SCM
Software for Chemistry & Materials

# Accelerating ReaxFF Simulation of Alkane Pyrolysis using CVHD

Tomáš Trnka
support@scm.com

October 25, 2018

# Motivation

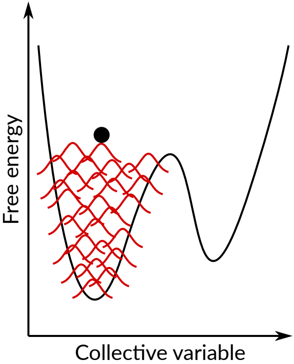
- Many interesting problems driven by rare events (ms-h+ timescales)
    - Chemical reactions (catalysis, industry, enzymes)
    - Slow physical processes (diffusion, phase transitions)
- Brute-force MD not feasible:
    - Theoretical "wonder HW" capable of a timestep (1 fs) per clock cycle (1 ns): $10^6$ wall-s (12 days) per sim-s
- How to find the rare reactive trajectory?
    - Lower the effective barrier – biased simulation
    - Cherry-pick interesting trajectories

# Introducing Enhanced Sampling Methods

- Biased MD philosophy
  - Simulate using a modified potential or equations of motion
  - Remove the effect of the bias from results
- Free-energy methods (Potential of Mean Force) – need to record the bias for unbiasing
  - Need to define suitable collective variables (CVs) to reduce dimensionality
  - Up to 2 (3) CVs feasible (Umbrella, Metadynamics, ABF ...), or sampling along a predefined reaction path (String Method, Nudged Elastic Band ...)
  - “Good” CVs crucial for usable results
    - Needs to unambiguously define the state of the system (different states must not overlap)
    - All orthogonal processes need to average out
    - Needs to be smooth and continuous
    - Not usable for studying many independent simultaneous processes

# Example Free Energy Sampling Method: Metadynamics

- $V_{\text{bias}}(\xi) = \sum_{i}^{N_{\text{biaspts}}} g(\xi - \xi_i)$
- New bias function $g(\xi)$ (usually Gaussian) periodically deposited at the current CV value $\xi_i$
- Incrementally fills energy valleys by bias until there are no barriers left (diffusive regime)
    - Free energy = negative of converged $V_{\text{bias}}$



# Alternatives to Free Energy Sampling Methods

- “(Almost-)Memory-less” methods: Derivatives of Hyperdynamics
  - No free energy information
  - “Unbias” on the fly to recover time
- Biased sampling of trajectories: Transition Path Sampling, Transition Interface Sampling

# Hyperdynamics

- Simulate on a modified potential $V'(x) = V(x) + V_{\text{bias}}(x)$
- Advance "hypertime clock" by $\Delta t_{\text{hyper}} = he^{V_{\text{bias}}(x)/(k_B T)}$
  - Hypertime = "How long would it take to get here in an unbiased simulation?"
- Transition state regions must be bias-free
- Suitable choice of $V_{\text{bias}}(x)$ crucial, hard to set manually
- Self-learning bias: CV-based Hyperdynamics (CVHD)
  - $V_{\text{bias}}$ built on the fly using Metadynamics with a suitable CV $\eta$
  - Bias reset once a transition is detected – CV only needs to be valid for a single transition

# CVHD in ReaxFF

- Global CV $\eta$ composed of primitive local CVs $\chi_i$
  - $\eta$ is dominated by the bond with the highest $\chi_i$

$$\eta = \frac{1}{2}\left(1 - \cos\left(\pi \chi_{\text{tot}}^2\right)\right) \qquad \chi_{\text{tot}} = \left(\sum_i^N \chi_i^p\right)^{1/p}$$

- Only a bond-breaking local CV implemented (Bal and Neyts, *J. Chem. Theory Comput.* 2015, doi:10.1021/acs.jctc.5b00597)

$$\chi_i = \frac{r_i - r_i^{\text{min}}}{r_i^{\text{max}} - r_i^{\text{min}}} \qquad 0 \leq \chi_i \leq 1$$

- Bias is applied to bonds approaching a transition – distance between $r_i^{\text{min}}$ and $r_i^{\text{max}}$
- Only bonds with $BO \geq cutoff$ at the beginning of the run are considered
- Transition is assumed when $\eta = 1$ for a specified time

