First Passage Time Boxed Molecular Dynamics

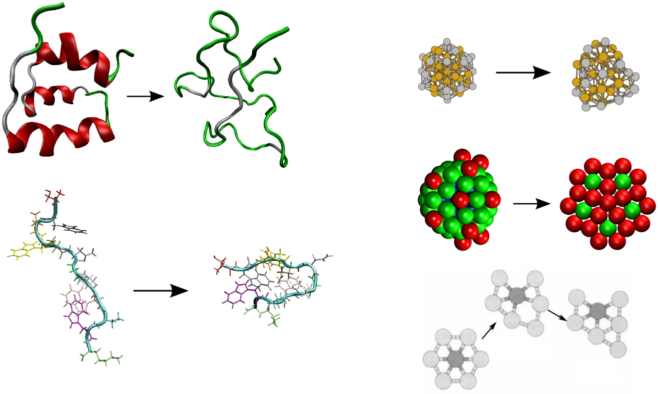
Boris Fačkovec

Wales group
Department of Chemistry
University of Cambridge

CPGS talk
13th December 2013

Chemical dynamics

- states = boxes in configuration space



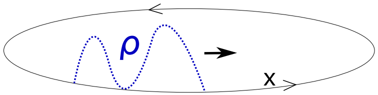
- general problem in chemical dynamics - given species A, what will the population of species be after time t?

Liouville equation is the answer

- density ρ at phase space point $(\mathbf{p}, \mathbf{q})$ and time t

$$\frac{\partial \rho(\mathbf{p}, \mathbf{q}, t)}{\partial t} = \sum_{i=1}^{N} \left(\frac{\partial H(\mathbf{p}, \mathbf{p})}{\partial q_i} \frac{\partial \rho(\mathbf{p}, \mathbf{q}, t)}{\partial p_i} - \frac{\partial H(\mathbf{p}, \mathbf{q})}{\partial p_i} \frac{\partial \rho(\mathbf{p}, \mathbf{q}, t)}{\partial q_i} \right)$$

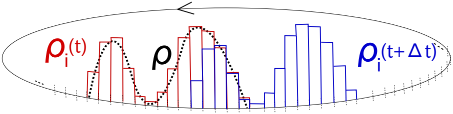
- limit of Newton equations of motion for infinite number of particles
- basic equation of non-equilibrium statistical mechanics
- example - motion on a circle with $V \equiv 0$



$$\frac{\partial \rho(x, t)}{\partial t} = -\dot{x} \frac{\partial \rho(x, t)}{\partial x}$$

Discretisation of the Liouville equation

- partial differential equation - can be discretised on a mesh



- dynamics are given by the transition matrix

$$\boldsymbol{\rho}(n\Delta t) = \mathbf{T}^n(\Delta t)\ \boldsymbol{\rho}(0)$$

- transition matrix can be estimated from MD
- not directly applicable for molecular (high-dimensional) systems

Master equation

- special form of the generalised master equation
- continuous limit of the description in terms of a transition matrix ($\mathbf{T}' = \lim_{\Delta t \to 0} \mathbf{T}(\Delta t)/\Delta t$)

$$\boldsymbol{\rho}(t) = e^{\mathbf{T}'t}\ \boldsymbol{\rho}(0)$$

- population P_{A} defined as

$$P_{\mathrm{A}}(t) = \int_{\mathrm{A}} \rho(\mathbf{p}, \mathbf{q}, t)\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}$$

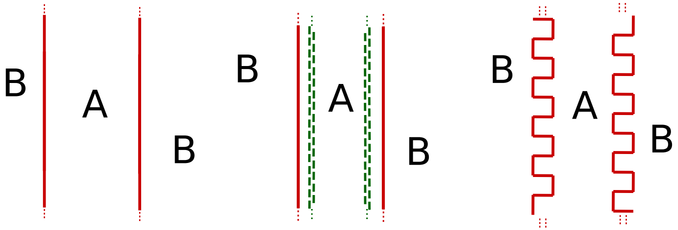
- assumption that the rate constant k_{A} is constant in time

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{\mathrm{A}}(t) = -\sum_{\mathrm{i}} k_{\mathrm{Ai}}\ P_{\mathrm{A}}(t) + \sum_{\mathrm{j}} k_{\mathrm{jA}} P_{\mathrm{j}}(t)$$

