4.2.1 Simulating thermodynamic processes

- This module will be a mini-lecture describing
  - importance sampling
  - umbrella sampling
  - thermodynamic processes
  - replica exchange
- At the end of this module, you should be able to answer the following questions:
  - What is a thermodynamic process?
  - What are the benefits of umbrella sampling?
  - What are the benefits of performing replica exchange?
  - How do you know if there are enough states along a thermodynamic process?

Importance sampling

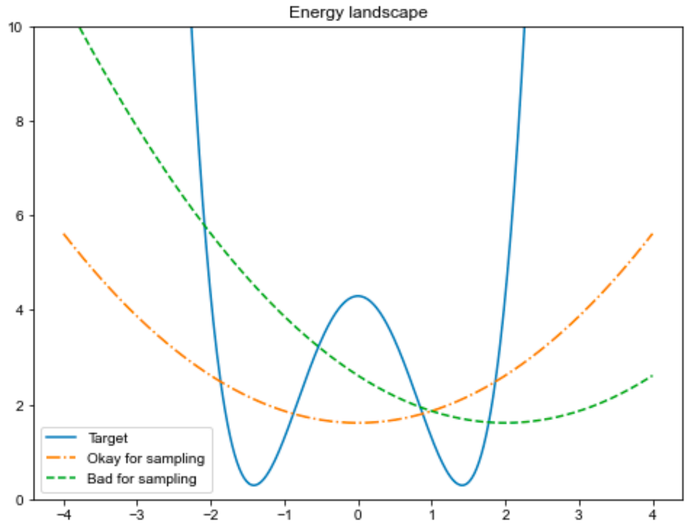
- Sampling from one distribution and estimating quantities in another
- $$\langle A \rangle_T = \int A(x) p_T(x) dx = \int A(x) \left(\frac{p_T(x)}{p_S(x)} \right) p_S(x) dx = \langle wA \rangle_S$$
  - $p_T(x)$ is the probability density in the target distribution
  - $p_S(x)$ is the probability density in the sampled distribution
  - $w = \left(\frac{p_T}{p_S} \right)$ is the ratio of weights in the two distributions

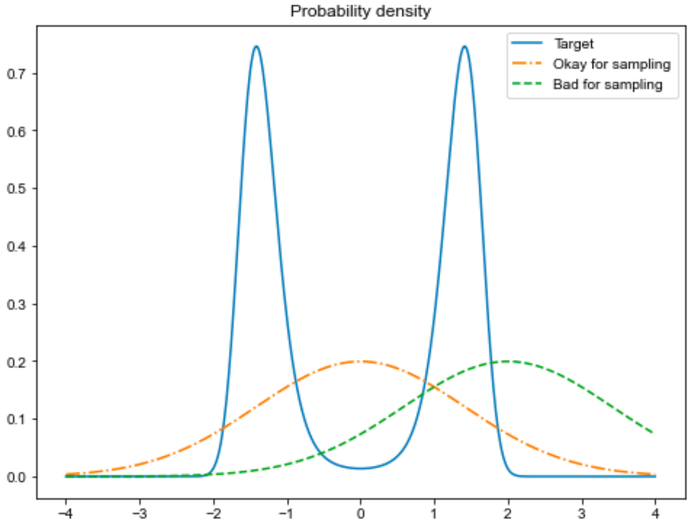
Why do importance sampling?

- Use one simulation to estimate quantities in multiple thermodynamic states, e.g. different temperatures
- Less computational expense to sample from one distribution, e.g. sample with molecular mechanics and calculate quantities in QM/MM distribution
- Sample from distribution with smaller configuration space, e.g. harmonic restraint towards a crystal structure

