Car–Parrinello Molecular Dynamics

Tutorial

Chemical Reactions

Jürg Hutter and Marcella Iannuzzi

University of Zurich
How does a Chemical Reaction Occur?

Standard MD samples equilibrium states
No statistically significant sampling of the transition surface
Coarse Grain Definition of a State

Map the $3N$ DoF onto a reduced configurational space.

Microstates $\Rightarrow$ Macrostates

No available equations of motion in the reduced space of macrostates
Constraints

One might want a constraint . . .

- as a reaction coordinate
- to prevent a reaction/change in system
- to increase time step (e.g. —CH₃ group)

Constraints are defined as implicit functions of the general degrees of freedom of the system.

$$\sigma(\mathbf{R}^N) = 0$$

The dependence on the electronic degrees of freedom will be considered later.
Lagrangian Formulation

Lagrangian (multipliers $\lambda$)

$$
\mathcal{L}'(\mathbf{R}^N, \mathbf{P}^N) = \mathcal{L}(\mathbf{R}^N, \mathbf{P}^N) - \sum_{\alpha} \lambda_{\alpha} \sigma_{\alpha}(\mathbf{R}^N)
$$

Equation of motion

$$
\frac{d}{dt} \frac{\partial \mathcal{L}'(\mathbf{R}^N, \mathbf{P}^N)}{\partial \mathbf{P}_I} = \frac{\partial \mathcal{L}'(\mathbf{R}^N, \mathbf{P}^N)}{\partial \mathbf{R}_I} - \sum_{\alpha} \lambda_{\alpha} \frac{\partial \sigma_{\alpha}}{\partial \mathbf{R}_I}
$$

$$
M_I \ddot{\mathbf{R}}_I = -\frac{\partial E}{\partial \mathbf{R}_I} - \sum_{\alpha} \lambda_{\alpha} \frac{\partial \sigma_{\alpha}}{\partial \mathbf{R}_I}
$$

$$
= \mathbf{F}_I + \sum_{\alpha} \mathbf{G}_{I}^{\alpha}
$$

where $\mathbf{G}_{I}^{\alpha} = \lambda_{\alpha} \frac{\partial \sigma_{\alpha}}{\partial \mathbf{R}_I} \text{ is the force from constraint } \alpha \text{ on atom } I.$
Simple Constraint Types

- **Distance:** $|\mathbf{R}_I - \mathbf{R}_J| - d_0$
  $d_0 = \text{target distance}$

- **Distance:** $|\mathbf{R}_I - \mathbf{R}_J|^2 - d_0^2$
  $d_0 = \text{target distance}$

- **Angle:** $\theta (\mathbf{R}_I, \mathbf{R}_J, \mathbf{R}_K) - \theta_0$
  $\theta_0 = \text{target angle}$

- **Dihedral angle:** $\Theta (\mathbf{R}_I, \mathbf{R}_J, \mathbf{R}_K, \mathbf{R}_L) - \Theta_0$
  $\Theta_0 = \text{target angle}$

- **Distance difference:** $|\mathbf{R}_I - \mathbf{R}_J| - |\mathbf{R}_J - \mathbf{R}_K| - d_0$
  $d_0 = \text{target difference}$
Coordination Constraints

- Coordination number

\[ n_I (R^N) = \sum_{J \neq I} S (|R_I - R_J|) \]

- Convolution

\[ S (r) = \frac{1}{\exp [\kappa (r - r_c)] + 1} \]

- \( \kappa^{-1} \) = width of transition region, \( r_c \) = coordination radius

- Force

\[ \tilde{g}_I = -\lambda \frac{\partial n_I}{\partial R_I} = \frac{1}{2} \frac{\lambda \kappa}{\cosh [\kappa (r - r_c)] + 1} \frac{R_{JI}}{R_{JI}} \]

acts only within the vicinity of \( r_c \)
Coordination number

A too narrow transition region $[r_c - \kappa^{-1}; r_c + \kappa^{-1}]$ causes sudden impulses and instabilities; if too large the coordination is not well determined.
Generalized Coordination Number

Select two arbitrary lists of atoms $L_1$ and $L_2$ and a reference distance $r_0$

$$C_{L_1L_2} = \frac{1}{N_{L_1}} \sum_{j=1}^{N_{L_1}} \left\{ \sum_{i=1}^{N_{L_2}} \frac{1 - \left( \frac{r_{ij}}{r_0} \right)^n}{1 - \left( \frac{r_{ij}}{r_0} \right)^m} \right\}$$
Closed Chain of H-Bonds