- idealisation of real systems

Rate constants for master equations

- definition - fits the evolution of population by an exponential
- all "reactive flux" methods are approximate!
  - Wigner's original transition state theory
  - Bennet-Chandler's method and all derived methods
- equilibrium box-to-box rate constants are not suitable for master equation
- box-to-box → the transmission coefficient is unity

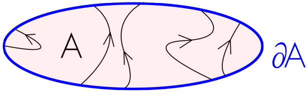


Boxed molecular dynamics (BXD)

- enhanced sampling by high number of short simulations
  - similar to milestoning and transition interface sampling
- BXD proposed by Glowacki, Shalashilin and Paci in 2009
  - peptide cyclisation
  - theory is based on previous methods of the authors
  - not benchmarked on simple systems
  - reaction coordinate required
  - Langevin dynamics
  - 2012 good agreement with experiment on ps to $\mu\mathrm{s}$ scale
- advantages over standard MD
  - exponential enhancement
  - diffusion
  - trivial parallelisation
- advantages over other enhanced sampling methods
  - natural connection to master equation (and NGT)
  - anharmonicity and diffusion effects included

Theory of chemical dynamics

single box



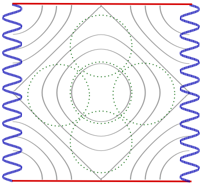
- $P_{\mathrm{A}}(t)$ - population in box A from equilibrium
  - absorbing boundary conditions
- $F_{\mathrm{A}}(t)$ - first passage time distribution
- $R_{\mathrm{A}}(t)$ - return time distribution - also distribution of trajectory lengths

$$F_{\mathrm{A}}(t) = -\frac{\mathrm{d}}{\mathrm{d}t} P_{\mathrm{A}}(t) \qquad R_{\mathrm{A}}(t) = \frac{-\frac{\mathrm{d}}{\mathrm{d}t} F_{\mathrm{A}}(t)}{F_{\mathrm{A}}(0)}$$

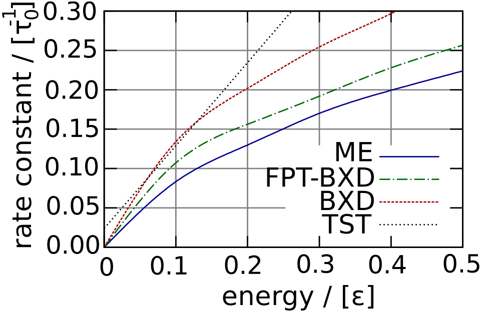
- mean values can be used as the rate constants

single box

- simple models - simulation of master equation possible



potential energy and
dividing surface



$$k^{\mathrm{FPT-BXD}}_{\mathrm{A}\to\partial\mathrm{A}} = \frac{1}{\mathrm{MFPT}} = \frac{1}{\langle F_{\mathrm{A}}(t)\rangle} \qquad k^{\mathrm{BXD}}_{\mathrm{A}\to\partial\mathrm{A}} = k^{\mathrm{TST}}_{\mathrm{A}\to\partial\mathrm{A}} = \frac{1}{\langle R_{\mathrm{A}}(t)\rangle}$$

Rate constants from MFPT I

two boxes

- MFPT to the boundaries is not reciprocal of the rate constant
  - also does not satisfy the equilibrium limit
- the whole process = exit A + penetration of B
- Bicout and Szabo 1997

$$\tau_{\mathrm{rx}}^{\mathrm{BS}} = \frac{K_{\mathrm{AB}} M_{\mathrm{A}} + M_{\mathrm{B}}}{K_{\mathrm{AB}} + 1}$$

- works only for strong diffusion
- does not agree with the transition state formula

$$\tau_{\mathrm{rx}}^{\mathrm{TST}} = \frac{\tau_{\mathrm{rx}}^{\mathrm{BS}}}{2}$$

