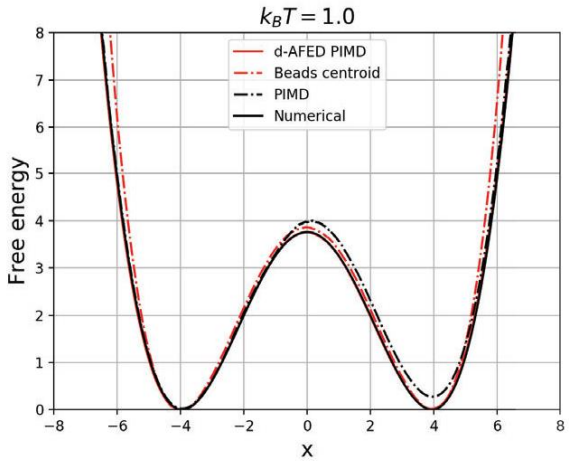
$$\left\langle \pi_b(\mathbf{r}_1,\ldots,\mathbf{r}_P;s)\right\rangle_{\mathbf{r}}(\sigma) = \frac{\left\langle \pi_c(\mathbf{r}_1,\ldots,\mathbf{r}_P;\sigma)\pi_b(\mathbf{r}_1,\ldots,\mathbf{r}_P;s)\right\rangle_{\mathbf{r}}}{\left\langle \pi_c(\mathbf{r}_1,\ldots,\mathbf{r}_P;\sigma)\right\rangle_{\mathbf{r}}}$$

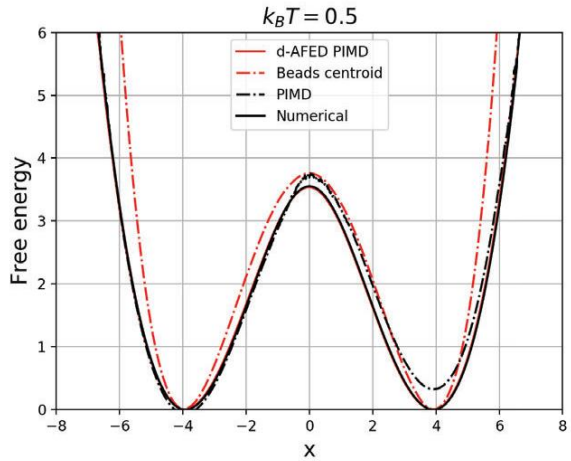
$$F_c(\sigma) = -\frac{1}{\beta}\ln\left\langle \pi_c(\mathbf{r}_1,\ldots,\mathbf{r}_P;\sigma)\right\rangle$$

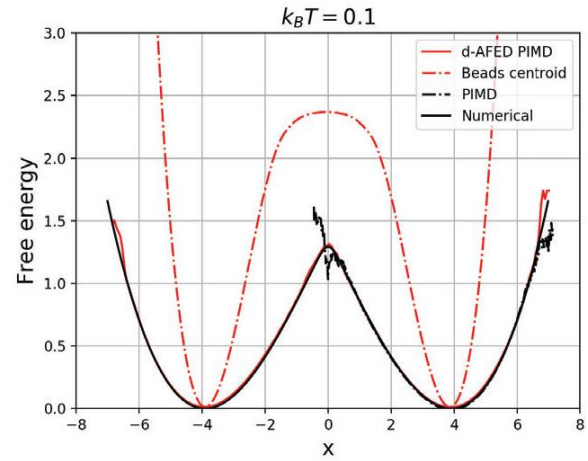
The centroid estimator $\pi_c(\mathbf{r}_1,\ldots,\mathbf{r}_P;\sigma)$ can be targeted via enhanced sampling methods such as metadynamics, adiabatic free energy dynamics, etc. to generate the true free energy profile from the expression at the top of the slide.

How different are these free energies?

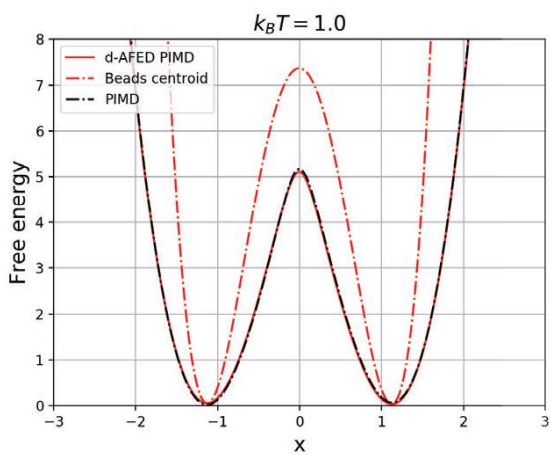
$$V(x) = \frac{1}{64}\left(x^2 - 16\right)^2$$