Define the presence of a hydrogen bond between two molecules \(I\) and \(J\) by

\[
f_{IJ} = \frac{1 - \left( \frac{r_{ij} - C}{r_{ij}^0} \right)^{p_C} \sum_{k=1}^{N_H} 1 - \left( \frac{r_{ik} + r_{jk} - r_{ij}}{r_{ik}^0 + r_{jk}^0 - r_{ij}^0} \right)^{p_H} }{1 - \left( \frac{r_{ij} - C}{r_{ij}^0} \right)^{q_C} q_C}
\]

Given the triplets \(T\) formed by one molecule \(I\) and pairs of its neighbors, and the closed paths \(P\) (length \(L_P\)) containing \(T\), the connectivity is computed by

\[
F_T^P = \prod_{IJ \in P} f_{IJ} \quad G_T = \lambda \left( \sum_P e^{\lambda F_T^P / L_P} \right)^{-1}
\]

\(G\) selects the shortest \(P\) and the number of \(n\)-membered rings is given by

\[
N_n = \sum_T \frac{1 - \left( \frac{n - G_T}{\sigma} \right)^p}{1 - \left( \frac{n - G_T}{\sigma} \right)^q}
\]
Hydronium

The presence of the $H_3O^+$ structure can be identified through the shape of the first coordination shell of the involved $O$.

$$S_{hyd}(R) = \frac{1}{\lambda} \ln \sum_{k=1}^{NO} \left[ \exp \left\{ \lambda f(n_{Hk}) n_{Ok} \right\} \right]$$

$$f(n_{Hk}) = 1 - \frac{1 - \left( \frac{n_{Hk}}{n^0_H} \right)^{p_f}}{1 - \left( \frac{n_{Hk}}{n^0_H} \right)^{q_f}}$$

$$n_{Hk} \sim 3 \text{ reference OH coordination number}$$

$$n_{Hk} = \sum_{j=1}^{NH} \frac{1 - \left( \frac{r_{jk}}{r^0_{OH}} \right)^{p_{OH}}}{1 - \left( \frac{r_{jk}}{r^0_{OH}} \right)^{q_{OH}}}$$

$$n_{Ok} = \sum_{i=1}^{NO} \frac{1 - \left( \frac{r_{ik}}{r^0_{OO}} \right)^{p_{OO}}}{1 - \left( \frac{r_{ik}}{r^0_{OO}} \right)^{q_{OO}}}$$
Proton exchange

To monitor the hopping of the protons along the chain we define a generalized relative displacement

\[
S_D(R) = \frac{1}{N_{\tilde{H}}} \sum_{i=1}^{N_{\tilde{H}}} d_i \cdot \hat{v}_{[001]} - \frac{1}{N_{\tilde{N}}} \sum_{j=1}^{N_{\tilde{N}}} d_j \cdot \hat{v}_{[001]}
\]

where \(i\) runs over the protons in the chain \((N_{\tilde{H}})\), \(j\) runs over the nitrogen in the chain \((N_{\tilde{N}})\), and \(d\) is the displacement wrt the initial position.
Constraints in Velocity Verlet

**Velocity Verlet**

\[
\dot{\mathbf{R}}_I = \dot{\mathbf{R}}_I(t) + \frac{\delta t}{2M_I} \mathbf{F}_I(t)
\]

\[
\ddot{R}_I = \mathbf{R}_I(t) + \delta t \ddot{R}_I
\]

\[
\mathbf{R}_I(t + \delta t) = \ddot{R}_I + \frac{\delta t^2}{2M_I} \mathbf{g}_p(t) \quad \text{(SHAKE)}
\]

calculate \quad \mathbf{F}_I(t + \delta t)

\[
\dot{\mathbf{R}}'_I = \dot{\mathbf{R}}_I(t) + \frac{\delta t}{2M_I} \mathbf{F}_I(t + \delta t)
\]

\[
\dot{\mathbf{R}}_I(t + \delta t) = \dot{\mathbf{R}}'_I + \frac{\delta t}{2M_I} \mathbf{g}_v(t + \delta t) \quad \text{(RATTLE)}
\]

**Constraint forces**

\[
\mathbf{g}_p(t) = - \sum_i \lambda^i_p \frac{\partial \sigma^{(i)}(\{\mathbf{R}_I(t)\})}{\partial \mathbf{R}_I}
\]

\[
\mathbf{g}_v(t) = - \sum_i \lambda^i_v \frac{\partial \sigma^{(i)}(\{\mathbf{R}_I(t)\})}{\partial \mathbf{R}_I}
\]
**SHAKE**