Caveat: Importance sampling

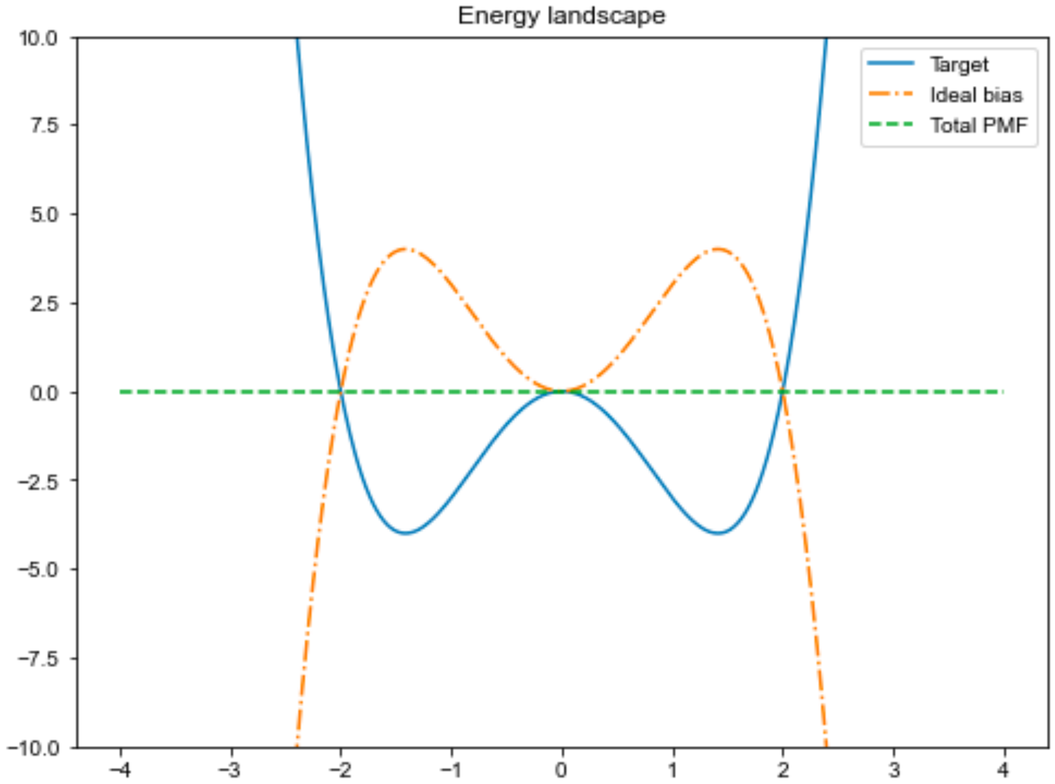
- The target and sampled distribution should be similar
  - If the *support* differs, important configurations may be missed
  - If the probability density significantly differs
    - the reweighing term can be noisy
    - thermodynamic expectations will require more samples to converge





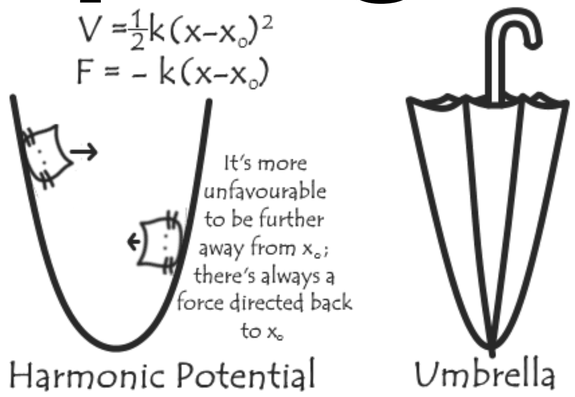
Idealized importance sampling

Covers the entire relevant configuration space along an order parameter

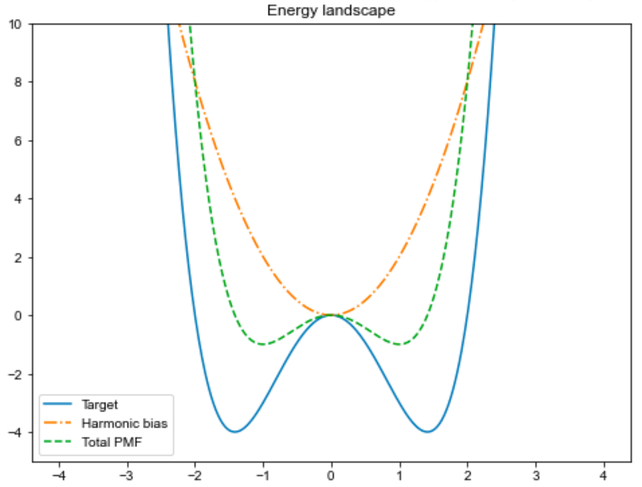


Harmonic umbrella sampling

- Sampling that *covers* the entire relevant configuration space along an order parameter
- Typically involves
  - a harmonic restraint towards a specific value of the order parameter, $U_b(x) = \frac{1}{2}k(z[x] - z_o)^2$
    - k is the spring constant
    - $z[x]$ is the value of the order parameter for configuration x
  - multiple simulations with different spring centers
    - the observed probability distributions should overlap

