Umbrella Sampling + WHAM

Henry Bloom, Arnav Brahmasandra

March 7, 2025

1 Motivation

Molecular systems often exhibit high free energy barriers that impede adequate sampling of configurational space in standard simulations. Umbrella sampling addresses this by introducing biasing potentials to enhance sampling in targeted regions, stitching the biased distributions together using WHAM, which provides a systematic framework to reconstruct unbiased free energy profiles from biased simulations.

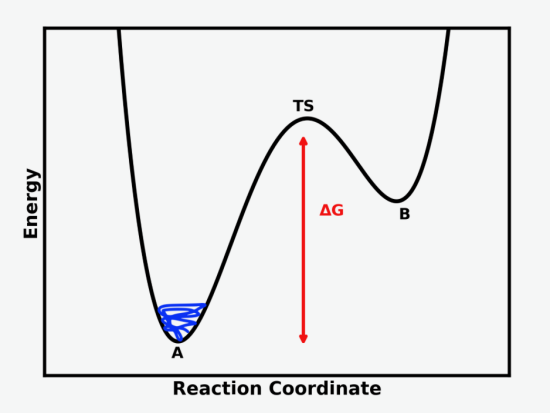


Figure 1: Free Energy Barrier [4]

2 Theory

2.1 Umbrella Sampling

Consider some system with a potential $U(r^N)$. Traditional sampling methods may suffer from poor ergodicity due to certain energy landscapes, particularly if the confifuration space has many high-energy barriers that may trap the trajectory in a local minimum. Introduced by Torrie and Valleau in 1977, umbrella

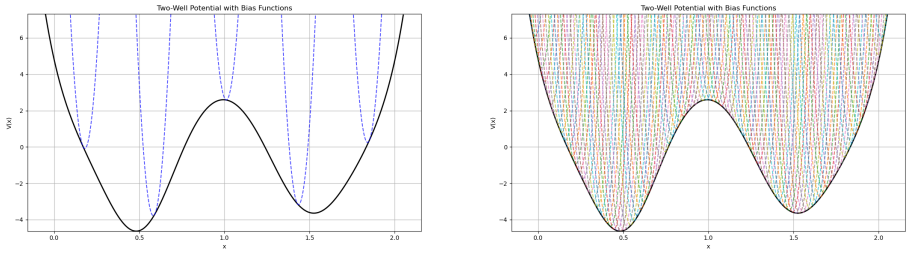
sampling uses bias potentials to encourage sampling of potentially inaccessible states. The biased distributions are then used to reconstruct an unbiased distribution and calculate quantities of interest, such as the free energy of the system.

2.1.1 Reaction Coordinate

We will introduce the notion of a reaction coordinate $q = f(r^N)$, an abstract typically one-dimensional parameter that unambiguously measures progress along a reaction pathway or transition between states. Some examples of a reaction coordinate, both geometric and non-geometric, include bond distance, bond angle, torsion, dihedral angle, bond order, Hydrogen bonds, and RMSD. The free energy of a system is often considered as a function of reaction coordinate to demonstrate in some schematic form the potential energy profile associated to the reaction.

2.1.2 Bias Potentials

Umbrella sampling introduces a window-specific soft restraint biasing potential $W_k(q)$, such that the total potential for window k becomes $U_k^{\text{biased}}(r^N) = U(r^N) + W_k(f(r^N))$. Often, the restraint is chosen to be a harmonic potential of the form $W_k(f(r^N)) = \frac{1}{2}\kappa(f(r^N) - s_k)^2$ centered around a chosen s_k. Some examples of bias functions added to a potential are shown below:



(a) Two-welled potential with selected bias functions

(b) Two-welled potential with 100 bias functions

Figure 2: Potentials with added bias

2.1.3 Simulations

In umbrella sampling, run a series of n simulations with different biasing potentials, collecting a biased distribution of values of q. Essentially, each simulation can be thought of as generating a histogram of values of q primarily located within the "window" corresponding to the bias potential chosen. The aim following the computation of these biased distributions is to first unbias them, and then to use the unbiased distributions to calculate the free energy of the

system. In figure 2, this corresponds to running one simulation with each of the bias potentials shown in (b).

2.2 WHAM

The weighted histogram analysis method (WHAM) is the method for unbiasing the biased distributions that is used in conjunction with umbrella sampling. We begin with biased probabilities $\tilde{P}_k(q)$ derived from the observed histograms from each biased simulation k. Our goal is to construct an unbiased global probability distribution $P(q)$ over the reaction coordinate by unbiasing the individual biased distributions and stitching them together to minimize the expected statistical error.