- **Position Constraint:** \( \sigma_i(R_I(t + \delta t)) = 0 \)
- System of nonlinear equations for Lagrange multipliers \( \lambda^i \).
- Solve by generalized Newton algorithm with convergence acceleration (DIIS).
- Error vector:
  \[
e_i(\lambda) = - \sum_j J_{ij}^{-1}(\lambda) \sigma^{(j)}(\lambda)
\]
- Jacobian (with \( f^c_I(\lambda) = \sum_i \lambda^i \partial \sigma^{(i)} / \partial R_I \))
  \[
  J_{ij}(\lambda) = \frac{\partial \sigma^{(i)}(\lambda)}{\partial \lambda^j} = \sum_I \frac{\partial \sigma^{(i)}(\lambda)}{\partial R_I(\lambda)} \frac{\partial R_I(\lambda)}{\partial \lambda^j} = - \sum_I \frac{\delta t^2}{2M_I} f^c_I(\lambda)f^c_I(0)
  \]
RATTLE

- **Velocity Constraint:** \( \dot{\sigma}_i(R_I(t + \delta t)) = 0 \)
- **System of linear equations:**

\[
A_{iI} = \frac{\partial \sigma^{(i)}}{\partial R_I}
\]

\[
- \sum_j \left( \sum_I \frac{\delta t^2}{2M_I} A_{iI} A_{jI} \right) \lambda_j^y = \sum_I A_{iI} \dot{R}_I'
\]
Electronic Constraints

Chemical reactions imply rearrangements of the electronic structure that can be described by electronic coordinates which more easily distinguish among the different reactive channels. In the “electron preceding” picture one could think to force the electronic cloud to certain directions, consequently pushing the ions to follow them.

- **Born-Oppenheimer MD**

\[ \sigma_i(R^N, \{\Phi_i(R^N)\}) = 0 \]

- Wavefunctions are implicit functions of atomic positions.
- Forces require wavefunction response \( \frac{\partial \Phi}{\partial R_i} \)

- **Car-Parrinello MD**

\[ \sigma_i(R^N, \{\Phi_i\}) = 0 \]

- Wavefunctions are independent degrees of freedom.
- Constraint forces may couple EOM for atomic positions and wavefunctions \( \rightarrow \) iterative solution for Verlet steps.
Constrain on the $KS$ Energies $\varepsilon_i$

A given linear combination of the $\varepsilon_i$ is constrained to a target value $E_T$

$$\sigma = \left( \sum_i w_i \varepsilon_i - E_T \right) = 0$$

$$\frac{\partial \sigma}{\partial R_I} = \sum_i w_i \frac{\partial \varepsilon_i}{\partial R_I}$$

where $\frac{\partial \varepsilon_i}{\partial R_I}$ can be evaluated as linear response.

Active perturbations of the orbital energies can be used to distort a molecule along its soft reacting modes, e.g. closure of the $\pi$ to $\pi^*$ gap in ethylene induces the rotation of the $CH_2$ group.
Restraints

- Similar to constraints, but harmonic and allows for deviations
- For example distance, \( E_{\text{restraint}} = k_r (|R_I - R_J| - d_0)^2 \), 
  \( d_0 \) = target distance, \( k_r \) = force constant of restraint
- Can be either attractive \( (k_r > 0) \) or repulsive \( (k_r < 0; \text{ e.g. in angles}) \)
- Does not in general allow a increase of the time step (rather would need a reduction), but can be used e.g. to hold atoms in a molecule together
- Simplifies treatment of electronic reaction coordinates
Bias on the $KS$ Energies $\varepsilon_i$ along $BOMD$

The system is driven by a bias potential along the desired reactive channel by variating smoothly the target value: $E_T(t) = E_T(0) + t\Delta E$.

$$E_{bias} = \frac{1}{2} \alpha \left( \sum_i w_i \varepsilon_i - E_T \right)^2$$

The correspondent forces on nuclei are

$$f_I = -\frac{\partial E_{bias}}{\partial R_I} = -\alpha \sum_i \left( \sum_j w_j \varepsilon_j - E_T \right) w_i \frac{\partial \varepsilon_i}{\partial R_I}$$

where $\frac{\partial \varepsilon_i}{\partial R_I}$ is calculated in a perturbative approach within linear response theory. When $i$ is $HOMO$ or $LUMO$, $\frac{\partial \varepsilon_i}{\partial R_I}$ is the nuclear Fukui function, i.e. the response of the force on the nucleus due to oxidation or reduction.