# Designing Advanced CVs

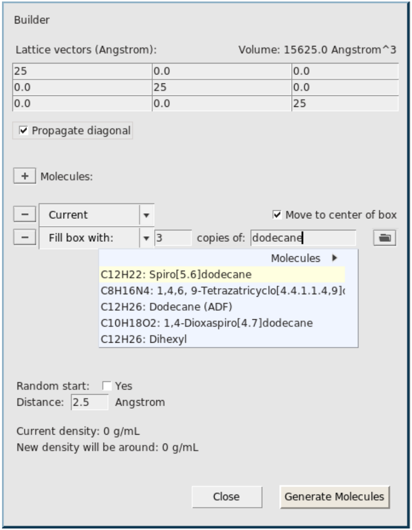
- Participating atoms defined in ADFinput as “Element-Region” pairs
  - Simplest case: Use the predefined “All” region (“C-All” means any carbon atom in the system)
  - Define multiple regions to distinguish chemically different atoms of the same element
- Exponent $p$ sets the degree of “locality” of $\eta$
  - Higher $p$ – focus on dissociating one bond (closest to transition – highest $\chi_i$)
  - Lower $p$ – bias several bonds at the same time (concerted dissociation)
  - Results commonly not very sensitive to the exact value used
- Only bond-breaking CVs currently supported
  - Other CV types (bond formation, angles/torsions etc.) planned
- Definition can be edited in **`cvhd.in`**
- See the online documentation for details

# Tutorial: Pyrolysis of Dodecane

- Following Bal and Neyts, *Chem. Sci.* 2016, doi:10.1039/c6sc00498a
- Pyrolysis of alkanes under realistic conditions (1000 K, about 50 kg m$^{-3}$)
- Some deviations from the paper to save time:
    - Integration timestep increased to 0.2 fs (lower accuracy)
    - Small system (only 114 atoms)
- We won't have enough data for proper statistics (rate constants)
    - We can still roughly estimate the timescales of different processes

# System Preparation

- Open ADFinput and switch to the ReaxFF module
- Open the Builder tool (Edit → Builder)
- Set up a 25 Å cubic box
- Add 3 molecules of dodecane (type into the field and click “Dodecane (ADF)” in the drop-down)
- Click “Generate molecules” and “Close”



Builder

Lattice vectors (Angstrom): Volume: 15625.0 Angstrom^3

| | | |
|---|---|---|
| 25 | 0.0 | 0.0 |
| 0.0 | 25 | 0.0 |
| 0.0 | 0.0 | 25 |

☑ Propagate diagonal

Molecules:

Current ☑ Move to center of box

Fill box with: 3 copies of: dodecane

C12H22: Spiro[5.6]dodecane
C8H16N4: 1,4,6, 9-Tetrazatricyclo[4.4.1.1.4,9]
C12H26: Dodecane (ADF)
C10H18O2: 1,4-Dioxaspiro[4.7]dodecane
C12H26: Dihexyl

Random start: ☐ Yes

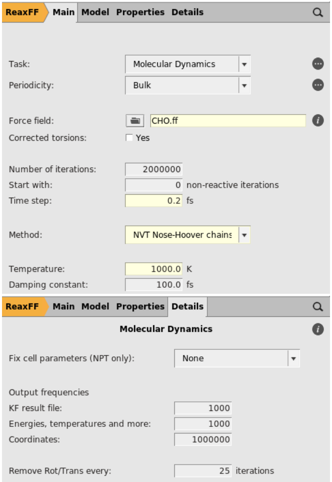
Distance: 2.5 Angstrom

Current density: 0 g/mL

New density will be around: 0 g/mL

# Simulation Setup

- Select the CHO.ff force field
- Set the timestep to 0.2 fs
- Turn on the Nosé-Hoover thermostat and set the temperature to 1000 K
- Set “number of iterations” (MD steps) as high as time permits
  - The absolute minimum needed to see a reaction is one million steps (20–40 minutes depending on your CPU)
  - More steps strongly recommended (at least 1.5–2 million)
- Increase the output periods on the Details → Molecular Dynamics panel to 1000



Task: Molecular Dynamics
Periodicity: Bulk
Force field: CHO.ff
Corrected torsions: ☐ Yes
Number of iterations: 2000000
Start with: 0 non-reactive iterations
Time step: 0.2 fs
Method: NVT Nose-Hoover chains
Temperature: 1000.0 K
Damping constant: 100.0 fs