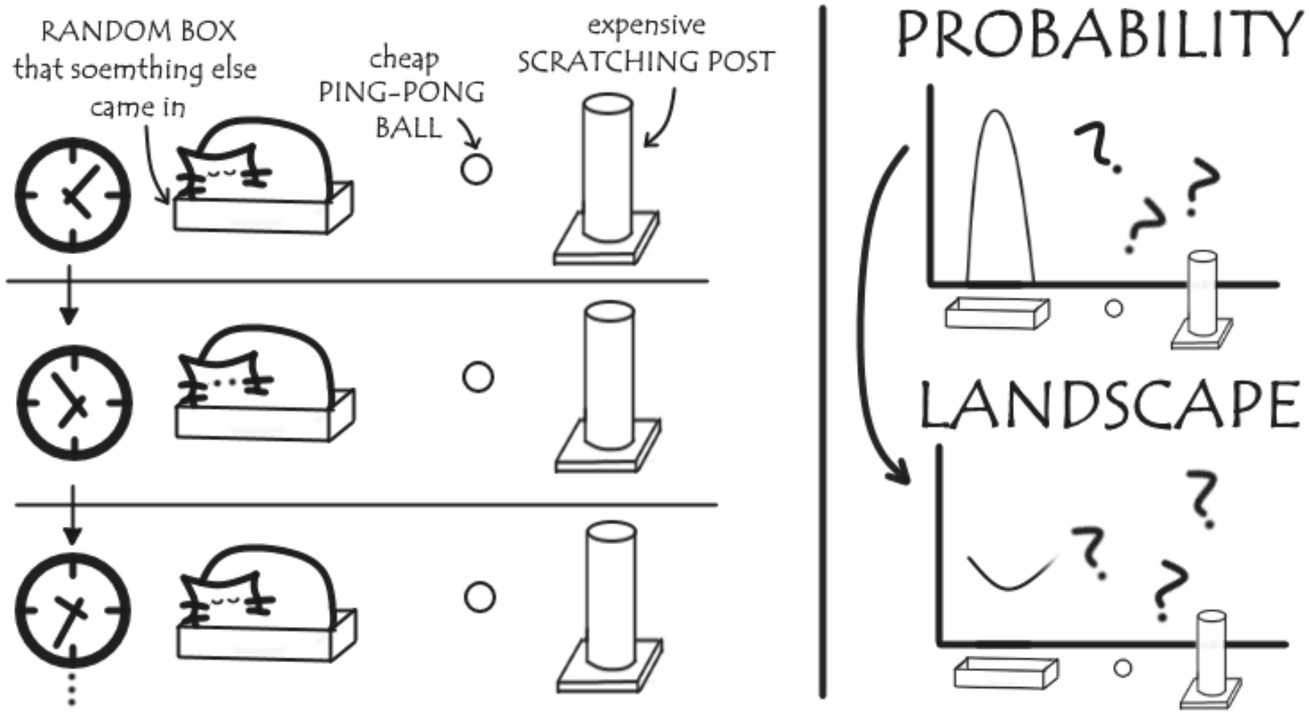


Used with permission from Fiona Naughton: http://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway



A cat analogy of umbrella sampling

Which toy does the cat like best?
To quantify this, we can look at the fraction of time they spend near a specific toy.

