Bounding Free Energy Difference with Flow Matching

Lu Zhao(赵辂)[1,2] and Lei Wang(王磊)[1,3*]

[1]*Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*University of Chinese Academy of Sciences, Beijing 100049, China*
[3]*Songshan Lake Materials Laboratory, Dongguan 523808, China*

(Received 28 September 2023; accepted manuscript online 23 November 2023)

We introduce a method for computing the Helmholtz free energy using the flow matching technique. Unlike previous work that utilized flow-based models for variational free energy calculations, this method provides bounds for free energy estimation based on targeted free energy perturbation by performing calculations on samples from both ends of the mapping. We demonstrate applications of the present method by estimating the free energy of a classical Coulomb gas in a harmonic trap.

DOI: 10.1088/0256-307X/40/12/120201

1. Introduction. Free energy calculations hold significant importance in statistical physics, chemistry, and biology. By evaluating the free energy landscape, one can gain a quantitative understanding of thermodynamic properties. These calculations enable prediction, optimization, and a deeper comprehension of complex phenomena in diverse scientific fields.[1]

Although molecular dynamics and Monte Carlo simulations allow for large-scale and high-precision sampling, computation of the partition function poses a formidable challenge, rendering the calculation of free energy using these methods an arduous endeavor.[2,3] In general, to compute the free energy one typically embarks upon the quest of identifying a reference system with a known free energy, thereby establishing a connection between the target system and the reference system. Based on this roadmap, scientists have developed two classes of computational methods. In one class of methods there is an assumption that the system under investigation remains in thermodynamic equilibrium or, at least, experiences slow temporal changes. The difference in free energy between two systems corresponds to the work done along the switching pathway. Researchers calculate this difference by performing thermodynamic integration[2] through a series of simulations along the pathway. The second category of methods is based on the non-equilibrium equality proposed by Jarzynski.[4] The free energy evaluation using annealed importance sampling[5] is essentially equivalent to the Jarzynski equality. On the other hand, free energy perturbation (FEP)[6] can be seen as an extreme form in the non-equilibrium regime when the time approaches zero.

The development of FEP has led to the targeted free energy perturbation (TFEP)[7,8] method. Given an invertible mapping, TFEP provides a bound on free energy as well as a unique free energy estimator by separately computing the generalized work on both sides. However, the true power of the TFEP method is challenging to unleash using manually designed mapping expressions, as the selection of the mapping is crucial. Nevertheless, with the explosive development in the field of deep learning, a method for learning invertible mappings, called normalizing flow,[9–12] has emerged. Leveraging the advances in flow-based models, we can enable the computer to autonomously discover a mapping with sufficient overlap. While it is beneficial to allow the computer train and select the mapping, incorporating certain human considerations can also be advantageous.

In previous work[13–15] that combined discrete flow-based models with TFEP, however, designing the structure of the flow to incorporate symmetry into the reference distribution was a challenging task. In physical systems, symmetry plays a crucial role, and imbuing the reference distribution with such considerations to narrow down its variational space can be valuable. Although continuous normalizing flows[16,17] (CNFs), which employ the construction of a velocity field to establish an invertible mapping between two spaces through ordinary differential equation (ODE) integration, enable us to conveniently incorporate symmetry into the constructed velocity field, their training method based on ODE integration results in suboptimal computational efficiency and accuracy. Advances in the field of normalizing flows have introduced the flow matching method.[18–20] Flow matching not only retains the concise and elegant characteristics of CNFs, but also demonstrates remarkable precision and efficiency due to circumventing the need for ODE integration during training, and has been showcased in the fields related to free energy.[21]

It is worth noting that in many previous works using flow-based models for physical problems, researchers trained the models variationally without heavy reliance

*Corresponding author. Email: wanglei@iphy.ac.cn
© 2023 Chinese Physical Society and IOP Publishing Ltd

on data, thus enabling the widespread adoption of this elegant method across various domains in recent years, for example lattice models,[22,23] molecules,[24] and hydrogen atoms.[25–27] However, variational free energy methods, in principle, only provide an upper bound estimation of free energy and do not achieve exact computation.