2.2.1 Unbiasing the Probabilities

We aim to unbias the probability distributions $\tilde{P}_k(q)$, which can be done as follows:

$$P_k(q) = e^{-\beta(A_k - A_0)} e^{\beta W_k(q)} \tilde{P}_k(q)$$

In this formulation, A_k represents the free energy due to the kth bias potential and A_0 is an arbitrary parameter that stems from the fact we do not know the absolute free energies of any of our biased systems, only their relative free energies. Recall that W_k is the kth bias potential. The first weighting term, $e^{-\beta(A_k - A_0)}$, corrects for the relative free energy shift introduced by each bias potential. It ensures that the biased distributions are properly scaled relative to each other, allowing them to be combined into a single, unbiased distribution. The second weighting term, $e^{\beta W_k(q)}$, directly counteracts the effect of the bias potential, effectively unweighting the biased distribution. This makes intuitive sense as, for example, any observations where the bias potential was very high should be considered to be more significant as it is indicative that the underlying potential must favor that state.

2.2.2 Reconstructing the Global Distribution

Our aim is to use the k unbiased probability functions from the simulations to construct the global distribution $P(q)$ while minimizing statistical error. We can set this up as follows:

$$P(q) = \sum_{k=1}^{n} C_k(q) P_k(q)$$

$$\text{subject to} \sum_{k=1}^{n} C_k(q) = 1$$

The aim is to optimize the statistical error of $P(q)$ while satisfying the given constriant.

The total statistical error in the global distribution will be:

$$\sigma^2 = \sum_{k=1}^{n} C_k^2(q)\sigma_k^2$$

Where σ_k^2 is the statistical error in the kth unbiased distribution. Through error propagation, this can be related to the statistical error in the kth biased distribution, which in turn depends on the number of samples taken, the width of the intervals used for analyzing q, and the values of $\tilde{H}_k$.

We can use Lagrange multipliers to perform this constrained minimization by instead minimizing the following:

$$\Sigma^2 = \sum_{k=1}^{n} C_k^2(q)\sigma_k^2 - \lambda\left(\sum_{k=1}^{n} C_k(q) - 1\right)$$

After solving for λ and $C_k(q)$ for all k by setting $\frac{\partial \Sigma^2}{\partial C_k(q)} = 0$, we can plug the $C_k(q)$ coefficients back into the equation for $P(q)$, which provides us with a set of equations to solve iteratively for P and A_k:

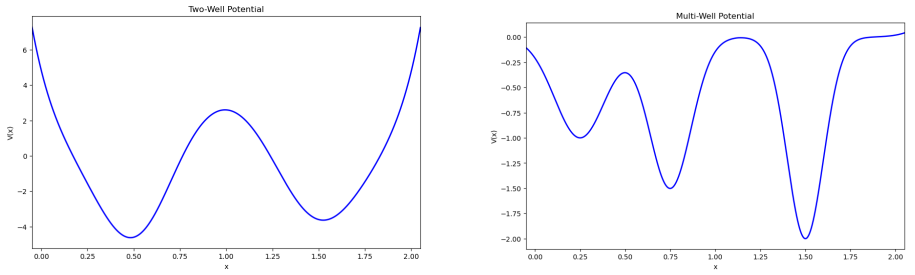
$$P(q) = \frac{\sum_{k=1}^{n} n_k \tilde{P}_k(q)}{\sum_{k=1}^{n} n_k e^{\beta(A_k - A_0)} e^{-\beta W_k(q)}}$$
$$e^{-\beta(A_k - A_0)} = \int \mathrm{d}q \; P(q) e^{-\beta W_k(q)}$$

Following the iterative solving of these equations, we can obtain the free energy from $P(q)$ through the following formula:

$$A(q) = -\frac{1}{\beta} \ln P(q) + A_0$$

3 Systems

In this project, we apply umbrella sampling and WHAM to three main systems: a one-dimensional two-welled potential, a one-dimensional multi-welled potential, and **a two-dimensional Muller-Brown potential**, which are shown in figure 3. These energy landscapes are chosen to demonstrate the effectiveness of umbrella sampling and WHAM when applied to systems with energy barriers that impede adequate sampling of configurational space in standard simulations. In all of these systems, we consider the only *one* particle in the system, and the reaction coordinate is the position of this particle. Additionally, we consider the system to be in thermal equilibrium at a temperature of $T = 50K$, to demonstrate the effectiveness of the methods in sampling the potential energy surface at a low temperature.