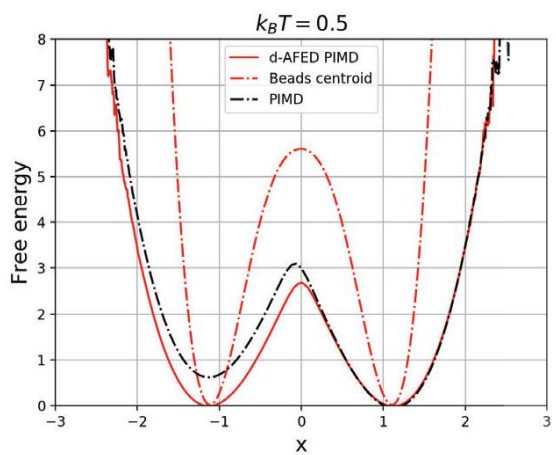


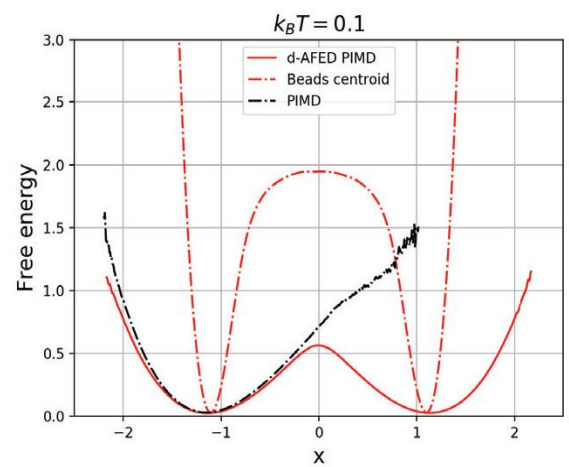




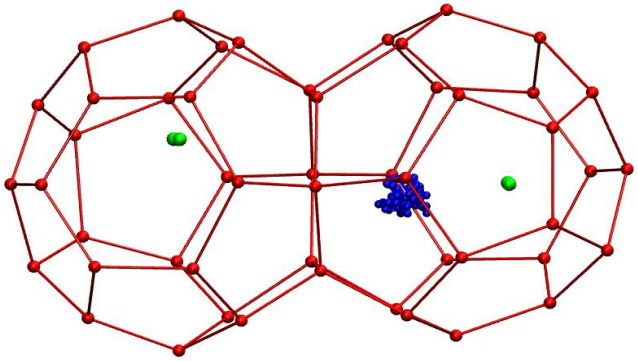
$$V(x,y) = 5\left(x^2 - 1\right)^2 + \frac{y^2}{2} + 2.878xy$$

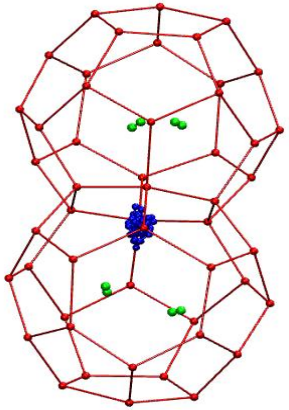






Sampling the hydrogen cage-hopping events at 25 K

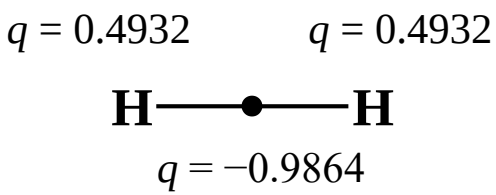


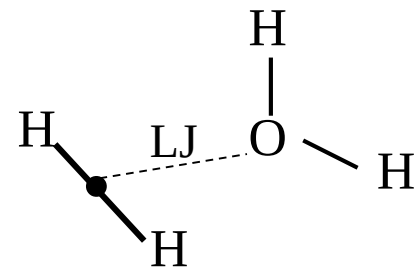


Hydrogen@clathrate potential

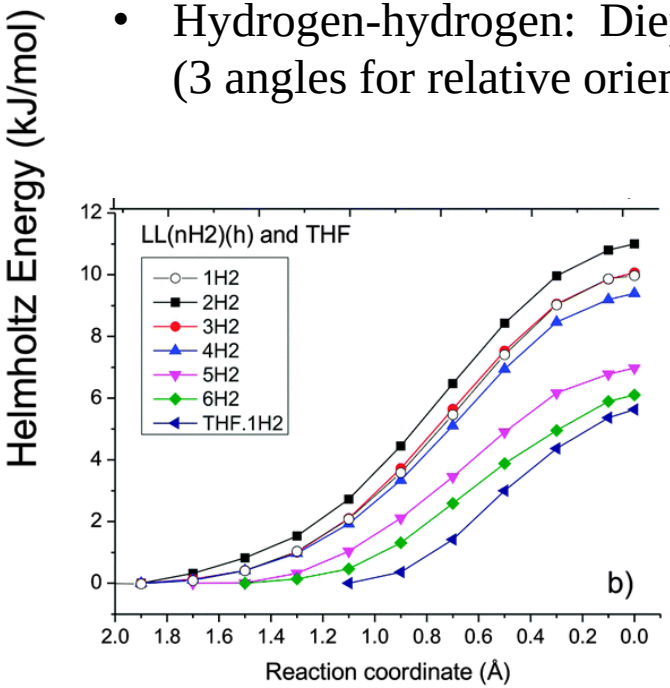
- Water-water potential: q-TIP4P/F [Habershon *et al. JCP* **131**, 024501 (2009)]

- Hydrogen-water potential: LJ + Coulomb [S.Alavi *et al*. *JCP* **123**, 024507 (2005)]