In this study we demonstrate the application of the targeted free energy perturbation with flow matching to the classical Coulomb gas in a harmonic trap,[28] providing bounds for the free energy. Notably, by plotting a frequency histogram of 'forward work' and 'reverse work', we can observe that the overlap of the two frequencies occurs precisely within the bounds.

2. Methods

2.1 Free Energy Estimation. We employ A and B to denote two thermodynamic equilibrium states, with their respective densities being $\rho_A(x) = e^{-\beta H_A(x)}/Z_A$ and $\rho_B(x) = e^{-\beta H_B(x)}/Z_B$, where x denotes a point in the configuration space, $Z = \int \exp[-\beta H(x)]dx$ is the partition function and β is the inverse temperature.

Given an invertible mapping $\mathcal{M}$: $A \to A'$, we can map A to a new state A', transforming configurations x sampled from A into new configurations $y = \mathcal{M}(x)$. Similarly, we also consider the reverse case where configurations are drawn from B and mapped to B' via the inverse $\mathcal{M}^{-1}$: $B \to B'$. We refer to this pair of prescriptions as the 'forward' and 'reverse' processes, while employing the symbols $\to$ and $\leftarrow$ to represent them, respectively. For each process, we denote generalized energy differences as

$$\begin{aligned}\Phi_{\to}(x) &= H_B(\mathcal{M}(x)) - H_A(x) - \beta^{-1}\log|J_{\mathcal{M}}(x)|,\\ \Phi_{\leftarrow}(x) &= H_A(\mathcal{M}^{-1}(x)) - H_B(x)\\ &\quad - \beta^{-1}\log|J_{\mathcal{M}^{-1}}(x)|,\end{aligned} \tag{1}$$

where $J_{\mathcal{M}}$ and $J_{\mathcal{M}^{-1}}$ are the Jacobian determinants associated with the mappings. We refer to the generalized energy differences as 'forward work' and 'reverse work', respectively.

By applying the Kullback–Leibler divergence to A' and B and considering the non-negative property, we can obtain

$$\int_{-\infty}^{\infty} \rho_{A'}(x)\ln\Big[\frac{\rho_{A'}(x)}{\rho_B(x)}\Big]dx \geq 0. \tag{2}$$

We can derive the following inequality:

$$\Delta F \leq \langle \Phi_{\to} \rangle_A. \tag{3}$$

Here, $\Delta F = -\frac{1}{\beta}(\ln Z_B - \ln Z_A)$ is defined as the free energy difference between A and B. The symbol $\langle\cdot\rangle$ represents the ensemble average under the specified state. Equation (3) represents the principle of variational free energy, providing an upper bound on the free energy difference by computing the ensemble average of the forward work $\Phi_{\to}$ under the state A. When the free energy of one state is known (e.g. Gaussian), this becomes a way to estimate the absolute free energy of the other state. By interchanging the probability distributions within the Kullback–Leibler divergence, we can derive the following constraints:

$$\langle -\Phi_{\leftarrow} \rangle_B \leq \Delta F \leq \langle \Phi_{\to} \rangle_A. \tag{4}$$

Therefore, we can establish the upper and lower bound of free energy by estimating the respective expectations of $\Phi_{\to}$ and $-\Phi_{\leftarrow}$ on the two distributions, respectively. It is evident that as the Kullback–Leibler divergence between the two distributions $\rho_{A'}$ and ρ_B (or $\rho_{B'}$ and ρ_A) approaches zero, the bounds become constricting. When the two distributions are perfectly identical, both inequalities in the equation hold true.