Rate constants from MFPT II

two boxes

- description for all FPT distributions - numerical integration of the system

$$\frac{\partial}{\partial t}F_{\mathrm{A}}(t,s) = \frac{\partial}{\partial s}F_{\mathrm{A}}(t,s) + F_{\mathrm{B}}(t,0)\ R_{\mathrm{A}}(s)$$
$$\frac{\partial}{\partial t}F_{\mathrm{B}}(t,s) = \frac{\partial}{\partial s}F_{\mathrm{B}}(t,s) + F_{\mathrm{A}}(t,0)\ R_{\mathrm{B}}(s)$$

with boundary conditions

$$F_{\mathrm{A}}(0,s) = F_{\mathrm{A}}(s)$$
$$F_{\mathrm{B}}(0,s) \equiv 0.$$
$$\lim_{t\to\infty} F_{\mathrm{A}}(t,s) = \lim_{t\to\infty} F_{\mathrm{B}}(t,s) = 0.$$

where $R_{\mathrm{B}}(s)$ is obtained directly from a BXD simulation

Optimised simulation protocol

multiple boxes

- obtaining the initial structures
- Metropolis Monte Carlo $\rightarrow$ canonical distribution of starting structures
- MD simulation starting from the MC initial structures, terminated at box boundaries

- assumption of total decorrelation of input and exit trajectories is too strong
  - downhill passage of the box (equilibrium)
  - recrossing (non-equilibrium)
- assumption of ergodic sampling
- boundaries with other boxes can be replaced by hard walls

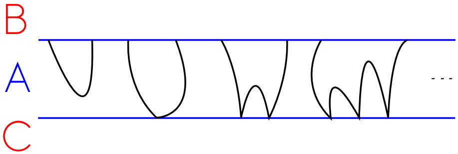
Equilibrium fluxes

multiple boxes

- Average length of the trajectories with endpoints i and j:

$$L_{\mathrm{ij|A}} = \frac{n_{\mathrm{ij}}}{\sum_{l=1}^{n_{\mathrm{ij}}} \frac{1}{\tau_{\mathrm{ij}}^{l}}}$$

analogously: probability of endpoints for a trajectory,



the equilibrium (TST) rate constant

$$k_{\mathrm{AB}}^{\mathrm{eq}} = \frac{1}{n_{\mathrm{BB}} + n_{\mathrm{BC}} + n_{\mathrm{CC}}} \left(\sum_{l=1}^{n_{\mathrm{BB}}} \frac{1}{\tau_{\mathrm{BB}}^{l}} + \frac{1}{2} \sum_{l=1}^{n_{\mathrm{BC}}} \frac{1}{\tau_{\mathrm{BC}}^{l}} \right)$$

Rate constants

multiple boxes

Discretisation of system from slide 11 for multiple boxes:

$$F^{\mathrm{i}}_{\mathrm{BB|A}}(t) = \int_{(\mathrm{i}-1)\Delta t}^{\mathrm{i}\ \Delta t} F_{\mathrm{BB|A}}(t,s)\ \mathrm{d}s$$

leads to

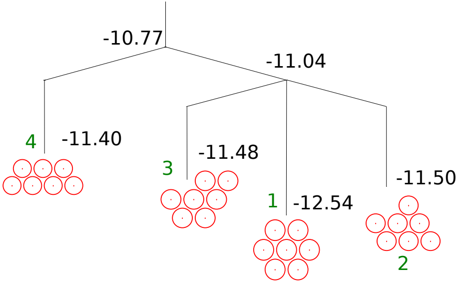
$$F^{\mathrm{i}}_{\mathrm{BB|A}}(t+\Delta t) = F^{\mathrm{i}+1}_{\mathrm{BB|A}}(t) + \left[F^{1}_{\mathrm{AA|B}}(t) + F^{1}_{\mathrm{AD|B}}(t)\right] P_{\mathrm{BB|A}} R^{\mathrm{i}}_{\mathrm{BB|A}}$$

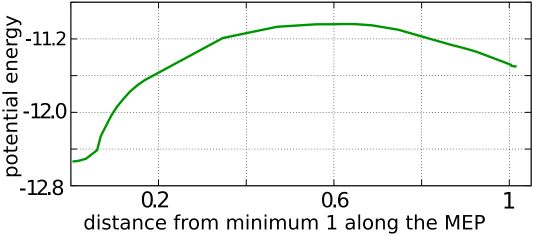
and 7 more analogous equations. Boundaries $\partial\mathrm{AC}$ and $\partial\mathrm{BD}$ are replaced by randomising hard walls. Population

$$P_{\mathrm{A}}(t) = \sum_{\mathrm{i}} \left(F^{\mathrm{i}}_{\mathrm{BB|A}}(t) + F^{\mathrm{i}}_{\mathrm{BC|A}}(t) + F^{\mathrm{i}}_{\mathrm{CB|A}}(t) + F^{\mathrm{i}}_{\mathrm{CC|A}}(t)\right)$$