$$\varphi_I = \frac{\partial F_I}{\partial N}$$

i.e. response of the force on the nucleus due to oxidation or reduction.
Example: \( HOMO - LUMO \) gap

In the cyclization of \textit{cis}-butadiene, the disrotatory symmetry forbidden pathway involves the correlation of \( \pi \)-type HOMO of butadiene correlates to antibonding \( \pi^* \) LUMO of cyclobutadiene \( \Rightarrow \) orbital crossing and a closure of the HOMO-LUMO gap.

\[
E_{bias} = \frac{1}{2} \alpha (\varepsilon_{LUMO} - \varepsilon_{HOMO} - E_T) \quad - E_{gap} \geq E_T \leq E_{gap}
\]
Rare Event: Barrier Crossing

It takes a very long time to cross barriers if the height is much larger than $k_B T$

$$\tau \sim \tau_{mol} e^{\frac{F}{K_B T}}$$
Multidimensional Reaction Coordinate

- What happens if a single reaction coordinate is not enough?
- The low-energy path might not be captured
Unforeseen Stationary States

- ... or what if the potential energy landscape is very rugged?
- Dynamical bottleneck and diffusive barriers
- Many irrelevant saddle points
Rare events

- MD is a real time method, with a time step of the order of 0.1–1 fs
- however, in Nature many effects occur at time scales much longer than the realistic times (e.g. in biology seconds)
- due to high energy barriers or improbable location in phase space (in Arrhenius rate of reaction low prefactor)
- Ways to direct reactions:
  - high temperature
  - constraints
  - bias potentials
  - metadynamics
Some Literature

- **Blue Moon ensemble theory:** G. Ciccotti et al. J. Chem. Phys. 109, 7737 (1998) and ref. therein
- **Hessian based methods:** eigenvalue following, dimer method.
- **Adaptive bias potential:** van Gunsteren et al. J. Comput. 8, 695 (1994)
- **Accelerated dynamics:** A. Voter, J. Chem. Phys. 106 (11), 4665 (1997)
- **Flooding potential:** H. Grubmueller Phys. Rev. E 52, 2893 (1997)
Partitioned Rational Function Optimizer

Assuming that the TS is a saddle point, the transition surface can be identified from the gradients of the PES.

- Select the mode along which the energy is maximized. Switching the mode along the optimization is possible (overlap).
- Minimize the energy along all the other modes
- Update the inverse Hessian applied to a gradient by extrapolation (L-BFGS)
- Split the system in core and environment

$$G = \frac{dE_{AB}}{dAB} = \frac{\partial E_A}{\partial A} + \frac{\partial E_B}{\partial B} + \frac{\partial E_B}{\partial A} + \frac{\partial E_A}{\partial B}$$

if \( G_B = 0 \) \( \Rightarrow \) \( G = \frac{\partial E_A}{\partial A} + \frac{\partial E_B}{\partial A} \)

- Environment relaxed with fixed core, before each optimization step for the core. If \( A' = A + D \) and \( B' \) gives \( \partial E_{A'B'}/\partial B' = 0 \) then

$$\frac{d^2E_{AB}}{dAB^2} \simeq \left[ \frac{\partial E_A}{\partial A'} + \frac{\partial E_B}{\partial A'} - \frac{\partial E_A}{\partial A} - \frac{\partial E_B}{\partial A} \right]/D$$

Hessian only needed for the core
Enzyme Modelling: LDH

Grid scanning to locate possible TS: $C_{pyr} - H'$ and $C_{pyr} - H''$. Refinement by PRFO.
Reaction Coordinates

The microstate is defined by a set of independent variables: Degrees of Freedom

\[ \{ \mathbf{R}_I \} + \{ \phi_i \} + \{ h_{pq} \} \]

We assume that some known functions of the independent variables are necessary and sufficient to describe the process we are interested in

\[ S_\alpha(\{ \mathbf{R}_I \}, \{ \phi_i \}, \mathbf{h}) \quad \alpha = 1, ..., N_S < \text{DoF} \]