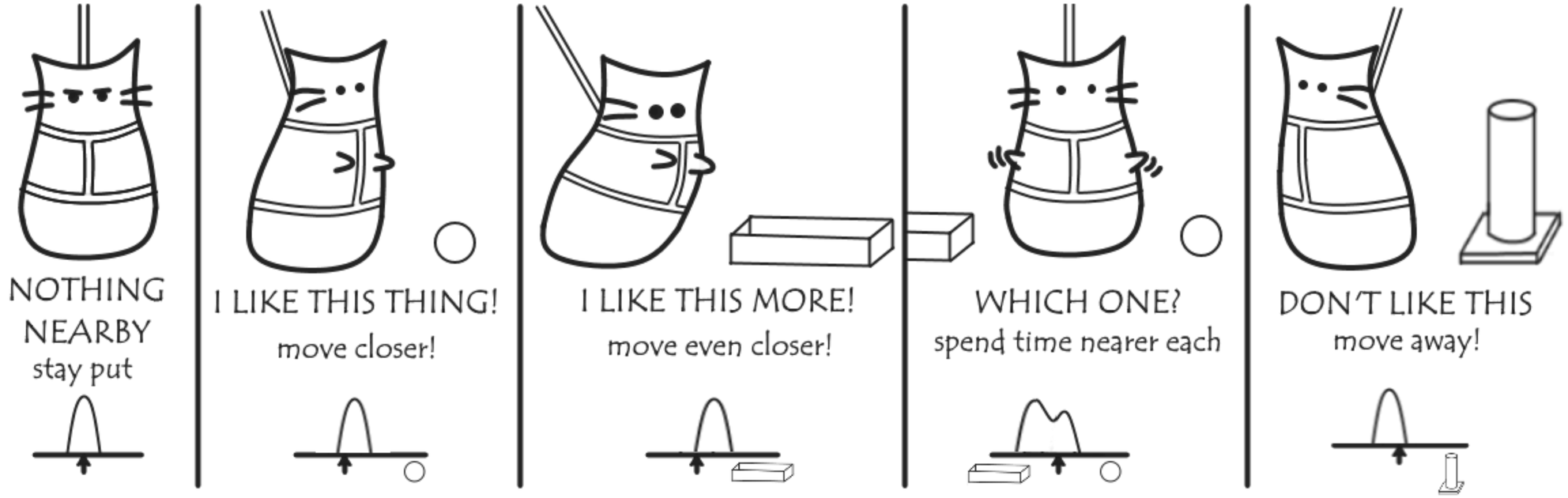


But we can be watching the cat for a long time…

Used with permission from Fiona Naughton: http://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway

Umbrella sampling with a cat

If we put a restraint on the cat, we can determine what they prefer in a smaller area. Overall, we don't need to watch as long.



Used with permission from Fiona Naughton: http://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway