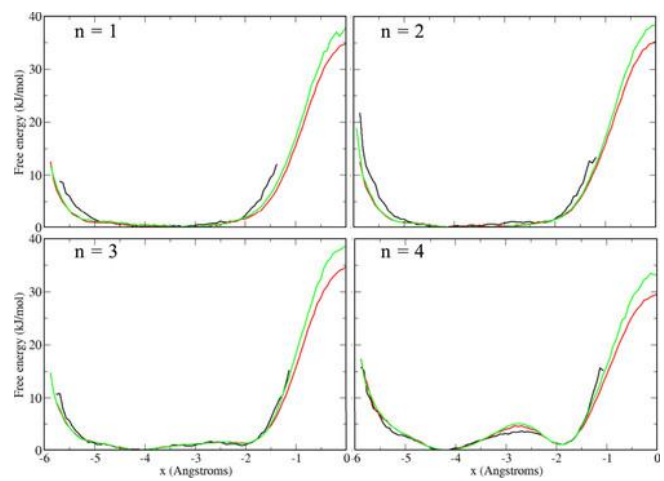




- Hydrogen-hydrogen: Diep-Johnson 4D potential surface [*JCP* (2000] (3 angles for relative orientation + COM-COM distance).

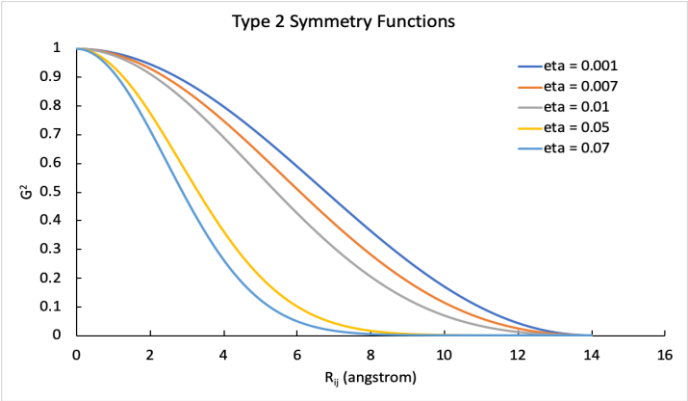


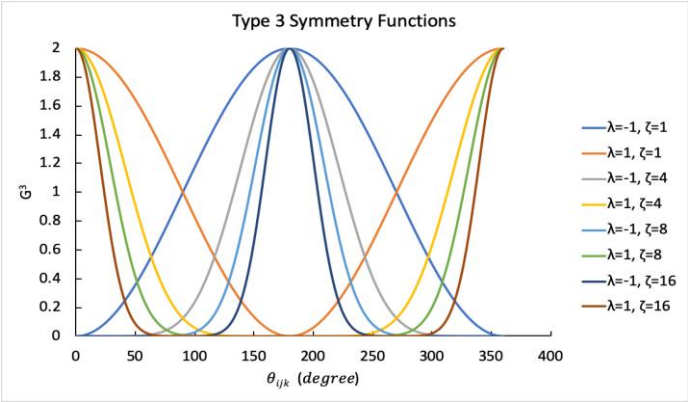
BLYP+D2: Trinh et al. PCCP (2015)

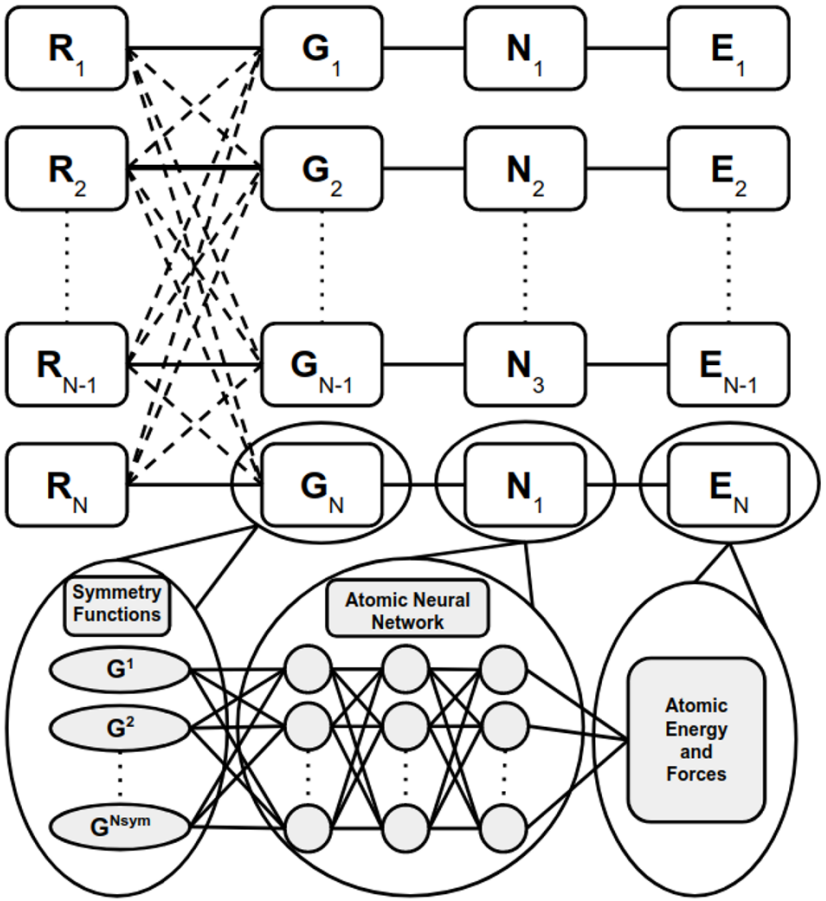


Force matching to DFT-DRSLL vdW functional
Burnham *et al*. *JCP* (2016).