Moreover, the fluctuation theorem[8,29] exists between the two processes,

$$\frac{p_{\to}(\phi)}{p_{\leftarrow}(-\phi)} = e^{\beta(\phi - \Delta F)}, \tag{5}$$

where

$$p_{\to}(\phi) = \int \delta(\phi - \Phi_{\to}(x))\rho_A(x)dx, \tag{6}$$

$$p_{\leftarrow}(\phi) = \int \delta(\phi - \Phi_{\leftarrow}(x))\rho_B(x)dx \tag{7}$$

are generalized work distributions with δ being the Dirac delta function. This means that, when we present a probability distribution in a graph, the precise value of free energy is inevitably situated at the intersection of the two probability distribution graphs.

2.2 Flow Matching. In order to make the bounds 4 narrower and the overlap between the two distributions $p_{\to}(\phi)$ and $p_{\leftarrow}(-\phi)$ more pronounced, we require an invertible mapping $\mathcal{M}$ that allows $\rho_{A'}(x)$ to accurately approximate $\rho_B(x)$. To calculate generalized energy differences Φ according to Eq. (1), we should not only be able to sample from $\rho_{A'}(x)$ but also possess knowledge about the specific values of the probability density corresponding to each sample. This, indeed, is the very essence of the normalizing flow's characteristic.

A flow model is a probability distribution $p_{\mathcal{X}}$: $\mathbb{R}^d_{\mathcal{X}} \to \mathbb{R}_{>0}$ defined as the pushforward of a base distribution $p_{\mathcal{Z}}$: $\mathbb{R}^d_{\mathcal{Z}} \to \mathbb{R}_{>0}$ through a flexible diffeomorphism f: $\mathbb{R}^d_{\mathcal{Z}} \to \mathbb{R}^d_{\mathcal{X}}$, typically parameterized by neural networks. When points $x \in \mathbb{R}^d_{\mathcal{X}}$ and $z \in \mathbb{R}^d_{\mathcal{X}}$ satisfy $x = f(z)$, their corresponding probability distributions satisfy $\log p_x(x) = \log p_z(z) - |\det J_f(z)|$.

Neural ordinary differential equations[17] can be seen as the continuous version of residual flows, as shown in Fig. 1. Instead of specifying a discrete sequence of hidden layers, we parameterize the derivative of the hidden state using a neural network, then the log-likelihood can be calculated by

$$\frac{dx}{dt} = v, \qquad \frac{d\ln p}{dt} = -\nabla \cdot v. \tag{8}$$

In order to endow $p(x)$ with a certain symmetry, specifically $p(x) = p(\mathcal{P}x)$, it can be observed that, in the context of CNFs, what we aim to do is to construct a velocity field that possesses the corresponding equivariance

$$v_t(\mathcal{P}x) = \mathcal{P}v_t(x). \tag{9}$$

For optimizing the parameterized velocity field, we use a brand new method called flow matching. Instead of negative log-likelihood, flow matching minimizes the velocity directly. By introducing a time-differentiable interpolant I_t: $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, such that

$$I_{t=0}(x_0, x_1) = x_0, \quad I_{t=1}(x_0, x_1) = x_1,$$

the velocity $v_t(x)$ that satisfies the continuity equation with the probability density $p_t(x)$ is the unique minimizer over $v_{\theta,t}(x)$ of the objective [20]

$$\mathcal{G} = \mathbb{E}_{x_0}\mathbb{E}_{x_1}\|v_{\theta,t}(I_t(x_0, x_1)) - \partial_t I_t(x_0, x_1)\|^2. \tag{10}$$

Once we have selected an interpolant that satisfies the boundary conditions, here we choose

$$I_t(x_0, x_1) = (1-t)x_0 + tx_1, \tag{11}$$

and parameterized the velocity field, we can proceed with the training based on Eq. (10).

