

Endpoint reweighting identities with overlap diagnostics for free-energy differences in strongly coupled open systems

Mohammad Rahbar 

*Technical University of Munich, TUM School of Natural Sciences, Department of Chemistry,
Lichtenbergstr. 4, D-85748 Garching, Germany*

Christopher J. Stein 

*Technical University of Munich, TUM School of Natural Sciences, Department of Chemistry, Catalysis Research Center,
Atomistic Modeling Center, Munich Data Science Institute, Lichtenbergstr. 4, D-85748 Garching, Germany*



(Received 12 December 2025; revised 10 April 2026; accepted 18 June 2026; published 9 July 2026)

We develop a reweighting framework to quantify the free energy difference between two equilibrium states of a strongly coupled open system at fixed temperature. For an open system described by the Hamiltonian of mean force (HMF), we show that the equilibrium free energy difference between two canonical endpoints can be written as exponential averages of the HMF shift, divided by an explicit factor built from the chi-squared divergence between the initial and final system marginals. These relations hold at the endpoint level and, under an explicit asymptotic-equilibration postulate, admit trajectory representations for protocols whose evolved system marginal reaches the prescribed final canonical endpoint. In this trajectory representation, the dynamics provide a route for generating samples from the final canonical endpoint. In the frozen-driving regime with a noninteracting reference, the equalities reduce to FEP-like expressions involving an environment functional and an explicit overlap diagnostic, with the Zwanzig formula recovered as a limiting case. We validate the approach on an open system coupled to an environment and evolved under overdamped Langevin dynamics, where conventional Zwanzig FEP suffers from poor phase-space overlap and slow numerical convergence. The finite-sampling tests show that the overlap factor can be reconstructed from the same samples entering the estimator, making the overlap burden directly observable. A multistage coupling ladder then decomposes this burden into local factors and provides direct comparison with recursive FEP, BAR, and MBAR.

DOI: [10.1103/3twj-4279](https://doi.org/10.1103/3twj-4279)

I. INTRODUCTION

Free energy differences are key quantities in statistical mechanics and molecular simulation [1]. They decide which chemical state is favored, how strongly molecules bind [2,3], when phases change [4], and how fast reactions proceed [5]. Because of this, many simulation methods in chemistry, biology, and materials science are built around finding these differences with high accuracy [6–9]. In most cases, one does not jump directly from one state to another. Instead, one builds a path that connects two thermodynamic states, either by slowly changing interactions or by moving an external control parameter [1,10–12]. Within this landscape, Zwanzig’s free energy perturbation theory [13,14] holds a special place. It writes the free energy difference between two canonical ensembles as an exponential average of the potential-energy difference, evaluated in the initial or final

ensemble. In this way, a simple microscopic quantity, the interaction change between two states, is turned into a macroscopic free energy difference. This idea is both elegant and powerful, and it underpins much of modern free energy calculations. However, the success of exponential averaging depends on subtle statistical conditions. Good estimates require strong phase-space overlap between reference and target ensembles [15,16]. When the overlap is poor, the average develops a strong dependence on rare configurations [1,11]. Then even simple-looking transformations become hard to sample, and straightforward FEP can demand very long simulations or fail to converge in practice. These well-known issues have motivated a wide range of improvements such as enhanced sampling schemes [10,17–28], finely spaced intermediate states [29,30], and more advanced estimators such as overlap sampling and Bennett-type methods [31–34]. All of them are, at their core, different ways of dealing with the same challenge: how to recover reliable free energy differences when the important regions of phase space are only weakly shared between the ensembles one wants to compare. A further layer of difficulty emerges when the system of interest is strongly coupled to its environment. In this regime, the interaction energy between system $\mathcal{S}$ and environment $\mathcal{E}$ is comparable to the bare system energy, and the standard weak coupling picture, where the bath is treated

*Contact author: christopher.stein@tum.de

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

as a simple background that only sets the temperature is no longer adequate. The Hamiltonian of mean force (HMF) is the natural tool to describe such open systems [12,35–43]. However, the HMF is difficult to construct explicitly, and its use raises conceptual issues, including ambiguities in the definition of internal energy, entropy, free energy and heat at strong coupling [37,38,44–47]. These challenges complicate the formulation of free energy methods for open systems where strong system-environment interactions are not a small correction but an essential part of the physics. In our recent works [48,49], we tackled this strong-coupling problem by placing the system-environment interaction energy at the center of the description. We derived exact endpoint identities that split both free and internal energies into two parts: one coming from the energetic interaction itself, and one coming from how the system’s probability distribution is reshaped [48]. The latter is captured by a chi-squared divergence that measures how far the coupled and uncoupled endpoint ensembles are from each other. The present article builds on this perspective and takes the next step. We ask whether one can recover the free energy difference between two canonical endpoint marginals of a strongly coupled open system, while keeping the overlap burden explicit. In the present work, we focus on the frozen-driving branch, where the endpoint identities connect naturally to FEP-type reweighting. The resulting relations should be read as equilibrium endpoint reweighting identities, not as forward-reverse trajectory fluctuation theorems in the standard path-ensemble sense. Their trajectory forms enter only after an explicit asymptotic-equilibration postulate, where the dynamics provide a route for generating samples from the final canonical endpoint. We begin by formulating the evolution of the composite $\mathcal{S} + \mathcal{E}$ in terms of an abstract trajectory map $\mathcal{T}_t$ acting on the full phase space (Definition 1), and by specifying the total Hamiltonian along trajectories, with two independently controlled protocols $\lambda(t)$ and $C(t)$ that encode driving and coupling (Definition 2). On this basis, we introduce the canonical ensembles of the composite $(\mathcal{S} + \mathcal{E})$ and the marginal equilibrium distributions of the system $\mathcal{S}$, expressed in terms of the HMF (Definition 3). The same trajectory map is then used to represent the time-dependent system marginal as a pushforward of the initial equilibrium state [Eq. (18)], which will be the key ingredient for the trajectory counterparts of our endpoint relations. With these kinematic and statistical ingredients in place, Theorem 1 establishes endpoint reweighting identities that express the open system free energy difference $\Delta F_{\mathcal{S}}^*(\beta)$ in terms of the HMF shift $\Delta \mathcal{H}_{\beta}^*(X_{\mathcal{S}})$ and the chi-squared divergence $\chi^2(P_{t_{\text{eq}}}^{\mathcal{S}} \parallel P_0^{\mathcal{S}})$ between the initial and final system marginals. The resulting relations depend only on the canonical endpoints and are equilibrium change-of-measure identities between the corresponding HMF marginals. Invoking asymptotic equilibration of the full composite system, Theorem 2 then rewrites these endpoint relations in trajectory form, rewriting the same free energy difference as a trajectory-ensemble average over initial conditions $X_0 \sim P_0$ whose evolved system coordinates sample the final canonical endpoint. In this representation, the overlap factor $1 + \chi^2(P_{t_{\text{eq}}}^{\mathcal{S}} \parallel P_0^{\mathcal{S}})$ remains explicitly present, and the trajectory language refers to the sampling route that produces the evolved configurations $\mathcal{T}_{t_{\text{eq}}}^{\mathcal{S}}(X_0)$. An optional thermodynamic reading of the trajectory HMF increment is given in

the Appendix. This material is interpretive bookkeeping and is not part of the endpoint estimator or the sample-based overlap reconstruction. Up to this point, the endpoint and trajectory equalities are formulated in full generality, without committing to a particular protocol. To connect this structure to concrete free energy schemes, it is useful to reorganize it in terms of two natural protocol views. In the frozen-coupling view, the coupling $C(t)$ is held fixed while the driving control $\lambda(t)$ carries the system between the equilibrium endpoints. In the frozen-driving view, $\lambda(t)$ is held fixed and the transition is instead generated by switching the coupling $C(t)$. These complementary perspectives identify which control parameter moves the system between endpoints and provide the bridge to conventional FEP and JE constructions. In Ref. [50] we address the frozen-coupling view in full detail and its connection to the JE. In the remainder of the article, we focus on the frozen-driving viewpoint, $\lambda(t) \equiv \lambda(0)$, and analyze coupling-changing protocols that mirror the construction of FEP. In this regime, Corollaries 1 and 2 show that the HMF shift reduces to a logarithm of a functional $\mathcal{M}$, so that both the endpoint and trajectory equalities can be written entirely in terms of ratios of $\mathcal{M}$ evaluated at $C(0)$ and $C(t_{\text{eq}})$, divided by the overlap factor $1 + \chi^2(P_{t_{\text{eq}}}^{\mathcal{S}} \parallel P_0^{\mathcal{S}})$. In Sec. V, by exploiting the central identity relating Eqs. (44) and (47), we recover the Zwanzig FEP relation in the HMF language [Eq. (64)]. Section VI then shows that conventional FEP is a particular realization of the more general endpoint and trajectory equalities derived here, and introduces a trajectory-based FEP-like estimator that contains an explicit overlap factor. Finally, Sec. VII validates this construction on an analytically tractable system coupled to an environment and evolved under overdamped Langevin dynamics. There we compare the exact HMF reference, the Zwanzig estimator, and the trajectory form of the endpoint reweighting identity, and we show that the degradation of one-step FEP at stronger coupling is directly tied to the growth of the endpoint overlap factor. In addition to the direct validation of the trajectory representation, we also examine its finite-sampling structure. First, we show that in the one-step frozen-driving setting, the overlap factor can be reconstructed directly from the same trajectory sample that enters the estimator, so that the statistical source of degradation becomes observable rather than hidden inside an exponential average. Second, we formulate a recursive implementation along a coupling ladder, where the global free energy difference is written as a sum of local contributions and the global overlap burden is decomposed into explicit local factors. This multistage representation clarifies how the present framework connects to standard stratification ideas, and it provides a natural setting for comparison with recursive FEP, BAR, and MBAR without claiming a universal numerical advantage over these estimators.

II. PRELIMINARY DEFINITIONS

Definition 1 (Microscopic trajectory of the composite system). Here, we formulate the evolution of the composite system in an abstract manner, remaining agnostic to the details of its underlying dynamics. We consider a composite system $\mathcal{S} + \mathcal{E}$, consisting of a system of interest $\mathcal{S}$ and its environment $\mathcal{E}$. The total phase space of the composite at time

t is denoted by Γ_t , and a single microscopic state by $X \in \Gamma_t$. Each microstate contains the full phase-space coordinates of both parts, $X = (X_S, X_E)$, where X_S and X_E collect the positions and momenta of particles in S and E , respectively. We introduce the trajectory map $\mathcal{T}_t : \Gamma_0 \rightarrow \Gamma_t$, which assigns to each initial state $X_0 \in \Gamma_0$ its evolved image $X(t|X_0) = \mathcal{T}_t(X_0)$ at time t . The map $\mathcal{T}_t$ is a purely kinematic construct and does not rely on any specific dynamical generator. It may represent Hamiltonian or stochastic evolution, provided the pushforward is well defined and the evolved system marginal relaxes to the prescribed canonical endpoint. Since $\mathcal{T}_t$ acts on the composite microstates $X_0 \in \Gamma_0$, no assumption is made that the system and environment trajectories can be defined independently, a requirement intrinsic to any open-system description [38]. The microscopic state of the composite system along a trajectory is therefore written as

$$X(t|X_0) = (X_S(t|X_0) = \mathcal{T}_t^S(X_0), X_E(t|X_0) = \mathcal{T}_t^E(X_0)), \quad (1)$$

where $\mathcal{T}_t^S$ and $\mathcal{T}_t^E$ denote the respective projections of the composite map $\mathcal{T}_t$ onto the system and environment coordinates. Next, we specify the Hamiltonian structure along the established trajectory.

Definition 2 (Total Hamiltonian along a trajectory). The instantaneous total Hamiltonian of the composite system $S + E$ evaluated along the trajectory $\mathcal{T}_t(X_0)$ is defined as

$$\begin{aligned} \mathcal{H}_{S+E}(\mathcal{T}_t(X_0), \lambda(t), C(t)) \\ = \mathcal{H}_S(\mathcal{T}_t^S(X_0), \lambda(t)) + \mathcal{H}_E(\mathcal{T}_t^E(X_0)) \\ + \mathcal{V}_{SE}(\mathcal{T}_t^S(X_0), \mathcal{T}_t^E(X_0), C(t)). \end{aligned} \quad (2)$$

Here $\mathcal{H}_S$ and $\mathcal{H}_E$ are the bare Hamiltonians of the system and environment, respectively, and $\mathcal{V}_{SE}$ is their interaction potential. Two externally controlled parameters appear: $\lambda(t)$ represents the driving protocol acting on S (for example, a control parameter or mechanical coordinate), and $C(t)$ controls the coupling strength or interaction channel between S and E . By allowing $\lambda(t)$ and $C(t)$ to vary independently, we retain the most general energetic structure of an open system subject to simultaneously applied driving and coupling controls. Each protocol may correspond to a sudden quench or to a continuous ramp in time. We denote by t_λ and t_c the times at which the corresponding protocols become constant,

$$\lambda(t) = \text{const} \quad \text{for } t \geq t_\lambda, \quad (3)$$

$$C(t) = \text{const} \quad \text{for } t \geq t_c. \quad (4)$$

The composite system subsequently relaxes to equilibrium at a later time t_{eq} . In general,

$$0 \leq t_\lambda, t_c \leq t_{\text{eq}}, \quad t_\lambda \neq t_c, \quad (5)$$

so the completion of driving, coupling, and equilibration need not coincide. This distinction will later be crucial when we analyze the frozen-coupling or frozen-driving limit in which $C(t) \equiv C(0)$ or $\lambda(t) \equiv \lambda(0)$, respectively. Now, we introduce the canonical ensembles for the composite system and the system of interest S . These equilibrium measures serve as statistical reference points for the initial and final endpoints of the process (see Fig. 1).