Creating a neural network potential [Behler and Parrinello *PRL* (2007)]







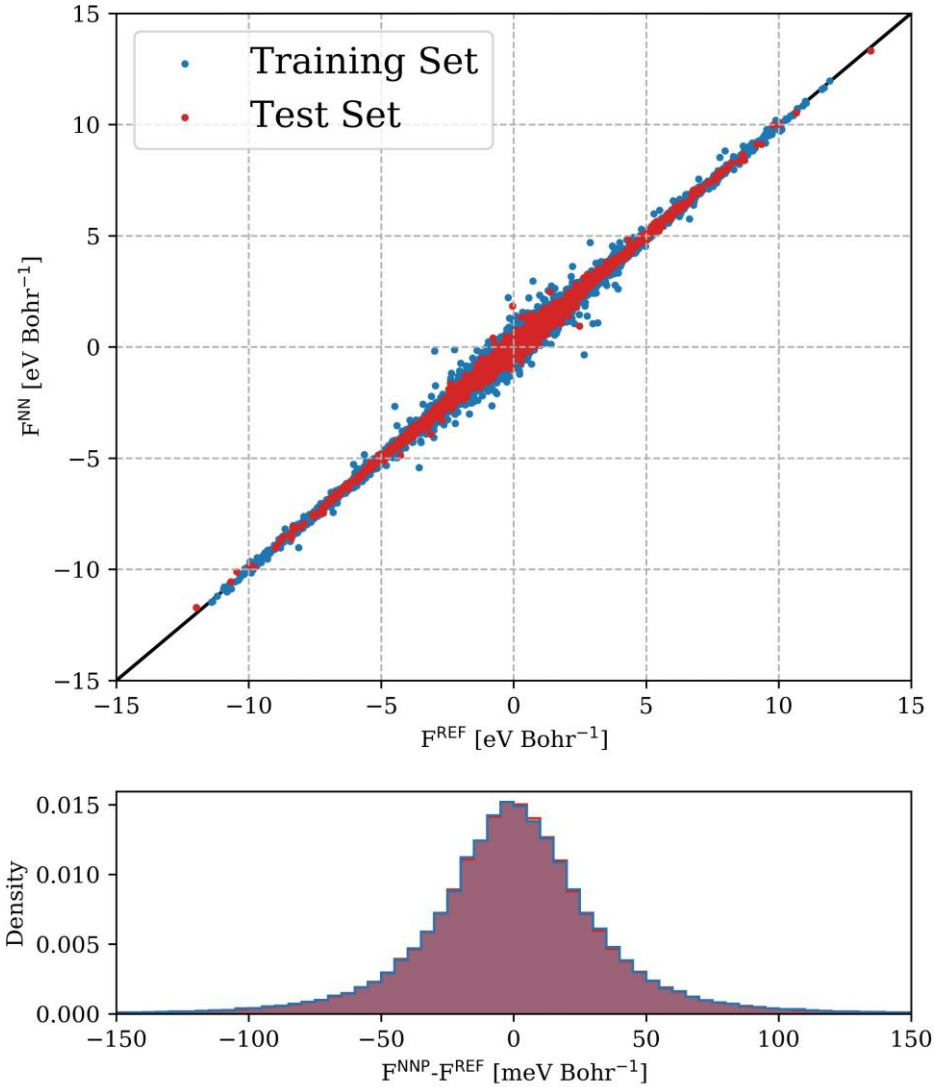
$$G_i^2 = \sum_{j=1}^{N_{\text{atom}}} e^{-\eta\left(R_{ij}-R_s\right)^2} f_c(R_{ij})$$

3 Networks: O_w, H_w, H_2

Trained to revPBE0 + D3 dispersion

$$G_i^3 = 2^{1-\zeta} \sum_{j\neq k} \sum_{k\neq i,j} \left[\left(1+\lambda\cos\theta_{ijk}\right)^{\zeta} e^{-\eta\left(R_{ij}^2+R_{ik}^2+R_{jk}^2\right)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}) \right]$$

Creating a neural network potential [Behler and Parrinello *PRL* (2007)]

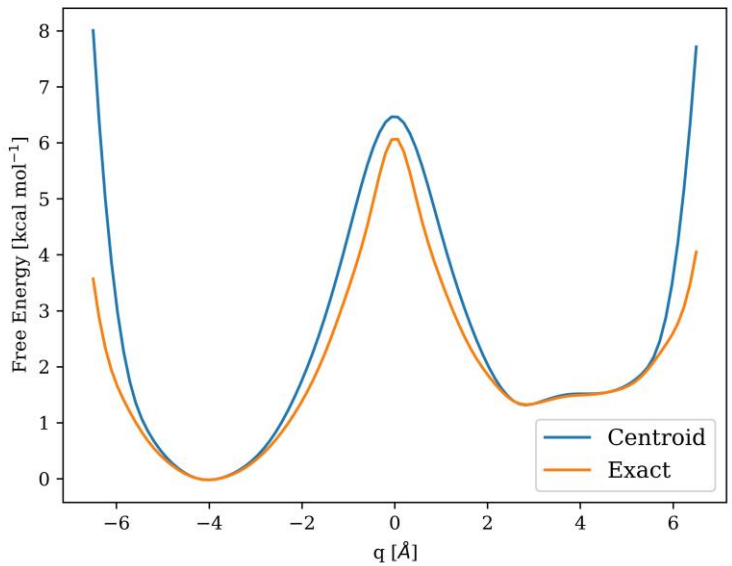


Training data:

