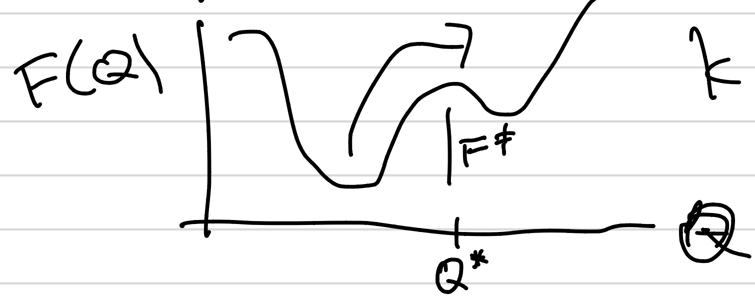
Enhaned Sampling:

Umbrella Sampling, Whem

Reminder

If we knew $F(Q)$



$k_{L \to R} \propto e^{-\beta F^{\ddagger}}$

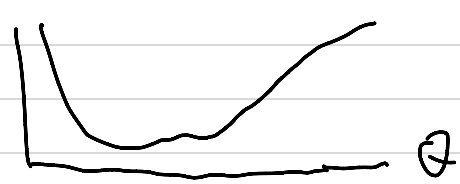
This means prob of seeing Q in sim $P(Q) \propto e^{-\beta F(Q)}$

Then could run simulation with

$\mathcal{H}_1 = \mathcal{H}_0(p, q) + V(Q)$ where

and so $P_1(Q) \propto e^{-\beta F(Q)} \, e^{-\beta V(Q)}$

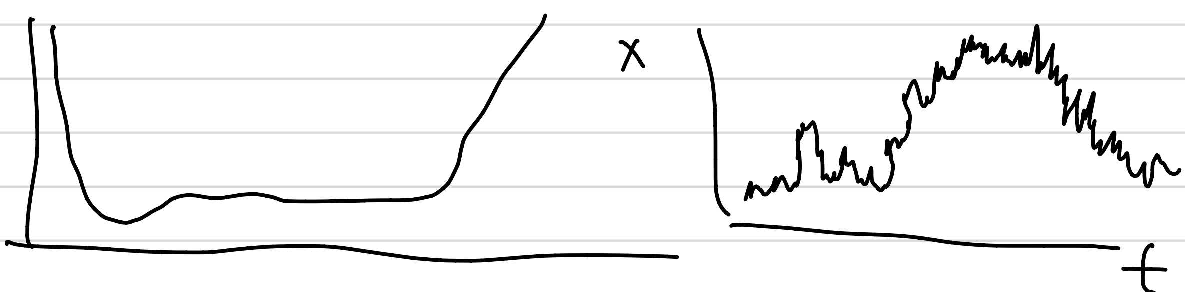
or $F_1(Q)$



or $V(Q) = -F(Q)$, "ideal"?

This may work well for a simple 1d barrier crossing, but it does not necessarily make the simulation averages converge very fast. I said for the average to be computed well, you should see every state multiple times

with perturb



now, diffusion

Diffusion $\langle \delta x^2 \rangle \propto Dt$

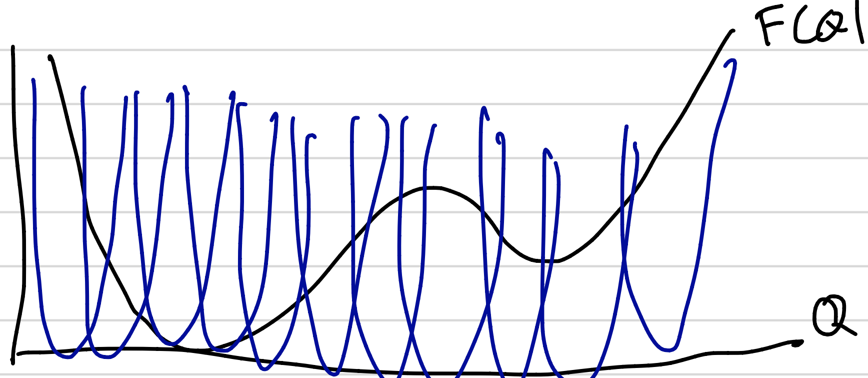
so to explore whole space, $t \sim L^2/D$

if D is small or space is big will take a long time

Later idea: Umbrella Sampling (umbrella, cover all the space)

Many biased simulations with different ω' to make sure all space is covered

So Eg:



Common to use $U_i'(Q) = \frac{1}{2} k (Q - Q_i)^2$

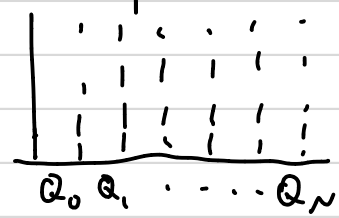
Goal is to use these sims to learn $F(Q)$ b/c then we know $P(Q)$ and can compute $\langle A \rangle = \int dQ \, A(Q) P(Q)$ and E.g. compare to experiment

How do we do this? Reminder how $F(Q)$ is defined,

$$F(Q) = -k_B T \log Z(Q)$$

$$Z(Q) = \int dp^{3N} dq^{3N} \frac{e^{-\beta H(p,q)}}{Z} \delta(Q(\vec{p},\vec{q}) - Q)$$

What would this mean in practice, break up Q range into Q_i, ΔQ bins →



Run simulation and count when $Q(p,q)$ falls in bin i, <u>histogram</u>

(replace $\delta(Q(p,q)-Q)$ w/ $\chi_i(Q) = \begin{cases} 1 & Q_i < Q < Q_{i+1} \\ 0 & \text{otherwise} \end{cases}$

to get $F(Q_i)$)