Definition 3 (Canonical ensembles of the composite system and marginal distributions of the system). The equilibrium probability distribution of the composite system $S + E$ at temperature β^{-1}/k_B is defined in canonical form as

$$P(X, \beta) = \frac{e^{-\beta \mathcal{H}_{S+E}(X, \lambda, C)}}{\mathcal{Z}_{S+E}(\lambda, C, \beta)}, \quad (6)$$

where $\mathcal{H}_{S+E}$ is the total Hamiltonian given in Eq. (2), and $\mathcal{Z}_{S+E}$ denotes the partition function of the composite system

$$\mathcal{Z}_{S+E}(\lambda, C, \beta) = \int dX e^{-\beta \mathcal{H}_{S+E}(X, \lambda, C)}. \quad (7)$$

Two specific equilibrium ensembles are relevant:

$$P_0(X, \beta) = \frac{e^{-\beta \mathcal{H}_{S+E}(X, \lambda(0), C(0))}}{\mathcal{Z}_{S+E}(\lambda(0), C(0), \beta)}, \quad (8)$$

$$P_{\text{eq}}(X, \beta) = \frac{e^{-\beta \mathcal{H}_{S+E}(X, \lambda(t_{\text{eq}}), C(t_{\text{eq}}))}}{\mathcal{Z}_{S+E}(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta)}. \quad (9)$$

The distributions P_0 and P_{eq} describe the initial and final equilibrium states of the composite system at the same bath temperature β^{-1}/k_B . The equilibrium marginal distribution of the system S is obtained by integrating over the environmental coordinates

$$P^S(X_S, \beta) = \int dX_E P(X_S, X_E, \beta). \quad (10)$$

This probability can be expressed in canonical form through the HMF, defined as [37,38]

$$\begin{aligned} \mathcal{H}_\beta^*(X_S, \lambda, C) = \mathcal{H}_S(X_S, \lambda) \\ - \frac{1}{\beta} \ln \int dX_E \frac{1}{\mathcal{Z}_E} e^{-\beta [\mathcal{H}_E(X_E) + \mathcal{V}_{SE}(X_S, X_E, C)]}, \end{aligned} \quad (11)$$

where

$$\mathcal{Z}_E = \int dX_E e^{-\beta \mathcal{H}_E(X_E)}. \quad (12)$$

Equation (11) defines an effective Hamiltonian that incorporates the influence of the environment on S at fixed (λ, C, β) . The corresponding partition function of the system is

$$\mathcal{Z}_S^*(\lambda, C, \beta) = \int dX_S e^{-\beta \mathcal{H}_\beta^*(X_S, \lambda, C)}. \quad (13)$$

Hence the equilibrium marginal distribution of the system reads

$$P^S(X_S, \beta) = \frac{e^{-\beta \mathcal{H}_\beta^*(X_S, \lambda, C)}}{\mathcal{Z}_S^*(\lambda, C, \beta)}. \quad (14)$$

Applying this to the two equilibrium endpoints yields

$$P_0^S(X_S, \beta) = \frac{e^{-\beta \mathcal{H}_\beta^*(X_S, \lambda(0), C(0))}}{\mathcal{Z}_S^*(\lambda(0), C(0), \beta)}, \quad (15)$$

$$P_{\text{eq}}^S(X_S, \beta) = \frac{e^{-\beta \mathcal{H}_\beta^*(X_S, \lambda(t_{\text{eq}}), C(t_{\text{eq}}))}}{\mathcal{Z}_S^*(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta)}. \quad (16)$$

Equations (15) and (16) define the initial and final equilibrium probability distributions of the system in terms of its HMF, corresponding to the protocol endpoints $(\lambda(0), C(0))$ and $(\lambda(t_{\text{eq}}), C(t_{\text{eq}}))$. In Theorem 1, we show the exact connection between the system free energy difference at two

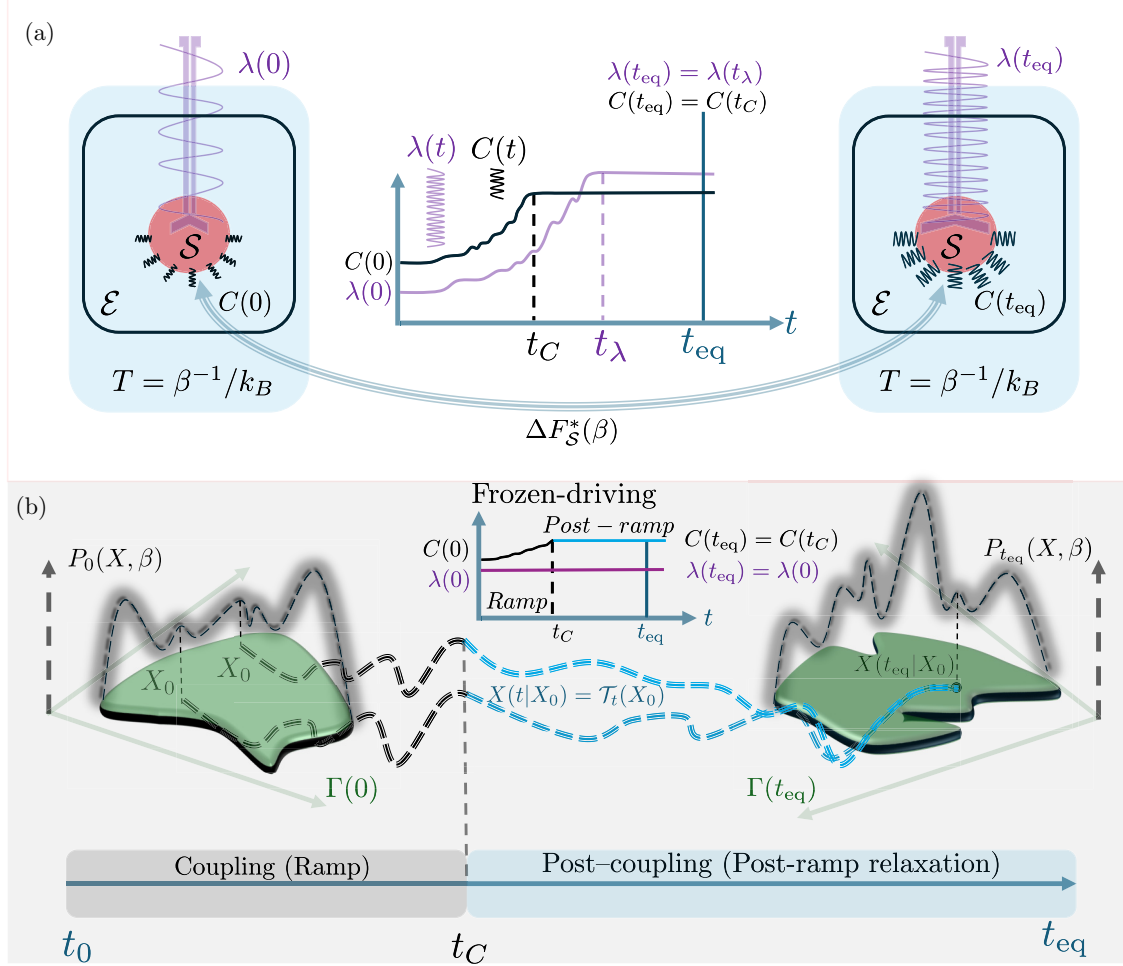


FIG. 1. (a) Schematic of the composite $\mathcal{S} + \mathcal{E}$ at a fixed bath temperature $T = \beta^{-1}/k_B$. The control $\lambda(t)$ (schematic purple spring) and the coupling $C(t)$ (schematic black spring) are driven independently and become constant at times t_λ and t_C , respectively; the composite system then relaxes to equilibrium at t_{eq} . The system free energy difference $\Delta F_S^*(\beta)$ refers to the canonical endpoints $(\lambda(0), C(0)) \rightarrow (\lambda(t_{eq}), C(t_{eq}))$ and is defined through the HMF partition functions. (b) An initial microstate $X_0 \sim P_0(X, \beta)$ on $\Gamma(0)$ evolves under a chosen dynamics whose long-time pushforward yields the canonical endpoint density $P_{eq}(X, \beta)$ on $\Gamma(t_{eq})$. The frozen driving case, $\lambda(t) \equiv \lambda(0)$, is the regime used in our validation.

endpoints and the probability distributions, together with their associated HMFs.

Definition 4 (Probability along the trajectory). Using the trajectory map $\mathcal{T}_t : \Gamma_0 \rightarrow \Gamma_t$ defined in Definition 1, the time evolution of the probability density of the composite system is

$$P_t(X, \beta) = \int_{\Gamma_0} dX_0 \delta(X - \mathcal{T}_t(X_0)) P_0(X_0, \beta), \quad (17)$$

where $P_0(X_0, \beta)$ is the initial canonical distribution given in Eq. (8). Integrating over the environmental degrees of freedom

yields the evolution of the system marginal

$$P_t^{\mathcal{S}}(X_S, \beta) = \int_{\Gamma_0} dX_0 P_0(X_0, \beta) \delta(X_S - \mathcal{T}_t^{\mathcal{S}}(X_0)). \quad (18)$$

Proof of Eq. (18). Starting from Eq. (17),

$$P_t(X_S, X_E, \beta) = \int_{\Gamma_0} dX_0 \delta((X_S, X_E) - \mathcal{T}_t(X_0)) P_0(X_0, \beta), \quad (19)$$

with $X = (X_S, X_E)$, the system marginal follows by integrating over X_E . Exchanging the order of integration and applying the property of the Dirac-delta function gives

$$\begin{aligned} P_t^{\mathcal{S}}(X_S, \beta) &= \int dX_E P_t(X_S, X_E, \beta) = \int dX_E \int_{\Gamma_0} dX_0 \underbrace{\delta(X_S - \mathcal{T}_t^{\mathcal{S}}(X_0)) \delta(X_E - \mathcal{T}_t^{\mathcal{E}}(X_0))}_{\delta((X_S, X_E) - \mathcal{T}_t(X_0))} P_0(X_0, \beta) \\ &= \int_{\Gamma_0} dX_0 P_0(X_0, \beta) \delta(X_S - \mathcal{T}_t^{\mathcal{S}}(X_0)) \underbrace{\int dX_E \delta(X_E - \mathcal{T}_t^{\mathcal{E}}(X_0))}_1 = \int_{\Gamma_0} dX_0 P_0(X_0, \beta) \delta(X_S - \mathcal{T}_t^{\mathcal{S}}(X_0)), \end{aligned} \quad (20)$$

yielding Eq. (18). ■

III. ENDPOINT REWEIGHTING IDENTITIES AND TRAJECTORY REPRESENTATIONS

Theorem 1 (Endpoint reweighting identities for free energy differences). For two equilibrium endpoints of the composite system $\mathcal{S} + \mathcal{E}$ prepared at the same temperature β^{-1}/k_B , we have

$$e^{-\beta \Delta F_S^*(\beta)} = \frac{\langle e^{-\beta \Delta \mathcal{H}_\beta^*(X_S)} \rangle_S}{1 + \chi^2(P_{t_{\text{eq}}}^S \parallel P_0^S)} \quad (21)$$

and

$$e^{+\beta \Delta F_S^*(\beta)} = \langle e^{+\beta \Delta \mathcal{H}_\beta^*(X_S)} \rangle_S. \quad (22)$$

The free energy difference is

$$\Delta F_S^*(\beta) = F_S^*(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta) - F_S^*(\lambda(0), C(0), \beta), \quad (23)$$

where $F_S^*(\lambda, C, \beta) = -\beta^{-1} \ln \mathcal{Z}_S^*(\lambda, C, \beta)$, and the HMF shift is

$$\Delta \mathcal{H}_\beta^*(X_S) = \mathcal{H}_\beta^*(X_S, \lambda(t_{\text{eq}}), C(t_{\text{eq}})) - \mathcal{H}_\beta^*(X_S, \lambda(0), C(0)). \quad (24)$$

Averages are taken over the final equilibrium ensemble,

$$\langle \bullet \rangle_S = \int dX_S \bullet P_{t_{\text{eq}}}^S(X_S, \beta), \quad (25)$$

and the chi-squared divergence between the endpoint marginals is

$$1 + \chi^2(P_{t_{\text{eq}}}^S \parallel P_0^S) = \int dX_S \frac{(P_{t_{\text{eq}}}^S(X_S, \beta))^2}{P_0^S(X_S, \beta)}. \quad (26)$$

The result established here coincides with the endpoint identity derived in [48], where the protocol is implicitly present. As seen from the derivation of Eqs. (21) and (22), the result is an equilibrium change-of-measure identity between the two HMF endpoint marginals. The reasoning relies only on the existence of well-defined equilibrium endpoints, which guarantees that the corresponding marginal distributions and free energy difference are statistically and thermodynamically meaningful. No forward-reverse path-ensemble relation is used in this derivation. The overlap factor in Eq. (26) is therefore a property of the endpoint marginals, not a dynamical correction.

Proof of Theorem 1. From Eqs. (15), (16), and (24), we have

$$e^{-\beta \Delta \mathcal{H}_\beta^*(X_S)} = \frac{P_{t_{\text{eq}}}^S(X_S, \beta) \mathcal{Z}_S^*(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta)}{P_0^S(X_S, \beta) \mathcal{Z}_S^*(\lambda(0), C(0), \beta)}, \quad (27)$$

and averaging Eq. (27) over the final ensemble, Eq. (25), yields

$$\begin{aligned} \langle e^{-\beta \Delta \mathcal{H}_\beta^*} \rangle_S &= \frac{\mathcal{Z}_S^*(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta)}{\mathcal{Z}_S^*(\lambda(0), C(0), \beta)} \\ &\times \int dX_S \frac{(P_{t_{\text{eq}}}^S(X_S, \beta))^2}{P_0^S(X_S, \beta)}. \end{aligned} \quad (28)$$

Using Eq. (23) for the partition function ratio and Eq. (26) for the divergence, Eq. (28) reduces to

$$\langle e^{-\beta \Delta \mathcal{H}_\beta^*} \rangle_S = e^{-\beta \Delta F_S^*(\beta)} [1 + \chi^2(P_{t_{\text{eq}}}^S \parallel P_0^S)], \quad (29)$$

which rearranges to Eq. (21). For the positive exponential, Eqs. (15), (16), and (24) give

$$e^{+\beta \Delta \mathcal{H}_\beta^*(X_S)} = \frac{P_0^S(X_S, \beta) \mathcal{Z}_S^*(\lambda(0), C(0), \beta)}{P_{t_{\text{eq}}}^S(X_S, \beta) \mathcal{Z}_S^*(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta)}, \quad (30)$$

and averaging Eq. (30) with Eq. (25) cancels the denominator and yields

$$\langle e^{+\beta \Delta \mathcal{H}_\beta^*} \rangle_S = \frac{\mathcal{Z}_S^*(\lambda(0), C(0), \beta)}{\mathcal{Z}_S^*(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta)} = e^{+\beta \Delta F_S^*(\beta)}, \quad (31)$$

reproducing Eq. (22). $\blacksquare$

Before introducing the trajectory representations, it is essential to clarify the assumption underlying their derivation. We postulate *asymptotic equilibration* of the full probability distribution along the trajectory. This assumption expresses the requirement that the evolved system marginal coincides with the final canonical endpoint marginal at t_{eq} . The trajectory construction is therefore a sampling representation of the endpoint identity. It does not introduce a forward-reverse trajectory fluctuation theorem or a separate path-probability statement.

Theorem 2 (Trajectory representation under asymptotic equilibration). We postulate asymptotic equilibration, which implies

$$\lim_{t \rightarrow t_{\text{eq}}} P_t^S(X_S, \beta) = P_{t_{\text{eq}}}^S(X_S, \beta). \quad (32)$$

Under this condition, the ensemble averages in Theorem 1 admit exact trajectory representations, because the evolved system coordinates sample the final endpoint marginal, where

$$\langle \bullet \rangle_{X_0} = \int dX_0 \bullet P_0(X_0, \beta). \quad (33)$$

Thus, the endpoint reweighting identities (21) and (22) take the trajectory form

$$e^{-\beta \Delta F_S^*(\beta)} = \frac{\langle e^{-\beta \Delta \mathcal{H}_\beta^*(\mathcal{T}_{t_{\text{eq}}}^S(X_0))} \rangle_{X_0}}{1 + \chi^2(P_{t_{\text{eq}}}^S \parallel P_0^S)} \quad (34)$$

and

$$e^{+\beta \Delta F_S^*(\beta)} = \langle e^{+\beta \Delta \mathcal{H}_\beta^*(\mathcal{T}_{t_{\text{eq}}}^S(X_0))} \rangle_{X_0}, \quad (35)$$

where

$$\begin{aligned} \Delta \mathcal{H}_\beta^*(\mathcal{T}_{t_{\text{eq}}}^S(X_0)) &= \mathcal{H}_\beta^*(\mathcal{T}_{t_{\text{eq}}}^S(X_0), \lambda(t_{\text{eq}}), C(t_{\text{eq}})) \\ &- \mathcal{H}_\beta^*(\mathcal{T}_{t_{\text{eq}}}^S(X_0), \lambda(0), C(0)). \end{aligned} \quad (36)$$

Proof of the trajectory representation of the numerator in Eq. (34). Starting from the numerator of Eq. (21) and using Eqs. (25), (32), and (18),

$$\begin{aligned} \langle e^{-\beta \Delta \mathcal{H}_\beta^*(X_S)} \rangle_S &= \int dX_S e^{-\beta \Delta \mathcal{H}_\beta^*(X_S)} P_{t_{\text{eq}}}^S(X_S, \beta) = \lim_{t \rightarrow t_{\text{eq}}} \int dX_S e^{-\beta \Delta \mathcal{H}_\beta^*(X_S)} P_t^S(X_S, \beta) \\ &= \lim_{t \rightarrow t_{\text{eq}}} \int dX_S e^{-\beta \Delta \mathcal{H}_\beta^*(X_S)} \int_{\Gamma_0} dX_0 P_0(X_0, \beta) \delta(X_S - \mathcal{T}_t^S(X_0)) \end{aligned}$$

$$\begin{aligned}
&= \int_{\Gamma_0} dX_0 P_0(X_0, \beta) \lim_{t \rightarrow t_{\text{eq}}} \underbrace{\int dX_S e^{-\beta \Delta \mathcal{H}_\beta^*(X_S)} \delta(X_S - \mathcal{T}_t^S(X_0))}_{e^{-\beta \Delta \mathcal{H}_\beta^*(\mathcal{T}_{t_{\text{eq}}}^S(X_0))}} \\
&= \int_{\Gamma_0} dX_0 P_0(X_0, \beta) e^{-\beta \Delta \mathcal{H}_\beta^*(\mathcal{T}_{t_{\text{eq}}}^S(X_0))} = \langle e^{-\beta \Delta \mathcal{H}_\beta^*(\mathcal{T}_{t_{\text{eq}}}^S(X_0))} \rangle_{X_0}. \quad (37)
\end{aligned}$$

This confirms the numerator in Eq. (34). Applying the same steps yields Eq. (35). ■

The Appendix gives an optional decomposition of the HMF increment along trajectory-generated endpoint samples. This decomposition is not used to estimate ΔF_S^* and is not required for the overlap diagnostics developed below.

