Journal Pre-proof

Ergodicity of the infinite swapping algorithm at low temperature

Georg Menz, André Schlichting, Wenpin Tang, Tianqi Wu

PII: S0304-4149(22)00151-X
DOI: https://doi.org/10.1016/j.spa.2022.06.015
Reference: SPA 4011

To appear in: *Stochastic Processes and their Applications*

Received date : 18 September 2021
Revised date : 14 June 2022
Accepted date : 17 June 2022

Please cite this article as: G. Menz, A. Schlichting, W. Tang et al., Ergodicity of the infinite swapping algorithm at low temperature, *Stochastic Processes and their Applications* (2022), doi: https://doi.org/10.1016/j.spa.2022.06.015.

ERGODICITY OF THE INFINITE SWAPPING ALGORITHM AT LOW TEMPERATURE

GEORG MENZ, ANDRÉ SCHLICHTING, WENPIN TANG, AND TIANQI WU

Abstract. Sampling Gibbs measures at low temperatures is an important but computationally challenging task. Numerical evidence suggests that the infinite-swapping algorithm (isa) is a promising method. The isa can be seen as an improvement of the parallel tempering replica method. We rigorously analyze the ergodic properties of the isa in the low temperature regime, deducing asymptotic estimates for the spectral gap (or Poincaré constant), optimal in dimension one, and an estimate for the log-Sobolev constant. Our main results indicate that the effective energy barrier can be reduced drastically using the isa compared to the classical over-damped Langevin dynamics. As a corollary, we derive a concentration inequality showing that sampling is also improved by an exponential factor. Finally, we study simulated annealing for the isa and prove that the isa again outperforms the over-damped Langevin dynamics.

Key words: Sampling, low-temperature, simulated annealing, infinite swapping, parallel tempering, replica exchange, Poincaré inequality, spectral gap, log-Sobolev inequality, Eyring-Kramers formula.

AMS 2010 Mathematics Subject Classification: 60J60, 39B62.

1. Introduction

Sampling from Gibbs measures at low temperatures is important in science and engineering. It has a variety of applications including molecular dynamics [And80, CS11] and Bayesian inference [RC04, GCS+14]. Usually, sampling at low temperatures is slow due to the fact that at low temperatures energy barriers in the underlying energy landscape are large. This traps the stochastic sampling process and slows down sampling.

One popular way to sample Gibbs measures is to run the over-damped Langevin equation or its various discretization schemes for approximation, see e.g. [RT96, Dal17, DM17, DCWY19]. A lot of efforts have been made to accelerate sampling at low temperatures and there are many competing methods. One of them is the replica exchange method which is also known as parallel tempering. In the simplest version of a replica exchange method, one considers two particles governed by independent copies of the underlying dynamics, for instance, the over-damped Langevin equation.

Date: June 14, 2022.

One particle evolves at the desired low temperature $\tau_1 > 0$, and the other particle evolves at a higher temperature $\tau_2 > 0$ with $\tau_1 \ll \tau_2 \ll 1$. At some random times, the positions of both particles are swapped. This approach has the advantage that the particle at the lower temperature correctly samples the low-temperature Gibbs measure whereas the particle at the higher temperature can explore the full state space, and discover the relevant states of the system efficiently.

Replica exchange methods or parallel tempering have been successfully applied in many different scenarios, and they seem to accelerate sampling in low-temperature situations quite well. As far as we are concerned, almost all evaluations of the performance of those methods are empirical. In an attempt to study the sampling performance of parallel tempering, it was discovered in [DLPD12] that the large deviation rate function for time-averaged empirical measures of parallel tempering is a monotone function of the swapping rate. It implies that sampling only improves at a faster swapping rate.

This led to the question of a suitable limiting process as the swapping rate goes to infinity. Since the number of jumps of the particles would grow to infinity in any bounded time-interval, the authors in [DLPD12] suggest the infinite swapping algorithm/process (isa), a procedure that can be interpreted as the limit of parallel tempering, where instead of the *particle positions*, the *particle temperatures* are swapped at an infinitely fast rate (see Section 2.1 for a review).

To be more precise, let $H : \mathbb{R}^n \to \mathbb{R}$ be the underlying energy landscape and the goal is to sample the Gibbs measure with density $\nu^{\tau_1}(x) := \frac{1}{Z^{\tau_1}} \exp\left(-\frac{H(x)}{\tau_1}\right)$ where Z^{τ_1} is the normalizing constant. Formally, given two different temperatures $0 < \tau_1 \ll \tau_2$, the isa is defined as the evolution of two particles $X_1 = (X_1(t), t \geq 0)$ and $X_2 = (X_2(t), t \geq 0)$ governed by the stochastic differential equations (SDEs):

$$\begin{cases} dX_1 = -\nabla H(X_1)\, dt + \sqrt{2\tau_1 \rho(X_1, X_2) + 2\tau_2 \rho(X_2, X_1)}\, dB_1, \\ dX_2 = -\nabla H(X_2)\, dt + \sqrt{2\tau_2 \rho(X_1, X_2) + 2\tau_1 \rho(X_2, X_1)}\, dB_2, \end{cases} \tag{1.1}$$

where (B_1, B_2) are independent Brownian motions in $\mathbb{R}^n$, and

$$\rho(x_1, x_2) := \frac{\pi(x_1, x_2)}{\pi(x_1, x_2) + \pi(x_2, x_1)} \quad \text{and} \quad \pi(x_1, x_2) := \nu^{\tau_1}(x_1)\nu^{\tau_2}(x_2). \tag{1.2}$$

Since $\tau_1 \neq \tau_2$, we have that $\pi(x_1, x_2) \neq \pi(x_2, x_1)$, and thus $\rho(x_1, x_2) \neq \rho(x_2, x_1)$. The functions $\rho(x_1, x_2), \rho(x_2, x_1)$ are relative weights assigned to the two configurations (x_1, x_2), (x_2, x_1) based on π. At each moment, this essentially assigns the higher temperature τ_2 to the particle whose potential energy H is higher at that moment (see also [DDN18, Section 3.2]).

The crucial feature of the dynamics (1.1) is that the empirical measure

$$\eta_t := \frac{1}{t} \int_0^t \rho(X_1, X_2)\delta_{(X_1,X_2)} + \rho(X_2, X_1)\delta_{(X_2,X_1)} ds$$

converges weakly to the product measure π as $t \to \infty$ by the ergodic theorem. In particular, by restricting to the first coordinate, the measure $\frac{1}{t}\int_0^t \rho(X_1, X_2)\delta_{X_1} +$

$\rho(X_2, X_1)\delta_{X_2} ds$ approximates the Gibbs measure ν^{τ_1} for t large enough. In [DLPD12], a large deviation principle was established for the measure η_t. However, it is not clear how the rate function depends on the temperatures (τ_1, τ_2), so it is less obvious why the higher temperature τ_2 may be helpful. Further numerical and heuristic studies in [DDN18] indicate that there is an exponential gain when using the isa for sampling in comparison with the classical over-damped Langevin dynamics. Let us also point out a connection to integrated tempering enhanced sampling method [Gao08], which has a reformulation as an infinite switching limit of simulated tempering method over a mixed potential [LVE13, LVE17, YLLG18, MLLVE19]. Recently the isa was applied to training restricted Boltzmann machines [HNR20], and was shown to be competitive empirically. But no rigorous result has been established so far on how well the isa accelerates sampling at low temperatures.

In this article we take the analysis of [DLPD12, DDN18] to the next level through a functional inequality approach. We carry out the first rigorous study of the ergodic properties of the isa at low temperatures by quantifying its convergence in terms of the temperatures (τ_1, τ_2). Under standard nondegeneracy assumptions, we deduce the low-temperature asymptotics for the Poincaré and the log-Sobolev constant of the isa, see Theorem 2.8 and Theorem 2.9 below. In the context of metastability, these formulas are also known as Eyring-Kramers formulas (see [Ber13] for background). Comparing our results to the Eyring-Kramers formulas for the over-damped Langevin equation (e.g. see [BEGK04, BGK05, MS14]), we have an exponential gain: the effective energy barrier of the underlying energy landscape H only sees the higher temperature τ_2. We also give indications that our results are optimal.

To the best of our knowledge, this is the first time an Eyring-Kramers formula was derived for inhomogeneous diffusions, for which the stationary and ergodic distribution is generally unknown. By construction, however, the isa (1.1) has an explicit stationary distribution μ given by $\mu(x_1, x_2) = \frac{1}{2}\left(\pi(x_1, x_2) + \pi(x_2, x_1)\right)$, where $\pi(\cdot, \cdot)$ is defined by (1.2). This makes a rigorous analysis of (1.1) feasible. For the proof of our main results, Theorem 2.8 and Theorem 2.9, we follow the transportation approach of [MS14]. The idea is to identify the right "paths" of transport which give the leading order term in the Poincaré and the log-Sobolev constant of the isa. In the case of the Langevin diffusion process those paths can be obtained from mountain pass paths between local minima of the energy H. Since the isa is a process on $\mathbb{R}^n \times \mathbb{R}^n$ swapping the two particle temperatures, it requires analyzing transport in a planar network obtained from the product structure of two energies, and so is more involved.

There are several other methods which could be used to deduce the Eyring-Kramers formula for the Poincaré constant. For instance, one could consider adapting the potential theoretic approach (see [BEGK04, BGK05]), or the semiclassical analysis (see [HKN04, HN05, HN06]), or the approach using quasi-stationarity (see [BR16, GLPN16, LLPN19, GLPN19]). We adopt the approach of [MS14], which is robust enough to deduce the Eyring-Kramers formula for the log-Sobolev constant in the setting of an inhomogeneous diffusion coefficient. The rate of convergence in relative entropy obtained from the log-Sobolev constant is important for our applications to sampling and simulated annealing.

In the first application, we apply the main results to study the sampling properties of the isa and compare it to the over-damped Langevin dynamics. It is well known that the Poincaré and the log-Sobolev constants characterize the rate of convergence to equilibrium of the underlying process. It is also known that Poincaré and log-Sobolev inequalities yield non-asymptotic concentration/deviation inequalities (see [CG08, WY08] and references therein). Hence, our main results yield a quantitative control in terms of the temperatures (τ_1, τ_2) on the rate of convergence of the time average to the ensemble average, quantifying the ergodic theorem. Let us note in comparison that the precise dependence on (τ_1, τ_2) is missing in the large deviation estimates for the isa in [DLPD12]. As a byproduct of our analysis, we find a condition on (τ_1, τ_2) under which sampling at low temperatures using the isa is exponentially faster than using the over-damped Langevin dynamics. This provides a guidance on the choice of the higher temperature τ_2 for the isa, which is the condition (2.18) in Corollary 2.10.

In the second application, we study the isa for simulated annealing and compare it to simulated annealing adapted to the over-damped Langevin dynamics. Simulated annealing (SA) is an umbrella term denoting a particular set of stochastic optimization methods. SA can be used to find the global extremum of a function $H : \mathbb{R}^n \to \mathbb{R}$, in particular when H is non-convex. Those methods have many applications in different fields, for example in physics, chemistry and operations research (see e.g. [vLA87, KAJ94, Nar99]). The name and inspiration comes from annealing in metallurgy, a process that aims to increase the size of the crystals by heating and controlled cooling. The SA mimics this procedure mathematically. The stochastic version of SA was independently described by Kirkpatrick, Gelatt and Vecchi [KGV83] and Černý [Č85]. See Section 2.7 for details on simulated annealing.

Replica exchange methods or parallel tempering have been successfully applied to nonconvex optimization (see e.g. [CCD+19, DT21]) and simulated annealing (see e.g. [KZ09, LPA+09]). Because the isa has better ergodic properties than parallel tempering, there is big hope that the isa can produce even better results. Additionally, our main results show that the isa mixes much faster than the over-damped Langevin dynamics. Therefore, one expects that the isa also outperforms the over-damped Langevin dynamics for simulated annealing. In this article, we show that this is indeed the case. From a computational point of view, one has to investigate the trade-off between the theoretical improvement and the cost of doubling the dimension of the underlying state space. In this regard, one might also investigate whether the use of a ladder of increasing temperature as described in [MLLVE19] is even more beneficial for the sampling versus the computational costs of higher and higher dimensions. Hence, further studies on the computational costs are needed to decide whether isa could practically compete with state-of-the-art methods for simulated annealing, e.g. methods based on Lévy flights [Pav07] or Cuckoo's search [YD09].

There are a few directions to extend this work. From the Eyring-Kramers formulas for the isa, we obtain deviation upper bounds for the convergence to equilibrium at low temperatures. It is interesting to know whether these upper bounds are optimal, and to derive matching lower bounds. Also, we plan to extend the study of the isa to the underdamped Langevin dynamics, for which the Eyring-Kramers formula of the Poincaré constant was established in [HHS11]. Furthermore, one could also extend the isa to Lévy flights and apply it to simulated annealing for even better performance.

Organization of the paper: In Section 2, we provide background, derive the isa, present the main results and apply these results to sampling and simulated annealing. In Section 3, we give proofs of the results stated in Section 2.

2. Setting, main results and applications

In this section, we start by discussing how the isa emerges as the weak limit from parallel tempering. Then we introduce the precise setting and assumptions. After this we present the main results of this article, the Eyring-Kramers formula for the Poincaré constant and an estimate of the log-Sobolev constant for the isa. We also give indications that they are optimal. We close this section by discussing two applications: sampling Gibbs measures at low temperatures and simulated annealing.

2.1. **ISA as the weak limit of parallel tempering.** Before describing parallel tempering and isa, let us first consider the over-damped Langevin equation which is a single diffusion specified by a sufficiently smooth, non-convex energy landscape $H : \mathbb{R}^n \to \mathbb{R}$ and a temperature $\tau > 0$. It is governed by the SDE:

$$d\xi_t = -\nabla H(\xi_t)dt + \sqrt{2\tau}dB_t, \tag{2.1}$$

where $(B_t,\, t \geq 0)$ is standard Brownian motion in $\mathbb{R}^n$. The infinitesimal generator of the diffusion process (2.1) is

$$L_\tau := \tau\Delta - \nabla H \cdot \nabla.$$

Under some growth assumptions on H (e.g. those of [MS14, Section 1.2]), the over-damped Langevin equation (2.1) has a unique invariant measure with density

$$\nu^\tau(x) := \frac{1}{Z^\tau} \exp\left(-\frac{H(x)}{\tau}\right),$$

where Z^τ is the normalizing constant. This probability measure is known as the Gibbs measure with energy landscape H and temperature τ. The Dirichlet form associated with the Gibbs measure ν^τ is defined for any suitable test function $f : \mathbb{R}^d \to \mathbb{R}$ by

$$\mathcal{E}_{\nu^\tau}(f) := \int_{\mathbb{R}^n} (-L_\tau f) f d\nu^\tau = \int_{\mathbb{R}^n} \tau |\nabla f|^2 d\nu^\tau.$$

For general non-convex energy landscape H, the over-damped Langevin equation shows metastable behavior at low temperatures τ in the sense of a separation of time scales:

- In the short run, the process converges fast to a local minimum of the energy landscape H;
- In the long run, the process stays near a local minimum for exponentially long time before it jumps to another local minimum.