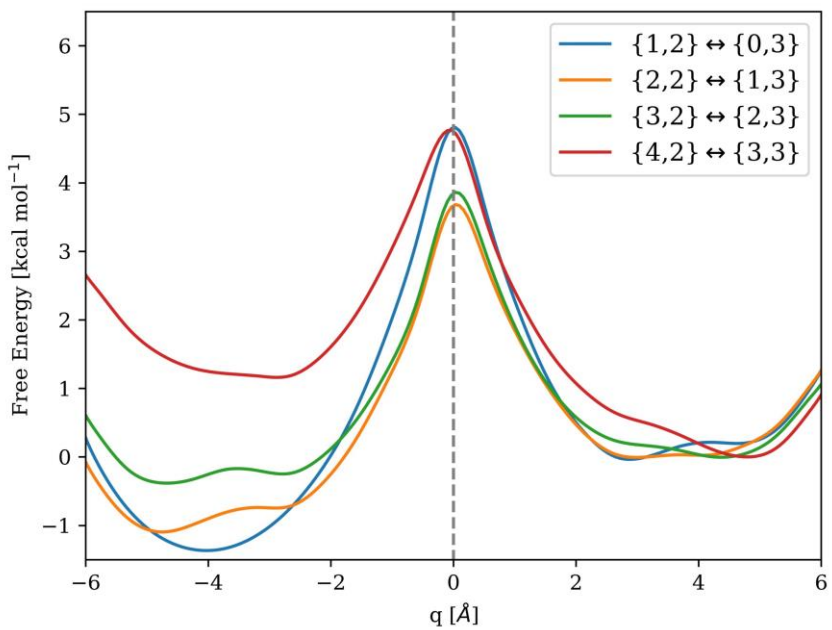
- 1,600 configurations of a water box with 64 molecules
- 3,000 configurations of a H_2 box with 56 molecules
- 500 configurations of an H_2 + H_2O box of different compositions.
- 500 actual H_2+clathrate configurations.

First results at 120 K

Solvation Process	Chemical Potential [kcal mol^{-1}]
$\{0\} \rightarrow \{1\}$	-5.6 ($\pm$ 0.01)
$\{1\} \rightarrow \{2\}$	-4.1 ($\pm$ 0.02)
$\{2\} \rightarrow \{3\}$	-3.9 ($\pm$ 0.02)
$\{3\} \rightarrow \{4\}$	-4.2 ($\pm$ 0.02)



Experimental barrier estimate from NMR: 3.8 kcal/mol.
[Senadheera, Conradi *JPCB* (2007)]