IV. FROZEN-DRIVING (COUPLING) VIEW

Although Eqs. (21) and (22) hold for arbitrary choices of the two canonical endpoints, and their trajectory representations hold when the evolved system marginal reaches the prescribed final canonical endpoint, their practical meaning becomes clearer through two protocol perspectives. A first perspective is the frozen-coupling (driving) view, in which $C(t) \equiv C(0)$ is held fixed while $\lambda(t)$ is varied, as in Jarzynski-type protocols. A second, complementary perspective is the frozen-driving (coupling) view, in which $\lambda(t) \equiv \lambda(0)$ is fixed and the interaction $C(t)$ is changed (see Fig. 1), mirroring

the structure of conventional free energy perturbation (FEP). When the endpoint equalities are interpreted through these lenses, their thermodynamic content becomes more transparent. In the present work, we adopt the frozen-driving viewpoint to analyze interaction-changing protocols and to clarify their connection to coupling-based free energy estimators such as FEP. In Ref. [50], we develop the complementary frozen-coupling viewpoint, showing how the same framework recovers Jarzynski-type relations in the appropriate limit and extends them to strong system-environment coupling even when the dynamics violate Liouvillian or detailed-balance constraints. It is important to note that both Jarzynski's equality and FEP formulas inherit the protocol-agnostic character. However, their standard derivations and applications are organized around these two aforementioned scenarios.

Corollary 1 (Frozen driving: endpoint forms in terms of system-environment interaction). For a fixed driving protocol $\lambda(t) \equiv \lambda(0)$, the HMF shift between the two equilibrium endpoints becomes

$$\Delta \mathcal{H}_\beta^*(X_S) = \mathcal{H}_\beta^*(X_S, \lambda(0), C(t_{\text{eq}})) - \mathcal{H}_\beta^*(X_S, \lambda(0), C(0)) = -\frac{1}{\beta} \ln \frac{\mathcal{M}(X_S, C(t_{\text{eq}}), \beta)}{\mathcal{M}(X_S, C(0), \beta)}. \quad (38)$$

The cancellation of the bare system Hamiltonian in Eq. (11) leaves only the environment-interaction contribution $\mathcal{M}$. Inserting Eq. (38) into Eqs. (21) and (22) and using the average in Eq. (25) gives

$$e^{-\beta \Delta F_S^*(\beta)} = \frac{\left\langle \frac{\mathcal{M}(X_S, C(t_{\text{eq}}), \beta)}{\mathcal{M}(X_S, C(0), \beta)} \right\rangle_S}{1 + \chi^2(P_{t_{\text{eq}}}^S \parallel P_0^S)}, \quad (39)$$

and the complementary form

$$e^{+\beta \Delta F_S^*(\beta)} = \left\langle \frac{\mathcal{M}(X_S, C(0), \beta)}{\mathcal{M}(X_S, C(t_{\text{eq}}), \beta)} \right\rangle_S. \quad (40)$$

In this regime, the free energy change is governed solely by ratios of the environment-interaction functional $\mathcal{M}$ evaluated at $C(0)$ and $C(t_{\text{eq}})$.

Corollary 2 (Frozen driving: trajectory forms in terms of system-environment interaction). When $\lambda(t) \equiv \lambda(0)$, the trajectory version of the HMF increment reads

$$\Delta \mathcal{H}_\beta^*(\mathcal{T}_{t_{\text{eq}}}^S(X_0)) = -\frac{1}{\beta} \ln \frac{\mathcal{M}(\mathcal{T}_{t_{\text{eq}}}^S(X_0), C(t_{\text{eq}}), \beta)}{\mathcal{M}(\mathcal{T}_{t_{\text{eq}}}^S(X_0), C(0), \beta)}. \quad (41)$$

Substituting Eq. (41) into the trajectory equalities (34) and (35) yields

$$e^{-\beta \Delta F_S^*(\beta)} = \frac{\left\langle \frac{\mathcal{M}(\mathcal{T}_{t_{\text{eq}}}^S(X_0), C(t_{\text{eq}}), \beta)}{\mathcal{M}(\mathcal{T}_{t_{\text{eq}}}^S(X_0), C(0), \beta)} \right\rangle_{X_0}}{1 + \chi^2(P_{t_{\text{eq}}}^S \parallel P_0^S)} \quad (42)$$

and

$$e^{+\beta \Delta F_S^*(\beta)} = \left\langle \frac{\mathcal{M}(\mathcal{T}_{t_{\text{eq}}}^S(X_0), C(0), \beta)}{\mathcal{M}(\mathcal{T}_{t_{\text{eq}}}^S(X_0), C(t_{\text{eq}}), \beta)} \right\rangle_{X_0}. \quad (43)$$

If $\lambda(t)$ becomes time dependent, the simplification in Eqs. (38) and (41) no longer holds and the full HMF structure must be used.

V. UNDERLYING IDENTITY AND FEP

The central identity underlying all our derivations is

$$e^{-\beta \Delta F_S^*(\beta)} = \frac{\mathcal{Z}_S^*(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta)}{\mathcal{Z}_S^*(\lambda(0), C(0), \beta)}, \quad (44)$$

where $\mathcal{Z}_S^*(\lambda, C, \beta)$ is the HMF partition function defined by Eq. (13). By construction, the composite partition function factorizes as [38]

$$\mathcal{Z}_{S+\mathcal{E}}(\lambda, C, \beta) = \mathcal{Z}_S^*(\lambda, C, \beta) \mathcal{Z}_\mathcal{E}(\beta), \quad (45)$$

where $\mathcal{Z}_\mathcal{E}(\beta)$ depends only on the temperature. Since $\mathcal{Z}_\mathcal{E}(\beta)$ cancels in ratios, Eq. (44) is equivalent to the composite equilibrium ratio

$$\frac{\mathcal{Z}_\mathcal{S}^*(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta)}{\mathcal{Z}_\mathcal{S}^*(\lambda(0), C(0), \beta)} = \frac{\mathcal{Z}_{\mathcal{S}+\mathcal{E}}(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta)}{\mathcal{Z}_{\mathcal{S}+\mathcal{E}}(\lambda(0), C(0), \beta)}. \quad (46)$$

Writing Eq. (46) in terms of composite free energies yields

$$e^{-\beta \Delta F_\mathcal{S}^*(\beta)} = e^{-\beta \Delta F_{\mathcal{S}+\mathcal{E}}(\beta)}, \quad (47)$$

with

$$\Delta F_{\mathcal{S}+\mathcal{E}}(\beta) = F_{\mathcal{S}+\mathcal{E}}(\lambda(t_{\text{eq}}), C(t_{\text{eq}}), \beta) - F_{\mathcal{S}+\mathcal{E}}(\lambda(0), C(0), \beta). \quad (48)$$

Thus, the open-system free energy difference encoded by the HMF coincides with the free energy difference of the composite system ($\mathcal{S} + \mathcal{E}$) evaluated at the same pair of protocol endpoints $(\lambda(0), C(0))$ and $(\lambda(t_{\text{eq}}), C(t_{\text{eq}}))$. In conventional FEP and most of its applications, the two endpoint equilibrium states of the composite $\mathcal{S} + \mathcal{E}$ are chosen such that the initial state is noninteracting, while the final state incorporates the full system-environment interaction [1,13]. To represent this standard situation, we consider an interaction potential of the form

$$\mathcal{V}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E}, C(t)) = C(t) \mathcal{U}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E}), \quad (49)$$

so that the coupling protocol $C(t)$ linearly controls a fixed interaction channel $\mathcal{U}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E})$. Although the FEP derivation is formally agnostic with respect to the precise choice of protocol [1], it is usually presented in the frozen-driving viewpoint, where

$$(\lambda(0), C(0)) = (\lambda_0, 0), \quad (\lambda(t_{\text{eq}}), C(t_{\text{eq}})) = (\lambda_0, C(t_{\text{eq}})), \quad (50)$$

so that the driving protocol $\lambda(t)$ is held fixed at λ_0 and only the coupling protocol is changed from 0 to $C(t_{\text{eq}})$ such that the interaction energy $\mathcal{V}_{\mathcal{SE}}$ is zero at $t = 0$. The resulting free energy difference is then expressed as an exponential average over the reference (noninteracting) ensemble [13]. We now derive this relation explicitly within our notation and connect it to Eq. (47). For the endpoint choice (50), the composite Hamiltonians read

$$\begin{aligned} \mathcal{H}_0(X) &\equiv \mathcal{H}_{\mathcal{S}+\mathcal{E}}(X, \lambda_0, C(0) = 0) \\ &= \mathcal{H}_\mathcal{S}(X_\mathcal{S}, \lambda_0) + \mathcal{H}_\mathcal{E}(X_\mathcal{E}) + \underbrace{C(0) \mathcal{U}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E})}_0, \end{aligned} \quad (51)$$

$$\begin{aligned} \mathcal{H}_1(X) &\equiv \mathcal{H}_{\mathcal{S}+\mathcal{E}}(X, \lambda_0, C(t_{\text{eq}})) \\ &= \mathcal{H}_\mathcal{S}(X_\mathcal{S}, \lambda_0) + \mathcal{H}_\mathcal{E}(X_\mathcal{E}) + \underbrace{C(t_{\text{eq}}) \mathcal{U}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E})}_{\mathcal{V}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E}, C(t_{\text{eq}}))}, \end{aligned} \quad (52)$$

so that their difference is purely given by the interaction,

$$\mathcal{H}_1(X) - \mathcal{H}_0(X) = \mathcal{V}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E}, C(t_{\text{eq}})). \quad (53)$$

The associated composite partition functions are

$$\mathcal{Z}_0 \equiv \mathcal{Z}_{\mathcal{S}+\mathcal{E}}(\lambda_0, C(0), \beta) = \int dX e^{-\beta \mathcal{H}_0(X)}, \quad (54)$$

$$\mathcal{Z}_1 \equiv \mathcal{Z}_{\mathcal{S}+\mathcal{E}}(\lambda_0, C(t_{\text{eq}}), \beta) = \int dX e^{-\beta \mathcal{H}_1(X)}, \quad (55)$$

and define the initial and final composite free energies,

$$F_0(\beta) \equiv F_{\mathcal{S}+\mathcal{E}}(\lambda_0, C(0), \beta) = -\frac{1}{\beta} \ln \mathcal{Z}_0, \quad (56)$$

$$F_1(\beta) \equiv F_{\mathcal{S}+\mathcal{E}}(\lambda_0, C(t_{\text{eq}}), \beta) = -\frac{1}{\beta} \ln \mathcal{Z}_1. \quad (57)$$

The composite free energy difference is

$$\Delta F_{\mathcal{S}+\mathcal{E}}(\beta) = F_1(\beta) - F_0(\beta) = -\frac{1}{\beta} \ln \frac{\mathcal{Z}_1}{\mathcal{Z}_0}. \quad (58)$$

To obtain the FEP identity, we now express the ratio $\mathcal{Z}_1/\mathcal{Z}_0$ as an average over the initial noninteracting equilibrium state. Starting from the definition of $\mathcal{Z}_1$ and inserting the identity $1 = e^{-\beta \mathcal{H}_0(X)}/e^{-\beta \mathcal{H}_0(X)}$ inside the integral, we find

$$\begin{aligned} \frac{\mathcal{Z}_1}{\mathcal{Z}_0} &= \frac{1}{\mathcal{Z}_0} \int dX e^{-\beta \mathcal{H}_1(X)} \\ &= \int dX e^{-\beta [\mathcal{H}_1(X) - \mathcal{H}_0(X)]} \frac{e^{-\beta \mathcal{H}_0(X)}}{\mathcal{Z}_0} \\ &= \int dX e^{-\beta [\mathcal{H}_1(X) - \mathcal{H}_0(X)]} P_0(X, \beta), \end{aligned} \quad (59)$$

where $P_0(X, \beta)$ is the canonical equilibrium distribution of the uncoupled composite

$$P_0(X, \beta) = \frac{e^{-\beta \mathcal{H}_0(X)}}{\mathcal{Z}_0}. \quad (60)$$

Using Eq. (53) we obtain

$$\begin{aligned} \frac{\mathcal{Z}_1}{\mathcal{Z}_0} &= \int dX e^{-\beta \mathcal{V}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E}, C(t_{\text{eq}}))} P_0(X, \beta) \\ &= \langle e^{-\beta \mathcal{V}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E}, C(t_{\text{eq}}))} \rangle_0, \end{aligned} \quad (61)$$

where

$$\langle \bullet \rangle_0 \equiv \int dX \bullet P_0(X, \beta) \quad (62)$$

denotes an average over the noninteracting reference ensemble. Combining Eqs. (58) and (61) leads to the standard Zwanzig FEP identity [1],

$$e^{-\beta \Delta F_{\mathcal{S}+\mathcal{E}}(\beta)} = \langle e^{-\beta \mathcal{V}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E}, C(t_{\text{eq}}))} \rangle_0. \quad (63)$$

Using the equivalence (47), this can be written in the HMF language as

$$e^{-\beta \Delta F_\mathcal{S}^*(\beta)} = \langle e^{-\beta \mathcal{V}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E}, C(t_{\text{eq}}))} \rangle_0. \quad (64)$$

VI. FEP-TYPE ENDPOINT REWEIGHTING WITH EXPLICIT OVERLAP DIAGNOSTICS

Conventional FEP [Eq. (64)] is recovered as the standard noninteracting-reference limit of Eqs. (21) and (34), obtained by fixing $\lambda(t) \equiv \lambda_0$, choosing a noninteracting reference state $\mathcal{V}_{\mathcal{SE}}(X_\mathcal{S}, X_\mathcal{E}, C(0)) = 0$, and interpreting the coupling change as an instantaneous quench [51]. In this setting the interaction is switched from zero to its final value while the driving parameter remains fixed. The endpoint identities used here should therefore be viewed as FEP-type reweighting relations written at the level of HMF marginals, with the overlap factor made explicit. This framing does not remove the usual overlap limitations of exponential reweighting; rather, it exposes them