(a) Two-Welled Potential (b) Multi-Welled Potential

Figure 3: Potentials Used

3.1 1D Two-Welled Potential

We first consider a simple one-dimensional two-welled potential of the form:

$$V(x) = 5(x-1)^8 - \epsilon_0 e^{-\frac{\epsilon_0(x-0.5)^2}{\sigma^2}} + \epsilon_1 e^{-\frac{\epsilon_1(x-1.0)^2}{\sigma^2}} - \epsilon_2 e^{-\frac{\epsilon_2(x-1.5)^2}{\sigma^2}}$$

Where $\epsilon_0 = 5$, $\epsilon_1 = 3$, $\epsilon_2 = 4$, and $\sigma = 0.6$. This potential has two wells at $x = 0.5$ and $x = 1.5$, and a barrier at $x = 1.0$, as seen in Figure 3

3.2 1D Multi-Welled Potential

We then consider a one-dimensional multi-welled potential of the form:

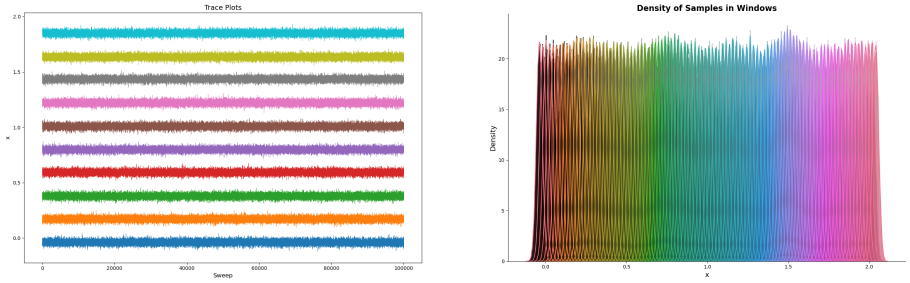
$$V(x) = 5 \min_i \left((x - x_i)^8\right) - \sum_{i=1}^{3} \epsilon_i e^{-\frac{\epsilon_i(x-x_i)^2}{\sigma^2}}$$

Where $x_1 = 0.25$, $x_2 = 0.75$, $x_3 = 1.5$, $\epsilon_0 = 1.0$, $\epsilon_1 = 1.5$, $\epsilon_2 = 2.0$, and $\sigma = 0.2$. This potential has three wells at $x = 0.25$, $x = 0.75$, and $x = 1.5$, as seen in Figure 3.

4 Results

4.1 1D Two-Welled Potential

We divided our one-dimensional two-welled potential into 100 windows, each with a biasing potential of the form $W_k(x) = \frac{1}{2}\kappa(x - s_k)^2$ with $\kappa = 2000$, where s_k is the center of the kth window, evenly spaced out across the range of x. We then ran MCMC simulations in each window, obtaining 100,000 samples for each window, resulting in biased distributions of x.



(a) Trace Plot (b) Sampled Densities

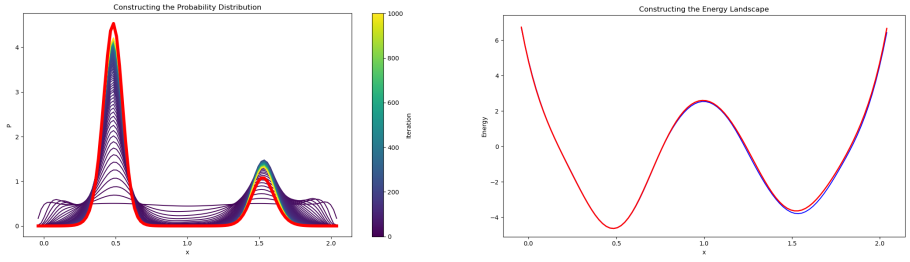
Figure 4: Confirmation of Window Sampling Success

4.1.1 Sampling

In Figure 4, we plot the trace plot of the biased distributions of x for 10 windows. We can see that the samples are stable and oscillate around the center of the window, indicating that the biasing potential has effectively encouraged sampling in the regions of the potential that would otherwise be difficult to sample. This gives us confirmation that the biasing potential is working as intended.