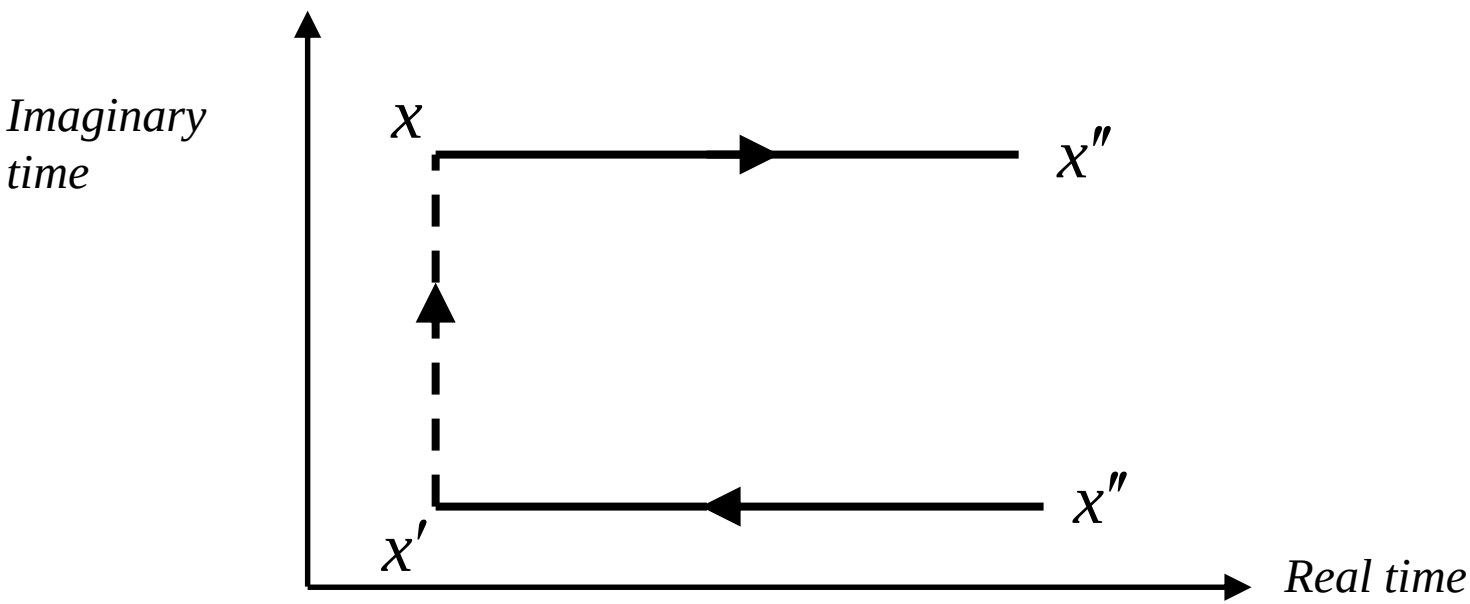
Quantum time correlation functions

Standard correlation function $$C_{\mathrm{AB}}^{\beta}(t) = \frac{1}{Z(\beta)} \mathrm{Tr}\left[e^{-\beta \hat{\mathrm{H}}} \hat{\mathrm{A}} e^{i\hat{\mathrm{H}}t/\hbar} \hat{\mathrm{B}} e^{-i\hat{\mathrm{H}}t/\hbar} \right]$$

If $\hat{\mathrm{A}} = \hat{\mathrm{A}}(\hat{\mathrm{x}}), \hat{\mathrm{B}} = \hat{\mathrm{B}}(\hat{\mathrm{x}})$

$$C_{\mathrm{AB}}^{\beta}(t) = \frac{1}{Z(\beta)} \int dx \; \langle x | \, e^{-\beta \hat{H}} \, \hat{\mathrm{A}} \, e^{i\hat{H}t/\hbar} \; \hat{\mathrm{B}} \, e^{-i\hat{H}t/\hbar} \, | x \rangle$$

$$C_{\mathrm{AB}}^{\beta}(t) = \frac{1}{Z(\beta)} \int dx \; dx' \; dx'' \; \langle x | \, e^{-\beta \hat{H}} \, | x' \rangle a(x') \langle x' | e^{i\hat{H}t/\hbar} | x'' \rangle b(x'') \langle x'' | e^{-i\hat{H}t/\hbar} | x \rangle$$



Symmetric time correlation functions

Symmetric correlation function

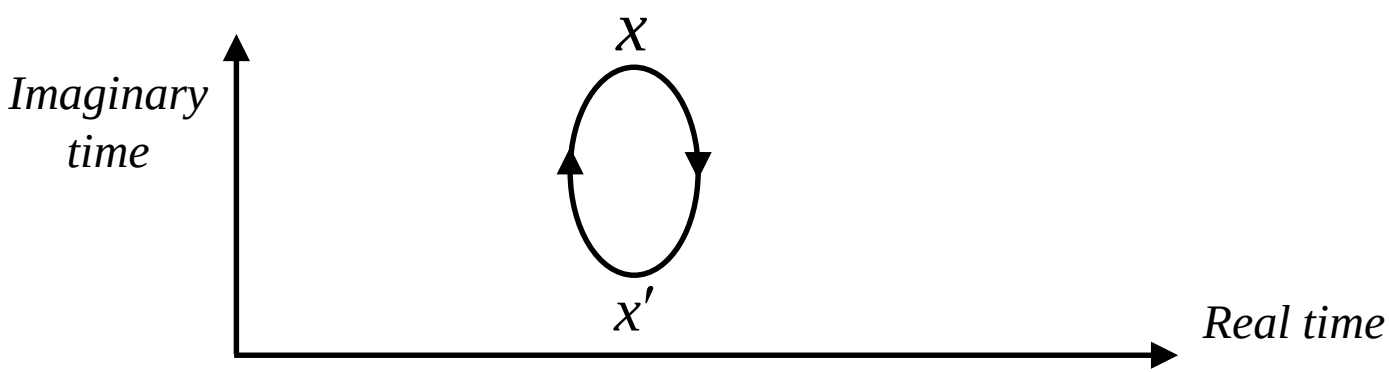
$$G_{\mathrm{AB}}^{\beta}(t) = C_{AB}^{\beta}(\tau^*) = \frac{1}{Z(\beta)}\mathrm{Tr}\left[\hat{\mathrm{A}}e^{i\hat{\mathrm{H}}\tau^*/\hbar}\hat{\mathrm{B}}e^{-i\hat{\mathrm{H}}\tau/\hbar}\right], \qquad \tau = t - \frac{i\beta\hbar}{2}$$

Connections:

$$\tilde{C}_{\mathrm{AB}}^{\beta}(\omega) = e^{\beta\hbar\omega/2}\tilde{G}_{\mathrm{AB}}^{\beta}(\omega), \qquad \tilde{G}_{\mathrm{AB}}^{\beta}(\omega) = \frac{\beta\hbar\omega}{2\sinh(\beta\hbar\omega/2)}\tilde{K}_{AB}^{\beta}(\omega)$$