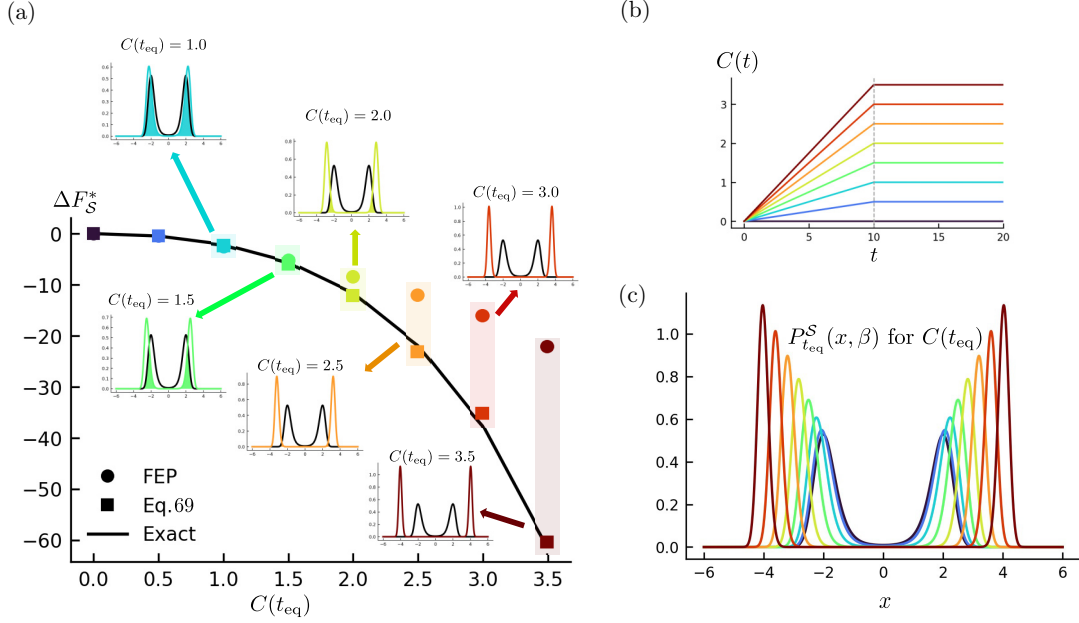


FIG. 2. (a) Open-system free energy differences $\Delta F_S^*(\beta)$ as a function of the final coupling $C(t_{eq})$. The solid line denotes the exact HMF reference computed from Eq. (44). Circles show the conventional Zwanzig FEP estimator [Eq. (64)]. Squares show the trajectory representation of the overlap-explicit endpoint identity [Eq. (69)]. The small insets display the endpoint system marginals $P_{t_{eq}}^S(x, \beta)$ for representative coupling values, illustrating how increasing $C(t_{eq})$ affects probabilities overlap. (b) Coupling protocol $C(t)$ used in the frozen-driving regime: a linear ramp from $C(0) = 0$ to $C(t_{eq})$ followed by a relaxation stage, as in Eqs. (86) and (87). (c) Endpoint system marginals $P_{t_{eq}}^S(x, \beta)$ for each final couplings. Here we demonstrate that, the degradation of one-step FEP at moderate and strong coupling is accompanied by the growth of the endpoint overlap burden, while Eq. (69) provides the corresponding overlap-explicit endpoint reweighting identity.

through $1 + \chi^2(P_{t_{eq}}^S \parallel P_0^S)$. To make this connection explicit, we first revisit the environment functional $\mathcal{M}$ appearing in the HMF definition (11),

$$\mathcal{M}(X_S, C, \beta) = \frac{1}{\mathcal{Z}_E} \int dX_E e^{-\beta[\mathcal{H}_E(X_E) + \mathcal{V}_{SE}(X_S, X_E, C)]}. \quad (65)$$

For the conventional FEP setting, the interaction is of the form $\mathcal{V}_{SE}(X_S, X_E, C(t)) = C(t)\mathcal{U}_{SE}(X_S, X_E)$ with $C(0) = 0$. At $t = 0$ the composite is uncoupled, $\mathcal{V}_{SE} = 0$, so the integral reproduces the environment partition function. Consequently,

$$\mathcal{M}(X_S, C(0) = 0, \beta) = \frac{1}{\mathcal{Z}_E} \int dX_E e^{-\beta\mathcal{H}_E(X_E)} = 1. \quad (66)$$

In the frozen-driving regime, Corollary 1 gives the pair of endpoint relations

$$e^{-\beta\Delta F_S^*(\beta)} = \frac{\langle \mathcal{M}(X_S, C(t_{eq}), \beta) \rangle_S}{1 + \chi^2(P_{t_{eq}}^S \parallel P_0^S)} \quad (67)$$

and

$$e^{+\beta\Delta F_S^*(\beta)} = \langle \mathcal{M}(X_S, C(t_{eq}), \beta)^{-1} \rangle_S. \quad (68)$$

The frozen-driving trajectory identities in Corollary 2 simplify in the same way. Under the asymptotic-equilibration postulate, the corresponding trajectory representations of Eqs. (67) and (68) become

$$e^{-\beta\Delta F_S^*(\beta)} = \frac{\langle \mathcal{M}(\mathcal{T}_{t_{eq}}^S(X_0), C(t_{eq}), \beta) \rangle_{X_0}}{1 + \chi^2(P_{t_{eq}}^S \parallel P_0^S)}, \quad (69)$$

$$e^{+\beta\Delta F_S^*(\beta)} = \langle \mathcal{M}(\mathcal{T}_{t_{eq}}^S(X_0), C(t_{eq}), \beta)^{-1} \rangle_{X_0}, \quad (70)$$

where the averages are taken over initial conditions $X_0 \sim P_0(X, \beta)$. Equations (69) and (70) therefore provide trajectory representations of the FEP-type endpoint reweighting identities. The interaction enters through $\mathcal{M}$ evaluated at the dynamically reached endpoint configurations $\mathcal{T}_{t_{eq}}^S(X_0)$, while the overlap factor retains its static definition via the chi-squared divergence of the endpoint marginals. Embedding the standard noninteracting reference into this structure clarifies how the usual FEP expression sits inside the HMF endpoint reweighting identities. First, the role of phase-space overlap, which in standard FEP is implicit but remains a serious numerical concern, appears here as an explicit multiplicative factor involving $\chi^2(P_{t_{eq}}^S \parallel P_0^S)$ that directly diagnoses when estimators are expected to be unreliable. Second, the free energy difference can be expressed as an average over equilibrated trajectory endpoints generated by a chosen ramp-relax protocol. Taken together, Eqs. (67)–(70) recast FEP-type reweighting in a form that keeps endpoint overlap explicit, while keeping phase-space overlap and strong-coupling effects explicitly visible at the level of the HMF marginals.

VII. VALIDATION

A. Validation setup for the frozen-driving protocol

We now validate the trajectory representation of the endpoint reweighting identity in Eq. (69) (see Fig. 2). The test is organized in the frozen-driving regime of Sec. IV, with a noninteracting reference state as in Eq. (50). This setup allows

a direct comparison between the Zwanzig identity (63) and its overlap-explicit HMF endpoint form (69).

Model specification. We consider a one-dimensional system $\mathcal{S}$ in a quartic double-well potential [5,52–55] coupled linearly to a harmonic environment $\mathcal{E}$. The bare Hamiltonians are

$$\mathcal{H}_{\mathcal{S}}(x, p_x; \lambda) = \frac{p_x^2}{2m} + U_{\mathcal{S}}(x; \lambda), \quad (71)$$

with

$$U_{\mathcal{S}}(x; \lambda) = \frac{1}{4}(x^2 - \lambda)^2 \quad (72)$$

and

$$\mathcal{H}_{\mathcal{E}}(y, p_y) = \frac{p_y^2}{2m} + \frac{1}{2}\omega^2 y^2. \quad (73)$$

The interaction channel is

$$\mathcal{V}_{\mathcal{SE}}(x, y; C) = Cxy. \quad (74)$$

The total Hamiltonian follows Definition 2, and we choose reduced units with $m = 1$.

Derivation of the HMF and system marginal distribution. We now derive the explicit HMF for the system. For fixed (λ_0, C) , the full Hamiltonian reads

$$\mathcal{H}_{\mathcal{S}+\mathcal{E}}(x, p_x, y, p_y; \lambda_0, C) = \mathcal{H}_{\mathcal{S}}(x, p_x; \lambda_0) + \mathcal{H}_{\mathcal{E}}(y, p_y) + \mathcal{V}_{\mathcal{SE}}(x, y; C), \quad (75)$$

with $\mathcal{H}_{\mathcal{S}}$, $\mathcal{H}_{\mathcal{E}}$, and $\mathcal{V}_{\mathcal{SE}}$ given by Eqs. (71)–(74). According to the HMF definition in Eq. (11) and restricting to the configurational part (the momentum factors are independent of x and can be absorbed into $\mathcal{Z}_{\mathcal{S}}^*$), we can write

$$e^{-\beta \mathcal{H}_{\beta}^*(x; \lambda_0, C)} = e^{-\beta U_{\mathcal{S}}(x; \lambda_0)} \underbrace{\frac{1}{\mathcal{Z}_{\mathcal{E}}(\beta)} \int dy dp_y \exp\left(-\beta\left(\frac{p_y^2}{2} + \frac{1}{2}\omega^2 y^2 + Cxy\right)\right)}_A. \quad (76)$$

The environment partition function in the normalization of Eq. (76) factorizes into momentum and configuration contributions, as in Eq. (12),

$$\mathcal{Z}_{\mathcal{E}}(\beta) = \int dp_y e^{-\beta p_y^2/2} \int dy e^{-\beta \omega^2 y^2/2}. \quad (77)$$

The p_y integral cancels between numerator and denominator. Introducing the configurational environment partition function

$$\mathcal{Z}_{\mathcal{E}}^{(y)}(\beta) = \int dy \exp\left(-\frac{\beta \omega^2 y^2}{2}\right), \quad (78)$$

we can write

$$A = \frac{1}{\mathcal{Z}_{\mathcal{E}}^{(y)}(\beta)} \int dy \exp\left[-\beta\left(\frac{1}{2}\omega^2 y^2 + Cxy\right)\right]. \quad (79)$$

To evaluate Eq. (79), we complete the square in y . The quadratic form in the exponent can be written as

$$\begin{aligned} \frac{1}{2}\omega^2 y^2 + Cxy &= \frac{1}{2}\omega^2 \left(y^2 + 2\frac{Cx}{\omega^2}y\right) \\ &= \frac{1}{2}\omega^2 ((y+a)^2 - a^2), \end{aligned} \quad (80)$$

where $a \equiv \frac{Cx}{\omega^2}$. Hence, Eq. (79) gives

$$A = \exp\left(\frac{\beta C^2 x^2}{2\omega^2}\right) \frac{1}{\mathcal{Z}_{\mathcal{E}}^{(y)}(\beta)} \int dy \exp\left(-\beta \frac{1}{2}\omega^2 (y+a)^2\right). \quad (81)$$

The integral in Eq. (81) is invariant under the shift $y \mapsto y - a$ and therefore equals $\mathcal{Z}_{\mathcal{E}}^{(y)}(\beta)$ defined in Eq. (78). We thus obtain

$$A = \exp\left(\frac{\beta C^2 x^2}{2\omega^2}\right). \quad (82)$$

Inserting Eq. (82) into Eq. (76) yields

$$e^{-\beta \mathcal{H}_{\beta}^*(x; \lambda_0, C)} = \exp\left[-\beta\left(U_{\mathcal{S}}(x; \lambda_0) - \frac{C^2 x^2}{2\omega^2}\right)\right]. \quad (83)$$

Taking the logarithm gives the HMF

$$\mathcal{H}_{\beta}^*(x; \lambda_0, C) = U_{\mathcal{S}}(x; \lambda_0) - \frac{C^2 x^2}{2\omega^2}. \quad (84)$$

Inserting Eq. (84) into Eq. (13) and Eq. (14) provides the marginals $P_0^{\mathcal{S}}$ and $P_{t_{\text{eq}}}^{\mathcal{S}}$ at the endpoints $(\lambda(0), C(0))$ and $(\lambda(t_{\text{eq}}), C(t_{\text{eq}}))$.

Frozen-driving protocol and dynamics. We adopt the frozen-driving viewpoint of Sec. IV by fixing

$$\lambda(t) \equiv \lambda_0, \quad C(0) = 0, \quad C(t_{\text{eq}}) = C_f, \quad (85)$$

so that the system is transported between equilibrium states only by changing the coupling. During a ramp stage $0 \leq t \leq t_{\text{ramp}}$, the coupling increases linearly from 0 to $C(t_{\text{eq}})$,

$$C(t) = C(t_{\text{eq}}) \frac{t}{t_{\text{ramp}}}, \quad 0 \leq t \leq t_{\text{ramp}}, \quad (86)$$

followed by a relaxation stage

$$C(t) \equiv C(t_{\text{eq}}), \quad t_{\text{ramp}} \leq t \leq t_{\text{eq}}. \quad (87)$$

The resulting coupling schedule and relaxation window used in the simulations are illustrated in Fig. 2(b). The composite evolves under overdamped Langevin dynamics [56] consistent with the canonical measures in Eqs. (8) and (9). Writing the total potential as

$$U_{\text{tot}}(x, y; C) = U_{\mathcal{S}}(x; \lambda_0) + \frac{1}{2}\omega^2 y^2 + Cxy, \quad (88)$$

the equations of motion are

$$\dot{x} = -\mu \partial_x U_{\text{tot}}(x, y; C(t)) + \sqrt{\frac{2\mu}{\beta}} \xi_x(t), \quad (89)$$

$$\dot{y} = -\mu \partial_y U_{\text{tot}}(x, y; C(t)) + \sqrt{\frac{2\mu}{\beta}} \xi_y(t), \quad (90)$$

where μ is a mobility [57] and $\xi_i(t)$ are independent unit white noises. For each fixed pair (λ_0, C) this dynamics satisfies the fluctuation-dissipation relation and converges to the equilibrium state associated with $\mathcal{H}_{S+\mathcal{E}}(x, y; \lambda_0, C)$. The corresponding endpoint system marginals for different final couplings are shown in Fig. 2(c). Choosing t_{eq} large compared to the relaxation time enforces the asymptotic equilibration condition in Eq. (32), so that the trajectory map $\mathcal{T}_{t_{\text{eq}}}$ realizes the endpoint marginal $P_{t_{\text{eq}}}^S$ in Eq. (16). Initial conditions $X_0 = (x_0, y_0)$ are sampled from the uncoupled composite equilibrium in Eq. (8) with $(\lambda(0), C(0)) = (\lambda_0, 0)$. At $C = 0$ the composite density factorizes, so we draw

$$x_0 \sim \frac{1}{\mathcal{Z}_S(\lambda_0, \beta)} \exp(-\beta U_S(x; \lambda_0)) \quad (91)$$

and

$$y_0 \sim \frac{1}{\mathcal{Z}_\mathcal{E}(\beta)} \exp\left(-\frac{\beta \omega^2 y^2}{2}\right), \quad (92)$$

which realizes the noninteracting reference of Eq. (50), where the Zwanzig identity (63) applies.

Estimators for $\Delta F_S^*(\beta)$. For each final coupling $C(t_{\text{eq}})$ we compute the open-system free energy difference

$$\Delta F_S^*(\beta) = F_S^*(\lambda_0, C(t_{\text{eq}}), \beta) - F_S^*(\lambda_0, 0, \beta), \quad (93)$$

using three constructions.

(1) *Exact HMF reference.* Using the central identity in Eq. (44), we evaluate the HMF partition functions $\mathcal{Z}_S^*(\lambda_0, C, \beta)$ by numerical quadrature of Eq. (13) with the model HMF in Eq. (84).