The Free Energy Hyper-surface in the reduced space defined by \( S \)

\[ \mathcal{F}(S') = -K_B T \ln \mathcal{P}(S') \]

\[ \mathcal{P}(S') = \frac{1}{Q} \int_{\Omega} d\Omega \ e^{\mathcal{H}(\{ \mathbf{R}_I \}, \{ \phi_i \}, \mathbf{h})/K_B T} \delta(\mathbf{S}(\{ \mathbf{R}_I \}, \{ \phi_i \}, \mathbf{h}) - S') \]

The integral can be evaluated by MD
Thermodynamic Integration

\[ \mathcal{F}(\xi_2) - \mathcal{F}(\xi_1) = \int_{\xi_1}^{\xi_2} dS' \left\langle \frac{\partial \mathcal{H}}{\partial S} \right\rangle_{s'}^{\text{cond.}} \]

where \( \mathcal{F} \) is the free energy and \( S'(...) \) a one–dimensional reaction coordinate, \( \mathcal{H} \) the Hamiltonian of the system and \( \langle \cdots \rangle_{s'}^{\text{cond.}} \) the conditioned average in the constraint ensemble.

By way of the Blue Moon Ensemble, the statistical average is replaced by a time average over a constrained trajectory with the reaction coordinate fixed at a special values,

\[ S(\{R_I\}, \{\phi_i\}, h) = \xi \]

and

\[ \dot{S}(\{R_I\}, \{\phi_i\}, h, \{\dot{R}_I\}, \{\dot{\phi}_i\}, \dot{h}) = 0 \]

The average is corrected for the bias introduced in momentum space: coordinate dependence of the kinetic energy.
The force $f_S$ drives the dynamical survey of the FES. A number of independent replica for each value of the constraint can be generated and an ensemble average over the replica can be used to improve the statistics.
Thermodynamic Integration

The quantity to evaluate is the mean force

\[
\frac{d\mathcal{F}}{dS} = \frac{\langle Z^{-1/2} [-\lambda + k_BT G] \rangle_{S_{MDn}}}{\langle Z^{-1/2} \rangle_{S_{MDn}}}
\]

\(\lambda\) is the Lagrange multiplier of the constraint,

\[
Z = \sum_I \frac{1}{M_I} \left( \frac{\partial S}{\partial \mathbf{R}_I} \right)^2
\]

and

\[
G = \frac{1}{Z^2} \sum_{I,J} \frac{1}{M_IM_J} \frac{\partial S}{\partial \mathbf{R}_I} \cdot \frac{\partial^2 S}{\partial \mathbf{R}_I \partial \mathbf{R}_J} \cdot \frac{\partial S}{\partial \mathbf{R}_J}
\]

where \(\langle \cdots \rangle_{\xi'}\) is the unconditioned average, obtained from a constrained MD and

\[
\mathcal{F}(\xi_2) - \mathcal{F}(\xi_1) = \int_{\xi_1}^{\xi_2} dS_{MDn} \left. \frac{d\mathcal{F}}{dS} \right|_{MDn}
\]

For the special case of a simple distance constraint \(\xi(\mathbf{R}) = |\mathbf{R}_I - \mathbf{R}_J|\) the parameter \(Z\) is a constant and \(G = 0\).
$S_{N2}$ reaction $F^- + CH_3Cl \rightarrow CH_3F + Cl^-$

Mugnai, Cardini & Schettino, *JCP* 2003

- Left: (Free) energy profile along the reaction paths at 0 and 300 K
- Right: Dipole moment $CH_3X$ and $Y^-$ along the 0 K reaction path
Autoionization of Liquid Water

Contraint: $O^* - H^*$ distance

Fig. 1. Free energy profile along the reaction coordinate. The dotted lines in the schematics represent the constraint and the dashed line, a hydrogen bond.
Fig. 2. Selected average bond distances as a function of the reaction coordinate. The non-starred H and O represent the nearest non-constrained atoms to the O* and H*, respectively.
Figure 3. Average hydrogen coordination number on O* and O'.
Autoionization of Liquid Water

Coordination Constraint treats all protons equally

In order to tune $r_c$ and $\kappa$, overlap the $OH$ radial distribution function and the weight function.

$\rho_{OH}(r)$, $S(r)$

$r_c = 1.38 \text{ Å} \quad \kappa^{-1} = 0.1 \text{ Å} \quad < n_H >= 1.98$
Averages over a trajectory of more than 2 ps.