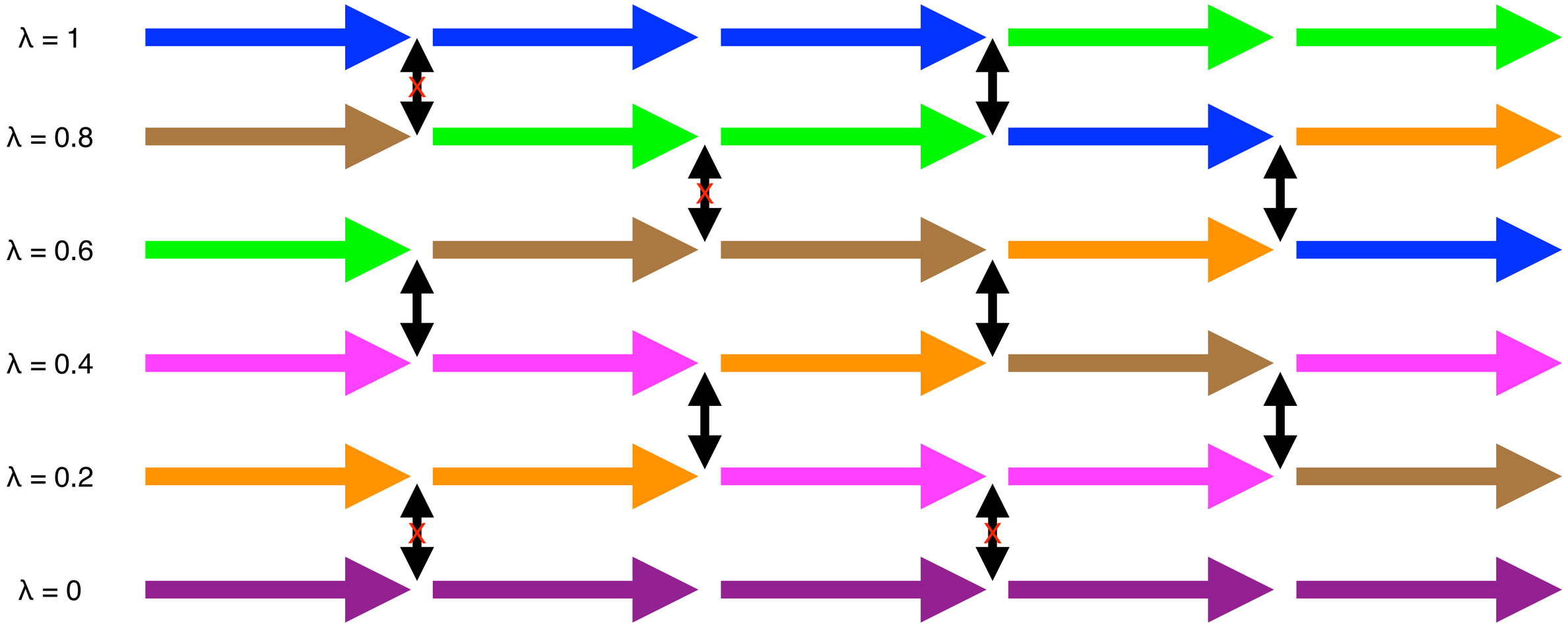
Thermodynamic processes

- A thermodynamic process involves a change in one or more variables that specify a thermodynamic state
- In general and physical chemistry, you learn about specifying a state with
  - temperature
  - pressure/volume
  - number of particles/chemical potential
- Processes include
  - isothermal expansion - a change in volume but not temperature
  - adiabatic expansion - a change in volume without heat transfer in and out of a system
- In molecular simulations, additional variables can define the potential energy
  - harmonic spring constant and center (umbrella sampling)
  - alchemical parameter for
    - transforming one substituent into another
    - decoupling a molecule from solvent

Replica exchange

- Simulations of multiple thermodynamic states with different parameters
  - originally, variation in temperature
  - often used for alchemical coupling
- Configurations from thermodynamic states are periodically swapped
  - Equivalently, thermodynamic parameters are swapped
  - Swapping satisfies detailed balance; both states sample from respective Boltzmann distributions

Swapping

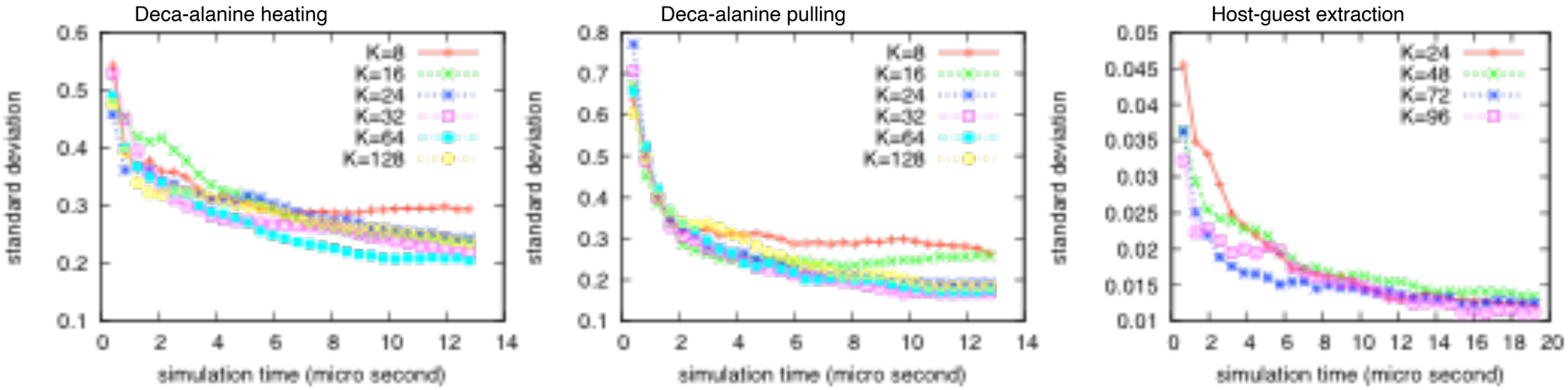


Why perform replica exchange?

- Improve mixing of MCMC chains; higher-entropy states help sample configurations in lower-entropy states
  - At high temperature, energetic barriers are crossed more quickly than at low temperature
  - Decoupled ligands move freely compared to bound ligands. See http://mypages.iit.edu/~dminh/images/pubs/dock_1hnn.gif.
- Minimal added computational expense

Selecting thermodynamic states

- Thermodynamic state selection has been thought to be a tricky optimization problem
- In Nguyen and Minh (2016)
  - processes were simulated 100x each for each number of states, K
  - the standard deviation of the free energy was evaluated as a function of the total simulation time



- If there are not enough states, the convergence curve levels off
- If there are enough states, the standard deviation of free energy estimates depends on the aggregate simulation time and is insensitive to the number of states.

Review

- What is a thermodynamic process?
- What are the benefits of umbrella sampling?
- What are the benefits of performing replica exchange?
- How do you know if there are enough states along a thermodynamic process?

References

- Nguyen, T. H.; Minh, D. D. L. Intermediate Thermodynamic States Contribute Equally to Free Energy Convergence: A Demonstration with Replica Exchange. Journal of Chemical Theory and Computation 2016, 12 (5), 2154–2161. https://doi.org/10.1021/acs.jctc.6b00060.