(2) *Zwanzig FEP.* We evaluate the interaction $\mathcal{V}_{S\mathcal{E}}(X_S, X_\mathcal{E})$ on the uncoupled configurations $X_0 \sim P_0(X, \beta)$ and estimate

$$\langle e^{-\beta \mathcal{V}_{S\mathcal{E}}(X_S, X_\mathcal{E}; C(t_{\text{eq}}))} \rangle_0. \quad (94)$$

This provides a Zwanzig estimator, Eq. (64), for calculating the free energy differences.

(3) *Trajectory representation with explicit overlap factor.* The main test concerns the trajectory representation of the FEP-type endpoint identity, Eq. (69). We evaluate $\mathcal{M}$ at the endpoint positions $x_{\text{eq}} = \mathcal{T}_{t_{\text{eq}}}^S(X_0)$ for each trajectory. Subsequently, we compute the sample average $\langle \mathcal{M}(\mathcal{T}_{t_{\text{eq}}}^S(X_0), C(t_{\text{eq}}), \beta) \rangle_{X_0}$ and divide it by the overlap factor $1 + \chi^2(P_{t_{\text{eq}}}^S \parallel P_0^S)$, obtained from Eq. (26). The exact HMF reference, the Zwanzig estimator, and the trajectory representation of the endpoint estimator are compared as a function of the final coupling in Fig. 2(a). For completeness, the numerical values of all simulation parameters used in the validation of Eq. (69) are reported in Table I.

TABLE I. Numerical inputs for the validation runs (frozen driving; overdamped Langevin dynamics with linear coupling ramp). All quantities are in reduced, dimensionless units.

Quantity	Value
Temperature β^{-1}/k_B	1.0
Double-well control parameter λ_0	4.0
Environment frequency ω	1.0
Mobility μ	1.0
Initial coupling C_0	0.0
Ramp duration t_{ramp}	10.0
Relaxation duration t_{relax}	10.0
Total protocol time t_{eq}	20.0
Time step Δt	10^{-2}
Trajectory ensemble size N_{traj}	40000
FEP ensemble size N_{fep}	40000

B. Sample-based overlap reconstruction and multistage comparison

In the validation above, the estimator in Eq. (69) was obtained in a mixed way: The trajectory average in the numerator was estimated from sampled trajectories, whereas the denominator was evaluated from the endpoint distributions through Eq. (26). This is sufficient to verify the exact identity in the present model, but it is not the appropriate practical finite-sampling test because it uses analytic information in the denominator. The relevant practical question is therefore whether the overlap factor can be reconstructed from the same sampled information that enters the endpoint estimator. That question becomes even more difficult in many-particle systems, where the endpoint distributions are rarely available in explicit form. The present framework is useful precisely because it turns this issue into a concrete statistical problem. We therefore reformulate the estimator so that every quantity is obtained from samples, while the exact HMF free energy is retained only as a reference curve.

Direct sampled reconstruction of the overlap factor. We first consider the one-step frozen-driving protocol connecting the uncoupled reference state to the final coupled state. Initial conditions are drawn from the equilibrium distribution of the initial composite and propagated under overdamped Langevin dynamics up to t_{eq} . The resulting system endpoints are

$$x_{\text{eq}}^{(i)} = \mathcal{T}_{t_{\text{eq}}}^S(X_0^{(i)}). \quad (95)$$

These propagated endpoints are then inserted directly into the frozen-driving trajectory equalities. Combining Eqs. (69) and (70) gives

$$1 + \chi^2(P_{t_{\text{eq}}}^S \parallel P_0^S) = \langle \mathcal{M}(\mathcal{T}_{t_{\text{eq}}}^S(X_0), C_f, \beta) \rangle_{X_0} \times \langle \mathcal{M}(\mathcal{T}_{t_{\text{eq}}}^S(X_0), C_f, \beta)^{-1} \rangle_{X_0}. \quad (96)$$

Equation (96) shows that the overlap factor can be written entirely in terms of trajectory averages, without direct evaluation of the endpoint distributions. For a finite sample of propagated trajectories, we therefore define

$$\hat{A}(C_f) = \frac{1}{N} \sum_{i=1}^N \mathcal{M}(x_{\text{eq}}^{(i)}, C_f, \beta) \quad (97)$$

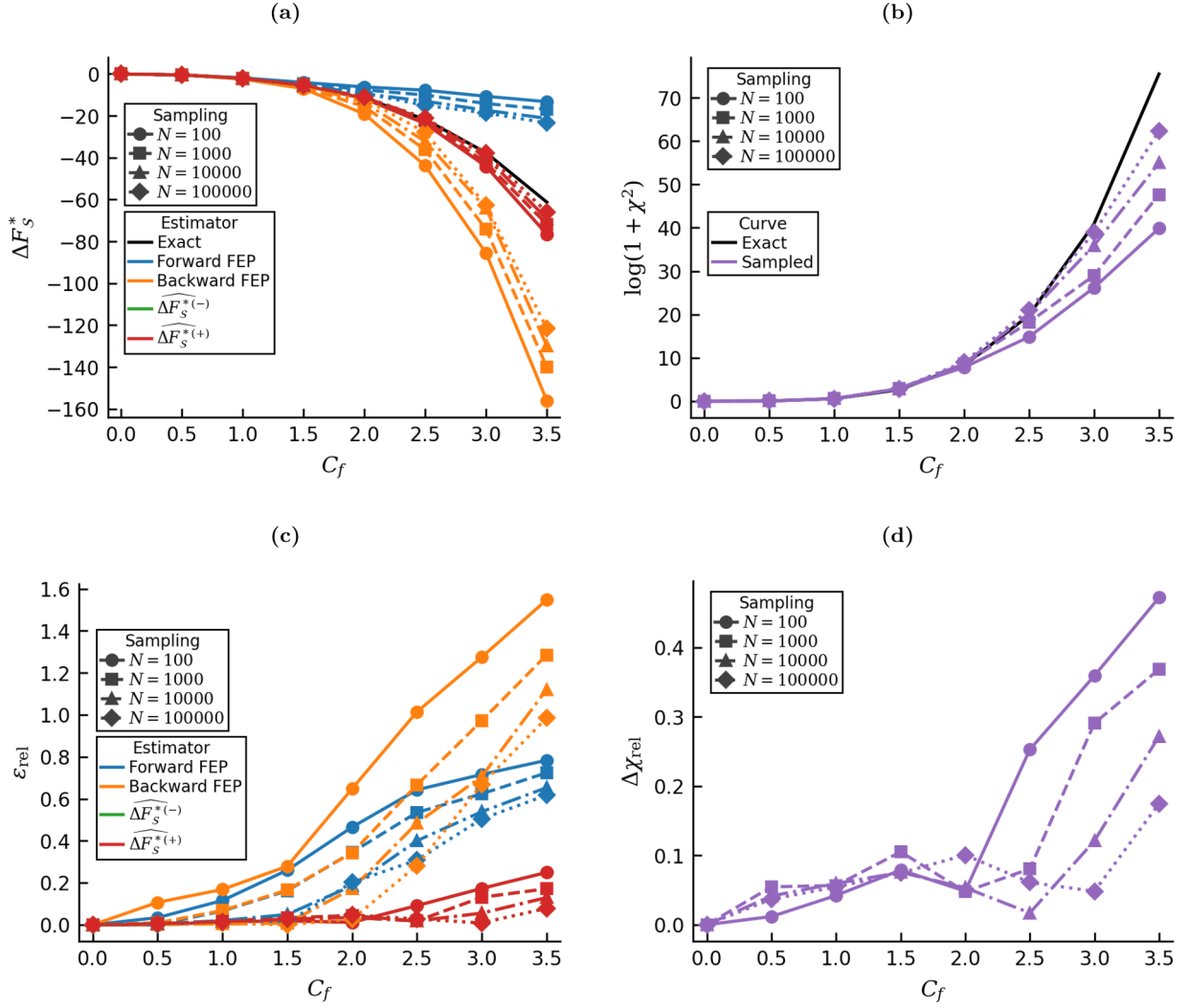


FIG. 3. Finite-sampling validation of the one-step frozen-driving formulation. (a) ΔF_S^* as a function of C_f . The solid black line shows the exact HMF reference. Blue and orange curves show forward and backward Zwanzig FEP. Green and red curves show Eqs. (100) and (101) and are practically indistinguishable. Different symbols and line styles correspond to different sample sizes N . (b) Exact and sampled $\log(1 + \chi^2)$ as a function of C_f , with sampled values obtained from Eq. (99). (c) Relative free energy error, ϵ_{rel} , for the curves in panel (a). (d) Relative error of the sampled overlap factor, $\Delta\chi_{\text{rel}}$, obtained from the data in panel (b). The numerical values of the model, protocol, and sampling parameters used for the one-step finite-sampling validation are listed in Table II.

and

$$\widehat{B}(C_f) = \frac{1}{N} \sum_{i=1}^N \mathcal{M}(x_{\text{eq}}^{(i)}, C_f, \beta)^{-1}. \quad (98)$$

The sampled overlap factor then follows as

$$1 + \chi^2(C_f) = \widehat{A}(C_f) \widehat{B}(C_f). \quad (99)$$

The corresponding finite-sample estimators are

$$\widehat{\Delta F_S^{*(-)}}(C_f) = -\frac{1}{\beta} \ln \left[\frac{\widehat{A}(C_f)}{1 + \chi^2(C_f)} \right] \quad (100)$$

and

$$\widehat{\Delta F_S^{*(+)}}(C_f) = \frac{1}{\beta} \ln \widehat{B}(C_f). \quad (101)$$

The numerical behavior of this one-step sampled formulation is summarized in Fig. 3. Figure 3(a) shows the free energy

estimates as functions of C_f for several sample sizes. Equations (100) (green) and (101) (red) track the exact HMF reference across the tested coupling range for the sample sizes shown, whereas forward and backward Zwanzig FEP develop the expected deviations as the final state moves away from the initial reference. The deterioration is strongly asymmetric: forward FEP becomes biased toward less negative values, whereas backward FEP overshoots in the opposite direction and fails more severely at large C_f . This is the usual signature of poor overlap in exponential reweighting. Figure 3(b) shows the sampled reconstruction of $\log(1 + \chi^2)$. At weak and intermediate coupling, the sampled quantity follows the exact trend closely even for moderate sample sizes. At stronger coupling, it begins to underestimate the exact curve when N is small. This is the expected finite-sampling regime: once the endpoint distributions become well separated, the overlap factor is controlled by rare events and is therefore resolved only with sufficiently large samples. This becomes clearer

when Figs. 3(c) and 3(d) are read together. Figure 3(c) shows the relative free energy error, and Fig. 3(d) shows the relative error of the sampled overlap factor. Both deteriorate in the same coupling regime. The overlap factor is therefore not an auxiliary quantity appended after the free energy estimate is produced. It is the quantity that makes the statistical source of the degradation explicit. In standard FEP the same difficulty is present, but it remains hidden inside the exponential average. Here it becomes directly observable.

A useful point follows from the relation between Eqs. (100), (101), and (96). Equation (100) displays the overlap factor explicitly through Eq. (99), whereas in Eq. (101) the same information is carried implicitly by the inverse average. Equation (96) makes this connection exact. This yields a useful numerical viewpoint: One may evaluate the free energy through Eq. (101) and use Eq. (99) in parallel to monitor the statistical quality of the path. Large values of $1 + \chi^2$ identify poor overlap and therefore indicate where the protocol should be refined, shortened, or redistributed. The framework thus provides a free energy estimator, together with a diagnostic criterion for identifying where the calculation would benefit from additional intermediate states or a refined path. A further point should be noted. If the denominator in Eq. (100) is replaced by the product estimator in Eq. (99), the result reduces identically to Eq. (101). In that implementation, the overlap factor is already encoded in Eq. (101), and the two sampled estimators coincide. This coincidence is specific to the choice of reconstructing the denominator from the same trajectory sample. It does not remove the independent numerical role of Eq. (100). Indeed, with $\widehat{D}(C_f) = \widehat{1 + \chi^2}(C_f)$, Eq. (69) implies that

$$\widehat{\Delta F_{S,D}^*}^{(-)}(C_f) = -\frac{1}{\beta} \ln \left[\frac{\widehat{A}(C_f)}{\widehat{D}(C_f)} \right] \quad (102)$$

defines a distinct estimator whenever the overlap factor is supplied by an alternative approximation $\widehat{D}(C_f)$ that is not constrained to equal $\widehat{A}(C_f)\widehat{B}(C_f)$. In that case, Eqs. (100) and (101) become numerically different realizations of the same exact identity. The former requires an external estimate of the divergence, whereas the latter carries the overlap information implicitly through the inverse trajectory average. This opens a broader numerical perspective for higher-dimensional systems, where direct access to the full endpoint distributions is usually impossible. One route is to estimate the ratio $(P_{\text{eq}}^S)^2/P_0^S$ directly from samples by density-ratio methods [58,59], such as classifier-based estimators, and then reconstruct $1 + \chi^2$ from that ratio. A second route is to estimate the endpoint densities in a reduced representation, for example in terms of dominant collective variables [60,61] or slow modes [62], and evaluate the divergence there. A third route is to employ flexible parametric surrogates for the endpoint marginals and evaluate the divergence from the learned representation [63–65].

Recursive implementation along a coupling ladder. We next consider a multistage formulation. Instead of a single jump from the initial to the final coupling, we introduce the ladder

$$0 = C_0 < C_1 < \dots < C_K \quad (103)$$

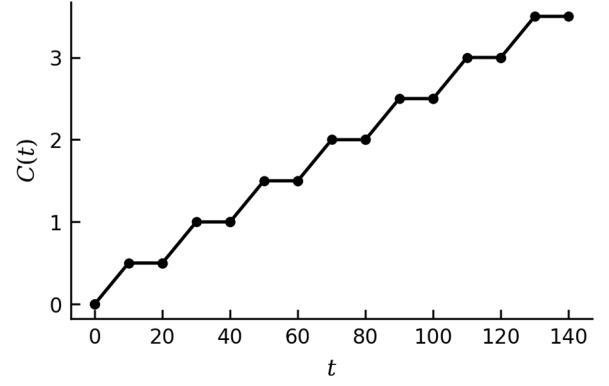


FIG. 4. Recursive frozen-driving protocol along the ladder $0 = C_0 < C_1 < \dots < C_K$. Each segment consists of a linear ramp from C_{k-1} to C_k followed by a relaxation stage at fixed C_k .

and treat each neighboring pair (C_{k-1}, C_k) as a local frozen-driving segment. As illustrated schematically in Fig. 4, the recursive protocol resolves the full coupling transformation into a sequence of local ramp–relax segments. For each segment k , initial composite configurations are sampled from the equilibrium distribution at coupling C_{k-1} , $X_{k-1} \sim P_{C_{k-1}}(X, \beta)$, and are then propagated under the local ramp–relax protocol from C_{k-1} to C_k . In each segment, the coupling is increased linearly from C_{k-1} to C_k , followed by relaxation at fixed C_k , so that the equilibrated endpoint distribution of one segment provides the reference distribution for the next. The total free energy change can be written as

$$\Delta F_S^*(C_f = C_K) = \sum_{k=1}^K \Delta F_{S,k}^*. \quad (104)$$

In the multistage setting, the local reference at step k is no longer an uncoupled state. Each segment connects two already coupled equilibrium states, C_{k-1} and C_k . The correct local relations must therefore be taken from Eqs. (42) and (43), rather than from the special uncoupled case. For the step $C_{k-1} \rightarrow C_k$, this gives

$$e^{-\beta \Delta F_{S,k}^*} = \frac{\left\langle \frac{\mathcal{M}(\mathcal{T}_{\text{eq}}^S(X_{k-1}), C_k, \beta)}{\mathcal{M}(\mathcal{T}_{\text{eq}}^S(X_{k-1}), C_{k-1}, \beta)} \right\rangle_{X_{k-1}}}{1 + \chi_k^2} \quad (105)$$

and

$$e^{+\beta \Delta F_{S,k}^*} = \left\langle \frac{\mathcal{M}(\mathcal{T}_{\text{eq}}^S(X_{k-1}), C_{k-1}, \beta)}{\mathcal{M}(\mathcal{T}_{\text{eq}}^S(X_{k-1}), C_k, \beta)} \right\rangle_{X_{k-1}}. \quad (106)$$