In the previous work of [MS14], this behavior is captured by explicit, low-temperature asymptotic formulas (known as Eyring-Kramers formulas) for the two constants $\rho, \alpha > 0$ appearing in the following two functional inequalities for the invariant measure ν^τ: the Poincaré inequality (PI(ρ))

$$\operatorname{Var}_{\nu^\tau}(f) := \int \Big(f - \int f d\nu^\tau\Big)^2 d\nu^\tau \le \frac{1}{\rho}\mathcal{E}_{\nu^\tau}(f) \tag{2.2}$$

and the log-Sobolev inequality (LSI(α))

$$\operatorname{Ent}_{\nu^\tau}(f^2) := \int f^2 \ln \frac{f^2}{\int f^2 d\nu^\tau} d\nu^\tau \le \frac{2}{\alpha}\mathcal{E}_{\nu^\tau}(f) \tag{2.3}$$

holding for all sufficiently smooth test functions $f : \mathbb{R}^n \to \mathbb{R}$.

It is understood that for larger constants $\rho, \alpha > 0$, the diffusion process tends faster to equilibrium. More precisely, the constants ρ and α are the exponential rate of relaxation to equilibrium measured in variance or relative entropy, respectively. Thus, it is useful to obtain lower bounds on the constants ρ, α, or equivalently upper bounds on their inverse ρ^{-1}, α^{-1}. Also note that the Poincaré and the log-Sobolev inequalities (2.2)–(2.3) are defined slightly different from those in [MS14], where $\mathcal{E}_{\nu^\tau}(f)$ is replaced with $\int |\nabla f|^2 d\nu^\tau$ on the right side. Thus, the constants ρ, α defined by (2.2)–(2.3) differ from those in [MS14] up to a factor of τ.

In the present work, we extend these results to an inhomogeneous diffusion, the "infinite swapping process". It arises from parallel tempering by swapping particle temperatures, which we now introduce. Given two temperatures $0 < \tau_1 < \tau_2 \ll 1$, $\tau_2 > K\tau_1$ for some $K > 1$, define two product measures on $\mathbb{R}^n \times \mathbb{R}^n$:

$$\pi^+(x_1, x_2) := \nu^{\tau_1}(x_1)\nu^{\tau_2}(x_2), \qquad \pi^-(x_1, x_2) := \nu^{\tau_2}(x_1)\nu^{\tau_1}(x_2).$$

Identify the symbols $\sigma = +, -$ with the identity and the swap permutation on $\{1, 2\}$, respectively. Then π^σ is the invariant measure of the following simple product SDE:

$$\begin{cases} dX_1 = -\nabla H(X_1)\,dt + \sqrt{2\tau_{\sigma(1)}}\,dB_1\,, \\ dX_2 = -\nabla H(X_2)\,dt + \sqrt{2\tau_{\sigma(2)}}\,dB_2\,, \end{cases}$$

where $B := (B_1, B_2)$ is standard Brownian motion in $\mathbb{R}^n \times \mathbb{R}^n$. Its infinitesimal generator consists of the two infinitesimal generators of the marginals

$$L_\sigma := L^{x_1}_{\tau_{\sigma(1)}} + L^{x_2}_{\tau_{\sigma(2)}},$$

where the superscripts indicate the variable the generators are acting on. By construction L_σ is reversible with respect to π^σ and its associated Dirichlet form is

$$\mathcal{E}_{\pi^\sigma}(f) := \int_{\mathbb{R}^n\times\mathbb{R}^n} (-L_\sigma f) f d\pi^\sigma = \mathbb{E}_{\pi^\sigma}(\tau_{\sigma(1)}|\nabla_{x_1} f|^2 + \tau_{\sigma(2)}|\nabla_{x_2} f|^2).$$

The idea of parallel tempering is to swap between the positions of X_1 and X_2. At some random times, X_1 is moved to the position of X_2 and vice-versa, so the resulting process is a Markov process with jumps. To guarantee that the invariant measure remains the same, the jump intensity is of the Metropolis form $a\, g(x_1, x_2)$, where the constant 'a' is the swapping rate of parallel tempering, and $g = \min(1, \pi^-/\pi^+)$. The resulting process is denoted by $(X_1^a(t), X_2^a(t))$.

Intuitively, larger values of 'a' lead to faster convergence to equilibrium. However, the process $(X_1^a(t), X_2^a(t))$ is not tight so it does not converge weakly as $a \to \infty$. The key idea of [DLPD12] is to swap the *temperatures* of (X_1, X_2) instead of swapping the *positions*. More precisely, they consider the following process

$$\begin{cases} d\overline{X}_1^a = -\nabla H(X_1)\,dt + \sqrt{2\tau_1 \mathbb{1}_{Z^a=0} + 2\tau_2 \mathbb{1}_{Z^a=1}}\,dB_1\,, \\ d\overline{X}_2 = -\nabla H(X_2)\,dt + \sqrt{2\tau_2 \mathbb{1}_{Z^a=0} + 2\tau_1 \mathbb{1}_{Z^a=1}}\,dB_2\,, \end{cases}$$

where Z^a is a jump process which switches from state 0 to state 1 with intensity $a\, g(\overline{X}_1^a, \overline{X}_2^a)$, and from state 1 to state 0 with intensity $a\, g(\overline{X}_2^a, \overline{X}_1^a)$. It was shown in [DLPD12] that as $a \to \infty$, the process $(\overline{X}_1^a(t), \overline{X}_2^a(t))$ converges weakly to the isa, whose dynamics is governed by the SDE (1.1). We rewrite it as

$$\begin{cases} dX_1 = -\nabla H(X_1)\,dt + \sqrt{2a_1(X_1, X_2)}\,dB_1\,, \\ dX_2 = -\nabla H(X_2)\,dt + \sqrt{2a_2(X_1, X_2)}\,dB_2\,, \end{cases} \tag{2.4}$$

where the state-dependent diffusion coefficients $a_1, a_2 : \mathbb{R}^n \times \mathbb{R}^n \to [\tau_1, \tau_2]$ are given by

$$a_1 := \tau_1\rho^+ + \tau_2\rho^- \quad \text{and} \quad a_2 := \tau_2\rho^+ + \tau_1\rho^-$$
$$\text{with} \quad \rho^+ := \frac{\pi^+}{\pi^+ + \pi^-} \quad \text{and} \quad \rho^- := \frac{\pi^-}{\pi^+ + \pi^-}.$$

The infinitesimal generator of the isa (2.4) is

$$\mathcal{L} := \rho^+ L_+ + \rho^- L_- = -\nabla H(x_1)\cdot\nabla_{x_1} - \nabla H(x_2)\cdot\nabla_{x_2} + a_1\Delta_{x_1} + a_2\Delta_{x_2},$$

which is no longer the sum of two one-particle generators due to the full-space dependent diffusion coefficients a_1, a_2. A short calculations shows that $\mathcal{L}$ is self-adjoint with respect to the invariant symmetric measure

$$\mu := \tfrac{1}{2}(\pi^+ + \pi^-). \tag{2.5}$$

Let us note that the measure μ in (2.5) is generally not of product form, which contributes to the effectiveness of the sampling, at the expense of certain complications in our analysis. The Dirichlet form associated with μ is given by

$$\mathcal{E}_\mu(f) := \int (-\mathcal{L}f) f d\mu = \frac{1}{2}\mathcal{E}_{\pi^+}(f) + \frac{1}{2}\mathcal{E}_{\pi^-}(f) = \int (a_1|\nabla_{x_1} f|^2 + a_2|\nabla_{x_2} f|^2)d\mu.$$

We also define the Fisher information

$$\mathcal{I}_\mu(f^2) := 2\mathcal{E}_\mu(f). \tag{2.6}$$

2.2. **Growth and nondegeneracy assumptions.** We adopt the same assumptions on the energy landscape H as in [MS14, Section 1.2]. These assumptions are standard in the study of metastability (see e.g. [BEGK04, BGK05]).

Definition 2.1 (Morse function). *A smooth function $H : \mathbb{R}^n \to \mathbb{R}$ is a* Morse function *if the Hessian $\nabla^2 H$ of H is nondegenerate on the set of critical points. That implies, for some $1 \le C_H < \infty$ holds*

$$\forall x \in \mathcal{S} := \{z \in \mathbb{R}^n : \nabla H(z) = 0\} \quad : \qquad \frac{|\xi|}{C_H} \le |\nabla^2 H(x)\xi| \le C_H|\xi|. \tag{2.7}$$

We also make the following growth assumptions on the potential H to ensure the existence of PI and LSI.

Assumption 2.2 (PI). *$H \in C^3(\mathbb{R}^n, \mathbb{R})$ is a nonnegative Morse function, such that for some constants $C_H > 0$ and $K_H \ge 0$ holds*

$$\liminf_{|x|\to\infty} |\nabla H(x)| \ge C_H, \tag{2.8}$$

$$\liminf_{|x|\to\infty} \big(|\nabla H(x)|^2 - \Delta H(x)\big) \ge -K_H. \tag{2.9}$$

Assumption 2.3 (LSI). *$H \in C^3(\mathbb{R}^n, \mathbb{R})$ is a nonnegative Morse function, such that for some constants $C_H > 0$ and $K_H \ge 0$ holds*

$$\liminf_{|x|\to\infty} \frac{|\nabla H(x)|^2 - \Delta H(x)}{|x|^2} \ge C_H,$$

$$\inf_x \nabla^2 H(x) \ge -K_H \,\mathrm{Id}\,.$$

Remark 2.4. Assumption 2.2 has the following consequences for the energy landscape H:

- The condition (2.8) and $H(x) \ge 0$ ensures that $e^{-\frac{H}{\tau}}$ is integrable and can be normalized to a probability measure on $\mathbb{R}^n$ (see [MS14, Lemma 3.14]). Hence, the probability measures ν^τ (and therefore π^+, π^- and μ) are well-defined.
- The Morse condition (2.7) together with the growth condition (2.8) ensures that the set $\mathcal{S}$ of critical points is discrete and finite. In particular, it follows that the set of local minima is a finite set $\mathcal{M} = \{m_1, \dots, m_N\}$.
- Together with the rest of Assumption 2.2, the Lyapunov-type condition (2.9) leads to a local PI for the Gibbs measures ν^τ (see [MS14, Theorem 2.9]).

Similarly, Assumption 2.3 yields the following consequences for the energy landscape H.

- It leads to a local LSI for the Gibbs measures ν^τ (see [MS14, Theorem 2.10]).

- Assumption 2.3 implies Assumption 2.2, which is natural in light of the fact that LSI is stronger than PI.

To keep the presentation clear, we also make some nondegeneracy assumptions on the energy landscape H. First, to simplify some formulas, we assume without loss of generality throughout that

$$\min_{x\in\mathbb{R}^n} H(x) = 0.$$

The saddle height $\widehat{H}(m_i, m_j)$ between two local minima m_i, m_j is defined by

$$\widehat{H}(m_i, m_j) := \inf\left\{\max_{s\in[0,1]} H(\gamma(s)) : \gamma \in C([0,1], R^n),\ \gamma(0) = m_i,\ \gamma(1) = m_j\right\}.$$

Assumption 2.5. *Let $m_1, \cdots, m_N$ be the positions of the local minima of H.*

(i) m_1 is the unique global minimum of H, and $m_1, \ldots, m_N$ are ordered in the sense that there exists $\delta > 0$ such that

$$H(m_N) \geq H(m_{N-1}) \geq \cdots \geq H(m_2) \geq \delta > 0 = H(m_1). \tag{2.10}$$

(ii) For each $i, j \in [N] := \{1, \ldots, N\}$, the saddle height between m_i, m_j is attained at a unique critical point s_{ij} of index one. That is, $H(s_{ij}) = \widehat{H}(m_i, m_j)$, and if $\{\lambda_1, \ldots, \lambda_n\}$ are the eigenvalues of $\nabla^2 H(s_{ij})$, then $\lambda_1 =: \lambda^- < 0$ and $\lambda_i > 0$ for $i \in \{2, \ldots, n\}$. The point s_{ij} is called the communicating saddle point between the minima m_i and m_j.

(iii) There exists $p \in [N]$ such that the energy barrier $H(s_{p1}) - H(m_p)$ dominates all the others. That is, there exists $\delta > 0$ such that for all $i \in [N] \setminus \{p\}$,

$$E_* := H(s_{p1}) - H(m_p) \geq H(s_{i1}) - H(m_i) + \delta.$$

The dominating energy barrier E_* is called the critical depth.

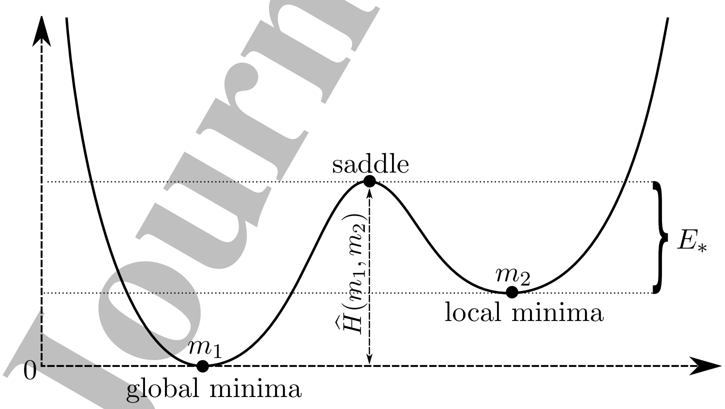

FIGURE 1. Illustration of the critical depth of a double-well function.

2.3. **The Eyring-Kramers formulas.** Our main results are the Eyring-Kramers formula for the Poincaré constant and a good estimate for log-Sobolev constant for the isa. Here a crucial new feature occurs in comparison to the over-damped Langevin dynamic: the lower temperature τ_1 cannot be arbitrarily smaller than the higher temperature τ_2 and there is an effective restriction on their ratio τ_1/τ_2. We comment on this observation in Subsection 2.4. For ease of comparison, we begin by recalling the Eyring-Kramers formulas for the Poincaré and log-Sobolev constants for the Gibbs measure ν^τ, which is the invariant measure of a single diffusion at temperature τ governed by the over-damped Langevin equation (2.1). To simplify the expression for these low-temperature asymptotic formulas, we introduce the following notation that will be used throughout the rest of this article:

$$\text{We write } A \lessapprox_\tau B \;\text{ if } A \leq B\left(1 + O(\sqrt{\tau}|\ln\tau|^{3/2})\right) \text{ as } \tau \to 0,$$

and $A \gtrapprox_\tau B$ if $B \lessapprox_\tau A$. If both $A \lessapprox_\tau B$ and $B \lessapprox_\tau A$, we write $A \approx_\tau B$.