If $\hat{\mathrm{A}} = \hat{\mathrm{A}}(\hat{\mathrm{x}})$, $\hat{\mathrm{B}} = \hat{\mathrm{B}}(\hat{\mathrm{x}})$

$$G_{\mathrm{AB}}^{\beta}(t) = \frac{1}{Z(\beta)}\int dx\, dx'\, \langle x|e^{iH\tau_c^*/\hbar}a(x)b(x')|x'\rangle\langle x'|e^{-iH\tau_c/\hbar}|x\rangle a(x)$$



Symmetric time correlation functions

Hamiltonian: $\hat{\mathrm{H}} = \dfrac{\hat{\mathrm{p}}^2}{2m} + \hat{\mathrm{U}}(\hat{\mathrm{x}})$

Complex time modulus: $|\tau|^2 = t^2 + \dfrac{\beta^2\hbar^2}{4}$

After application of Trotter splitting of complex-time propagators and insertion of resolution of identity [Krilov, *et al*. *J. Chem. Phys*. **114**, 1075 (2001)].

$$G^{\beta}_{\mathrm{AB},P}(t) = \frac{1}{Z(\beta)} \int \mathrm{d}\mathbf{x}\; a(x_1)\; b(x_{P+1})\; \rho(\mathbf{x};\beta,t)\; e^{i\varphi(\mathbf{x};\beta,t)}, \qquad \mathbf{x} = x_1, \ldots, x_{2P}$$

Density:

$$\rho(\mathbf{x};\beta,t) = \left(\frac{mP}{2\pi|\tau|\hbar}\right)^P \exp\left\{-\sum_{k=1}^{2P}\left[\frac{mP\beta}{4|\tau|^2}\left(x_{k+1}-x_k\right)^2 + \frac{\beta}{2P}U(x_k)\right]\right\}$$

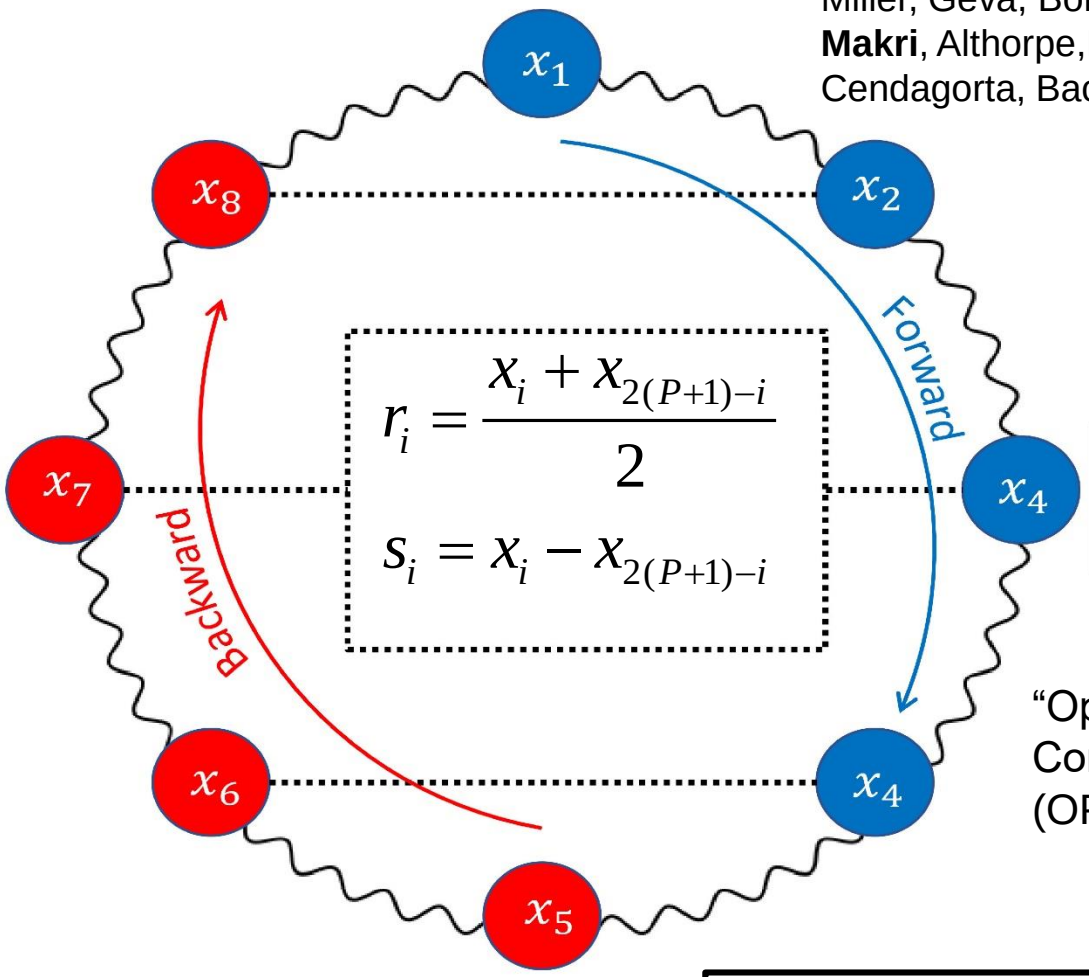
$$\boxed{x_{2P+1} = x_1}$$

Phase:

$$\varphi(\mathbf{x};\beta,t) = \frac{mPt}{2\hbar|\tau|^2}\left[\sum_{k=1}^{P}\left(x_{k+1}-x_k\right)^2 - \sum_{k=P+1}^{2P}\left(x_{k+1}-x_k\right)^2\right] - \frac{t}{P\hbar}\left[\sum_{k=2}^{P}U(x_k) - \sum_{k=P+2}^{2P}U(x_k)\right]$$