Multiplying Eqs. (105) and (106) yields

$$1 + \chi_k^2 = \left\langle \frac{\mathcal{M}(\mathcal{T}_{\text{eq}}^S(X_{k-1}), C_k, \beta)}{\mathcal{M}(\mathcal{T}_{\text{eq}}^S(X_{k-1}), C_{k-1}, \beta)} \right\rangle_{X_{k-1}} \times \left\langle \frac{\mathcal{M}(\mathcal{T}_{\text{eq}}^S(X_{k-1}), C_{k-1}, \beta)}{\mathcal{M}(\mathcal{T}_{\text{eq}}^S(X_{k-1}), C_k, \beta)} \right\rangle_{X_{k-1}}. \quad (107)$$

Equation (107) is the local counterpart of Eq. (96). It shows that the global overlap problem is decomposed into a sequence of local overlap factors. The corresponding sampled averages

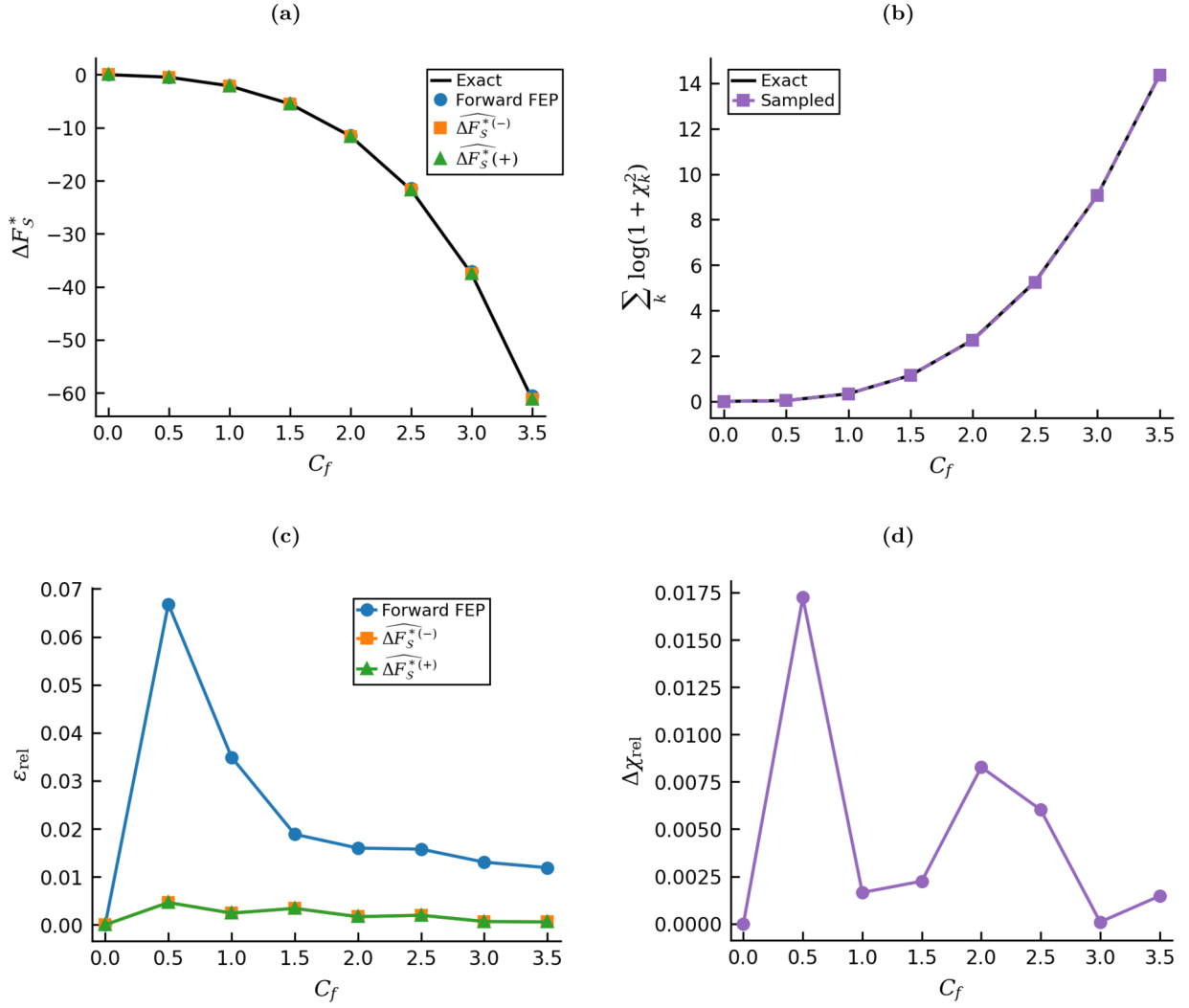


FIG. 5. Recursive implementation along the coupling ladder. (a) Cumulative ΔF_S^* as a function of C_f . The solid black line shows the exact HMF reference. Blue circles show recursive forward FEP. Orange squares and green triangles show Eqs. (113) and (114). (b) Exact and sampled cumulative overlap quantity, Eqs. (115) and (116). (c) Relative free energy error, ε_{rel} , for the curves in panel (a). (d) Relative error of the cumulative overlap quantity, $\Delta \chi_{\text{rel}}$, obtained from the data in panel (b). The numerical values of the model, ladder, protocol, and sampling parameters used for the recursive multistage calculations are summarized in Table III.

are

$$\hat{A}_k = \frac{1}{N} \sum_{i=1}^N \frac{\mathcal{M}(x_{k-1,\text{eq}}^{(i)}, C_k, \beta)}{\mathcal{M}(x_{k-1,\text{eq}}^{(i)}, C_{k-1}, \beta)} \quad (108)$$

and

$$\hat{B}_k = \frac{1}{N} \sum_{i=1}^N \frac{\mathcal{M}(x_{k-1,\text{eq}}^{(i)}, C_{k-1}, \beta)}{\mathcal{M}(x_{k-1,\text{eq}}^{(i)}, C_k, \beta)}, \quad (109)$$

from which the sampled local overlap factor follows as

$$\widehat{1 + \chi_k^2} = \hat{A}_k \hat{B}_k. \quad (110)$$

The local estimators are then

$$\widehat{\Delta F_{S,k}^{*(-)}} = -\frac{1}{\beta} \ln \left[\frac{\hat{A}_k}{\widehat{1 + \chi_k^2}} \right] \quad (111)$$

and

$$\widehat{\Delta F_{S,k}^{*(+)}} = \frac{1}{\beta} \ln \hat{B}_k. \quad (112)$$

The cumulative sampled free energy estimates along the ladder are therefore

$$\widehat{\Delta F_S^{*(-)}}(C_f = C_K) = \sum_{k=1}^K \widehat{\Delta F_{S,k}^{*(-)}} \quad (113)$$

and

$$\widehat{\Delta F_S^{*(+)}}(C_f = C_K) = \sum_{k=1}^K \widehat{\Delta F_{S,k}^{*(+)}}. \quad (114)$$

At the same time, the overlap burden accumulates additively as

$$\log \left[\prod_{k=1}^K (1 + \chi_k^2) \right] = \sum_{k=1}^K \log (1 + \chi_k^2), \quad (115)$$

with sampled counterpart

$$\sum_{k=1}^K \log(\widehat{1 + \chi_k^2}). \quad (116)$$

The recursive results are shown in Figs. 5 and 6. Figure 5(a) shows the cumulative estimates in Eqs. (113) and (114) together with the exact HMF reference. Both cumulative sampled forms remain essentially locked to the exact curve throughout the ladder. Recursive forward FEP also performs well in this setting, as expected for a sufficiently fine sequence of well-overlapped local steps, and the proposed estimators show comparable accuracy over the tested range. This is consistent with the logic of the ladder construction: the global transformation is replaced by a sequence of milder local estimations, each with a controlled overlap burden. This interpretation is confirmed by Fig. 5(b), which shows the cumulative overlap quantity in Eqs. (115) and (116). The sampled and exact curves are practically indistinguishable. The same conclusion appears in the error diagnostics. Figure 5(c) shows that the relative free energy error remains small along the ladder, while Fig. 5(d) shows that the error in the cumulative overlap factor also remains controlled. The point is important: once the transformation is decomposed into local equilibrated steps, the overlap factor becomes far easier to estimate than in the one-step calculation. The overlap structure identified by the theory therefore remains quantitative in the multistage regime. This is also where the present formulation differs from standard multistage FEP. A multistage FEP construction improves numerical stability by reducing the spacing between neighboring states, but it does

not by itself report how the statistical burden is distributed along the path. Here that information remains explicit through the local factors in Eq. (107) and the cumulative measure in Eq. (115). The method therefore yields a recursive estimate of the free energy together with an explicit overlap diagnostic. It identifies where the path is statistically easy and where it is costly. That feature is not explicit in ordinary multistage FEP, where the overlap problem is alleviated by construction but not monitored as an observable. This point also suggests a practical use. Because the local factors reveal where overlap is lost, they can serve as a guide for refining the ladder itself in an adaptive scheme. Large values of $1 + \chi_k^2$ indicate those segments where additional intermediate states are most valuable, whereas uniformly small values indicate that the path is already well resolved. The recursive formulation therefore supplies both an estimator and a criterion for improving the path on which that estimator is built. The numerical values of the model, protocol, and sampling parameters used for the one-step finite-sampling validation are listed in Table II.

Comparison with BAR and MBAR. The recursive ladder also provides the natural setting for comparison with standard equilibrium estimators. For this purpose, BAR and MBAR are applied to equilibrium samples of the composite system $\mathcal{S} + \mathcal{E}$ generated independently at the ladder states $C_0, \dots, C_K$. For the k th segment, BAR is applied to the two neighboring composite equilibrium states at C_{k-1} and C_k . Defining the reduced potentials

$$u_{k-1}(X) = \beta \mathcal{H}_{\mathcal{S}+\mathcal{E}}(X, \lambda_0, C_{k-1}) \quad (117)$$

and

$$u_k(X) = \beta \mathcal{H}_{\mathcal{S}+\mathcal{E}}(X, \lambda_0, C_k), \quad (118)$$

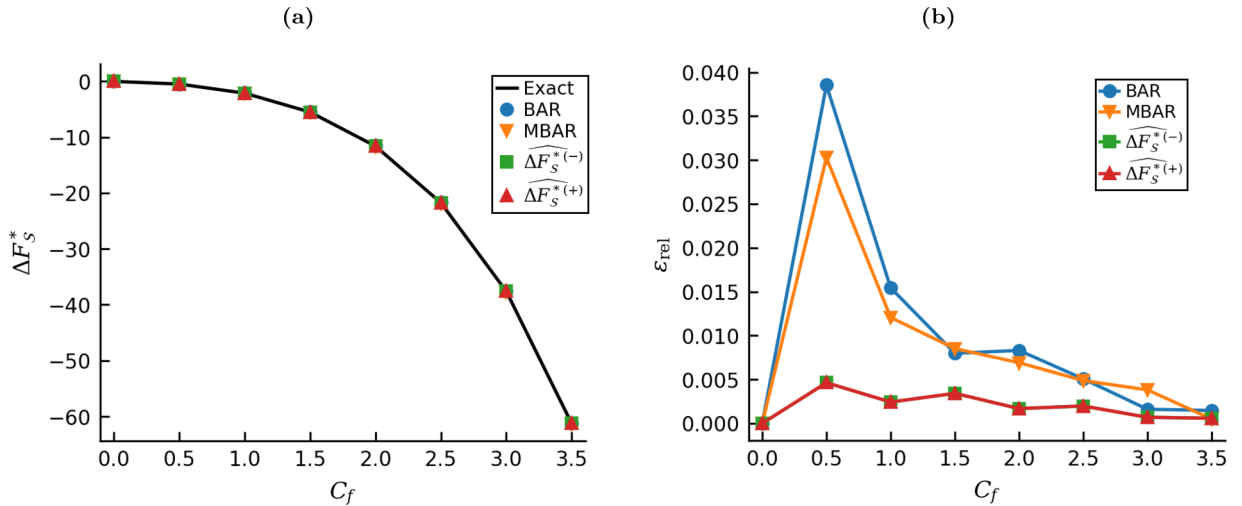


FIG. 6. Comparison along the coupling ladder. (a) Cumulative ΔF_S^* as a function of C_f . The solid black line shows the exact HMF reference. Blue circles and orange inverted triangles show BAR and MBAR. Green squares and red triangles show Eqs. (113) and (114). (b) Relative free energy error for the same curves. The numerical values of the model, ladder, protocol, and sampling parameters used for the recursive multistage calculations are summarized in Table III. The comparison is intended to position the overlap-diagnostic estimator relative to standard baselines, not to claim a universal numerical advantage over BAR or MBAR.

TABLE II. Numerical inputs for the one-step finite-sampling validation in Fig. 3. All quantities are given in reduced, dimensionless units.

Quantity	Value
Temperature β^{-1}/k_B	1.0
Double-well control parameter λ_0	4.0
Environment frequency ω	1.0
Mobility μ	1.0
Initial coupling $C(0)$	0.0
Ramp duration t_{ramp}	10.0
Total protocol time t_{eq}	20.0
Relaxation duration $t_{\text{eq}} - t_{\text{ramp}}$	10.0
Time step Δt	10^{-2}
Sample sizes N	$\{10^2, 10^3, 10^4, 10^5\}$
Quadrature tolerance for exact reference	10^{-4}

the local BAR free energy increment $\Delta f_k = \beta \Delta F_k$ is obtained from Ref. [31]

$$\begin{aligned} & \frac{1}{N_{k-1}} \sum_{n=1}^{N_{k-1}} \frac{1}{1 + \exp[u_k(X_{k-1}^{(n)}) - u_{k-1}(X_{k-1}^{(n)}) - \Delta f_k - c_k]} \\ &= \frac{1}{N_k} \sum_{m=1}^{N_k} \frac{1}{1 + \exp[u_{k-1}(X_k^{(m)}) - u_k(X_k^{(m)}) + \Delta f_k + c_k]}, \end{aligned} \quad (119)$$

with

$$c_k = \ln\left(\frac{N_k}{N_{k-1}}\right). \quad (120)$$

The cumulative BAR estimate is then

$$\Delta F_{\text{BAR}}(C_f = C_K) = \sum_{k=1}^K \Delta F_k^{\text{BAR}}. \quad (121)$$

MBAR [29] uses the equilibrium samples from all ladder states at once. Pooling the configurations from every state, the dimensionless free energies $f_k = \beta F_k$ satisfy

$$f_k = -\ln \sum_{n=1}^{N_{\text{tot}}} \frac{\exp[-u_k(X_n)]}{\sum_{\ell=0}^K N_{\ell} \exp[f_{\ell} - u_{\ell}(X_n)]}, \quad (122)$$

with reference

$$f_0 = 0. \quad (123)$$

The cumulative MBAR estimate is therefore

$$\Delta F_{\text{MBAR}}(C_f = C_K) = \frac{f_K - f_0}{\beta}. \quad (124)$$

The comparison with these equilibrium baselines is shown in Fig. 6. Figure 6(a) shows that recursive BAR, ladder MBAR, and the recursive estimators in Eqs. (113) and (114) all reproduce the exact HMF free energy curve with high accuracy over the full range of C_f . This agreement is expected for the sufficiently fine ladder construction chosen here, where neighboring equilibrium states remain well overlapped. The more relevant distinction is therefore structural rather than numerical. BAR and MBAR are built from equilibrium endpoint samples of the composite system alone. By contrast, the

TABLE III. Numerical inputs for the recursive multistage validation in Figs. 5 and 6. All quantities are given in reduced, dimensionless units.