Theorem 2.6 (Corollary 2.15 and 2.18 in [MS14])**.** *Assume* $0 < \tau \ll 1$. *Suppose that the energy landscape* H *satisfies Assumptions 2.2 and 2.5. Then the Gibbs measure* ν^τ *satisfies the Poincaré inequality* (2.2) *with the constant* ρ *satisfying*

$$\frac{1}{\rho} \lessapprox_\tau \frac{1}{\rho^\tau} := \frac{2\pi\sqrt{|\det \nabla^2 H(s_{p1})|}}{\sqrt{|\det \nabla^2 H(m_p)|}|\lambda^-(s_{p1})|} \exp\left(\frac{H(s_{p1}) - H(m_p)}{\tau}\right). \tag{2.11}$$

Here $\lambda^-(s_{p1})$ *is the negative eigenvalue of the Hessian* $\nabla^2 H(s_{p1})$ *at the communicating saddle point* s_{p1}.

Theorem 2.7 (Corollary 2.17 and 2.18 in [MS14])**.** *Assume* $0 < \tau \ll 1$. *Suppose that the energy landscape* H *satisfies Assumptions 2.3 and 2.5. Then the Gibbs measure* ν^τ *satisfies the log-Sobolev inequality* (2.3) *with the constant* α *satisfying*

$$\frac{2}{\alpha} \lessapprox_\tau \frac{2}{\alpha^\tau} := \left(\frac{H(m_p)}{\tau} + \ln\sqrt{\frac{|\det \nabla^2 H(m_1)|}{|\det \nabla^2 H(m_p)|}}\right)\frac{1}{\rho^\tau}, \tag{2.12}$$

where ρ^τ *is defined in* (2.11).

Now we are ready to state our main results.

Theorem 2.8 (Eyring-Kramers formula for the Poincaré constant for the isa)**.** *Assume that* $\tau_2 \geq K\tau_1$ *for some constant* $K > 1$. *Let* μ *be the invariant measure of the isa defined by* (2.5). *Suppose that the energy landscape* H *satisfies Assumptions 2.2 and 2.5. Then the measure* μ *satisfies the Poincaré inequality*

$$\operatorname{Var}_\mu(f) \leq \frac{1}{\rho}\mathcal{E}_\mu(f) \tag{2.13}$$

with the constant ρ *satisfying*

$$\frac{1}{\rho} \lessapprox_{\tau_2} \frac{1}{\rho^{PI}} := \frac{1}{\rho^{\tau_2}} + C\Phi_n\left(\frac{\tau_2}{\tau_1}\right). \tag{2.14}$$

Here ρ^{τ_2} is given by the formula (2.11) with $\tau = \tau_2$, C is a numerical constant independent of τ_1 and τ_2, and $\Phi_n : [1,\infty) \to [0,\infty)$ is the function

$$\Phi_n(x) = \begin{cases} 1 & \text{if } n = 1, \\ 1 + \ln x & \text{if } n = 2, \\ 1 + x^{(n-2)/2} & \text{if } n \geq 3. \end{cases} \tag{2.15}$$

Theorem 2.9 (Estimate for the log-Sobolev constant of the isa)**.** *Assume that $\tau_2 \geq K\tau_1$ for some constant $K > 1$. Let μ be the invariant measure of the isa defined by (2.5). Suppose that the energy landscape H satisfies Assumptions 2.3 and 2.5. Then the measure μ satisfies the log-Sobolev inequality*

$$\mathrm{Ent}_\mu(f) := \int f \ln f \, d\mu - \int f \, d\mu \ln \int f \, d\mu \leq \frac{1}{\alpha} \mathcal{I}_\mu(f), \tag{2.16}$$

so that $\mathrm{Ent}_\mu(f^2) \leq \frac{2}{\alpha}\mathcal{E}_\mu(f)$ with

$$\frac{2}{\alpha} \lessapprox_{\tau_2} \frac{2}{\alpha^{LSI}} := 2N^2 \left(\frac{H(m_p)}{\tau_1} + \frac{H(m_p)}{\tau_2} \right) \frac{1}{\rho^{\tau_2}} + \frac{C}{\tau_1} \Phi_n\left(\frac{\tau_2}{\tau_1}\right). \tag{2.17}$$

Here N is the number of local minima of H, ρ^{τ_2} is given by the formula (2.11) with $\tau = \tau_2$, C is a numerical constant independent of τ_1 and τ_2, and Φ_n is the function defined in (2.15).

A simple calculation shows that the terms involving Φ_n are asymptotically negligible compared to the rest of these formulas, provided τ_1 is not too small compared to τ_2:

Corollary 2.10. *Impose the condition that as $\tau_2 \to 0$,*

$$\frac{1}{\tau_1} = \begin{cases} \exp\left(o\left(\frac{1}{\tau_2}\right)\right) & \text{if } n \geq 3, \\ \exp\left(\exp\left(o\left(\frac{1}{\tau_2}\right)\right)\right) & \text{if } n = 2. \end{cases} \tag{2.18}$$

Then, with the assumptions of Theorem 2.8, the measure μ satisfies the Poincaré inequality (2.13) with the constant ρ satisfying

$$\frac{1}{\rho} \lessapprox_{\tau_2} \frac{1}{\rho^{\tau_2}}, \tag{2.19}$$

and with the assumptions of Theorem 2.9, the measure μ satisfies the log-Sobolev inequality (2.16) with the constant α satisfying

$$\frac{2}{\alpha} \lessapprox_{\tau_2} 2N^2 \left(\frac{H(m_p)}{\tau_1} + \frac{H(m_p)}{\tau_2} \right) \frac{1}{\rho^{\tau_2}}. \tag{2.20}$$

Here ρ^{τ_2} is given by the formula (2.11) with $\tau = \tau_2$.

Remark 2.11. Comparing the Eyring-Kramers formulas (2.19) and (2.20) for the isa at temperatures (τ_1, τ_2) to the corresponding formulas (2.11) and (2.12) derived for a single diffusion at the lower temperature τ_1, the main difference is that in the exponent $\frac{H(s_{p1}) - H(m_p)}{\tau_1}$, the lower temperature τ_1 is now replaced by the higher temperature τ_2, as long as $\frac{1}{\tau_1}$ grows sub-exponentially as $\frac{1}{\tau_2}$ in the limit $\tau_1, \tau_2 \to 0$.

Since we assume $\tau_2 \geq K\tau_1$ for some constant $K > 1$, this means the energy barrier $H(s_{p1}) - H(m_p)$ is effectually reduced by a factor of $K > 1$.

2.4. **Dependence on the ratio between temperatures.** The following proposition shows that the dependence on τ_2/τ_1 in the Poincaré and LSI constants of the isa is necessary and the function Φ_n that describes this dependence is nearly optimal.

Proposition 2.12. *If $\tau_2, \tau_1/\tau_2$ are sufficiently small, then there exists a constant $C > 0$ and for every $\eta > 0$, there exists a constant $C_\eta > 0$, such that*

$$\sup_{f \in H^1(\mu)} \frac{\operatorname{Var}_\mu(f)}{\mathcal{E}_\mu(f)} \geq \begin{cases} C_\eta (\tau_2/\tau_1)^{(1-\eta)(n-2)/2} & \text{for } n \geq 3, \\ C \ln(\tau_2/\tau_1) & \text{for } n = 2. \end{cases}$$

2.5. **Optimality of the Eyring-Kramers formulas in dimension one.** For the over-damped Langevin dynamics, the corresponding Eyring-Kramers formula for Poincaré inequality has been shown to be optimal. For the isa, the Poincaré constant of (2.14) is optimal in a generic one-dimensional case. This gives a strong indication of optimality in higher dimensions.

Proposition 2.13. *Assume $n = 1$, and H has three critical points: two minima $m_1 < m_2$ with $H(m_1) = 0 < \delta \leq H(m_2)$ and a local maximum s in between. Then*

$$\sup_{f \in H^1(\mu)} \frac{\operatorname{Var}_\mu(f)}{\mathcal{E}_\mu(f)} \gtrapprox_{\tau_2} \frac{1}{\rho^{\mathrm{PI}}},$$

where ρ^{PI} is given by the formula (2.14) and $H^1(\mu) := \{f : \int_{\mathbb{R}^n} |\nabla f|^2 d\mu < \infty\}$.

For the over-damped Langevin dynamics, the corresponding Eyring-Kramers formula for LSI inequality has been shown to be optimal in the one-dimensional case. For the isa, we do not expect the LSI constant of (2.17) to be optimal. However, up to the combinatorial pre-factor in the number of local minima N, it captures the asymptotic behavior for a generic one-dimensional case.

Proposition 2.14. *Assume $n = 1$, and H has three critical points: two minima $m_1 < m_2$ with $H(m_1) = 0 < \delta \leq H(m_2)$ and a local maximum s in between. Then*

$$\sup_{f \in H^1(\mu)} \frac{\operatorname{Ent}_\mu(f^2)}{\mathcal{I}_\mu(f^2)} \gtrapprox_{\tau_2} \frac{1}{\alpha^{\mathrm{LSI}}},$$

where α^{LSI} is given by the formula (2.17).

2.6. **Application to sampling.** It is well known that estimates on the Poincaré and the log-Sobolev constant yield estimates for the rate of convergence to equilibrium of the underlying process. Applying this to the isa, we obtain the following direct consequence of Theorem 2.8 and Theorem 2.9. We refer to [Sch12, Theorem 1.7] for a proof in the setting of the over-damped Langevin dynamics. The argument directly carries over to the isa.

Corollary 2.15. *Let f_t be the relative density of the isa (2.4) at time t with respect to the invariant measure μ.*

(i) Under the same assumptions as in Theorem 2.8, it holds that

$$\operatorname{Var}_\mu(f_t) \le e^{-2\rho t} \operatorname{Var}_\mu(f_0),$$

where ρ satisfies the estimate (2.14).

(ii) Under the same assumptions as in Theorem (2.9), it holds that

$$\operatorname{Ent}_\mu(f_t) \le e^{-2\alpha t} \operatorname{Ent}_\mu(f_0),$$

where α satisfies the estimate (2.17).

Another well-known consequence is that the Poincaré or log-Sobolev constant allows to quantify the ergodic theorem i.e. to estimate speed of convergence of the time average to the ensemble mean. See [CG08, Proposition 1.2] and [Wu00, Corollary 4] for a proof in the setting of the over-damped Langevin dynamics. The same argument carries over to the isa.

Corollary 2.16. *Let ν denote the initial law of the isa (2.4).*

(1) Under the same assumptions as in Theorem 2.8, it holds that for all functions $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that $\sup |f| = 1$, all $0 < R \le 1$ and all $t > 0$,

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t f(X_1(s), X_2(s))\, ds - \int f\, d\mu \ge R\right) \le \left\|\frac{d\nu}{d\mu}\right\|_{L^2} \exp\left(-\frac{tR^2\rho}{8\operatorname{Var}_\mu(f)}\right),$$

where ρ satisfies the estimate (2.14).

(2) Under the same assumptions as in Theorem 2.9, it holds that for all functions $f \in L^1(\mu)$ and all $R, t > 0$,

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t f(X_1(s), X_2(s))ds - \int f d\mu \ge R\right) \le \left\|\frac{d\nu}{d\mu}\right\|_{L^2} \exp\big(-t\alpha H^*(R)\big),$$

where α satisfies the estimate (2.17) and

$$H^*(R) := \sup_{\lambda\in\mathbb{R}}\left\{\lambda R - \ln\int \exp\left(\lambda\Big(f - \int f\, d\mu\Big)\right) d\mu\right\}.$$

Similar bounds hold for the negative deviation.

One consequence of Corollary 2.16 is that the isa has an exponential gain in comparison with the over-damped Langevin dynamics for sampling (see also Remark 2.11). The deviation bounds show an explicit dependence of the convergence on the temperatures, which is missing in the large deviation analysis in [DLPD12]. This justifies why the choice of a second higher temperature in the isa is useful, and shows how it increases the speed of convergence in the ergodic theorem.

2.7. **Application to simulated annealing.** Here we apply the log-Sobolev inequality in Theorem 2.9 to the isa for simulated annealing.

The goal of simulated annealing is to find the global minimum of a function $H : \mathbb{R}^n \to \mathbb{R}$ that is potentially non-convex. Let us explain the main idea of the stochastic version of simulated annealing. One considers a stochastic process on H subject to thermal noise. When simulating this process, one lowers the temperature slowly over time. Hereby, the stochastic process gets trapped. Now, the goal is to show that the trapped process converges to the global minimum of H with high probability. This is typically true if the temperature is lowered slowly enough. Hence, another goal is to find the best stochastic process with the fastest possible cooling schedule that still allows to approximate the global minimum.

Simulated annealing adapted to the over-damped Langevin dynamics was studied in [GH86, Mic92], see also [TZ21] for a review and results in discrete time. As we will see below, the cooling schedule has to be logarithmically slow. This implies long waiting time in order to reach the global minimum. There are many approaches to improve this behavior. Luckily, one has the freedom to choose the underlying stochastic process used for simulated annealing. One of the most efficient approach is called Cuckoo search and is based on Lévy flights (see [Pav07, YD09]). Those methods are able to find the global minimum in certain situations with a polynomial cooling schedule. An alternative is to use replica exchange or parallel tempering. As we know from [DLPD12], mixing only improves when particles are swapped faster, making the isa a natural candidate for accelerating simulated annealing.

In [Mic92], it was shown that for simulated annealing adapted to the over-damped Langevin dynamics, the fastest successful cooling schedule is characterized by the Eyring-Kramers formula for the log-Sobolev constant. However, no estimates on the associated log-Sobolev constant at low temperatures were known at that time. Hence, more sophisticated arguments were applied by [HKS89] to replace the log-Sobolev constant by the Poincaré constant showing that the fastest successful cooling schedule is characterized by the critical depth $E_* = H(s_{1p}) - H(m_p)$. Only in 2014, the Eyring-Kramers formula for the log-Sobolev constant was derived in [MS14] which leads to a more direct proof of the same result. This formula was then used by [Mon18] to study simulated annealing adapted to the underdamped Langevin dynamics, showing that it is at least as good as simulated annealing adapted to the over-damped Langevin dynamics. The main result of [HKS89, Mic92] is stated as follows.