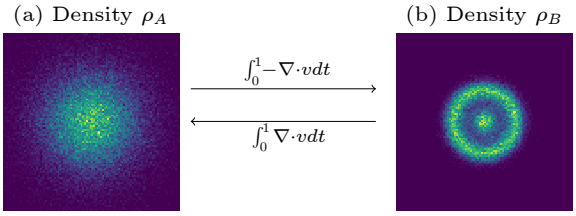

Fig. 1. The probability densities of the states A and B, which can be connected between the two states through the integration of the velocity field via Eq. (8): (a) a Gaussian distribution, (b) the distribution obtained through Markov chain Monte Carlo simulation of classical Coulomb gas in a harmonic trap.

2.3 Architectures. We opt for the transformer [30] architecture which satisfies the permutation symmetry of particles to parameterize a velocity field $v(x, t)$. In Fig. 2, we present a diagram of the network we used. In order to incorporate temporal dependence into the network, we input a sequence of length n, replicate the time variable tn times and concatenate it with each vector in the sequence, thereby augmenting its dimension to $d+1$, where d denotes the dimensionality of the physical system and n represents the number of particles.

Subsequently, we feed the sequence into the layer consisting of two residual connections. The first residual block comprises a multi-head attention layer, while the second one consists of a dense connection composed of two linear layers and a Gaussian error linear unit layer. After passing the sequence through the five aforementioned layers, we apply a linear transformation to the sequence, resulting in a sequence of length n, where each vector has a dimension of d. This sequence represents the velocity field we have obtained.

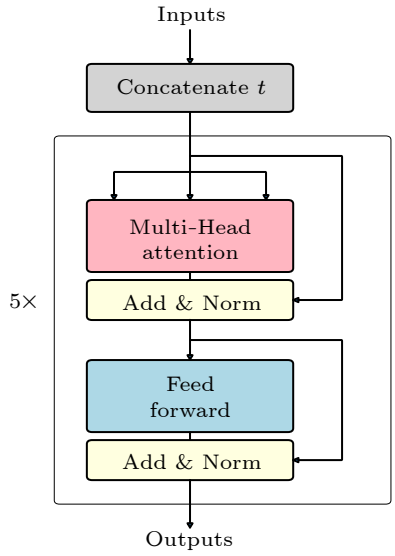

Fig. 2. The network architecture for velocity field parameterization.

3. Results. In this study, we have employed the flow matching method to investigate a classical Coulomb gas in a harmonic trap, where we consider the electrons as point charges and omit their kinetic energy. [28] The Hamiltonian reads

$$H = \sum_{i<j} \frac{1}{|x_i - x_j|} + \sum_i x_i^2, \tag{12}$$

with x_i representing the coordinate of the ith electron and $\sum_i x_i^2$ corresponding to the harmonic trapping potential eliminating the need to consider periodic boundary conditions.

We are now ready to present the compelling results obtained from our research. We conducted training on a system with a dimensionality of two and a particle count of six, employing the network (section 2.3) as the trainable velocity field where the key size and the number of heads are both 16 within the multi-head attention block.

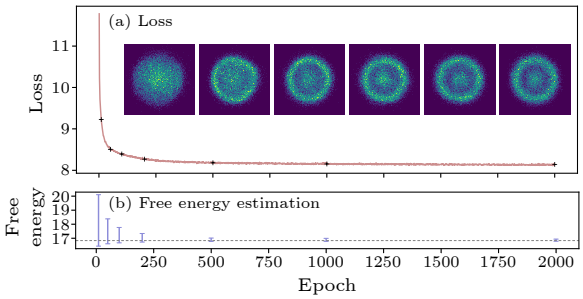

Fig. 3. (a) The evolution of the loss function (10) during training and an array of sample distributions generated by flow-matching models at various stages of the training process. The training epochs of these models correspond to 10, 50, 100, 200, 500, 1000, and 2000, respectively, marked on the plot with black crosses. (b) The upper and lower bounds of free energy, derived from estimations based on $\rho_{A'}(x)$ of the selected training stages. The dashed line represents the TFEP estimation of free energy.