Quantity	Value
Temperature β^{-1}/k_B	1.0
Double-well control parameter λ_0	4.0
Environment frequency ω	1.0
Mobility μ	1.0
Number of ladder states	8
Local ramp duration t_{ramp}	10.0
Local total segment time t_{eq}	20.0
Local relaxation duration $t_{\text{eq}} - t_{\text{ramp}}$	10.0
Time step Δt	10^{-2}
Quadrature tolerance for exact reference	10^{-4}
Sample size for recursive FEP comparison	$N = 1000$
Sample size for BAR/MBAR comparison	$N = 1000$

present recursive formulation is tied to the trajectory identities of Eqs. (69) and (70), applied locally through Eqs. (105) and (106). The estimator can therefore be constructed from trajectory samples generated by the ramp-relax protocol rather than only from independently sampled equilibrium endpoints. At the same time, the overlap burden remains explicit through the local factors in Eq. (107) and the cumulative quantity in Eq. (115). BAR and MBAR also rely on overlap, but they do not expose the same local χ^2 factors. The central point is therefore not a universal numerical improvement over BAR or MBAR, but the explicit diagnostic structure of the endpoint formulation. This is the main added value of the multistage formulation. Standard multistage FEP, BAR, and MBAR improve stability by replacing a difficult global transformation with a sequence of better-overlapped local steps. The present approach does the same within a framework that also preserves the trajectory representation and keeps the overlap structure visible at every stage through Eqs. (107) and (115). The numerical values of the model, ladder, protocol, and sampling parameters used for the recursive multistage calculations are summarized in Table III.

VIII. CONCLUSIONS AND OUTLOOK

We have presented an endpoint reweighting framework for classical open systems that expresses HMF free energy differences through canonical endpoint marginals and an explicit overlap diagnostic. A central feature of this framework is the systematic use of the HMF, which provides the correct thermodynamic potential for a system that interacts with its environment at finite strength. The necessity of using the HMF in place of the bare system Hamiltonian is a central theme in modern thermodynamics [38], and the present work builds directly upon this principle. The main result should therefore be read as a conceptual and diagnostic reformulation of HMF free energy reweighting. It makes the endpoint overlap burden explicit under canonical endpoints and, in the trajectory form, under the asymptotic-equilibration postulate. It does not define a new forward-reverse fluctuation theorem and does not remove the usual overlap limitations of exponential reweighting. Our starting point is a fully gen-

eral description of the composite $\mathcal{S} + \mathcal{E}$, formulated through an abstract trajectory map $\mathcal{T}_t$ that does not presuppose any constraints on dynamics. Within this setting, the canonical endpoints of the evolution are taken as bona fide equilibrium states of the composite, and all thermodynamic quantities are defined through the HMF. Theorem 1 demonstrates that the open system free energy difference $\Delta F_{\mathcal{S}}^*(\beta)$ between two canonical endpoint marginals admits exact representations in terms of the HMF shift and the chi-squared divergence between the endpoint system marginals. This chi-squared factor quantifies the statistical change in the system distribution induced by strong coupling, explicitly linking the free energy difference to the mismatch between endpoint ensembles. Such mismatches, and their impact on thermodynamic potentials and information measures, are widely emphasized in strong coupling formulations of open system thermodynamics [38,44] but typically appear only implicitly. Here they enter at the level of equilibrium endpoint reweighting identities, providing a clean and quantitative diagnostic. A notable feature of the trajectory construction is that it does not rely on a specific dynamical generator once the final endpoint marginal has equilibrated. Once the assumption of asymptotic equilibration is made, Theorem 2 rewrites the endpoint identities as trajectory sampling representations. This result establishes that the free energy difference between two canonical states can be computed from trajectory-generated endpoint samples, even when strong coupling prevents the existence of a clear separation between system and environment energies. The Appendix gives an optional decomposition of the trajectory HMF increment. This decomposition is not used in the estimator; it only provides a thermodynamic reading of the HMF increment when such an interpretation is desired.

A major conceptual outcome of this work is the identification of two complementary protocol views—frozen coupling and frozen driving—that divide the framework into two branches. In the frozen-driving view analyzed here, the driving parameter $\lambda(t)$ is held constant while the interaction $C(t)$ transports the system between endpoints. In this regime, the endpoint and trajectory identities simplify substantially, the bare system Hamiltonian cancels from the HMF, and the free energy difference does not depend on the system Hamiltonian directly but entirely on the environment functional $\mathcal{M}(X_{\mathcal{S}}, C, \beta)$. Corollaries 1 and 2 show that all exponential relations can be expressed as ratios of $\mathcal{M}$ evaluated at $C(0)$ and $C(t_{\text{eq}})$, together with the overlap factor. This yields FEP-type endpoint reweighting formulas with explicit overlap factors [Eqs. (67)–(68)], with trajectory versions obtained under asymptotic equilibration [Eqs. (69)–(70)]. Standard FEP emerges as the special case in which the initial composite is uncoupled and the interaction is applied instantaneously. The complementary frozen-coupling view is developed in a separate work [50]. Here we focus only on the frozen-driving, FEP-type branch and on the overlap diagnostics that follow from the endpoint marginals. The numerical validation in Sec. VII reveals the diagnostic role of the present framework. For a strongly coupled system evolving under overdamped Langevin dynamics, we compared the exact HMF reference, Zwanzig FEP, and the overlap-explicit endpoint estimator of $\Delta F_{\mathcal{S}}^*(\beta)$ across a range of coupling strengths. When the endpoint marginals exhibit substantial overlap, the

classical Zwanzig estimator, the trajectory representation of the endpoint estimator, and the exact HMF reference agree. At moderate and strong coupling, however, the distributions separate, the phase-space overlap deteriorates, and the classical FEP estimator shows the expected finite-sampling degradation. The overlap-explicit identity makes this degradation visible through the same endpoint mismatch that controls the chi-squared factor. These results demonstrate that the chi-squared factor supplies a principled diagnostic of the overlap burden; it does not remove the usual overlap limitations of exponential reweighting. They also illustrate how strong-coupling corrections manifest operationally in a model system, complementing the conceptual discussions in the strong-coupling literature [38].

A second practical result of this work is that the overlap structure is not only formal but numerically accessible. In the one-step sampled formulation, the factor $1 + \chi^2$ can be reconstructed directly from trajectory averages, which makes the onset of poor overlap visible in the same simulation used to estimate the free energy. In the recursive formulation, this structure becomes local: the full transformation is decomposed into a ladder of neighboring states, and the global overlap burden is resolved into a sum of local contributions. This gives the method a dual role. It acts as a free energy estimator, but it also acts as a diagnostic that identifies where a path is statistically well resolved and where additional intermediate states are needed. The BAR/MBAR comparison supports the same conclusion: with adequate staging, the standard baselines and the proposed staged construction all reproduce the exact reference in this model. The contribution is therefore structural and diagnostic: the endpoint formulation reports the local overlap burden, rather than claiming a universal numerical improvement over BAR or MBAR.

The broader implications of this work span both numerical methodology and strong-coupling thermodynamics. From a computational perspective, the identities derived here suggest new ways to design and optimize free energy calculations in open systems, because their explicit dependence on the chi-squared divergence identifies, in a quantitative manner, how protocol choices affect the overlap between endpoint marginals. This provides a natural conceptual bridge to long-standing strategies developed within the FEP community, where the reliability of a calculation is known to be controlled by the degree of ensemble overlap. In conventional FEP, overlap is often improved by inserting intermediate states and stratifying ensembles [10,17,29,30,33,34], or by applying enhanced sampling methods such as umbrella sampling, metadynamics, accelerated MD, and related biasing schemes [18–22,66–68]. Building on this foundation, recent progress in alchemical free energy calculations and advanced enhanced-sampling schemes for free energy landscapes [23,25–28] has made the notion of overlap even more explicit, with optimized coupling schedules and variational biases designed to flatten barriers and improve phase-space mixing. Adaptive schemes for constructing alchemical paths aim to enhance or optimize phase-space overlap directly [26,28], often selecting intermediate states to stabilize estimators or reduce rare-event contributions. In parallel, thermodynamically consistent open-system coupling procedures based on adaptive-resolution and particle-domain

methods [69,70] demonstrate that correct endpoint statistics and, in particular, matching the system marginal to a target ensemble are essential for reliable equilibrium and nonequilibrium characterizations. In adaptive-resolution simulations [69], a thermodynamic force or free energy compensation term is tuned so that density and structural properties in the atomistic region reproduce those of a fully atomistic reference, effectively enforcing the desired subsystem marginal. In particle-continuum couplings with fluctuating hydrodynamics [70], the exchange of matter and energy is constructed so that the particle domain exhibits the same thermodynamic fluctuations as a macroscopic reservoir, again using subsystem statistics as the standard of correctness. Seen from this perspective, the present framework provides a useful endpoint diagnostic for these algorithmic developments. By identifying the chi-squared divergence $\chi^2(P_{\text{eq}}^S \| P_0^S)$ as the exact measure of ensemble mismatch entering the HMF endpoint reweighting identity, it offers a direct and thermodynamically grounded criterion for choosing, optimizing, or adaptively refining intermediate states. Because χ^2 depends only on endpoint marginals, it becomes a practical diagnostic for assessing overlap and a principled target for adaptive schemes designed to maintain ensemble similarity along a chosen thermodynamic path. Another natural extension of this framework is its integration with machine-learned collective variables and unsupervised dimensionality-reduction methods. Recent studies have shown that low-dimensional representations can be constructed to preserve or enhance overlap between ensembles [23,24,71–73]. Within such approaches, the chi-squared divergence can provide a concrete objective for selecting or refining collective variables and for determining when enhanced sampling has reached sufficient coverage of the relevant configuration space. Similarly, multi-stage and replica-based strategies [10,19,21,22,27,74,75] can naturally be reinterpreted as sampling schemes that explore controlled sequences of effective Hamiltonians (or HMFs) along paths in (λ, C, β) . Rewriting these constructions in the HMF language can clarify when strong-coupling effects are absorbed into effective potentials and when they might be treated explicitly at the level of open-system energetics. On the conceptual side, the framework contributes to a growing effort to establish a consistent thermodynamic theory for strongly coupled open systems, complementing modern analyses that highlight the ambiguity of heat, work, entropy

production, and free energy for strongly coupled open system [38,44]. Expressing observables in terms of the HMF and monitoring ensemble discrepancies through χ^2 suggests a route by which such ambiguities can be reduced, while the optional decomposition gives a thermodynamic reading of energetic contributions in nonequilibrium processes that sits naturally alongside earlier formulations [38,76,77]. We expect that the endpoint identities and overlap diagnostics developed here, together with those presented in Ref. [50] and our recent related work [48,49], will help guide future free energy calculations in strongly coupled open systems.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy EXC 2089/2-390776260 (e-conversion). We thank the anonymous referees for their careful reading and constructive comments, which helped improve the clarity and scope of the manuscript.

DATA AVAILABILITY

The numerical data and source code required to reproduce all data and figures generated here are openly available [78].

APPENDIX: OPTIONAL THERMODYNAMIC STRUCTURE

Remark 1 (Worklike, heatlike, and reference-field decomposition). We now give an optional thermodynamic interpretation of the trajectory representation in Eqs. (34) and (35). This decomposition is not required for estimating ΔF_S^* from the endpoint reweighting identities, but it clarifies the energetic content of the HMF increment along sampled trajectories. In regimes where both driving and coupling are activated, energetic changes in $\mathcal{S}$ arise from (1) the manipulations of $[\lambda(t), C(t)]$ and (2) exchange with the environment through the microscopic evolution of $X_{\mathcal{E}}$. To inspect these contributions, we start from Eq. (36) and apply a simple algebraic insertion-subtraction of the HMF at fixed protocols, $\mathcal{H}_{\beta}^*(X_S(0|X_0), \lambda(0), C(0))$, which yields the following representation:

$$\begin{aligned} \Delta \mathcal{H}_{\beta}^*(\mathcal{T}_{\text{eq}}^S(X_0)) &= \underbrace{[\mathcal{H}_{\beta}^*(X_S(t_{\text{eq}}|X_0), \lambda(t_{\text{eq}}), C(t_{\text{eq}})) - \mathcal{H}_{\beta}^*(X_S(0|X_0), \lambda(0), C(0))]}_{I^*(t_{\text{eq}}|X_0)} \\ &\quad - \underbrace{[\mathcal{H}_{\beta}^*(X_S(t_{\text{eq}}|X_0), \lambda(0), C(0)) - \mathcal{H}_{\beta}^*(X_S(0|X_0), \lambda(0), C(0))]}_{II^*(t_{\text{eq}}|X_0)}. \end{aligned} \quad (\text{A1})$$

Term I^* . Along each realization, the system follows the trajectory $X_S(t|X_0)$ in its phase space. The total time variation of the HMF, $\mathcal{H}_{\beta}^*(X_S(t|X_0), \lambda(t), C(t))$, is obtained by applying the chain rule,

$$\frac{d}{dt} \mathcal{H}_{\beta}^*(X_S(t|X_0), \lambda(t), C(t)) = \underbrace{(\nabla_{X_S} \mathcal{H}_{\beta}^*) \cdot \dot{X}_S}_{\dot{Q}^*} + \underbrace{\frac{\partial \mathcal{H}_{\beta}^*}{\partial \lambda} \dot{\lambda} + \frac{\partial \mathcal{H}_{\beta}^*}{\partial C} \dot{C}}_{\dot{W}^*}. \quad (\text{A2})$$

For stochastic dynamics, the differential relation in Eq. (A2) is understood in the Stratonovich sense. For deterministic dynamics, it reduces to the ordinary chain rule. The first term describes the energy exchange with the environment along the actual phase-space trajectory and defines the instantaneous heatlike rate, $\dot{Q}^*$, whereas the sum of the second and third terms accounts for the parametric energy input due to the protocols and defines the instantaneous worklike rate, $\dot{W}^*$. Following the standard conventions of stochastic energetics [38,76], their time integrals give

$$W^*(t_{\text{eq}}|X_0) = \int_0^{t_{\text{eq}}} dt \frac{\partial \mathcal{H}_\beta^*}{\partial \lambda} \dot{\lambda}(t) + \int_0^{t_{\text{eq}}} dt \frac{\partial \mathcal{H}_\beta^*}{\partial C} \dot{C}(t), \quad (\text{A3})$$

$$Q^*(t_{\text{eq}}|X_0) = \int_0^{t_{\text{eq}}} dt (\nabla_{X_S} \mathcal{H}_\beta^* \cdot \dot{X}_S(t|X_0)), \quad (\text{A4})$$

so that integrating Eq. (A2) over $t \in [0, t_{\text{eq}}]$ yields the total increment

$$I^*(t_{\text{eq}}|X_0) = W^*(t_{\text{eq}}|X_0) + Q^*(t_{\text{eq}}|X_0). \quad (\text{A5})$$

Term I^* . For fixed controls $\lambda(0)$ and $C(0)$, $I^*(t_{\text{eq}}|X_0)$ can be written as

$$I^*(t_{\text{eq}}|X_0) = \int_0^{t_{\text{eq}}} dt (\nabla_{X_S} \mathcal{H}_\beta^*(X_S, \lambda(0), C(0))) \cdot \dot{X}_S. \quad (\text{A6})$$

Here the gradient is evaluated with respect to $\mathcal{H}_\beta^*$ frozen at $\lambda(0)$ and $C(0)$, whereas $\dot{X}_S$ belongs to the driven trajectory

generated by $\lambda(t)$ and $C(t)$. Consequently, I^* is a reference functional that projects the driven generalized velocity $\dot{X}_S$ onto the generalized force field $\nabla_{X_S} \mathcal{H}_\beta^*$ of the initial protocol. Combining Eqs. (A5) and (A6) with Eq. (A1), we obtain

$$\Delta \mathcal{H}_\beta^*(\mathcal{T}_{t_{\text{eq}}^S}(X_0)) = W^*(t_{\text{eq}}|X_0) + Q^*(t_{\text{eq}}|X_0) - I^*(t_{\text{eq}}|X_0). \quad (\text{A7})$$

Substituting Eq. (A7) into Eqs. (34) and (35) yields the compact protocol-exchange-reference-field representation of the endpoint reweighting identities,

$$e^{-\beta \Delta F_S^*(\beta)} = \frac{\langle e^{-\beta[W^* + Q^* - I^*]} \rangle_{X_0}}{1 + \chi^2(P_{t_{\text{eq}}}^S \parallel P_0^S)}, \quad (\text{A8})$$

$$e^{+\beta \Delta F_S^*(\beta)} = \langle e^{+\beta[W^* + Q^* - I^*]} \rangle_{X_0}. \quad (\text{A9})$$

The decomposition isolates three pathwise contributions: (1) W^* , a protocol contribution due to changes in $\lambda(t)$ and $C(t)$ (worklike); (2) Q^* , an exchange contribution associated with motion along the system trajectory (heatlike); and (3) I^* , a reference-field projection obtained by contracting the driven velocity ($\dot{X}_S$) with the $\lambda(0)$ and $C(0)$ force field ($\nabla_{X_S} \mathcal{H}_\beta^*$). These terms provide a thermodynamic reading of the trajectory representation but are not operational prerequisites for the endpoint reweighting estimator.