Theorem 2.17 ([HKS89, Mic92])**.** *Let $(X_t,\, t \geq 0)$ be the process of simulated annealing adapted to the over-damped Langevin dynamics:*

$$dX_t = -\nabla H(X_t)\, dt + \sqrt{2\tau(t)}\, dB_t. \tag{2.21}$$

Let $E_* := H(s_{p1}) - H(m_p)$ denote the critical depth of the energy landscape H. Then *(i) If* $E \le \liminf_{t\to\infty} \tau(t)\ln t \le \limsup_{t\to\infty}\tau(t)\ln t < \infty$ *with* $E > E_*$*, then for all* $\delta > 0$*,*

$$\mathbb{P}\big(H(X_t) \le H(m_1)+\delta\big) \to 1 \quad \text{as } t\to\infty.$$

(ii) If $\limsup_{t\to\infty}\tau(t)\ln t \le E$ *with* $0 < E < E_*$*, then for* δ *small enough,*

$$\limsup_{t\to\infty} \mathbb{P}\big(H(X_t) \le H(m_1)+\delta\big) < 1.$$

Applying the isa to simulated annealing yields:

$$\begin{cases} dX_1 = -\nabla H(X_1)\,dt + \sqrt{2\,\tau_1(t)\,\rho(X_1,X_2) + 2\,\tau_2(t)\,\rho(X_2,X_1)}\,dB_1, \\ dX_2 = -\nabla H(X_2)\,dt + \sqrt{2\,\tau_2(t)\,\rho(X_1,X_2) + 2\,\tau_1(t)\,\rho(X_2,X_1)}\,dB_2. \end{cases} \tag{2.22}$$

We require that for some fixed constant $K > 1$

$$\tau_2(t) = K\tau_1(t) \qquad \text{and} \qquad \tau_1(t) \downarrow 0\,.$$

In Theorem 2.8 and Theorem 2.9, we showed that the infinite swapping dynamics mixes faster than the over-damped Langevin dynamics. Choosing $\tau_2 = K\tau_1$, the effective critical depth of the potential H is $\frac{E_*}{K}$ compared to E_* for simulated annealing adapted to the over-damped Langevin dynamics given by (2.21). This indicates that the infinite swapping dynamics could outperform the over-damped Langevin dynamics for simulated annealing. The main result of this section shows that this is true.

Theorem 2.18. *Assume that the energy landscape* H *satisfies Assumptions 2.3 and 2.5. Let* $E_* := H(s_{p1}) - H(m_p)$ *be the critical depth of the energy landscape* H*. For* $E > \frac{E_*}{K}, K > 1$*, let*

$$\tau_1(t) = \frac{E}{\ln(2+t)} \quad \textit{and} \quad \tau_2(t) = \frac{KE}{\ln(2+t)}. \tag{2.23}$$

Let X_1, X_2 *be given by (2.22) with initial distribution* m*. Let* $m_t(x_1,x_2)$ *be the probability density of* $(X_1(t), X_2(t))$*. Assume the following moment condition for the initial distribution* m*: for every* $p \ge 1$*, there exists a constant* C_p *such that*

$$\int \big(H(x_1) + H(x_2)\big)^p dm(x_1,x_2) \le C_p. \tag{2.24}$$

Then for all $\delta > 0$*,* $\varepsilon > 0$

$$\mathbb{P}(\min\{H(X_1(t)), H(X_2(t))\} > \delta) \lesssim \left(\frac{1}{1+t}\right)^{\min\left(\frac{\delta}{E},\frac{1}{2}-\frac{E_*}{2KE}\right)-\varepsilon}. \tag{2.25}$$

3. Proofs

3.1. **Proof of Theorem 2.8 and Theorem 2.9.** As in [MS14], we decompose the state space $\mathbb{R}^n$ into an "admissible partition" of metastable regions $\{\Omega_i\}_{i=1}^N$, as defined below.

Definition 3.1 (Admissible partition). *The family $\{\Omega_i\}_{i=1}^N$ with Ω_i open and connected is called an admissible partition for H if*

(i) for each $i \in [N]$, the local minimum $m_i \in \Omega_i$,
(ii) $\{\Omega_i\}_{i=1}^N$ forms a partition of $\mathbb{R}^n$ up to sets of Lebesgue measure zero,
(iii) The partition sum of Ω_i is approximately Gaussian. That is, there exists $\tau_0 > 0$ such that for all $\tau < \tau_0$, for $i \in [N]$,

$$\nu^\tau(\Omega_i) Z^\tau = \int_{\Omega_i} \exp\Big(-\frac{H(x)}{\tau}\Big) dx \approx_\tau \frac{(2\pi\tau)^{n/2}}{\sqrt{\det \nabla^2 H(m_i)}} \exp\Big(-\frac{H(m_i)}{\tau}\Big). \tag{3.1}$$

Remark 3.2. A canonical way to obtain an admissible partition for H is to associate to each local minimum m_i for $i \in [N]$ its basin of attraction with respect to the gradient flow of H. That is,

$$\Omega_i = \Big\{ y \in \mathbb{R}^N : \lim_{t\to\infty} y_t = m_i,\ \frac{dy_t}{dt} = -\nabla H(y_t),\ y_0 = y \Big\}.$$

However, as in [MS14], to facilitate the proof, we choose instead the basins of attraction for the gradient flow of a suitable perturbation of H (see Section 3.3).

Suppose $\{\Omega_i\}_{i=1}^N$ is an admissible partition in the sense of Definition 3.1. Define local measures on $\mathbb{R}^n$

$$\nu_i^\tau(x) := \frac{1}{Z_i^\tau} \nu^\tau(x)|_{\Omega_i}, \tag{3.2}$$

$$Z_i^\tau := \nu^\tau(\Omega_i) \approx_\tau \frac{\sqrt{\det \nabla^2 H(m_1)}}{\sqrt{\det \nabla^2 H(m_i)}}.$$

This induces a decomposition of the measure μ on $\mathbb{R}^n \times \mathbb{R}^n$ as

$$\mu = \frac{1}{2}(\pi^+ + \pi^-) = \sum \frac{1}{2} Z_{ij}^+ \pi_{ij}^+ + \sum \frac{1}{2} Z_{ij}^- \pi_{ij}^-, \tag{3.3}$$

where for $1 \le i, j \le n$, $Z_{ij}^+ := Z_i^{\tau_1} Z_j^{\tau_2}$, $Z_{ij}^- := Z_i^{\tau_2} Z_j^{\tau_1}$ and

$$\pi_{ij}^+(x_1, x_2) := \frac{1}{Z_{ij}^+} \pi^+(x_1, x_2)|_{\Omega_i \times \Omega_j} = \nu_i^{\tau_1}(x_1) \nu_j^{\tau_2}(x_2),$$

$$\pi_{ij}^-(x_1, x_2) := \frac{1}{Z_{ij}^-} \pi^-(x_1, x_2)|_{\Omega_i \times \Omega_j} = \nu_i^{\tau_2}(x_1) \nu_j^{\tau_1}(x_2).$$

The following results are read from [MS14, Lemma 2.4 and Corollary 2.8].

Lemma 3.3 (Decomposition of variance). *For the mixture representation (3.3) of the Gibbs measure μ, and a smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, it holds*

$$\operatorname{Var}_\mu(f) = \frac{1}{2} \sum Z_{ij}^+ \operatorname{Var}_{\pi_{ij}^+}(f) + \frac{1}{2} \sum Z_{ij}^- \operatorname{Var}_{\pi_{ij}^-}(f) \tag{3.4}$$

$$+ \frac{1}{4} \sum_{\sigma \in \{-,+\}} \sum Z_{ij}^\sigma Z_{kl}^\sigma (\mathbb{E}_{\pi_{ij}^\sigma}(f) - \mathbb{E}_{\pi_{kl}^\sigma}(f))^2 \tag{3.5}$$

$$+\frac{1}{4}\sum Z_{ij}^{+}Z_{kl}^{-}(\mathbb{E}_{\pi_{ij}^{+}}(f)-\mathbb{E}_{\pi_{kl}^{-}}(f))^{2}. \tag{3.6}$$

where the second line is summing over unordered pairs $\{(i,j),(k,l)\}$ and the last line is summing over ordered pairs $((i,j),(k,l))$.

Lemma 3.4 (Decomposition of entropy)**.** *For the mixture representation (3.3) of the Gibbs measure μ, and a smooth function $f:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$, it holds*

$$\operatorname{Ent}_\mu(f^2)\le\frac{1}{2}\sum Z_{ij}^{+}\operatorname{Ent}_{\pi_{ij}^{+}}(f^2)+\frac{1}{2}\sum Z_{ij}^{-}\operatorname{Ent}_{\pi_{ij}^{-}}(f^2) \tag{3.7}$$

$$+\frac{1}{2}\sum_{(i,j)}\left(\sum_{(k,l)\neq(i,j)}\frac{Z_{kl}^{+}}{\Lambda(Z_{ij}^{+},Z_{kl}^{+})}+\sum_{(k,l)}\frac{Z_{kl}^{-}}{\Lambda(Z_{ij}^{+},Z_{kl}^{-})}\right)Z_{ij}^{+}\operatorname{Var}_{\pi_{ij}^{+}}(f) \tag{3.8}$$

$$+\frac{1}{2}\sum_{(i,j)}\left(\sum_{(k,l)\neq(i,j)}\frac{Z_{kl}^{-}}{\Lambda(Z_{ij}^{-},Z_{kl}^{-})}+\sum_{(k,l)}\frac{Z_{kl}^{+}}{\Lambda(Z_{ij}^{-},Z_{kl}^{+})}\right)Z_{ij}^{-}\operatorname{Var}_{\pi_{ij}^{-}}(f) \tag{3.9}$$

$$+\frac{1}{2}\sum_{\sigma\in\{-,+\}}\sum\frac{Z_{ij}^{\sigma}Z_{kl}^{\sigma}}{\Lambda(Z_{ij}^{\sigma},Z_{kl}^{\sigma})}(\mathbb{E}_{\pi_{ij}^{\sigma}}(f)-\mathbb{E}_{\pi_{kl}^{\sigma}}(f))^2 \tag{3.10}$$

$$+\frac{1}{2}\sum\frac{Z_{ij}^{+}Z_{kl}^{-}}{\Lambda(Z_{ij}^{+},Z_{kl}^{-})}(\mathbb{E}_{\pi_{ij}^{+}}(f)-\mathbb{E}_{\pi_{kl}^{-}}(f))^2, \tag{3.11}$$

where the second to last line is summing over unordered pairs $\{(i,j),(k,l)\}$ and the last line is summing over ordered pairs $((i,j),(k,l))$. Here the function $\Lambda:[0,\infty)\times[0,\infty)\to[0,infty)$ is the logarithmic mean defined by

$$\Lambda(a,b)=\int_0^1 a^{(1-s)}b^s\,ds=\begin{cases}\frac{a-b}{\ln a-\ln b}, & a\neq b;\\ a, & a=b.\end{cases}$$

The local variances appearing in (3.4), (3.8) and (3.9) and the local entropies appearing in (3.7) are treated by the Poincaré and the log-Sobolev inequalities for local product measures.

Lemma 3.5 (Local PI for π_{ij}^{σ})**.** *Under Assumption 2.2 and given τ_2 small enough, there exists an admissible partition $\{\Omega_i\}_{i=1}^N$ such that for all $\tau\le\tau_2$, , for all smooth functions $f:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$*

$$\operatorname{Var}_{\pi_{ij}^{\sigma}}(f)\overset{(3.22)}{\le}O(1)\,\mathbb{E}_{\pi_{ij}^{\sigma}}(\tau_{\sigma(1)}|\nabla_{x_1}f|^2+\tau_{\sigma(2)}|\nabla_{x_2}f|^2).$$

Lemma 3.6 (Local LSI for π_{ij}^{σ})**.** *Under Assumption 2.3, for all smooth functions $f:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$*

$$\operatorname{Ent}_{\pi_{ij}^{\sigma}}(f^2)\overset{(3.23)}{\le}O(1)\,\mathbb{E}_{\pi_{ij}^{\sigma}}(|\nabla_{x_1}f|^2+|\nabla_{x_2}f|^2).$$

We defer the details of the proof of Lemmas 3.5 and 3.6 to Section 3.3. They are based on the simple product structure of the measures π_{ij}^{σ} and an adaption of the local Poincaré inequality [MS14, Theorem 2.9] and the local LSI inequality [MS14, Theorem 2.10]. In the sequel, for a Dirichlet form $\mathcal{E}(f)$, we denote $\mathcal{E}(f)[\Omega]$ to be the Dirichlet integral with region of integration restricted to Ω. It follows that

$$Z_{ij}^{\sigma} \operatorname{Var}_{\pi_{ij}^{\sigma}}(f) \leq O(1)\mathcal{E}_{\pi^{\sigma}}(f)[\Omega_i \times \Omega_j], \tag{3.12}$$

$$Z_{ij}^{\sigma} \operatorname{Ent}_{\pi_{ij}^{\sigma}}(f) \leq O(\tau_1^{-1})\mathcal{E}_{\pi^{\sigma}}(f)[\Omega_i \times \Omega_j]. \tag{3.13}$$

To deal with the mean-differences appearing in (3.5) and (3.10), we will apply the mean-difference estimate from [MS14, Theorem 2.12], which allows us to transport in one of the variables x_1, x_2 at a time from one metastable region Ω_j to another metastable region Ω_k. In order to ensure that we only get exponential dependence on $1/\tau_2$ rather than $1/\tau_1$ in the Eyring-Kramers formulas, we only transport in the high-temperature variable, and not in the low-temperature variable. This allows us to deal with mean-differences of the type between π_{ij}^+ and π_{ik}^+, or the type between π_{ji}^- and π_{ki}^-.