Furthermore, we consider the physical system to be in

state B, while state A is chosen to follow a Gaussian distribution whose free energy can be obtained analytically. As shown in Fig. 3, as the number of epochs increases, we can observe a gradual decrease in the loss function, signifying the successful optimization of our model. Figure 3 also illustrates the gradual evolution of the distribution $\rho_{A'}(x)$ during the training process. It is evident that as the training deepens, the sample distribution derived from flow sampling gradually converges towards the actual distribution simulated by Fig. 1. Simultaneously, Fig. 3 shows that the estimates of the upper and lower bounds of free energy corresponding to state B also become increasingly constrained. Though not endowing the network with any spatial equivariance, the training results indicate an automatic restoration of rotational symmetry to some extent, as is evident from the images.

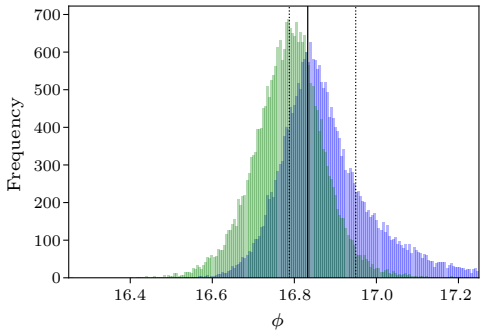

Fig. 4. The frequency histograms of distributions $p_{\rightarrow}(\phi)$ and $p_{\leftarrow}(-\phi)$, represented in blue and green, respectively. The solid line in the graph represents the TFEP estimation of free energy, while the two dashed lines to its left and right indicate the lower and upper bounds obtained from variational free energy estimations. It can be observed that the intersection of the histograms precisely aligns with the solid line.

Upon a well-trained distribution $\rho_{A'}(x)$, the estimations of the bounds as well as the distributions $p_{\rightarrow}(\phi)$ and $p_{\leftarrow}(-\phi)$ are locally displayed in Fig. 4. Evidently, the actual value of free energy is bounded by $\langle -\Phi_{\leftarrow}\rangle_B$ and $\langle \Phi_{\rightarrow}\rangle_A$. Furthermore, from the distribution perspective, the two distributions $p_{\rightarrow}(\phi)$ and $p_{\leftarrow}(-\phi)$ exhibit a certain degree of overlap, with the abscissa of the intersection point aligning with the TFEP estimation of free energy. These two observations align precisely with the theoretical framework expounded in section 2.1.

Overall, with the classical Coulomb gas in a harmonic trap as an exemplar, we have demonstrated the utilization of sample data to estimate the free energy through the method of flow matching. In this approach, the estimation of free energy has attained explicit upper and lower bounds, and we can also estimate the actual value of free energy from the overlap of the distribution plots. Moreover, a method for obtaining tighter bounds is demonstrated in the Supplemental Material.

4. Discussion. Machine learning and physics research share some commonalities in their methods and objectives.[31] Both disciplines focus on the process of gathering and analyzing data to devise models capable of predicting the behavior of intricate systems. However, in practical applications we typically utilize training and testing sets to assess the generalization performance of models. However, this approach merely provides an estimation of the model's performance on a given dataset without offering an exact estimation of the error on unknown data. Similarly, when employing the variational principle to solve problems, it is also challenging to obtain an exact error estimation for the predicted results. In this context, we present a scenario where deep learning can be employed to provide upper and lower bounds for free energy.

This methodology could be applied in more practical systems in the future, such as the field of molecular science. In intricate systems, the efficacy and precision of flow matching can be further leveraged to a greater extent. Also, combined with a direct generalization on the Jarzynski equality,[32] the method may potentially have wider impacts on non-equilibrium dynamics. In the experiments conducted in this work, we did not impose constraints on spatial symmetry. Although the experimental results indicate the restoration of spatial symmetry, in larger and more intricate systems, utilizing vector fields that consider spatial symmetry would likely have a greater impact on the training outcomes.

Acknowledgments. We thank Han Wang and Linfeng Zhang for the discussion. This project was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (Grant Nos. XDB0500000 and XDB30000000), and the National Natural Science Foundation of China (Grant Nos. 92270107, 12188101, 12122103, T2225018, and T2121001).