-
- [1] *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, edited by C. Chipot and A. Pohorille, Springer Series in Chemical Physics, Vol. 86 (Springer, Berlin, Heidelberg, 2007).
 - [2] D. Jiao, P. A. Golubkov, T. A. Darden, and P. Ren, Calculation of protein–ligand binding free energy by using a polarizable potential, *Proc. Natl. Acad. Sci. USA* **105**, 6290 (2008).
 - [3] M. K. Gilson and H.-X. Zhou, Calculation of protein-ligand binding affinities, *Annu. Rev. Biophys. Biomol. Struct.* **36**, 21 (2007).
 - [4] P. C. Hohenberg and A. P. Krekhov, An introduction to the Ginzburg–Landau theory of phase transitions and nonequilibrium patterns, *Phys. Rep.* **572**, 1 (2015).
 - [5] P. Hänggi, P. Talkner, and M. Borkovec, Reaction-rate theory: Fifty years after Kramers, *Rev. Mod. Phys.* **62**, 251 (1990).
 - [6] Z. Cournia and C. Chipot, Applications of free-energy calculations to biomolecular processes. a collection, *J. Phys. Chem. B* **128**, 3299 (2024).
 - [7] D. M. York, Modern alchemical free energy methods for drug discovery explained, *ACS Phys. Chem. Au* **3**, 478 (2023).
 - [8] A. S. Mey, B. K. Allen, H. E. B. Macdonald, J. D. Chodera, D. F. Hahn, M. Kuhn, J. Michel, D. L. Mobley, L. N. Naden, S. Prasad, *et al.*, Best practices for alchemical free energy calculations, *Live. J. Comput. Mol. Sci.* **2**, 18378 (2020).
 - [9] E. Duboué-Dijon and J. Hénin, Building intuition for binding free energy calculations: Bound state definition, restraints, and symmetry, *J. Chem. Phys.* **154**, 204101 (2021).
 - [10] G. M. Torrie and J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *J. Comput. Phys.* **23**, 187 (1977).
 - [11] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 3rd ed. (Academic Press, San Diego, 2023).
 - [12] J. G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* **3**, 300 (1935).
 - [13] R. W. Zwanzig, High-temperature equation of state by a perturbation method. I. Nonpolar gases, *J. Chem. Phys.* **22**, 1420 (1954).
 - [14] R. W. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).
 - [15] P. V. Klimovich, M. R. Shirts, and D. L. Mobley, Guidelines for the analysis of free energy calculations, *J. Comput. Aided Mol. Des.* **29**, 397 (2015).
 - [16] S. Y. Willow, L. Kang, and D. D. Minh, Learned mappings for targeted free energy perturbation between peptide conformations, *J. Chem. Phys.* **159**, 124104 (2023).
 - [17] G. M. Torrie and J. P. Valleau, Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid, *Chem. Phys. Lett.* **28**, 578 (1974).
 - [18] D. Hamelberg, J. Mongan, and J. A. McCammon, Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules, *J. Chem. Phys.* **120**, 11919 (2004).
 - [19] J. Kästner and W. Thiel, Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “Umbrella integration”, *J. Chem. Phys.* **123**, 144104 (2005).
 - [20] A. Barducci, G. Bussi, and M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method, *Phys. Rev. Lett.* **100**, 020603 (2008).

- [21] J. Kästner, Umbrella sampling, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 932 (2011).
- [22] N. V. Plotnikov, Computing the free energy barriers for less by sampling with a coarse reference potential while retaining accuracy of the target fine model, *J. Chem. Theory Comput.* **10**, 2987 (2014).
- [23] M. Invernizzi, P. M. Piaggi, and M. Parrinello, Unified approach to enhanced sampling, *Phys. Rev. X* **10**, 041034 (2020).
- [24] T. D. Swinburne and M.-C. Marinica, Unsupervised calculation of free energy barriers in large crystalline systems, *Phys. Rev. Lett.* **120**, 135503 (2018).
- [25] S. Falkner, A. Coretti, and C. Dellago, Enhanced sampling of configuration and path space in a generalized ensemble by shooting point exchange, *Phys. Rev. Lett.* **132**, 128001 (2024).
- [26] S. Zhang, T. J. Giese, T.-S. Lee, and D. M. York, Alchemical enhanced sampling with optimized phase space overlap, *J. Chem. Theory Comput.* **20**, 3935 (2024).
- [27] A. D. Wade, A. P. Bhati, S. Wan, and P. V. Coveney, Alchemical free energy estimators and molecular dynamics engines: Accuracy, precision, and reproducibility, *J. Chem. Theory Comput.* **18**, 3972 (2022).
- [28] A. Knirsch, B. N. Falcone, and J. D. Hirst, Practical guidelines for optimising free energy calculations using thermodynamic integration, *Chem. Phys. Lett.* **880**, 142395 (2025).
- [29] M. R. Shirts and J. D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states, *J. Chem. Phys.* **129**, 124105 (2008).
- [30] M. Reinhardt and H. Grubmueller, Determining free-energy differences through variationally derived intermediates, *J. Chem. Theory Comput.* **16**, 3504 (2020).
- [31] C. H. Bennett, Efficient estimation of free energy differences from Monte Carlo data, *J. Comput. Phys.* **22**, 245 (1976).
- [32] N. Lu, D. Wu, T. B. Woolf, and D. A. Kofke, Using overlap and funnel sampling to obtain accurate free energies from nonequilibrium work measurements, *Phys. Rev. E* **69**, 057702 (2004).
- [33] D. Wu and D. A. Kofke, Asymmetric bias in free-energy perturbation measurements using two Hamiltonian-based models, *Phys. Rev. E* **70**, 066702 (2004).
- [34] Y. Matsunaga, M. Kamiya, H. Oshima, J. Jung, S. Ito, and Y. Sugita, Use of multistate Bennett acceptance ratio method for free-energy calculations from enhanced sampling and free-energy perturbation, *Biophys. Rev.* **14**, 1503 (2022).
- [35] R. P. Feynman and F. L. Vernon Jr., The theory of a general quantum system interacting with a linear dissipative system, *Ann. Phys.* **281**, 547 (2000).
- [36] B. Roux and T. Simonson, Implicit solvent models, *Biophys. Chem.* **78**, 1 (1999).
- [37] P. Talkner and P. Hänggi, Open system trajectories specify fluctuating work but not heat, *Phys. Rev. E* **94**, 022143 (2016).
- [38] P. Talkner and P. Hänggi, *Colloquium*: Statistical mechanics and thermodynamics at strong coupling: Quantum and classical, *Rev. Mod. Phys.* **92**, 041002 (2020).
- [39] C. Jarzynski, Stochastic and macroscopic thermodynamics of strongly coupled systems, *Phys. Rev. X* **7**, 011008 (2017).
- [40] H. J. D. Miller and J. Anders, Energy-temperature uncertainty relation in quantum thermodynamics, *Nat. Commun.* **9**, 2203 (2018).
- [41] U. Seifert, First and second law of thermodynamics at strong coupling, *Phys. Rev. Lett.* **116**, 020601 (2016).
- [42] H. J. D. Miller and J. Anders, Entropy production and time asymmetry in the presence of strong interactions, *Phys. Rev. E* **95**, 062123 (2017).
- [43] N. Anto-Sztrikacs, A. Nazir, and D. Segal, Effective-Hamiltonian theory of open quantum systems at strong coupling, *PRX Quantum* **4**, 020307 (2023).
- [44] P. Strasberg and M. Esposito, Measurability of nonequilibrium thermodynamics in terms of the Hamiltonian of mean force, *Phys. Rev. E* **101**, 050101(R) (2020).
- [45] M. Campisi, P. Talkner, and P. Hänggi, Fluctuation theorem for arbitrary open quantum systems, *Phys. Rev. Lett.* **102**, 210401 (2009).
- [46] X. Xing and M. Ding, Thermodynamics and stochastic thermodynamics of strongly coupled systems, *Phys. Rev. E* **109**, 034105 (2024).
- [47] A. Colla, F. Hasse, D. Palani, T. Schaetz, H.-P. Breuer, and U. Warring, Observing time-dependent energy level renormalisation in an ultrastrongly coupled open system, *Nat. Commun.* **16**, 2502 (2025).
- [48] M. Rahbar and C. J. Stein, Thermodynamic potentials from a probabilistic view on the system-environment interaction energy, [arXiv:2505.00188](https://arxiv.org/abs/2505.00188).
- [49] M. Rahbar and C. J. Stein, A probabilistic approach to system-environment coupling, [arXiv:2505.00192](https://arxiv.org/abs/2505.00192).
- [50] M. Rahbar and C. J. Stein, Exact and variational identities for free energy differences in strongly coupled open systems, [arXiv:2511.10236](https://arxiv.org/abs/2511.10236).
- [51] Z. Davoudi, C. Jarzynski, N. Mueller, G. Oruganti, C. Powers, and N. Y. Halpern, Work and heat exchanged during sudden quenches of strongly coupled quantum systems, [arXiv:2502.19418](https://arxiv.org/abs/2502.19418).
- [52] S. Whitelam, Improving noisy free-energy measurements by adding more noise, *Phys. Rev. E* **112**, 014133 (2025).
- [53] G. Li and Z. C. Tu, Equilibrium free-energy differences from a linear nonequilibrium equality, *Phys. Rev. E* **103**, 032146 (2021).
- [54] Y. Saito, Relaxation in a bistable system, *J. Phys. Soc. Jpn.* **41**, 388 (1976).
- [55] S. X. Sun, Equilibrium free energies from path sampling of nonequilibrium trajectories, *J. Chem. Phys.* **118**, 5769 (2003).
- [56] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [57] J. Spiechowicz, I. G. Marchenko, P. Hänggi, and J. Łuczka, Diffusion coefficient of a Brownian particle in equilibrium and nonequilibrium: Einstein model and beyond, *Entropy* **25**, 42 (2022).
- [58] T. Kanamori, S. Hido, and M. Sugiyama, A least-squares approach to direct importance estimation, *J. Mach. Learn. Res.* **10**, 1391 (2009).
- [59] O. Thomas, R. Dutta, J. Corander, S. Kaski, and M. U. Gutmann, Likelihood-free inference by ratio estimation, *Bayesian Anal.* **17**, 1 (2022).
- [60] C. Hartmann, J. C. Latorre, and G. Ciccotti, On two possible definitions of the free energy for collective variables, *Eur. Phys. J. Spec. Top.* **200**, 73 (2011).
- [61] L. Bonati, E. Trizio, A. Rizzi, and M. Parrinello, A unified framework for machine learning collective variables for enhanced sampling simulations: *m1colvar*, *J. Chem. Phys.* **159**, 014801 (2023).

- [62] J. Rydzewski, Spectral map: Embedding slow kinetics in collective variables, *J. Phys. Chem. Lett.* **14**, 5216 (2023).
- [63] F. Noé, S. Olsson, J. Köhler, and H. Wu, Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning, *Science* **365**, eaaw1147 (2019).
- [64] J. Köhler, A. Krämer, and F. Noé, Smooth normalizing flows, *Adv. Neural Inf. Process. Syst.* **34**, 2796 (2021).
- [65] L. Klein and F. Noé, Transferable Boltzmann generators, *Adv. Neural Inf. Process. Syst.* **37**, 45281 (2024).
- [66] O. Valsson and M. Parrinello, Variational approach to enhanced sampling and free energy calculations, *Phys. Rev. Lett.* **113**, 090601 (2014).
- [67] R. C. Bernardi, M. C. Melo, and K. Schulten, Enhanced sampling techniques in molecular dynamics simulations of biological systems, *Biochim. Biophys. Acta Gen. Subj.* **1850**, 872 (2015).
- [68] A. Laio and F. L. Gervasio, Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science, *Rep. Prog. Phys.* **71**, 126601 (2008).
- [69] S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site, and K. Kremer, Adaptive resolution molecular dynamics simulation through coupling to an internal particle reservoir, *Phys. Rev. Lett.* **108**, 170602 (2012).
- [70] A. Gholami, R. Klein, and L. Delle Site, Simulation of a particle domain in a continuum, fluctuating hydrodynamics reservoir, *Phys. Rev. Lett.* **129**, 230603 (2022).
- [71] S. Mehdi, Z. Smith, L. Herron, Z. Zou, and P. Tiwary, Enhanced sampling with machine learning, *Annu. Rev. Phys. Chem.* **75**, 347 (2024).
- [72] H. Fu, H. Bian, X. Shao, and W. Cai, Collective variable-based enhanced sampling: From human learning to machine learning, *J. Phys. Chem. Lett.* **15**, 1774 (2024).
- [73] D. Mendels and J. J. de Pablo, Collective variables for free energy surface tailoring: Understanding and modifying functionality in systems dominated by rare events, *J. Phys. Chem. Lett.* **13**, 2830 (2022).
- [74] D. Pregeljč, R. J. Huegli, and S. Riniker, Efficient multistate free-energy calculations with QM/MM accuracy using replica-exchange enveloping distribution sampling, *J. Phys. Chem. B* **129**, 5948 (2025).
- [75] W.-T. Hsu and M. R. Shirts, Replica exchange of expanded ensembles: A generalized ensemble approach with enhanced flexibility and parallelizability, *J. Chem. Theory Comput.* **20**, 6062 (2024).
- [76] C. Jarzynski, Nonequilibrium work theorem for a system strongly coupled to a thermal environment, *J. Stat. Mech.* (2004) P09005.
- [77] T. Sagawa and M. Ueda, Generalized Jarzynski equality under nonequilibrium feedback control, *Phys. Rev. Lett.* **104**, 090602 (2010).
- [78] M. Rahbar and C. J. Stein, Endpoint reweighting identities with overlap diagnostics for free energy differences in strongly coupled open systems, Zenodo, 2026, <https://doi.org/10.5281/zenodo.20761318>.