Lemma 3.7 (Mean-difference estimates for π_{ij}^+, π_{ik}^+ and for π_{ji}^-, π_{ki}^-)**.** *Let*

$$C_{jk}^{\tau_2} := \frac{2\pi\sqrt{\det \nabla^2 H(s_{jk})}}{\sqrt{\det \nabla^2 H(m_k)}|\lambda^-(s_{jk})|} \exp\left(\frac{H(s_{jk}) - H(m_k)}{\tau_2}\right),$$

then

$$Z_{ik}^+(\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\pi_{ik}^+} f)^2 \lessapprox_{\tau_2} C_{jk}^{\tau_2} \cdot \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n], \tag{3.14}$$

$$Z_{ki}^-(\mathbb{E}_{\pi_{ji}^-} f - \mathbb{E}_{\pi_{ki}^-} f)^2 \lessapprox_{\tau_2} C_{jk}^{\tau_2} \cdot \mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_i]. \tag{3.15}$$

Proof. For the first estimate, applying Cauchy-Schwarz and [MS14, Theorem 2.12], we get

$$\begin{aligned}
Z_{ik}^+(\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\pi_{ik}^+} f)^2 &\leq Z_i^{\tau_1} Z_k^{\tau_2}\, \mathbb{E}_{\nu_i^{\tau_1}}(\mathbb{E}_{\nu_j^{\tau_2}} f - \mathbb{E}_{\nu_k^{\tau_2}} f)^2 \\
&\lessapprox_{\tau_2} Z_i^{\tau_1}\, \mathbb{E}_{\nu_i^{\tau_1}}\, C_{jk}^{\tau_2} \int \tau_2 |\nabla_{x_2} f|^2 d\nu^{\tau_2}(x_2) \\
&\leq C_{jk}^{\tau_2} \cdot \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n].
\end{aligned}$$

The second estimate is completely analogous. □

To deal with the mean-differences in (3.6) and (3.11), we have another move available, which is to swap the temperatures of the two variables, i.e. to swap between π_{ij}^+ and π_{ij}^-. This is the main new technical ingredient compared to [MS14], which comes at a cost of a term involving the ratio of the higher temperature to the lower temperature, τ_2/τ_1.

Lemma 3.8 (Mean-difference estimate for π_{ij}^+, π_{ij}^-)**.**

$$(\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\pi_{ij}^-} f)^2 \leq \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(\tau_2)(\mathbb{E}_{\pi_{ij}^+} |\nabla_{x_2} f|^2 + \mathbb{E}_{\pi_{ij}^-} |\nabla_{x_1} f|^2)$$

$$+\,\omega(\tau_2)\sum_{\sigma\in\{+,-\}}\mathbb{E}_{\pi^\sigma_{ij}}(\tau_{\sigma(1)}|\nabla_{x_1}f|^2+\tau_{\sigma(2)}|\nabla_{x_2}f|^2)$$

for any smooth function $f:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$*, where* Φ_n *is the function defined in equation* (2.15) *and* $\omega(\tau_2):=O(\sqrt{\tau_2}|\ln\tau_2|^{3/2})$.

We defer the proof of this lemma to Section 3.4. It follows that

$$\min(Z^+_{ij},Z^-_{ij})(\mathbb{E}_{\pi^+_{ij}}f-\mathbb{E}_{\pi^-_{ij}}f)^2\le\Phi_n\Big(\frac{\tau_2}{\tau_1}\Big)O(1)\mathcal{E}_\mu(f). \tag{3.16}$$

Using these estimates, we will show that the dominating terms in Lemma 3.3 are the mean-differences between π^+_{ip},π^+_{11} and between π^-_{pj},π^-_{11} where i,j are arbitrary and p is the local minimum with the dominating energy barrier.

Lemma 3.9. *Let* p *be the local minimum with the dominating energy barrier. Then for any* $i,j\in[N]$*, and* $\sigma\in\{+,-\}$

$$Z^+_{ip}Z^\sigma_{11}(\mathbb{E}_{\pi^+_{ip}}(f)-\mathbb{E}_{\pi^\sigma_{11}}(f))^2\lessapprox_{\tau_2}C^{\tau_2}_{1p}\cdot\mathcal{E}_{\pi^+}(f)[\Omega_i\times\mathbb{R}^n]+\Phi_n\Big(\frac{\tau_2}{\tau_1}\Big)O(1)\mathcal{E}_\mu(f),$$
$$Z^-_{pj}Z^\sigma_{11}(\mathbb{E}_{\pi^-_{pj}}(f)-\mathbb{E}_{\pi^\sigma_{11}}(f))^2\lessapprox_{\tau_2}C^{\tau_2}_{1p}\cdot\mathcal{E}_{\pi^-}(f)[\mathbb{R}^n\times\Omega_j]+\Phi_n\Big(\frac{\tau_2}{\tau_1}\Big)O(1)\mathcal{E}_\mu(f).$$

Moreover, if $\{(i,j)^{\sigma_1},(k,l)^{\sigma_2}\}$ *is one of the following forms*

$$\{(i,1)^+,(1,1)^+\},\{(1,j)^-,(1,1)^-\},\{(i,1)^+,(1,1)^-\},\{(1,1)^+,(1,l)^-\},$$

then

$$Z^{\sigma_1}_{ij}Z^{\sigma_2}_{kl}(\mathbb{E}_{\pi^{\sigma_1}_{ij}}(f)-\mathbb{E}_{\pi^{\sigma_2}_{kl}}(f))^2\le\Phi_n\Big(\frac{\tau_2}{\tau_1}\Big)O(1)\mathcal{E}_\mu(f).$$

Finally, for any other $\{(i,j)^{\sigma_1},(k,l)^{\sigma_2}\}$*, the term* $Z^{\sigma_1}_{ij}Z^{\sigma_2}_{kl}(\mathbb{E}_{\pi^{\sigma_1}_{ij}}(f)-\mathbb{E}_{\pi^{\sigma_2}_{kl}}(f))^2$ *is negligible in the sense of being exponentially smaller in* $1/\tau_2$ *compared to one of the terms above on the right hand side.*

Proof. Let Γ be the graph whose vertices are labelled $\cdot^\sigma_{ij}$ and have three kinds of edges:

- "vertical" edges between $\cdot^+_{ij},\cdot^+_{ik}$;
- "horizontal" edges between $\cdot^-_{ij},\cdot^-_{kj}$;
- "swapping" edges between $\cdot^+_{ij},\cdot^-_{ij}$.

We decompose the mean-difference between any two measures π^+_{ij},π^-_{kl} as a sum of mean-differences of the types in (3.14), (3.15), and (3.16), corresponding to a sequence of "moves" using the edges of the graph Γ. Given any sequence of moves $v_0\to v_1\to\cdots\to v_m$ on graph Γ, we assign to each move a positive weight $\omega_t>0$,

$1 \le t \le m$, with total sum $\sum_{t=1}^{m} \omega_t = 1$. Then we have

$$Z_{v_0} Z_{v_m} (\mathbb{E}_{\pi^{v_0}} f - \mathbb{E}_{\pi^{v_m}} f)^2 = Z_{v_0} Z_{v_m} \left(\sum_{t=1}^{m} \sqrt{\omega_t} \frac{1}{\sqrt{\omega_t}} (\mathbb{E}_{\pi^{v_{t-1}}} f - \mathbb{E}_{\pi^{v_t}} f) \right)^2$$
$$\le \sum_{t=1}^{m} \frac{1}{\omega_t} Z_{v_0} Z_{v_m} (\mathbb{E}_{\pi^{v_{t-1}}} f - \mathbb{E}_{\pi^{v_t}} f)^2. \tag{3.17}$$

After taking into account the weights Z_{ij}^+, Z_{kl}^-, this leads to the choice of the following three types of sequences of moves for the three types of mean-differences occurring in Lemma 3.3:

- Type I sequence: $\cdot_{ij}^+ \to \cdot_{i1}^+ \to \cdot_{i1}^- \to \cdot_{11}^- \to \cdot_{k1}^- \to \cdot_{k1}^+ \to \cdot_{kl}^+$;
- Type II sequence: $\cdot_{ij}^- \to \cdot_{1j}^- \to \cdot_{1j}^+ \to \cdot_{11}^+ \to \cdot_{1l}^+ \to \cdot_{1l}^- \to \cdot_{kl}^-$;
- Type III sequence: $\cdot_{ij}^+ \to \cdot_{i1}^+ \to \cdot_{i1}^- \to \cdot_{11}^- \to \cdot_{11}^+ \to \cdot_{1l}^+ \to \cdot_{1l}^- \to \cdot_{kl}^-$.

Let us first look at the decomposition (3.17) for a Type I sequence. For the 1st move,

$$Z_{ij}^+ Z_{kl}^+ (\mathbb{E}_{\pi_{ij}^+}(f) - \mathbb{E}_{\pi_{i1}^+}(f))^2 \lessapprox_{\tau_2} Z_{kl}^+ C_{j1}^{\tau_2} \cdot \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n],$$

which is negligible unless $j = p, k = l = 1$. For the 2nd move,

$$Z_{ij}^+ Z_{kl}^+ (\mathbb{E}_{\pi_{i1}^+}(f) - \mathbb{E}_{\pi_{i1}^-}(f))^2 \le Z_j^{\tau_2} Z_{kl}^+ \cdot \Phi_n \left(\frac{\tau_2}{\tau_1} \right) O(1) \mathcal{E}_\mu(f),$$

which is negligible unless $j = k = l = 1$. For the 3rd move,

$$Z_{ij}^+ Z_{kl}^+ (\mathbb{E}_{\pi_{i1}^-}(f) - \mathbb{E}_{\pi_{11}^-}(f))^2 \lessapprox_{\tau_2} e^{-H(m_i)\left(\frac{1}{\tau_1} - \frac{1}{\tau_2}\right)} Z_j^{\tau_2} Z_{kl} C_{1i}^{\tau_2} \cdot \mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_1],$$

which is always negligible. The analysis for the remaining three moves are completely symmetric: the 4th move is always negligible, the 5th move is negligible unless $i = j = l = 1$, and the 6th move is negligible unless $l = p, i = j = 1$.

Overall, if $(i, j), (k, l)$ is not one of the exceptions mentioned, we can just assign $\omega_1 = \omega_1 = \cdots = \omega_6 = 1/6$, then the overall sum is negligible. This choice of $(\omega_t)_{t=1}^6$ also works in the exceptional cases $k = j = l = 1$ and $i = j = l = 1$ (since we can afford to lose a constant factor because of the $O(1)$).

Lastly, in the exceptional case $j = p, k = l = 1$, we consider a shortened 2-move sequence $\cdot_{ip}^+ \to \cdot_{i1}^+ \to \cdot_{11}^+$. For the 1st move in this sequence,

$$Z_{ip}^+ Z_{11}^+ (\mathbb{E}_{\pi_{ij}^+}(f) - \mathbb{E}_{\pi_{i1}^+}(f))^2 \lessapprox_{\tau_2} C_{p1}^{\tau_2} \cdot \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n]$$

and for the 2nd move in this sequence,

$$Z_{ip}^+ Z_{11}^+ (\mathbb{E}_{\pi_{i1}^+}(f) - \mathbb{E}_{\pi_{11}^+}(f))^2 \approx_{\tau_2} Z_p^{\tau_2} \cdot Z_{i1}^+ Z_{11}^+ (\mathbb{E}_{\pi_{i1}^+}(f) - \mathbb{E}_{\pi_{11}^+}(f))^2$$
$$\lessapprox_{\tau_2} Z_p^{\tau_2} \cdot \Phi_n \left(\frac{\tau_2}{\tau_1} \right) O(1) \mathcal{E}_\mu(f).$$

Thus, for this sequence, we can assign $\omega_1 = 1 - Z_p^{\tau_2} \approx_{\tau_2} 1, \omega_2 = Z_p^{\tau_2}$, then the overall sum is as claimed. The exceptional case $l = p, i = j = 1$ is completely symmetric.

The analysis for Type II and Type III sequences are completely analogous. □

We can adapt this approach to estimate the terms in Lemma 3.4.

Lemma 3.10. *Let p be the local minimum with the dominating energy barrier. Then for $i, k, l \in [N]$ and $\sigma \in \{+, -\}$ such that*

$$H(m_i) < H(m_p) \text{ or } i = p, \text{ and } \frac{H(m_i)}{\tau_1} + \frac{H(m_p)}{\tau_2} \geq \frac{H(m_k)}{\tau_{\sigma(1)}} + \frac{H(m_l)}{\tau_{\sigma(2)}},$$

it holds that

$$\frac{Z_{ip}^+ Z_{kl}^\sigma}{\Lambda(Z_{ip}^+, Z_{kl}^\sigma)} (\mathbb{E}_{\pi_{ip}^+}(f) - \mathbb{E}_{\pi_{kl}^\sigma}(f))^2 \lessapprox_{\tau_2} \frac{1}{\Lambda(\frac{Z_{ip}^+}{Z_{kl}^\sigma}, 1)} \left(C_{1p}^{\tau_2} \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n] + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1) \mathcal{E}_\mu(f) \right),$$

$$\frac{Z_{pi}^- Z_{kl}^\sigma}{\Lambda(Z_{pi}^-, Z_{kl}^\sigma)} (\mathbb{E}_{\pi_{pi}^-}(f) - \mathbb{E}_{\pi_{kl}^\sigma}(f))^2 \lessapprox_{\tau_2} \frac{1}{\Lambda(\frac{Z_{pi}^-}{Z_{kl}^\sigma}, 1)} \left(C_{1p}^{\tau_2} \mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_i] + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1) \mathcal{E}_\mu(f) \right).$$

Finally, for any other $\{(i,j)^{\sigma_1}, (k,l)^{\sigma_2}\}$, the term $\dfrac{Z_{ij}^{\sigma_1} Z_{kl}^{\sigma_2}}{\Lambda(Z_{ij}^{\sigma_1}, Z_{kl}^{\sigma_2})} (\mathbb{E}_{\pi_{ij}^{\sigma_1}}(f) - \mathbb{E}_{\pi_{kl}^{\sigma_2}}(f))^2$ is negligible in the sense of being exponentially smaller in $1/\tau_2$ compared to one of the terms above on the right hand side.

Proof. The analysis is similar as in the previous lemma, but now we have to take into account the logarithmic mean, using the estimate

$$\frac{ab}{\Lambda(a,b)} = a \cdot \frac{b}{\Lambda(a/b, 1)} \lessapprox_\tau a \ln(1/a)$$

for $b \lessapprox_\tau 1, a \ll 1$. The main difference is that we now need to be more careful to show the transport from $\cdot_{ip}^+$ to $\cdot_{11}^+$ is negligible if $H(m_i) \geq H(m_p)$ and $i \neq p$ by choosing the alternative path: $\cdot_{ip}^+ \to \cdot_{ip}^- \to \cdot_{1p}^- \to \cdot_{1p}^+ \to \cdot_{11}^+$. □

Proof of Theorem 2.8. Combining Lemma 3.3, (3.12) and Lemma 3.9, we get

$$\begin{aligned}
\mathrm{Var}_\mu(f) \lessapprox_{\tau_2} & \frac{1}{2} \sum_{i,j} O(1) \mathcal{E}_{\pi^+}(f)[\Omega_i \times \Omega_j] + \frac{1}{2} \sum O(1) \mathcal{E}_{\pi^-}(f)[\Omega_i \times \Omega_j] \\
& + 2 \cdot \frac{1}{4} \sum_i C_{p1}^{\tau_2} \cdot \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n] + 2 \cdot \frac{1}{4} \sum_j C_{1p}^{\tau_2} \cdot \mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_j] \\
& + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1) \mathcal{E}_\mu(f)
\end{aligned}$$

$$\leq \left(O(1) + C_{1p}^{\tau_2} + \Phi_n\Big(\frac{\tau_2}{\tau_1}\Big)O(1)\right)\mathcal{E}_\mu(f),$$

as desired.