Distance of $O^*$ to the three nearest hydrogen atoms as a function of coordination number ($n_H = 1.0$ ($OH^-$); $n_H = 2.0$ ($H_2O$); $n_H = 3.0$ ($H_3O^+$))
Free energy along the reaction path

Mean force (closed symbol) computed from the bluemoon ensemble and $\Delta F$ (free energy difference) computed as numerical integral of $f_{n_H}$ wrt $n_H = 2$. 
Statistical error $\sim 2 \text{ Kcal/mol}$

$$f_{nH} = \frac{\langle Z^{-1/2}[\lambda - K_BTG] \rangle_{nH}}{\langle Z^{-1/2} \rangle_{nH}}$$

Time dependence of correction quantities at $n_H = 1.3$. Top curve: $Z^{-1/2}$; middle curve: $\lambda$; bottom curve: $-kTG$. The correction effects never exceed a few percent.
Comparison of different constraints

Coordination number for $O^*$ as a function of OH distance for different types of constraints: distance $O_D H_T$, coordination $n_H$ Asymmetry in $\Delta F$ between protonation and deprotonation
Adiabatic Free Energy Sampling

Basic Idea

- Separate system in two subsystems
  Reactive system or coordinates (S)
  Bath or environment system (E)

- Couple both systems to thermostats, keeping system R and B at temperatures $T^S$ and $T^E$

- Choose the effective mass of system S such that it moves slowly and system E can follow adiabatically

- Choose temperatures $T^S$ high enough, such that the full phase space of S is sampled.
Adiabatic Free Energy Sampling

Hamiltonian of the system

\[ H(q, p) = \sum_{i}^{N^S} \frac{(p_i)^2}{2M_i^S} + \sum_{j}^{N^E} \frac{(p_j^E)^2}{2M_j^E} + V(q^S, q^E) \]

The fast degrees of freedom \( q_j^E, p_j^E \) sample the canonical distribution at temperature \( T^E \) for every configuration \( q_j^S, p_j^S \) of the reactive system. The dynamics of the slow degrees of freedom approximates the one resulting from a Hamiltonian with the effectively averaged potential. This potential \( \bar{V}(q^S) \) is, for a full decoupling given by

\[
\exp \left[ -\beta^E \bar{V}(q^S) \right] = \frac{\int \exp \left[ -\beta^E V(q^S, q^E) \right] d{q^E}}{\int \exp \left[ -\beta^E V(q^S, q^E) \right] d{q^E} d{q^S}}
\]

This potential of mean force equals the physical free energy surface of the reactive system.
Adiabatic Free Energy Sampling

An enhanced canonical sampling of this free energy surface can be obtained by choosing $T^S > T^E$. Probability distribution of the full system in configuration space is given by

$$\rho \left( q^S, q^E \right) = \frac{\exp \left[ - (\beta^S - \beta^E) \bar{V} \left( q^S \right) \right]}{\int \exp \left[ - \beta^S \bar{V} \left( q^S \right) \right] dq^S} \times \frac{\exp \left[ - \beta^E V \left( q^S, q^E \right) \right]}{\int \exp \left[ - \beta^E V \left( q^S, q^E \right) \right] dq^E dq^S}$$

This probability distribution can be used to calculate all thermodynamic properties of the physical system from a trajectory generated within the adiabatically decoupled system. Efficiency and accuracy of this method depends on the choice of subsystems and the free parameters $M^S / M^E$ and the temperature $T^S$. 
Conversion of 2-Bromoethanol to Dibromoethane

The heavy atoms of the bromoethanol molecule are taken as the reactive sub-system (masses scaled by 100, coupled to a Nose-Hoover at $T=2000$ K). Solute and $H$ at 300 K, QM/MM setup

Free exploration of configurations

110 C-C-O
angle
Br-C-C
dissociation
and bromonium concerted
Metadynamics

Coarse Grain and Non-Markovian dynamics

- Choose a set of relevant collective variables, e.g. distances, coordination number, simulation cell parameters, ...

$$S_{\alpha}(\{R_i\}, \{\phi_i\}, h)$$

assuming that the process is well described in the reduced space of $$\{S_{\alpha}\}$$

$$\Sigma_{CV}$$

- Perform a MD run and map at each step the microstate in $$\Sigma_{CV}$$ projection of the MD trajectory on $$\Sigma_{CV}$$

- Enhance the exploration by adding a penalty potential that discourage the system to visit a region already explored and keeps memory of the trajectory
After a time $t$ the force driving the dynamics in $\Sigma_{CV}$ is

$$f_\alpha(t) = f_{FES}(S_\alpha(R(t), \phi(t), h(t))) = -\frac{\partial}{\partial S_\alpha} \int_0^t dt' W e\left(-\frac{(S_\alpha(t) - S_\alpha(t'))^2}{2\Delta^2}\right)$$
Extended Lagrangian Formalism