Additionally, we plot the sampled densities of x for all the windows in Figure 4. We can see that the sampling results in a relatively flat distribution, indicating that the biasing potential has effectively encouraged sampling in the regions of the potential that would otherwise be difficult to sample. In addition, the sampled densities overlap, indicating that our choice of κ, the biasing spring constant, was appropriate given our number of windows, and our computed free energy profile should be accurate and continuous.

4.1.2 Unbiasing and Free Energy Profile



(a) Unbiased Probabilities (b) Free Energy Profile

Figure 5: Unbiased Densities and Free Energy Profile

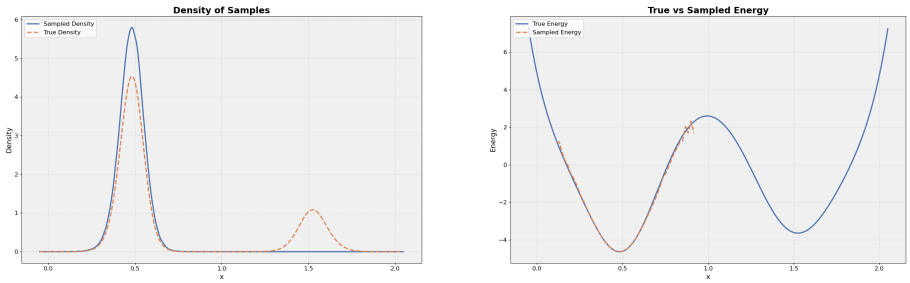
Next, we applied WHAM to unbias the biased distributions and reconstruct the global probability distribution across x. We then calculated the free energy

profile of the system.

In Figure 7, we plot the global unbiased probabilities of x for every iteration of WHAM. We can see that the probabilities converge to a stable distribution, closely approximating the true distribution of x in the system (the red curve), which is given by $P(x) \propto e^{-\beta V(x)}$. This indicates that WHAM has successfully unbiased the biased distributions. Similarly, we see that the computed free energy profile of the system as a function of x is accurate.

4.1.3 Comparison with Metropolis-Hastings

We also compared the energy profile obtained from umbrella sampling and WHAM with the energy profile obtained from a standard Metropolis-Hastings simulation. We ran a Metropolis-Hastings simulation for 1,000,000 steps, and calculated the free energy profile of the system as a function of x.



(a) Probability Distribution

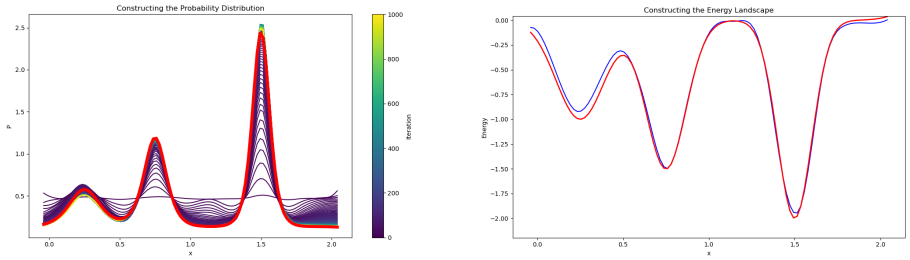
(b) Free Energy Profile

Figure 6: Metropolis-Hastings Simulation Results

In Figure 6, we plot the probability distribution across x obtained from the Metropolis-Hastings simulation and the potential energy profile of the system. We can see that because of the high energy barrier in the potential, as well as the low temperature of the system, the standard Metropolis-Hastings simulation is unable to sample the potential energy surface effectively, and gets stuck in the local minima. This results in a free energy profile doesn't fully capture the entire domain of the system. In contrast, the free energy profile obtained from umbrella sampling and WHAM captures the entire domain of the system, and is able to accurately sample the potential energy surface.

4.2 1D Multi-Welled Potential

Similarly, we divided our one-dimensional multi-welled potential into 100 windows, each with a biasing potential of the same form as the two-welled potential. We then ran MCMC simulations in each window, obtaining $100,000$ samples for each window, resulting in biased distributions of x. We show the evolution of the global unbiased probability distribution and the reconstructed free energy profile from the umbrella sampling below:



(a) Unbiased Probabilities (b) Free Energy Profile

Figure 7: Unbiased Densities and Free Energy Profile

As we can see from these plots, our method generalizes well to more complex potential functions, accurately reconstructing the expected probability distribution and the true free energy profile.