□

Proof of Theorem 2.9. Combining Lemma 3.4, (3.12), (3.13) and Lemma 3.10, we get

$$\begin{aligned}
\mathrm{Ent}_\mu(f) \lessapprox_{\tau_2} &\ \frac{1}{2}\sum_{i,j} O(\tau_1^{-1})\mathcal{E}_{\pi^+}(f)[\Omega_i\times\Omega_j] + \frac{1}{2}\sum O(\tau_1^{-1})\mathcal{E}_{\pi^-}(f)[\Omega_i\times\Omega_j] \\
&+ \frac{1}{2}\sum_{i,j} 2N^2 O(\tau_1^{-1})\cdot O(1)\mathcal{E}_{\pi^+}(f) + \frac{1}{2}\sum_{i,j} 2N^2 O(\tau_1^{-1})\cdot O(1)\mathcal{E}_{\pi^-}(f) \\
&+ \frac{1}{2}\sum_{i\leq p}\left(\sum_\sigma\sum_{(k,l)}\frac{1}{\Lambda(\frac{Z_{ip}^+}{Z_{kl}^\sigma},1)}\right)\left(C_{1p}^{\tau_2}\cdot\mathcal{E}_{\pi^+}(f)[\Omega_i\times\mathbb{R}^n] + \Phi_n\Big(\frac{\tau_2}{\tau_1}\Big)O(1)\mathcal{E}_\mu(f)\right) \\
&+ \frac{1}{2}\sum_{i\leq p}\left(\sum_\sigma\sum_{(k,l)}\frac{1}{\Lambda(\frac{Z_{pi}^-}{Z_{kl}^\sigma},1)}\right)\left(C_{1p}^{\tau_2}\cdot\mathcal{E}_{\pi^-}(f)[\mathbb{R}^n\times\Omega_j] + \Phi_n\Big(\frac{\tau_2}{\tau_1}\Big)O(1)\mathcal{E}_\mu(f)\right) \\
\leq &\ 2N^2\left(O(\tau_1^{-1}) + H(m_p)(\tau_1^{-1}+\tau_2^{-1})C_{1p}^{\tau_2} + O(\tau_1^{-1})\Phi_n\Big(\frac{\tau_2}{\tau_1}\Big)\right)\mathcal{E}_\mu(f),
\end{aligned}$$

as desired. □

3.2. **Proof of Theorem 2.18.** With the help of Theorem 2.9, i.e. the low-temperature asymptotics for the log-Sobolev constant, the proof of Theorem 2.18 follows the arguments in [Mic92, Mon18].

For each $t>0$, let μ_t be the probability measure given in (2.5) at temperatures $\tau_1=\tau_1(t), \tau_2=\tau_2(t)$ as defined in (2.23), i.e. $\mu_t(x_1,x_2)=\frac{1}{2}(\pi_t(x_1,x_2)+\pi_t(x_2,x_1))$, with

$$\pi_t(x_1,x_2) := \frac{1}{Z_t}\exp\left(-\frac{H(x_1)}{\tau_1(t)} - \frac{H(x_2)}{\tau_2(t)}\right),$$

where Z_t is the normalizing constant. Our first observation is that the mass of the instantaneous equilibrium μ_t concentrates around the global minimum $\min H = 0$ as $t\to\infty$.

Lemma 3.11. *If $(\tilde{X}_1(t),\tilde{X}_2(t))$ has law μ_t, then for every $0<\varepsilon<\delta$, there exists a constant $C>0$ such that*

$$\mathbb{P}(\min\{H(\tilde{X}_1(t)),H(\tilde{X}_2(t))\}>\delta) \leq Ce^{-\frac{\delta-\varepsilon}{\tau_1(t)}} \leq C(2+t)^{-\frac{\delta-\varepsilon}{E}}.$$

Proof. Since $\mu_t(x_1, x_2) = \frac{1}{2}(\pi_t(x_1, x_2) + \pi_t(x_2, x_1))$, and $\min(H(x_1), H(x_2))$ is symmetric,

$$\begin{aligned}\mathbb{P}(\min\{H(\tilde{X}_1(t)), H(\tilde{X}_2(t))\} > \delta) &= \mathbb{P}(\min\{H(\tilde{Y}_1), H(\tilde{Y}_2)\} > \delta)\\ &= \mathbb{P}(H(\tilde{Y}_1) > \delta)\mathbb{P}(H(\tilde{Y}_2) > \delta)\\ &\le \mathbb{P}(H(\tilde{Y}_1) > \delta),\end{aligned}$$

where $(\tilde{Y}_1, \tilde{Y}_2)$ has law π_t, and $\tilde{Y}_1, \tilde{Y}_2$ are independent. It remains to bound

$$\mathbb{P}(H(\tilde{Y}_1) > \delta) = \frac{\int_{H(x)>\delta} e^{-\frac{H(x)}{\tau_1}}\,dx}{\int e^{-\frac{H(x)}{\tau_1}}\,dx}.$$

Under Assumption 2.3, [MS14, Lemma 3.14] applies and shows H has linear growth at infinity. More specifically, there exists a constant C_H such that for all sufficiently large R,

$$H(x) \ge \min_{|z|=R} H(z) + C(|x| - R) \quad \text{for } |x| > R.$$

In the above, we can choose R large enough so that $\min_{|z|=R} H(z) > \delta$. Then

$$\begin{aligned}\int_{H(x)>\delta} e^{-\frac{H(x)}{\tau_1}}\,dx &= \int_{H(x)>\delta, |x|<R} e^{-\frac{H(x)}{\tau_1}}\,dx + \int_{|x|>R} e^{-\frac{H(x)}{\tau_1}}\,dx\\ &\le e^{-\frac{\delta}{\tau_1}}\left(|B_R(0)| + \int_{|x|>R} e^{-\frac{C(|x|-R)}{\tau_1}}\,dx\right)\\ &\le e^{-\frac{\delta}{\tau_1}}(|B_R(0)| + O(\tau_1)).\end{aligned}$$

On the other hand, there exists $r > 0$ such that $H(x) < \varepsilon$ when $|x| < r$. Then

$$\int e^{-\frac{H(x)}{\tau_1}}\,dx > \int_{|x|<r} e^{-\frac{H(x)}{\tau_1}}\,dx > e^{-\frac{\varepsilon}{\tau_1}}|B_r(0)|.$$

Combining these gives the desired estimate. □

Let $(\tilde{X}_1(t), \tilde{X}_2(t))$ be a random vector with law μ_t. By Lemma 3.11 and Pinsker's inequality, we have

$$\begin{aligned}\mathbb{P}(\min\{H(X_1(t)), H(X_2(t))\} > \delta) &\le \mathbb{P}(\min\{H(\tilde{X}_1(t)), H(\tilde{X}_2(t))\} > \delta) + d_{TV}(\mu_t, m_t)\\ &\le C(2+t)^{-\frac{\delta-\varepsilon}{E}} + \sqrt{2\,\mathrm{Ent}(m_t|\mu_t)},\end{aligned} \tag{3.18}$$

where

$$\mathrm{Ent}(m_t|\mu_t) := \int \frac{m_t}{\mu_t} \ln\left(\frac{m_t}{\mu_t}\right) d\mu_t$$

is the relative entropy of m_t with respect to μ_t. Thus, it remains to bound $\mathrm{Ent}(m_t|\mu_t)$. The following lemma gives an estimate of $\frac{d}{dt}\mathrm{Ent}(m_t|\mu_t)$, the proof of which is in the same spirit of [Mic92, Proposition 3].

Lemma 3.12. *It holds with $\mathcal{I}_\mu(\cdot)$ defined in (2.6) the estimate*

$$\begin{aligned}\frac{d}{dt}\operatorname{Ent}(m_t|\mu_t) \le -2\mathcal{I}_{\mu_t}\left(\frac{m_t}{\mu_t}\right) \\ + \frac{d}{dt}\left(\frac{1}{\tau_1(t)}+\frac{1}{\tau_2(t)}\right)\mathbb{E}[H(X_1(t))+H(X_2(t))].\end{aligned} \tag{3.19}$$

Proof. First note that

$$\begin{aligned}\frac{d}{dt}\operatorname{Ent}(m_t|\mu_t) &= \int \frac{dm_t}{dt}\ln\left(\frac{m_t}{\mu_t}\right)dx + \int m_t\frac{d}{dt}\ln\left(\frac{m_t}{\mu_t}\right)dx \\ &= \int \frac{dm_t}{dt}\ln\left(\frac{m_t}{\mu_t}\right)dx + \int \frac{dm_t}{dt}dx - \int \frac{m_t}{\mu_t}\frac{d\mu_t}{dt}dx \\ &= \int \frac{dm_t}{dt}\ln\left(\frac{m_t}{\mu_t}\right)dx - \int \frac{d\ln(\mu_t)}{dt}dm_t. \end{aligned} \tag{3.20}$$

We consider the first term in (3.20). Observe that m_t satisfies the Fokker-Planck equation

$$\frac{dm_t}{dt} = \nabla_{x_1}\cdot(m_t\nabla_{x_1}H) + \nabla_{x_2}\cdot(m_t\nabla_{x_2}H) + \Delta_{x_1}(a_1 m_t) + \Delta_{x_2}(a_2 m_t).$$

Combining this with the identity $\nabla_{x_i}(a_i\mu_t) = -\mu_t\nabla_{x_i}H$, we get

$$\frac{dm_t}{dt} = \nabla_{x_1}\cdot\left(a_1\mu_t\nabla_{x_1}\left(\frac{m_t}{\mu_t}\right)\right) + \nabla_{x_2}\cdot\left(a_2\mu_t\nabla_{x_2}\left(\frac{m_t}{\mu_t}\right)\right).$$

Integrating by parts, we have

$$\begin{aligned}\int \frac{dm_t}{dt}\ln\left(\frac{m_t}{\mu_t}\right)dx &= -\int\left(a_1\left|\nabla_{x_1}\left(\frac{m_t}{\mu_t}\right)\right|^2 + a_2\left|\nabla_{x_2}\left(\frac{m_t}{\mu_t}\right)\right|^2\right)\frac{\mu_t}{m_t}d\mu_t \\ &= -2\mathcal{I}_{\mu_t}\left(\frac{m_t}{\mu_t}\right),\end{aligned} \tag{3.21}$$

where $\mathcal{I}_{\mu_t}$ is the Fisher information defined in (2.6) for $\mu = \mu_t$. Next we consider the second term in (3.20). Using that $\min H = 0$ and that $\tau_1(t), \tau_2(t)$ are decreasing, direct calculation yields

$$\begin{aligned}-\frac{d\ln(\mu_t)}{dt} &\le \frac{d}{dt}\left(\frac{1}{\tau_1(t)}\right)\big(H(x_1)\rho(x_1,x_2) + H(x_2)\rho(x_2,x_1)\big) \\ &\quad + \frac{d}{dt}\left(\frac{1}{\tau_2(t)}\right)\big(H(x_1)\rho(x_2,x_1) + H(x_2)\rho(x_1,x_2)\big) \\ &\le \frac{d}{dt}\left(\frac{1}{\tau_1(t)}+\frac{1}{\tau_2(t)}\right)\big(H(x_1)+H(x_2)\big).\end{aligned}$$

Integrating this against dm_t and combining it with (3.21) yields (3.19). □

The second term on the right hand side of (3.19) are controlled via the following lemma.

Lemma 3.13. *For any $\varepsilon > 0$, there exists a constant C such that*
$$\mathbb{E}\big[H(X_1(t)) + H(X_2(t))\big] \le C(1+t)^{\varepsilon}.$$

We omit the proof of Lemma 3.13, which closely follows that of [Mic92, Lemma 2], using the moment assumptions on the initial distribution m given by (2.24) and growth assumptions on the energy landscape H in Assumption 2.3.

Lemma 3.14. *For any $\varepsilon > 0$, there exists C such that*
$$\operatorname{Ent}(m_t|\mu_t) \le C\left(\frac{1}{1+3t}\right)^{1-\frac{E^*}{KE}-\varepsilon}.$$

Proof. Using the log-Sobolev inequality in Theorem 2.9, the estimate (3.19) becomes
$$\frac{d}{dt}\operatorname{Ent}(m_t|\mu_t) \le -2\alpha_t \operatorname{Ent}(m_t|\mu_t) + \frac{2}{E}(2+t)^{-1}\,\mathbb{E}[H(X_1(t)) + H(X_2(t))],$$
where α_t is the LSI constant in (2.16) for $\mu = \mu_t$. From (2.17) we see that for any $\varepsilon > 0$, there exists $t_0 > 0$ and $C_1 > 0$ such that for $t > t_0$,
$$2\alpha_t \ge C_1(2+t)^{-\frac{E_*}{KE}-\varepsilon}.$$
Together with Lemma 3.13, we get that for $t > t_0$,
$$\frac{d}{dt}\operatorname{Ent}(m_t|\mu_t) \le -C_1(1+t)^{-\frac{E_*}{E}-\varepsilon}\operatorname{Ent}(m_t|\mu_t) + C_2(1+t)^{-1+\varepsilon}.$$
A standard Gronwall-type argument as in the proof of [Mon18, Lemma 19] then finishes off the estimate. For $0 < \varepsilon < \frac{1}{2}\left(1 - \frac{E_*}{KE}\right)$, let
$$Q(t) = \operatorname{Ent}(m_t|\mu_t) - \frac{2C_2}{C_1}(1+t)^{-1+\frac{E_*}{KE}+2\varepsilon}.$$
Then for t_0 large enough and $t > t_0$,
$$\frac{d}{dt}Q(t) \le -C_1(1+t)^{-\frac{E_*}{KE}-\varepsilon}Q(t),$$
$$Q(t) \le Q(t_0)\exp\left(-C_1\int_{t_0}^{t}(1+s)^{-\frac{E_*}{KE}+\varepsilon}\,ds\right),$$
$$\operatorname{Ent}(m_t|\mu_t) \le \frac{2C_2}{C_1}(1+t)^{-1+\frac{E_*}{KE}+2\varepsilon} + \operatorname{Ent}(m_{t_0}|\mu_{t_0})\exp\left(-\frac{C_1}{\nu}\big((1+t)^{\beta} - (1+t_0)^{\beta}\big)\right),$$
where $\beta := 1 - \frac{E_*}{KE} - \varepsilon > 0$, and the conclusion follows. $\square$

Combining (3.18) and Lemma 3.14, we get that for any $\delta > 0, \varepsilon > 0$, there exists a constant C such that
$$\mathbb{P}\Big(\min\{H(X_1(t)), H(X_2(t))\} > \delta\Big) \le C\left(\left(\frac{1}{1+t}\right)^{\frac{\delta-\varepsilon}{E}} + \left(\frac{1}{1+t}\right)^{\frac{1}{2}\left(1-\frac{E^*}{KE}-\varepsilon\right)}\right),$$
which implies (2.25).