**Molecular Dynamics**

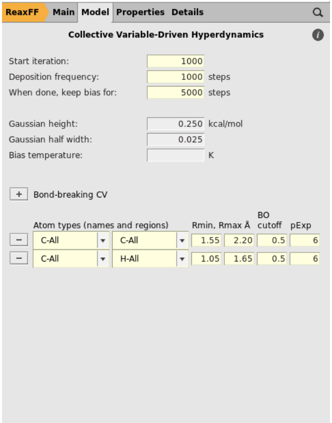
Fix cell parameters (NPT only): None

Output frequencies
KF result file: 1000
Energies, temperatures and more: 1000
Coordinates: 1000000

Remove Rot/Trans every: 25 iterations

# CVHD Setup

- Set the bias deposition frequency to 1000 steps (200 fs) after an initial delay of 1000 steps
- Set the waiting time for events to 5000 steps (1 ps)
- Using the “+” button next to “Bond-breaking CV”, define these two local collective variables (parameters taken from the Bal and Neyts paper):
  - C–C with Rmin 1.55 Å and Rmax 2.20 Å
  - C–H with Rmin 1.05 Å and Rmax 1.65 Å
- Now you can save and run the simulation!



**Collective Variable-Driven Hyperdynamics**

Start iteration: 1000
Deposition frequency: 1000 steps
When done, keep bias for: 5000 steps

Gaussian height: 0.250 kcal/mol
Gaussian half width: 0.025
Bias temperature: K

+ Bond-breaking CV

| Atom types (names and regions) | | Rmin, Rmax Å | | BO cutoff | pExp |
|---|---|---|---|---|---|
| C-All | C-All | 1.55 | 2.20 | 0.5 | 6 |
| C-All | H-All | 1.05 | 1.65 | 0.5 | 6 |

# Understanding the Logfile

```
Iteration Nmol   HyperTime(s)
        0      3  0.00000000E+00
     1000      3  2.00000000E-13
     2000      3  4.25989508E-13
     3000      3  6.76215653E-13
     4000      3  9.69874961E-13
     5000      3  1.30267608E-12
…
   100000      3  2.18783039E-09
…
   200000      3  2.05119905E-07
```

- Select “Logfile” from the SCM menu
- Hypertime is the “true” timescale after unbiasing (how long a process would take without CVHD)
- First 1000 steps unbiased (hypertime equals simulation time – 200 fs)
- Afterwards, bias starts building up, gradually accelerating the simulation
  - 30 % speedup already after 5000 steps (1 ps)
  - 2.2 ns effectively sampled in 20 ps
  - 205 ns sampled in 40 ps

# CVHD Events in the Logfile

```
Iteration Nmol   HyperTime(s)
  588000     3  9.23210824E-04
  589000     3  9.41709394E-04
  590000     3  9.41904071E-04
  591000     3  9.42270720E-04
  592000     4  9.42279978E-04
  593000     4  9.42279979E-04
  594000     4  9.42279979E-04
  595000     4  9.42279979E-04
  596000     4  9.42279979E-04
CVHD: the system has completed the
state change. Resetting CVHD.
  597000     4  9… -- CVHD event
  598000     4  9.42279980E-04
  599000     4  9.42279980E-04
  600000     4  9.42279980E-04
  601000     4  9.42279980E-04
```

- System approaches a transition between steps 589000 and 591000 (hypertime evolution slows down somewhat – not much bias in this region)
- A molecule dissociates around step 592000 and the CV reaches 1.0 (hypertime evolution slows down a lot – transition state unbiased)
- The algorithm waits for 5000 steps to see if the system recrosses back into the original state
- No recrossing occurs, so CVHD declares that an event (reaction) just took place and removes all bias
- New bias starts slowly building up from step 597000