5 Next Steps

In the future, we plan to apply umbrella sampling and WHAM to more complex systems, increasing both the dimensionality and the number of particles in the system. For example, it would be interesting to apply this methodology to the two-dimensional Muller-Brown potential, shown in Figure 8, which has three minima and two saddle points, and is often used for testing algorithms that explore transition states and barriers.

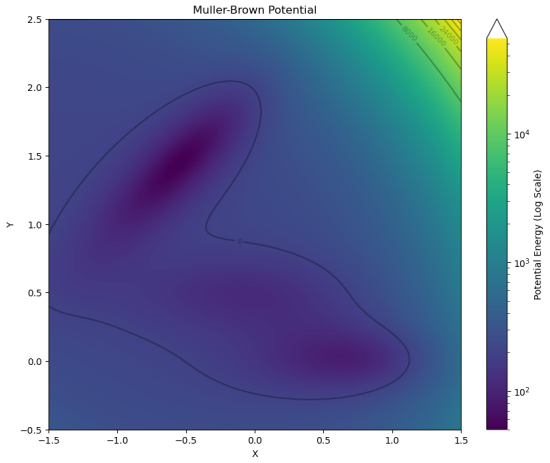
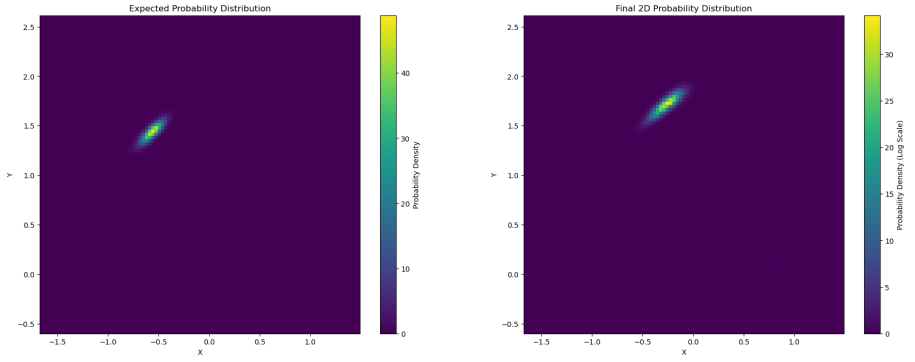


Figure 8: Muller-Brown Potential

We have been working on a multi-dimensional implementation of both our umbrella sampling and WHAM to handle this two-dimensional case. We have

validated its ability to recreate the results from our 1D-only WHAM model, and our preliminary results in 2D are shown below:



(a) Expected Probability Distribution (b) Calculated Probability Distribution

Figure 9: Expected and Calculated Probability Distributions from Muller-Brown Potential

In Figure 9, we plot the true probability distribution of the two dimensional Muller-Brown potential, as well as the reconstructed distribution from umbrella sampling and WHAM. We can see that while the distributions are not identical, they are quite similar, indicating that our method is effective in sampling the potential energy surface of the system.

Additionally, we could apply this methodology to many-particle systems such as liquid Argon to study the thermodynamics of the system.

6 Conclusion

In this project, we applied umbrella sampling and WHAM to three main systems: a one-dimensional two-welled potential, a one-dimensional multi-welled potential, and a two-dimensional Muller-Brown potential. We demonstrated that umbrella sampling and WHAM are effective methods for sampling the potential energy surface of systems with high energy barriers that impede adequate sampling of configurational space in standard simulations. This method is particularly useful for systems at low temperatures, where the system is more likely to get stuck in local minima. We showed that umbrella sampling and WHAM are able to accurately sample the potential energy surface of the system, and calculate the free energy profile of the system as a function of the reaction coordinate, whereas traditional markov-chain monte-carlo methods suffer from poor ergodicity and are unable to sample the potential energy surface effectively.

Such methodology can be applied to a wide range of physical systems, including molecular systems, materials, and biological systems, to sample the potential energy surface and calculate the free energy profile of the system. This

can be used to study the thermodynamics and kinetics of the system, and to understand the behavior of the system at different temperatures and pressures.

References

[1] Andrew Ferguson (2025), *MENG 25500, Classical Molecular and Materials Modeling*

[2] MoBioChem (2020), *Enhanced Sampling Methods - Chapter 2: Umbrella Sampling*

[3] Johannes Kastner (2011), *Umbrella sampling*

[4] CompChems (2024), *An Introduction to Enhanced Sampling and Metadynamics*