3.3. **Proof of Lemmas 3.5 and 3.6.** The following decomposition of variance and entropy for a product measure reduces proving Lemmas 3.5 and 3.6 to proving corresponding estimates for the component measures ν_i^τ.

Lemma 3.15 (Variance and entropy for product measure)**.** *Let $\pi = \nu_i \otimes \nu_j$ be a product of two probability measures on open subsets of $\mathbb{R}^n$. For any smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*

$$\begin{aligned}\mathrm{Var}_\pi(f) &= \mathbb{E}_{\nu_j}\big(\mathrm{Var}_{\nu_i}(f)\big) + \mathrm{Var}_{\nu_j}\big(\mathbb{E}_{\nu_i}(f)\big)\\ &\le \mathbb{E}_{\nu_j}\big(\mathrm{Var}_{\nu_i}(f)\big) + \mathbb{E}_{\nu_i}\big(\mathrm{Var}_{\nu_j}(f)\big).\end{aligned} \tag{3.22}$$

For any smooth function $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{>0}$,

$$\begin{aligned}\mathrm{Ent}_\pi(g) &= \mathbb{E}_{\nu_j}\big(\mathrm{Ent}_{\nu_i}(g)\big) + \mathrm{Ent}_{\nu_j}\big(\mathbb{E}_{\nu_i}(g)\big)\\ &\le \mathbb{E}_{\nu_j}\big(\mathrm{Ent}_{\nu_i}(g)\big) + \mathbb{E}_{\nu_i}\big(\mathrm{Ent}_{\nu_j}(g)\big).\end{aligned} \tag{3.23}$$

Definition 3.16 (Local PI and LSI for ν_i^τ)**.** *The local Gibbs measure ν_i^τ defined in (3.2) satisfies a Poincaré inequality with constant ρ if for all smooth functions $f : \mathbb{R}^n \to \mathbb{R}$*

$$\mathrm{Var}_{\nu_i^\tau}(f) \le \frac{1}{\rho}\,\mathbb{E}_{\nu_i^\tau}\,|\nabla f|^2,$$

which is denoted by PI(ρ). Likewise, ν_i^τ, defined in (3.2), satisfies a log-Sobolev inequality with constant α if for all smooth functions $f : \mathbb{R}^n \to \mathbb{R}$

$$\mathrm{Ent}_{\nu_i^\tau}(f^2) \le \frac{2}{\alpha}\,\mathbb{E}_{\nu_i^\tau}\,|\nabla f|^2,$$

which is denoted by LSI(α).

Lemma 3.17 (Local PI for ν_i^τ)**.** *Under Assumption 2.2, given τ_2 small enough, there exists an admissible partition $\{\Omega_i\}_{i=1}^N$ such that for all $\tau \le \tau_2$, the local Gibbs measures ν_i^τ satisfy PI(ρ) with $\rho^{-1} = O(\tau)$.*

Lemma 3.18 (Local LSI for ν_i^τ)**.** *Under Assumption 2.3, given τ_2 small enough, for the same admissible partition $\{\Omega_i\}_{i=1}^N$, for all $\tau \le \tau_2$, the local Gibbs measures ν_i^τ satisfy satisfy LSI(α) with $\alpha^{-1} = O(1)$.*

Lemmas 3.17 and 3.18 are very similar to [MS14, Theorem 2.9] and [MS14, Theorem 2.10], except now that we have two temperatures $\tau_1 < \tau_2$, we want the regions Ω_i in the admissible partition only depend on the higher temperature τ_2 but not the lower temperature τ_1, so that we can get PI and LSI for the local Gibbs measures $\nu_i^{\tau_1}, \nu_i^{\tau_2}$ at different temperatures **in the same regions** Ω_i.

This can be shown by making a small modification to the proof of [MS14, Theorem 2.9, 2.10], which is based on constructing a Lyapunov function. Let us recall the definition of a Lyapunov function and the criterion for PI based on it from [MS14].

Definition 3.19 (Lyapunov function, Definition 3.7 in [MS14])**.** *A smooth function $W_\tau : \Omega_i \to (0, \infty)$ is a Lyapunov function for ν_i^τ if the following hold for $L_\tau := \tau\Delta - \nabla H \cdot \nabla$:*

(i) There exists an open set $U_i \subset \Omega_i$ and constants $b > 0, \lambda > 0$ such that

$$\frac{L_\tau W_\tau}{W_\tau} \leq -\lambda + b\mathbb{1}_{U_i} \quad \forall x \in \Omega_i. \tag{3.24}$$

(ii) W_τ satisfies Neumann boundary condition on Ω_i in the sense that it satisfies the integration by parts formula

$$\int_{\Omega_i} (-L_\tau W_\tau) g d\nu_i^\tau = \int_{\Omega_i} \nabla g \cdot \nabla W_\tau d\nu_i^\tau. \tag{3.25}$$

Lemma 3.20 (Lyapunov condition for local PI, Theorem 3.8 in [MS14])**.** *If there exists a Lyapunov function for ν_i^τ in the sense of Definition 3.19 and that the truncated Gibbs measure $\nu_i^\tau|_{U_i}$ satisfies PI(ρ_{U_i}), then the local Gibbs measure ν_i^τ satisfies PI(ρ) with*

$$\rho^{-1} \leq \frac{b}{\lambda}\rho_{U_i}^{-1} + \frac{1}{\lambda}\tau.$$

We choose U_i to be a ball centered at the local minimum m_i with a small, fixed radius R_0 such that H is strongly convex on U_i. Then the Bakry-Émery criterion provides the following result.

Lemma 3.21 (PI for truncated Gibbs measure, Lemma 3.6 in [MS14])**.** *The measures $\nu_i^\tau|_{U_i}$ satisfy PI(ρ_{U_i}) with $\rho_{U_i}^{-1} = O(\tau)$.*

In [MS14], the candidate for the Lyapunov function is $W_\tau = \exp\left(\frac{H}{2\tau}\right)$, so that (see [MS14, equation (3.9)])

$$\frac{L_\tau W_\tau}{W_\tau} = \frac{1}{2}\Delta H(x) - \frac{1}{4\tau}|\nabla H(x)|^2.$$

In order to satisfy the condition (3.24), the Hamiltonian H was replaced by a perturbed one H_τ such that $\|H - H_\tau\|_\infty = O(\tau)$. In order to satisfy the condition (3.25), Ω_i is then chosen to be a basin of attraction with respect to the gradient flow of this perturbed Hamiltonian H_τ. Consequently, the local PI was first deduced for the perturbed Gibbs measure $\frac{1}{Z}\exp\frac{H_\tau}{2\tau}$ on Ω_i, which then implies PI for the original measure via Holley-Stroock perturbation principle. One side effect of this approach is that the region Ω_i depends on the temperature τ, which is unsuitable in our setting with two different temperatures.

We modify this approach as follows: instead of perturbing the Hamiltonian in the Gibbs measure, we only perturb the Hamiltonian in the Lyapunov function. Given $\tau_2 = \varepsilon$ small enough, we will choose a perturbation $H_\varepsilon = H + V_\varepsilon$ where $V_\varepsilon = O(\varepsilon)$, and choose Ω_i to be the basin of attraction with respect to the gradient flow of H_ε. Then, for every $\tau \leq \varepsilon$, we choose the Lyapunov function to be $W_\tau = \exp\frac{H_\varepsilon}{2\tau}$. Then (3.25) is satisfied by [MS14, Theorem B.1] and

$$\frac{L_\tau W_\tau}{W_\tau} = -\frac{\nabla H \cdot \nabla H_\varepsilon}{2\tau} + \tau\left(\frac{\Delta H_\varepsilon}{2\tau} + \frac{|\nabla H_\varepsilon|^2}{4\tau^2}\right)$$

$$= \frac{1}{2}\Delta H_\varepsilon - \frac{1}{4\tau}\left(|\nabla H|^2 - |\nabla V_\varepsilon|^2\right) \le \frac{L_\varepsilon W_\varepsilon}{W_\varepsilon},$$

where the last inequality holds as long as $|\nabla V_\varepsilon| \le |\nabla H|$. Then once (3.24) is verified for $\tau = \varepsilon$, PI for ν_i^τ follows for every $\tau \le \varepsilon$ on the same region Ω_i.

It turns out the same perturbation used in [MS14] works here. Let $\mathcal{S}$ be the set of critical points of H and $\mathcal{M} = \{m_1, m_2, \dots, m_N\}$ be the set of local minima of H.

Lemma 3.22 (ε-modification)**.** *Given a function H satisfying Assumption 2.2, there exist constants $\varepsilon_0, \lambda_0, a, C \in (0,\infty)$ and a family of C^3 functions $\{V_\varepsilon\}_{0<\varepsilon<\varepsilon_0}$ such that for $H_\varepsilon := H + V_\varepsilon$ it holds*

(i) V_ε is supported on $\bigcup_{s\in\mathcal{S}\setminus\mathcal{M}} B_{a\sqrt{\varepsilon}}(s)$ and $|V_\varepsilon(x)| \le C\varepsilon$ for all x.
(ii) Lyapunov-type condition: $|\nabla V_\varepsilon(x)| \le |\nabla H(x)|$ for all x and

$$\frac{1}{2}\Delta H_\varepsilon - \frac{1}{4\varepsilon}(|\nabla H|^2 - |\nabla V_\varepsilon|^2) \le -\lambda_0 \quad \text{for all } x \notin \bigcup_{m\in\mathcal{M}} B_{a\sqrt{\varepsilon}}(m).$$

We omit the proof of Lemma 3.22. It can be shown by carefully following the proof of [MS14, Lemma 3.12]; indeed, the perturbation V_ε can be taken to be the same one used there. It is easy to see that H_ε has the same local minima as H. For each local minimum m_i of H, let Ω_i be the associated basin of attraction w.r.t. the gradient flow defined by the τ_2-modified potential H_{τ_2}, that is

$$\Omega_i := \left\{ y \in \mathbb{R}^n : \lim_{t\to\infty} y_t = m_i,\ \frac{dy_t}{dt} = -\nabla H_{\tau_2}(y_t),\ y_0 = y \right\}.$$

Then $(\Omega_i)_{i=1}^N$ is an admissible partition in the sense of Definition 3.1. We omit the proof of this fact, which can be shown by slightly modifying the proof of [MS14, Lemma 3.12]. The preceding discussion shows ν_i^τ defined on Ω_i by (3.2) satisfies PI(ρ) with $\rho^{-1} = O(\tau)$ for all $\tau \le \tau_2$.

Equipped with the Poincaré inequality for ν_i^τ, the log-Sobolev inequality for ν_i^τ is now a simple consequence of the following criterion from [MS14].

Lemma 3.23 (Lyapunov condition for local LSI, Theorem 3.15 in [MS14])**.** *Assume the following hold:*

(i) There exists a smooth function $W_\tau : \Omega_i \to (0,\infty)$ and constants $\lambda, b > 0$ such that for $L_\tau := \tau\Delta - \nabla H \cdot \nabla$

$$\frac{L_\tau W_\tau}{W_\tau} \le -\lambda |x|^2 + b \quad \forall x \in \Omega_i.$$

(ii) $\nabla^2 H \ge -K_H$ for some $K_H > 0$ and ν_i^τ satisfies PI(ρ).
(iii) W_τ satisfies Neumann boundary condition on Ω_i (see (3.25)).

Then ν_i^τ satisfies LSI(α) with

$$\alpha^{-1} \leq 2\sqrt{\frac{\tau}{\lambda}\left(\frac{1}{2}+\frac{b+\lambda\nu_i^\tau(|x|^2)}{\rho\tau}\right)} + \frac{K_H}{\lambda}\left(\frac{1}{2}+\frac{b+\lambda\nu_i^\tau(|x|^2)}{\rho\tau}\right)+\frac{2}{\rho},$$

where $\nu_i^\tau(|x|^2)$ denotes the second moment of ν_i^τ.

Choosing W_τ to be the same Lyapunov function we chose for the PI, it is straightforward to check that, under Assumption 2.3, the conditions (i)-(iii) holds and the second moment $\nu_i^\tau(|x|^2)$ is uniformly bounded. We omit the proofs, which are virtually identical to their counterparts in [MS14] (see Lemmas 3.17-3.19). Finally, $\rho^{-1} = O(\tau)$ yields $\alpha^{-1} = O(1)$.

3.4. **Proof of Lemma 3.8.** In order to prove Lemma 3.8, we observe that the local Gibbs measures ν_i^τ are close to a class of truncated Gaussian measures in the sense of mean-difference, see [MS14, Lemma 4.6].

Definition 3.24 (Truncated Gaussian measure)**.** *Given $m \in \mathbb{R}^n$, Σ a symmetric positive definite $n \times n$ matrix, $R \geq 1$, consider the ellipsoid*

$$E^\tau := \{x \in \mathbb{R}^n : (x-m)\cdot\Sigma^{-1}(x-m) \leq R^2\tau\}.$$

The truncated Gaussian measure γ^τ at temperature τ with mean m and covariance Σ on scale R is defined to be

$$\gamma^\tau(x) := \frac{\exp\left(-\frac{1}{2\tau}(x-m)\cdot\Sigma^{-1}(x-m)\right)}{Z_R\sqrt{\tau}^n\sqrt{\det\Sigma}}\mathbb{1}_{E^\tau},$$

where Z_R is the constant needed to make this a probability density. More precisely,

$$Z_R := \int_{B_R(0)} \exp\left(-|x|^2/2\right)dx = \sqrt{2\pi}^n\left(1 - O(e^{-R^2}R^{n-2})\right).$$

Lemma 3.25 (Approximation by truncated Gaussian)**.** *For $\tau \leq \tau_2$, let γ_i^τ be the truncated Gaussian measure at temperature τ with mean m_i and covariance $\Sigma_i = (\nabla H^2(m_i))^{-1}$ on scale $R(\tau_2) = |\ln\tau_2|^{1/2}$. Then*

$$\frac{d\gamma_i^\tau}{d\nu_i^\tau}(x) \approx_{\tau_2} 1, \tag{3.26}$$

uniformly in the support of γ_i^τ, and for any smooth function $f : \mathbb{R}^n \to \mathbb{R}$

$$(\mathbb{E}_{\nu_i^\tau} f - \mathbb{E}_{\gamma_i^\tau} f)^2 \leq \mathrm{Var}_{\nu_i^\tau}\left(\frac{d\gamma_i^\tau}{d\nu_i^\tau}\right)\mathrm{Var}_{\nu_i^\tau}(f) = O(\sqrt{\tau_2}|\ln\tau_2|^{3/2})\cdot\tau\,\mathbb{E}_{\nu_i^\tau}|\nabla f|^2.$$

We omit the proof of Lemma 3.25, which is the same as [MS14, Lemma 4.6] with only minor changes.