# Monitoring Bias Deposition

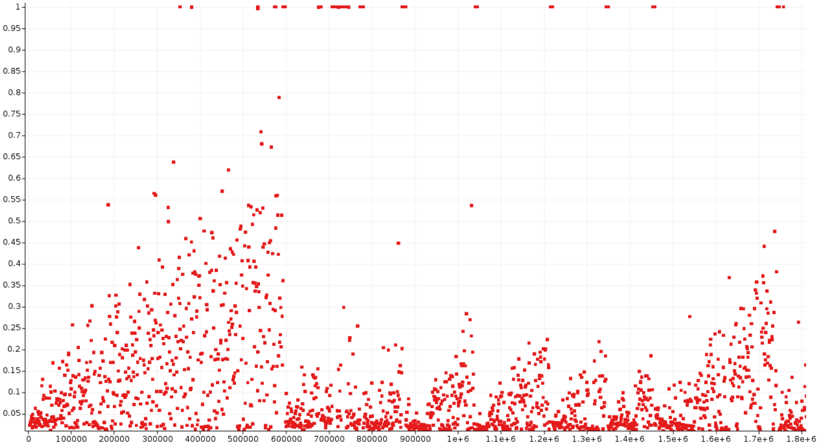
## The cvhd-hills Tool

- Run **`cvhd-hills`** to plot where biasing hills where deposited
    - Either **`cd`** into the **`.results`** directory (or **`tmp.*`** for running jobs), or pass the path to a **`fort.84`** file as an argument
    - Output file name can be set with a second optional argument
    - By default, output is written to **`cvhd-hills.csv`** and displayed using ADFgraphs
- ADFgraphs defaults to plotting using lines (not very useful in this case)
    - Open Plot → Options → Curves, uncheck “Curve” and check “Data Points”

# Monitoring Bias Deposition

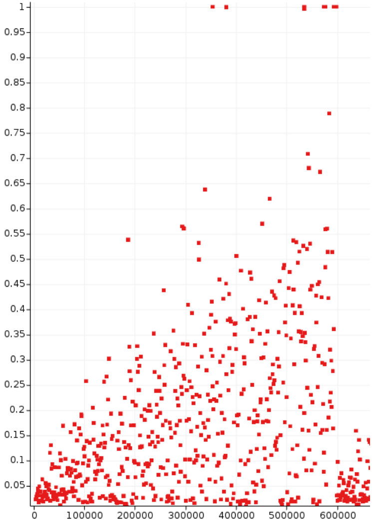
## Example Result

- Horizontal axis – MD step, vertical axis – CVHD global collective variable $\eta$
  - Some versions of ADFgraphs incorrectly label axes as “Energy” – ignore it
- Each point represent a single Gaussian bias hill



# Monitoring Bias Deposition

A Closer Look at One CVHD Event



- Initially, system stays around $\eta = 0$ (all bonds relaxed) and deposits bias there
- Bias gradually pushes $\eta$ to higher and higher values (more and more stretched bonds)
- Sometimes (between steps 300000–400000), $\eta$ reaches 1.0 but returns back to lower values (no transition occurs during specified wait time)
- Finally, around step 600000 $\eta$ stays at 1.0 long enough to indicate that a bond has dissociated

# Improving the CV using the Bias Deposition Plot

- Initial $\eta$ not close to zero – Rmin is set too low (CVHD “thinks” the equilibrium structure is already partly dissociated)
- $\eta$ doesn’t explore higher values (stays at zero) – Rmin set too high
- Too many recrossings ($\eta$ keeps hitting 1.0 but no true reaction occurs) – Rmax set too low (not close enough to a transition)
- System jumps from low $\eta$ directly to 1.0, triggering a reaction – Rmax set too high (past the transition state)
- It’s often useful to tune each local CV (bond type) separately before combining them in a single simulation
- Global CV contains a weighted sum of local CVs – optimal Rmin is system-size-dependent (large systems need somewhat higher Rmin)

# Analyzing Event Timescales

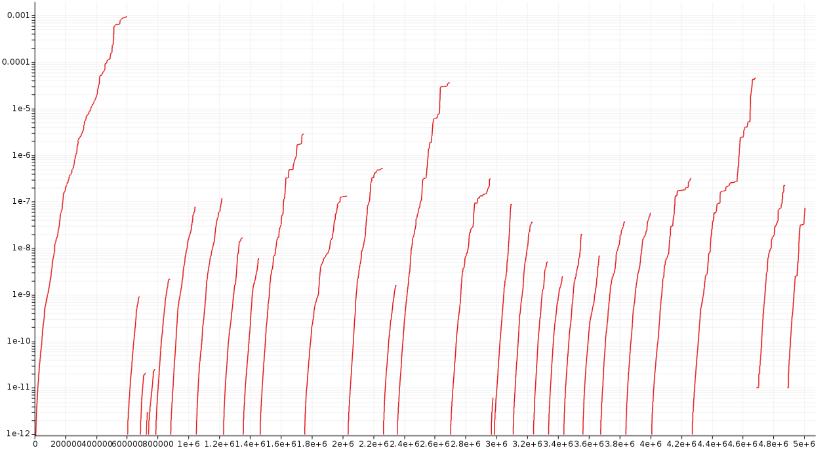
## The cvhd-hypertime Tool

- Run `cvhd-hypertime jobname.logfile` to plot the hypertime for each CVHD event
  - Output file name can be set with an optional second argument
  - By default, output is written to `cvhd-hypertime.csv` and displayed using ADFgraphs
- Switch the vertical axis to logarithmic scale for best results
  - Open Plot → Options → Left Y axis
  - Make sure that “Minimum value” is set to a positive number (for example 1e-12)
  - Check “Logscale”