References

[1] Kollman P 1993 *Chem. Rev.* **93** 2395
[2] Frenkel D and Smit B 2023 *Understanding Molecular Simulation: From Algorithms to Applications* (Amsterdam: Elsevier)
[3] Tuckerman M E 2023 *Statistical Mechanics: Theory and Molecular Simulation* (Oxford: Oxford University Press)
[4] Jarzynski C 1997 *Phys. Rev. Lett.* **78** 2690
[5] Neal R M 2001 *Stat. Comput.* **11** 125
[6] Zwanzig R W 1954 *J. Chem. Phys.* **22** 1420
[7] Jarzynski C 2002 *Phys. Rev. E* **65** 046122
[8] Hahn A M and Then H 2009 *Phys. Rev. E* **79** 011113
[9] Tabak E G and Vanden-Eijnden E 2010 *Commun. Math. Sci.* **8** 217
[10] Dinh L, Krueger D, and Bengio Y 2014 arXiv:1410.8516 [cs.LG]
[11] Rezende D and Mohamed S 2015 *Proceedings of Machine Learning Research (PMLR)* vol 37 p 1530
[12] Papamakarios G, Nalisnick E, Rezende D J, Mohamed S, and Lakshminarayanan B 2021 *J. Mach. Learn. Res.* **22** 2617
[13] Wirnsberger P, Ballard A J, Papamakarios G, Abercrombie S, Racanière S, Pritzel A, Rezende D J, and Blundell C 2020 *J. Chem. Phys.* **153** 144112
[14] Wirnsberger P, Papamakarios G, Ibarz B, Racanière S, Ballard A J, Pritzel A, and Blundell C 2022 *Mach. Learn.: Sci. Technol.* **3** 025009

[15] Caselle M, Cellini E, Nada A, and Panero M 2022 *J. High Energy Phys.* 2022 15
[16] Zhang L and Wang L *et al.* 2018 arXiv:1809.10188 [cs.LG]
[17] Chen R T Q, Rubanova Y, Bettencourt J, and Duvenaud D K 2018 *Advances in Neural Information Processing Systems (NeurIPS 2018)* vol 31
[18] Lipman Y, Chen R T, Ben-Hamu H, Nickel M, and Le M 2022 arXiv:2210.02747 [cs.LG]
[19] Liu X C, Gong C Y, and Liu Q 2022 arXiv:2209.03003 [cs.LG]
[20] Albergo M S and Vanden-Eijnden E 2022 arXiv:2209.15571 [cs.LG]
[21] Klein L, Krämer A, and Noé F 2023 arXiv:2306.15030 [cs.LG]
[22] Li S H and Wang L 2018 *Phys. Rev. Lett.* **121** 260601
[23] Nicoli K A, Nakajima S, Strodthoff N, Samek W, Müller K R, and Kessel P 2020 *Phys. Rev. E* **101** 023304
[24] Li S H, Dong C X, Zhang L, and Wang L 2020 *Phys. Rev. X* **10** 021020
[25] Xie H, Zhang L, and Wang L 2022 arXiv:2105.08644 [cond-mat.str-el]
[26] Xie H, Zhang L, and Wang L 2023 *SciPost Phys.* **14** 154
[27] Xie H, Li Z H, Wang H, Zhang L, and Wang L 2023 *Phys. Rev. Lett.* **131** 126501
[28] Bolton F and Rössler U 1993 *Superlattices Microstruct.* **13** 139
[29] Crooks G E 2000 *Phys. Rev. E* **61** 2361
[30] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł, and Polosukhin I 2017 *Advances in Neural Information Processing Systems (NIPS 2017)* vol 30
[31] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L, and Zdeborová L 2019 *Rev. Mod. Phys.* **91** 045002
[32] Tang Y, Yuan R, Chen J, and Ao P 2015 *Phys. Rev. E* **91** 042108