Define new independent variables that describe a trajectory in $\Sigma_{CV}$

$$s_\alpha \Rightarrow S_\alpha(R, \Phi, h)$$

The dynamics of the $s_\alpha$ is coupled to the dynamics of the system by

$$V_{harm}(S, s) = \sum_{\alpha} \frac{1}{2} k_\alpha (S_\alpha(R, \Phi, h) - s_\alpha)^2$$

The Penalty potential is a function of the $\{s_\alpha\}$ only

$$\mathcal{L}_{MTD} = \mathcal{L}_{CP} + \sum_{\alpha} \frac{1}{2} M_\alpha s_\alpha^2 - V_{harm}(S, s) - V_{MTD}(s, t)$$
Monitoring the Local Minimum

The fictitious kinetic energy controls the dynamics of the new variables. The limit $M_\alpha \Rightarrow \infty$ is constraint dynamics.
Slow variables are recommended for a better sampling of the FES.
To monitor the two coupled trajectories at equilibrium set $V_{MTD} = 0$. 

\[
\frac{\partial F}{\partial s_\alpha} \simeq \langle k_\alpha (S_\alpha(R, \Phi, h) - s_\alpha) \rangle
\]

\[
\Delta s
\]

Estimate well width
Forces on All the Variables

Smooth and continuous trajectory by integrating the equations of motions

Force on $s_\alpha$

$$f_{s_\alpha}(t) = -\sum_\alpha k_\alpha [S_\alpha(R(t)) - s_\alpha(t)] + \frac{\partial}{\partial s_\alpha} V_{MTD}(s, t)$$

on ion I

$$f_I(t) = f_I^{CP}(t) + \sum_\alpha k_\alpha [S_\alpha(R(t)) - s_\alpha(t)] \frac{\partial S_\alpha}{\partial R_I}$$

$V_{harm}$ and $V_{MTD}$ generate opposite contributions. Progressive growth of $V_{MTD}$ favors a thorough sampling of the well, which follows the minimum energy path as close as possible.
Shape of the Potential

Bead by bead while the trajectory proceeds

\[ V(s, t) = \int_0^t W(t') \ e\left(-\frac{|s-s(t')|}{2(\Delta s_{\perp})^2}\right) \delta\left(\frac{\dot{s}(t')}{|\dot{s}(t')|} \cdot (s - s(t'))\right) |\dot{s}(t')| \, dt' \]

Each bead is localized on already visited configurations \( s(t') \)
The volume occupied by \( V_{MTD} \) should correspond to the space actually explored, i.e. the size of each bead is determined by the displacement \( |\dot{s}(t')| \, dt' \)
Discretized Form used in CPMD

\[ V(s, t) = \sum_{t_i < t} W_i \exp \left[ -\frac{(s - s^i)^2}{2(\Delta^\perp)^2} \right] \exp \left[ -\frac{((s^{(i+1)} - s^i) \cdot (s - s^i))^2}{2(\Delta^\parallel_i)^4} \right] \]

The \( \delta(x) \) has been expressed in the approximate Gaussian form

\[ \frac{1}{\beta\sqrt{2\pi}} e^{-\frac{x^2}{\beta^2}} \]

The discretization of MTD defines the deposition rate of beads while building \( V_{MTD} \)

\[ \delta t_{MD} \ll \Delta t_{MTD} \ll \omega_s^{-1} \]
$N_\alpha$-dimensional Gaussian Tube

Accumulate slices along the trajectory

$\Delta s_{\perp}$ (input)

$\Delta s_{\parallel}$
Some Warnings

- Parameters \((M, k, \Delta^\perp, \Delta^\parallel, W, ..)\) are system dependent
- The selected CV must discriminate among the relevant states along the pathway (reactants, products, TS)
- The number of hills required to fill the well is proportional to \(1/((\Delta s)^{N_{CV}})\)
- The sampling of floppy modes is expensive: large variations of the CV over almost flat energy regions.
- The MTD does not generate a true dynamics. Reaction rates have to be derived a posteriori from the computed FES
- The analysis of the trajectory is needed to isolate the TS
- The MTD trajectory describes the most probable pathway in the actual thermodynamic conditions, taking into account also possible kinetic effects (lager and shallower channels are preferred)
- The accuracy in the evaluation of the FES depends on hills’ shape and size, and on the deposition rate. The ideal coverage by \(V_{MTD}\) is achieved when the system finally moves on a flat surface: barrier-less recrossing
Si$_x$H$_y$ Clusters

Given the Stoichiometry which is the optimal structure? Common guesses are based on chemical intuition, by taking into account symmetry issues and even distribution of the protons among the Si atoms. MC sampling or simulated annealing are typically used to optimize the structures.