Coordinate transformation

Miller, Geva, Bonella, Coker, Ciccotti,
Makri, Althorpe,Huo, and many others!
Cendagorta, Bacic, MET *JCP* **148**, 102340 (2018)



$$r_i = \frac{x_i + x_{2(P+1)-i}}{2}$$

$$s_i = x_i - x_{2(P+1)-i}$$

Integrate over s_i

$$W(r) = U(r) + V(r)$$

"Open Path Symmetrized Correlation Function" (OPSCF)

$$r_{P+1} \neq r_1$$

$$x_1 = r_1, \qquad x_{P+1} = r_{P+1}$$

$$x_i = r_i + \frac{1}{2} s_i, \qquad x_{2(P+1)-i} = r_i - \frac{1}{2} s_i$$

Formally exact sampling scheme

Integration over $s_2,\ldots,s_P$

$$I(r) = \int ds_2 \cdots ds_P \; f(r,s) e^{i\chi(r,s)}$$

$$f(r,s) = \exp\left\{ -\alpha \sum_{i=2}^{P-1} \left(s_{i+1} - s\right)^2 - \alpha\left(s_2^2 + s_P^2\right) - \frac{\beta}{2P} \sum_{i=1}^{P} \left[U\left(r_i + \frac{1}{2}s_i\right) + U\left(r_i - \frac{1}{2}s_i\right) \right] \right\}$$

$$\alpha = \frac{mP\beta}{8\left|\tau_c\right|^2},$$

Formally exact sampling scheme

$$G_{\mathrm{AB}}^{\beta}(t) = \frac{1}{Z(\beta)} \int dx\, dx'\, a(x) b(x') \left| \left\langle x' \middle| e^{-iH\tau_c/\hbar} \middle| x \right\rangle \right|^2$$

Clearly,

$$\left| \left\langle x' \middle| e^{-iH\tau_c/\hbar} \middle| x \right\rangle \right|^2 \geq 0$$

After the transformation,

$$G_{AB,P}^{\beta}(t) = \int dr\; a(r_1) b(r_{P+1}) I(r;t,\beta)\; e^{-4\alpha \sum_{i=1}^{P} (r_{i+1} - r_i)^2 - \frac{\beta}{2P}\left(U(r_1) + U(r_{P+1})\right)}$$

Since the exponential term is positive definite, it follows that $I(r)$ must be positive definite, and we can write it like a PMF:

$$I(r;t,\beta) = e^{-\beta \tilde{V}(r;t,\beta)}$$

***Therefore, a formally exact sampling scheme, i.e., sampling of the r-integral, exists for the calculation of the symmetrized time correlation function.*

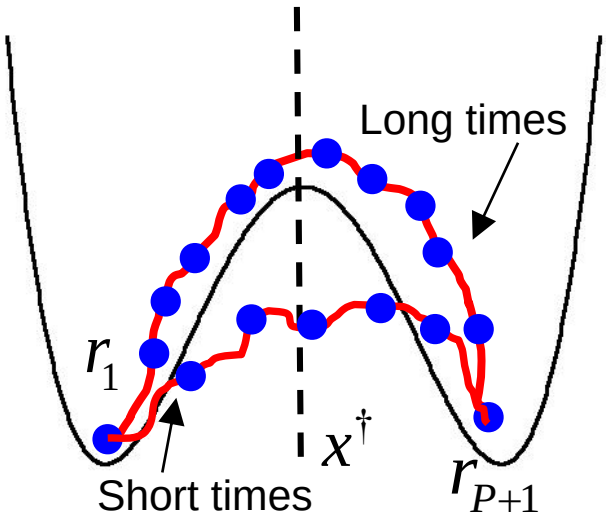
Formally exact sampling scheme

Rate constant in terms of side-side correlation function

$$k(\beta) = \frac{1}{Z_r(\beta)} \lim_{t\to\infty} \frac{d}{dt} G_{\mathrm{ss}}(t)$$

$$G_{ss}^{\beta}(t) \;=\; \mathrm{Tr}\left[\theta(\hat{q}(\hat{x}) - q^{\dagger} I) e^{i\tau^* \hat{H}/\hbar} \left(1 - \theta(\hat{q}(\hat{x}) - q^{\dagger} I)\right) e^{-i\tau \hat{H}/\hbar}\right]$$

$$G_{ss,P}^{\beta}(t) = \int dr\; \theta(q(r_1) - q^{\dagger} I)\left(1 - \theta(q(r_{P+1}) - q^{\dagger} I)\right) I(r;t,\beta)\; e^{-\beta u(r_1,\ldots,r_{P+1};t,\beta)}$$



Acknowledgments

- **Joseph Cendagorta** (NYU)
- **Ondrej Marsalek** (Prague)
- **Alejandro Pérez** (Spain)
- **Zlatko Bačić** (NYU)
- **Timothy Hele** (Cornell)
- **Magali Benoit** (CEMES, Toulouse)
- **Merlin Méheut** (University of Toulouse)
- **Romain Dupuis** (Spain)