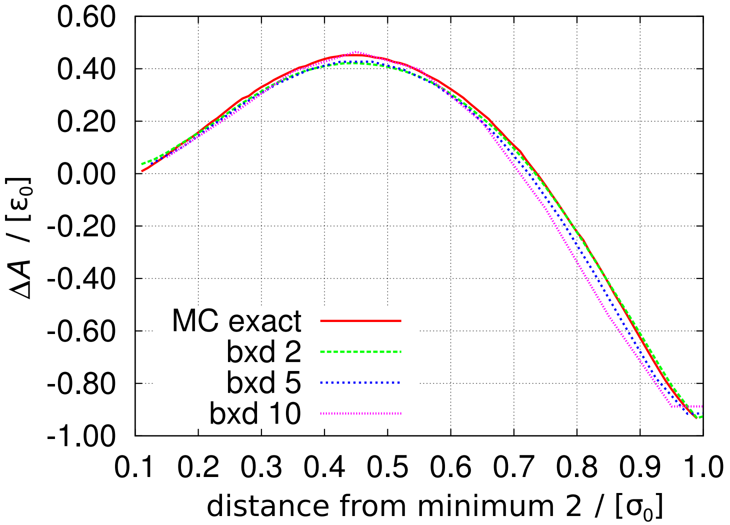
can be fitted to an exponential.

Test system - LJ_7^{2D}

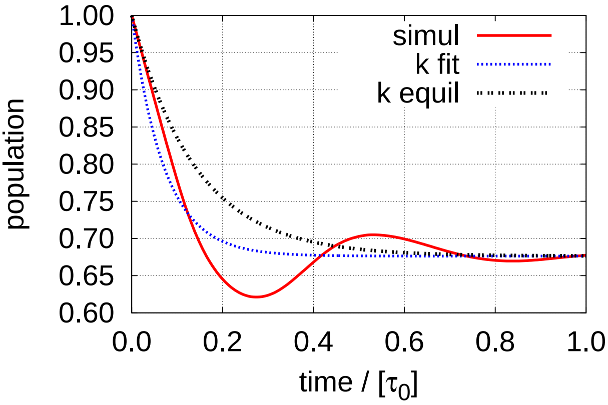




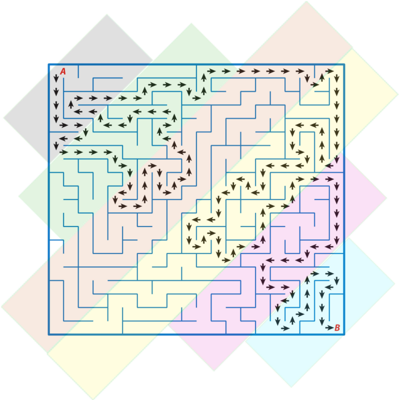
Free energy profile

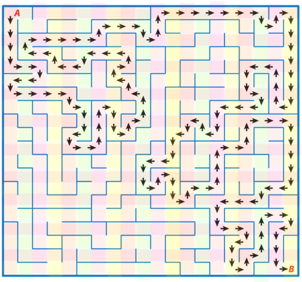


	box 2	box 5	box 10	TPS	DPS
$\ln(k_{12}/\tau_0)$	-28.9 $\pm$ 0.3	-29.6 $\pm$ 0.5	-30.5 $\pm$ 1.0	-28.7	-26.9
$\ln(k_{21}/\tau_0)$	-9.3 $\pm$ 0.2	-9.4 $\pm$ 0.3	-9.7 $\pm$ 0.7	-7.2	-9.2



Limitations for boxes





- a problem is indicated by an inequality:

$$p_{\mathrm{BB|A}} \neq \left(2p_{\mathrm{BB|A}} + \sum_{\mathrm{B}\neq\mathrm{i}} p_{\mathrm{Bi|A}}\right)^2$$

FPT-BXD - current state and future work

- enhanced sampling
- general definition of states (weighted Voronoi construction, slicing along a collective coordinate)
- deterministic dynamics
- works for non-exponential FPT distributions

- larger LJ clusters
  - benchmarking
  - minima lumping error studies
- simplistic polymers
- alanine dipeptide
- towards large proteins
  - hyperdynamics
  - local rigidification
  - systematic coarse-graining

Acknowledgement

Nadace
Zdeňka Bakaly