**Si$_6$H$_8$**

Known minimum tetrahedric
$Si_6H_8$ Clusters
Unknown Structures Found by Metadynamics

Predictive Power
Free Energy Reconstruction

The fast degrees of freedom are integrated out by the MD sampling if the sampling of $\Sigma_{CV}$ is properly adiabatically separated. The probability to visit a given configuration in $\Sigma_{CV}$ is measured by $V_{MTD}$:

$$P(\tilde{s}_1, \tilde{s}_2, \ldots) = \frac{1}{Q} \int_{\Sigma} \exp \left[ -\frac{V_{MTD}(s_1, s_2, \ldots)}{K_BT} \right] \delta(s_1 - \tilde{s}_1)ds_1 \delta(s_2 - \tilde{s}_2)ds_2\ldots$$

where $Q$ is the partition function $\int_{\Sigma} \exp \left[ -\frac{V_{MTD}}{K_BT} \right] d\Sigma$. The FES can be reconstructed if the wells have been completely explored, up to a flattening of the total surface $FES + V_{MTD}$.

$$\lim_{t \to \infty} [-KT \ln(P(S))] = F(S)$$
Free Energy Surface for $Si_8H_{y-2x} + xH_2$
Influence of the Parameters

- **Adiabatic separation** Slow and heavy $s_α$ moving over the rugged free energy landscape, while the relatively fast molecular motions sample the perpendicular directions. The frequency of $s_α$ is given by

$$\omega_s = \sqrt{\frac{k_α}{M_α}}$$

- Coupling with electrons is to be avoided, but a small amount of non-adiabaticity is accepted for the sake of efficiency.

- Common control of temperature to minimize the tendency of the $s_{alpha}$ to speed up due to the down hill effect.

- Rate of deposition $\Delta t$, width of the hills $\Delta_\perp$ and height $W$ cannot be chosen independently. They control efficiency and accuracy in convergence towards the FES.
  - Large $\Delta_\perp$ and small $\Delta t$ : hill-surfing
  - Ratio between $W$ and $\Delta_\perp$ determines the steepness: force
  - The error in the evaluation of the FES is proportional to $W$
By standard MD transitions are observable only under extreme conditions: heating and overpressurization. In this conditions the observed mechanisms are not reliable.
Solid State Transition

The cell parameters are used as CV and the Parrinello-Rahman scheme to perform MD simulations at constant pressure.
Si : diamond, \( \beta \)-tin, simple-hex.
Transition Path Sampling

Throwing ropes over rough mountains passes, in the dark

- Known $A$ and $B$ states, but nothing needs be presupposed about the pathway (no reaction coordinate)
- Initial path connecting $A$ and $B$, that can be a very crude approximation, $\Xi(t)$
- Choose a time slice $\tau$ and alterate the momentum of the trajectory by $\delta$: a new trajectory is generated propagating forwards and backwards, $\chi'(t)$.
- The step from $\chi$ to $\chi'$ is accepted with the probability

$$\omega(\chi, \tau, \delta) h_A(\chi(0)) \rho(\chi(T)) h_B(\chi(T))/h_A(\chi'(0)) \rho(\chi'(0)) h_B(\chi(0)) \omega(\chi', \tau, -\delta)$$

where $\rho(\chi(0))$ is the unconstrained distribution of the initial phase space point $\chi(0)$.
- Iterate the procedure from the new trajectory
• **Importance sampling** of trajectory space, based on MC techniques. The statistical weight of a trajectory $x(t)$ with $t \in (0, T)$ and starting from $x_0 \in A$ and ending in $x(T) \in B$ is proportional to

$$\exp \left\{ -\frac{1}{4} \int_0^T |\dot{x} + \nabla V|^2 \right\}$$

• This way new reaction paths can be found, with a lower reaction barrier: Starting from a point “p” we might find another path with a lower barrier

Reaction rates can be obtained, but with a large amount of statistics (different MD runs)
Transition path sampling

Example: 7-atom Lennard-Jones clusters

David Chandler & co-workers

- Move the central (red) atom to the boundary; the central atom in the final state random (blue)

- Difficult, correlated “reaction” mechanism