# Analyzing Event Timescales

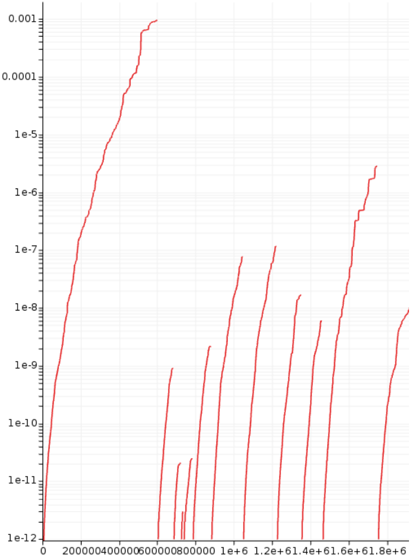
## Example Result

- X axis – MD step, Y axis – hypertime in seconds since previous event



# Analyzing Event Timescales

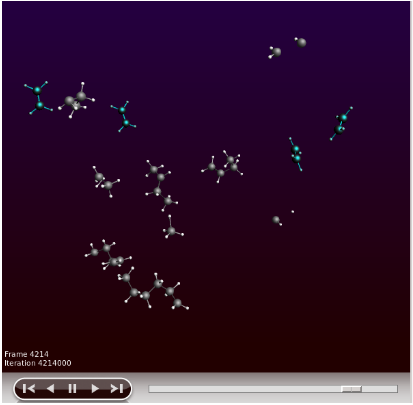
A Closer Look

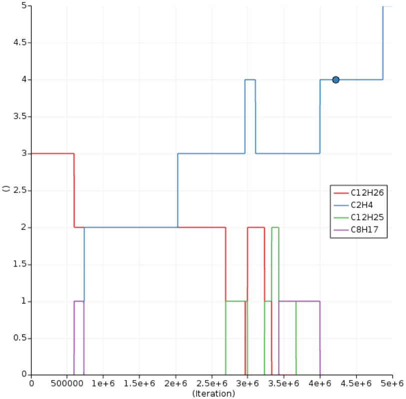


- Each curve shows the gradual acceleration of time as the bias evolves until an event (bond dissociation) is detected
- Classes of processes corresponding to different timescales are visible:
    - Initiation (dodecane chain splitting) at a (sub-)millisecond timescale
    - Propagation steps at ns–µs timescales
- Curves should be (mostly) smooth
    - Jagged staircase-like curves indicate issues with CVHD setup (Rmin/Rmax, deposition rate etc.)

# Analyzing System Composition

- Open ADFmovie and use Properties → Molecule Fractions
  - Check “Graph” to plot the number of particular molecules/fragments
  - Click on a curve to highlight corresponding molecules
- Example: dodecane consumption (red), ethylene production (blue), initiation by C-C splitting (purple) or H abstraction (green)





# Discussion

- Not enough data for reasonable statistics or rate constants
  - Most elementary reactions observed only once, some probably completely missed
  - Only rough order-of-magnitude estimate of the timescale from a single observed event time
  - Large differences in results between multiple trajectories
  - Bigger system and longer simulation would be necessary for usable results
- Integration with the 0.2 fs timestep possibly too inaccurate during transitions
  - Results should be verified using a shorter timestep (0.1 fs)
- Bias deposition plot shows cases of $\eta$ going from low values straight to dissociation (1.0)
  - Rmin/Rmax for C–H bonds may need to be tuned

# CVHD Summary

- Suitable for accelerating bond dissociation
  - Any process that starts with bond dissociation can be accelerated
- Relatively low setup effort
  - Only Rmin/Rmax distances needed for each bond – can be estimated and tuned in a few short testing runs
  - Hill height and deposition rate may need to be adjusted depending on expected barrier height
- Works best for moderately-sized systems
  - CVs comprising many thousands of bonds trigger events too often, limiting bias buildup and accelerations
  - Only the number of biased bonds matters (not the overall system size) – can be limited using suitable regions

# Current Software Limitations and TODOs

- ChemTraYzer (“Reaction Event Detection” in the GUI) not yet fully compatible with CVHD
  - Hypertime not considered – calculated reaction rates incorrect
  - Reaction network analysis and visualization works
- CVHD not yet ported to the new AMS MD driver
  - New implementation will write all data to the RKF file – bias/hypertime visualization through the GUI instead of separate helper scripts