To see how this might work w/ multiple sims, simulate w/ $\mathcal{H}_1(p,q) = \mathcal{H}_0(p,q) + V_j(Q)$

where $V_j(Q)$ is instead

$$V_j(Q) = \begin{cases} 0 & |Q - Q_j| < a \\ \infty & \text{otherwise} \end{cases}$$

so $\mathcal{H}_1(p,q) = \mathcal{H}(p,q)$ in the box

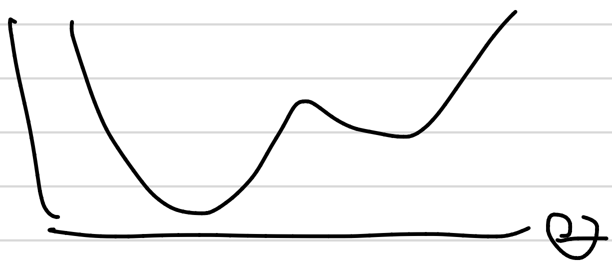
In box j

$$F_1(Q_i) = -k_BT\log\left[\frac{\int dp\,dq\; e^{-\beta(\mathcal{H}(p,q)+V_j(Q_i))}\,\chi_i(Q(p,q))}{\int dp\,dq\; e^{-\beta(\mathcal{H}(p,q)+V_j(Q_i))}}\right] \qquad \chi_i(Q(p,q)) = 1 \text{ in box}$$

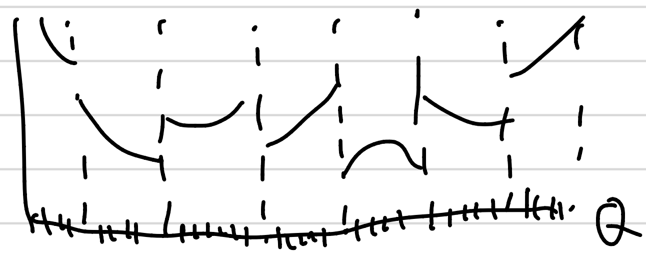
$$= \underbrace{-k_BT\log\left[\int dp\,dq\; e^{-\beta\mathcal{H}(p,q)}\,\chi_i(Q)\right]}_{F_0(Q_i)} - k_BT\log Z_j$$

Exact $F_0(Q_i)$ but unknown offset

So if true $F(Q)$



then separate sims might be



and have to match adjacent windows

Can do the same thing w/ $V_j(Q) = \frac{1}{2}k(Q - Q_j)^2$

but because it changes stats unlike last time, have to use our perturbation formula

$$P_j(Q_i) \propto \langle \chi_i \, e^{+\beta V_j(Q)} \rangle_1 \, / \, \langle e^{+\beta V_j(Q)} \rangle_1$$

$$F_j(Q_i) = -k_B T \log P_j(Q_i) - k_B T \log Z_j$$

how can we combine this info?

match $F_j(Q_i) \; \forall_j$ by fitting the Z_j to give best agreement, least squares fit

Or by matching single points adjacent windows

Most common way uses info from all windows and combines → weighted histogram analysis method (WHAM). [Other ideas exist and are better, eg MBAR, EMUS ...]

How does WHAM work? Pretend know Z_i's (guess) and combine histogram from all windows using reweighting. Then use this to get new Z_i's & repeat to convergence

Result turns out to be

$$P(Q_i) = \frac{\sum_{j=1}^{N} n_j(Q_i)\, P_j(Q_i)}{\sum_{j=1}^{N} n_j(Q_i)\, e^{-\beta(V_j(Q_i) - F_j)}}$$

$$F_j = -k_B T \log\left(\sum_{i=1}^{N} e^{-\beta V_j(Q_i)} P(Q_i)\right)$$

Tuckerman Chapter 8, can prove minimizes squared error

Can do a similar thing w/ eg parallel tempering

Eg: T.I. Want $F(\text{state}_1) - F(\text{state}2)$

suppose state1 has $\mathcal{H}_1 = KE + U_1(x)$

and $H_2 = KE + U_2(x)$

Then $H(\lambda) = KE + (1-\lambda)U_1(x) + \lambda U_2(x)$

can be simulated to try to get free energy between states

$$Q(\lambda) = C \int dX\, e^{-\beta \mathcal{H}(x,\lambda)}$$

$$A = -k_B T \log Q(\lambda) = -k_B T \log Z(\lambda) + KE.$$

$$\frac{\partial A}{\partial \lambda} = -\frac{k_B T}{Q}\frac{\partial Q}{\partial \lambda} = -\frac{k_B T}{Z}\frac{\partial Z}{\partial \lambda}$$

$$\partial Z/\partial \lambda = \int dq^{3N} -\beta \frac{\partial U}{\partial \lambda} e^{-\beta U(q,\lambda)}$$

so $$\frac{\partial A}{\partial \lambda} = \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$$

if linear

$$\Delta A = \int_0^1 d\lambda \frac{\partial A}{\partial \lambda} = \int_0^1 d\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda = \int_0^1 \langle U_2 - U_1 \rangle_\lambda d\lambda$$

In contrast, direct/end point swiching
(Free Energy Perturbation)

$$A_2 - A_1 = -k_B T \log Z_2 - k_B T \log Z_1$$

$$= -k_B T \log Z_2 / Z_1$$

$$= -k_B T \log \left(\left\langle e^{-\beta(U_2 - U_1)} \right\rangle_1 \right)$$

(see last time)