Corollary 3.26. *For any smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*

$$\left(\mathbb{E}_{\pi_{ij}^\sigma} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}}\otimes\gamma_j^{\tau_{\sigma(2)}}} f\right)^2 = O(\sqrt{\tau_2}|\ln\tau_2|^{3/2})\cdot\mathbb{E}_{\pi_{ij}^\sigma}\left(\tau_{\sigma(1)}|\nabla_{x_1}f|^2 + \tau_{\sigma(2)}|\nabla_{x_2}f|^2\right).$$

Proof. This follows from the previous lemma by writing

$$\mathbb{E}_{\pi^\sigma_{ij}} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}} \otimes \gamma_j^{\tau_{\sigma(2)}}} f = \big(\mathbb{E}_{\nu_i^{\tau_{\sigma(1)}} \otimes \nu_j^{\tau_{\sigma(2)}}} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}} \otimes \nu_j^{\tau_{\sigma(2)}}} f\big) \\ + \big(\mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}} \otimes \nu_j^{\tau_{\sigma(2)}}} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}} \otimes \gamma_j^{\tau_{\sigma(2)}}} f\big).$$

□

This reduces our task to proving mean-difference estimate for truncated Gaussian.

Lemma 3.27 (Mean-difference estimate for truncated Gaussians at two temperatures)**.** *For any smooth function* $f : \mathbb{R}^n \to \mathbb{R}$

$$(\mathbb{E}_{\gamma_i^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_1}} f)^2 \le C_n \|\Sigma_i\| \left(1 + \Phi_n\left(\frac{\tau_2}{\tau_1}\right)\right) \tau_2 \, \mathbb{E}_{\gamma_i^{\tau_2}} |\nabla f|^2,$$

where the function Φ_n *is given by* (2.15)*, and* C_n *is a constant only depending on* n*.*

Proof. By change of variables, it suffices to show the first inequality for $m_i = 0, \Sigma_i = \mathrm{Id}$. From the Cauchy-Schwarz inequality and the fundamental theorem of calculus, we can deduce

$$\begin{aligned}
(\mathbb{E}_{\gamma_i^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_1}} f)^2 &\le \mathbb{E}_{\gamma_i^1} \big(f(\sqrt{\tau_2} X) - f(\sqrt{\tau_1} X)\big)^2 \\
&\le \int_{S^{n-1}} d\omega \int_0^R \left(\int_{\sqrt{\tau_1} r}^{\sqrt{\tau_2} r} |\nabla f(s\omega)| ds\right)^2 \frac{e^{-\frac{r^2}{2}}}{Z_R} r^{n-1} dr \\
&\le 2(I_1 + I_2),
\end{aligned}$$

where, we recall that $R \ge 1$ from Definition 3.24

$$\begin{aligned}
I_1 &:= \int_{S^{n-1}} d\omega \int_0^R \left(\int_{\sqrt{\tau_1} r}^{\sqrt{\tau_2} r} |\nabla f(s\omega)| \mathbb{1}_{s \le \sqrt{\tau_2}} ds\right)^2 \frac{e^{-\frac{r^2}{2}}}{Z_R} r^{n-1} dr, \\
I_2 &:= \int_{S^{n-1}} d\omega \int_0^R \left(\int_{\sqrt{\tau_1} r}^{\sqrt{\tau_2} r} |\nabla f(s\omega)| \mathbb{1}_{s > \sqrt{\tau_2}} ds\right)^2 \frac{e^{-\frac{r^2}{2}}}{Z_R} r^{n-1} dr.
\end{aligned}$$

Estimate for I_2: By Cauchy-Schwarz,

$$\begin{aligned}
I_2 &\le \int_{S^{n-1}} d\omega \int_0^R (\sqrt{\tau_2} r - \sqrt{\tau_1} r) \left(\int_{\sqrt{\tau_2}}^{R\sqrt{\tau_2}} |\nabla f(s\omega)|^2 \mathbb{1}_{s \le r\sqrt{\tau_2}} ds\right) \frac{e^{-\frac{r^2}{2}}}{Z_R} r^{n-1} dr \\
&\le \sqrt{\tau_2} \int_{S^{n-1}} d\omega \int_{\sqrt{\tau_2}}^{R\sqrt{\tau_2}} |\nabla f(s\omega)|^2 \left(\int_{\frac{s}{\sqrt{\tau_2}}}^R \frac{e^{-\frac{r^2}{2}}}{Z_R} r^n dr\right) ds.
\end{aligned}$$

Using integration by parts and standard Gaussian tail bound, for $s \ge \sqrt{\tau_2}$,

$$\int_{\frac{s}{\sqrt{\tau_2}}}^R e^{-\frac{r^2}{2}} r^n dr \le C_n e^{-\frac{s^2}{2\tau_2}} \left(\frac{s^2}{\tau_2}\right)^{\frac{n-1}{2}},$$

where C_n is a constant only depending on n. This gives

$$I_2 \le C_n \tau_2 \, \mathbb{E}_{\gamma_i^{\tau_2}} |\nabla f|^2.$$

Estimate for I_1: By Cauchy-Schwarz

$$\begin{aligned} I_1 &\le \int_{S^{n-1}} d\omega \int_0^R \left(\int_0^{\sqrt{\tau_2}} |\nabla f(s\omega)|^2 s^{n-1} ds \right) \left(\int_{\sqrt{\tau_1} r}^{\sqrt{\tau_2} r} s^{-(n-1)} ds \right) \frac{e^{-\frac{r^2}{2}}}{Z_R} r^{n-1} dr \\ &= \frac{1}{Z_R} \|\nabla f\|^2_{L^2(B_{\sqrt{\tau_2}}(0))} \int_0^R \left(\int_{\sqrt{\tau_1}}^{\sqrt{\tau_2}} u^{-(n-1)} du \right) r e^{-\frac{r^2}{2}} dr \\ &\le C_n e^{\frac{1}{2}} \tau_2 \, \mathbb{E}_{\gamma_i^{\tau_2}} |\nabla f|^2 \cdot \Phi_n\left(\frac{\tau_2}{\tau_1}\right), \end{aligned}$$

where C_n is a constant only depending on n. □

Corollary 3.28. *For any smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*

$$(\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_2} \otimes \gamma_j^{\tau_1}} f)^2 \le \left(1 + \Phi_n\left(\frac{\tau_2}{\tau_1}\right)\right) O(\tau_2) \big(\mathbb{E}_{\pi_{ij}^+} |\nabla_{x_2} f|^2 + \mathbb{E}_{\pi_{ij}^-} |\nabla_{x_1} f|^2 \big).$$

Proof. This follows from the previous lemma and (3.26) by writing

$$\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_2} \otimes \gamma_j^{\tau_1}} f = (\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_1}} f) + (\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_1}} f - \mathbb{E}_{\gamma_i^{\tau_2} \otimes \gamma_j^{\tau_1}} f).$$

□

Lemma 3.8 follows from Corollary 3.26 and 3.28.

Remark 3.29. *One can show a weaker version of Lemma 3.8 by a simpler approach: First we split the mean-difference as*

$$(\mathbb{E}_{\pi^+} f - \mathbb{E}_{\pi^-} f)^2 \le 2 \, \mathbb{E}_{\nu_i^{\tau_1}} (\mathbb{E}_{\nu_j^{\tau_2}} f - \mathbb{E}_{\nu_j^{\tau_1}} f)^2 + 2 \, \mathbb{E}_{\nu_j^{\tau_1}} (\mathbb{E}_{\nu_i^{\tau_1}} f - \mathbb{E}_{\nu_i^{\tau_2}} f)^2.$$

Now, using the covariance representation of mean-difference and Cauchy-Schwarz, we have

$$(\mathbb{E}_{\nu_j^{\tau_2}} f - \mathbb{E}_{\nu_j^{\tau_1}} f)^2 \le \mathrm{Var}_{\nu_j^{\tau_2}}(f) \, \mathrm{Var}_{\nu_j^{\tau_2}}\left(\frac{d\nu_j^{\tau_1}}{d\nu_j^{\tau_2}}\right) \le O(\tau_2) \, \mathbb{E}_{\nu_j^{\tau_2}} |\nabla_{x_2} f|^2 \, \mathbb{E}_{\nu_j^{\tau_1}} \left(\frac{d\nu_j^{\tau_1}}{d\nu_j^{\tau_2}}\right).$$

Finally, using the partition size given in (3.1), we have

$$\frac{d\nu_j^{\tau_1}}{d\nu_j^{\tau_2}} = \frac{\nu_j^{\tau_2}(\Omega_j)}{\nu_j^{\tau_1}(\Omega_j)} e^{-H(x)(\tau_1^{-1} - \tau_2^{-1})} \le \frac{\nu_j^{\tau_2}(\Omega_j)}{\nu_j^{\tau_1}(\Omega_j)} \lessapprox_{\tau_2} \left(\frac{\tau_2}{\tau_1}\right)^{\frac{n}{2}}.$$

3.5. **Proof of Proposition 2.12.** It suffices to consider test functions of the form $f(x, y) = f(x)$. This is equivalent to replacing μ by its first marginal, which is $\bar{\mu} = \frac{1}{2}(\nu^{\tau_1} + \nu^{\tau_2})$. In this case, $\mathrm{Var}_\mu(f)$ and $\mathcal{E}_\mu(f)$ reduces to

$$\begin{aligned} \mathrm{Var}_{\bar{\mu}}(f) &= \frac{1}{2}(\mathrm{Var}_{\nu^{\tau_1}}(f) + \mathrm{Var}_{\nu^{\tau_2}}(f)) + \frac{1}{4}(\mathbb{E}_{\nu^{\tau_1}} f - \mathbb{E}_{\nu^{\tau_2}} f)^2, \\ \mathcal{E}_{\bar{\mu}}(f) &= \frac{1}{2}(\tau_1 \, \mathbb{E}_{\nu^{\tau_1}} |\nabla f|^2 + \tau_2 \, \mathbb{E}_{\nu^{\tau_2}} |\nabla f|^2). \end{aligned}$$

We further restrict f to $C_c(\Omega_1)$. By (3.1) and (2.10), $\nu^{\tau_1}(\Omega_1), \nu^{\tau_2}(\Omega_1) \approx 1$ once $\tau_1 < \tau_2$ are small enough, so that $\frac{d\nu_1^{\tau_1}}{d\nu^{\tau_1}}, \frac{d\nu_1^{\tau_2}}{d\nu^{\tau_2}} \approx 1$ on Ω_1 (see equation (3.2)). A crude application of Young's inequality then yields

$$\mathrm{Var}_{\bar\mu}(f) \gtrsim (\mathbb{E}_{\nu^{\tau_1}} f)^2 - 4(\mathbb{E}_{\nu^{\tau_2}} f)^2 \gtrsim (\mathbb{E}_{\nu_1^{\tau_1}} f)^2 - 5(\mathbb{E}_{\nu_1^{\tau_2}} f)^2,$$
$$\mathcal{E}_{\bar\mu}(f) \lesssim \tau_1 \, \mathbb{E}_{\nu_1^{\tau_1}} |\nabla f|^2 + \tau_2 \, \mathbb{E}_{\nu_1^{\tau_2}} |\nabla f|^2,$$

where $\lesssim$ means $\le$ up to a multiplicative constant. By change of variables, we may assume $m_1 = 0, \Sigma_1 = (\nabla^2 H(m_1))^{-1} = \mathrm{Id}$. We consider a test function of the form

$$f(x) = f_\varepsilon(x) = h(|x|/\sqrt{\varepsilon}),$$

where $h \ge 0$ is a compactly supported, absolutely continuous function and $\varepsilon \in [\tau_1, \tau_2]$ is a scaling parameter, both to be specified later. As in the proof of Lemma 3.8, we will approximate by truncated Gaussian measures (see Definition 3.24). Since $\varepsilon \le \tau_2$, f_ε is supported in the support of $\gamma_1^{\tau_2}$. By Lemma 3.25,

$$\mathrm{Var}_{\bar\mu}(f) \gtrsim (\mathbb{E}_{\gamma_1^{\tau_1}} f_\varepsilon)^2 - 6(\mathbb{E}_{\gamma_1^{\tau_2}} f_\varepsilon)^2, \tag{3.27}$$
$$\mathcal{E}_{\bar\mu}(f) \lesssim \tau_1 \, \mathbb{E}_{\nu_1^{\tau_1}} |\nabla f_\varepsilon|^2 + \tau_2 \, \mathbb{E}_{\gamma_1^{\tau_2}} |\nabla f_\varepsilon|^2, \tag{3.28}$$

if τ_2 is small enough. By rescaling, we have:

$$\tau_1 \, \mathbb{E}_{\nu_1^{\tau_1}} |\nabla f_\varepsilon|^2 = \frac{\tau_1}{\varepsilon} \mathbb{E}_{\nu_1^{\frac{\tau_1}{\varepsilon}}} |\nabla f_1|^2, \tag{3.29}$$
$$\tau_2 \, \mathbb{E}_{\gamma_1^{\tau_2}} |\nabla f_\varepsilon|^2 = \frac{\tau_2}{\varepsilon} \mathbb{E}_{\gamma_1^{\frac{\tau_2}{\varepsilon}}} |\nabla f_1|^2 \le \frac{1}{\sqrt{2\pi}^n} (\varepsilon/\tau_2)^{(n-2)/2} \|\nabla f_1\|_{L^2}^2, \tag{3.30}$$
$$\mathbb{E}_{\gamma_1^{\tau_2}} f_\varepsilon = \mathbb{E}_{\gamma_1^{\frac{\tau_2}{\varepsilon}}} f_1 \le \frac{1}{\sqrt{2\pi}^n} (\varepsilon/\tau_2)^{n/2} \|f_1\|_{L^1}, \tag{3.31}$$

and for any $r \ge 0$,

$$\mathbb{E}_{\gamma_1^{\tau_1}} f_\varepsilon = \mathbb{E}_{\gamma_1^{\frac{\tau_1}{\varepsilon}}} f_1 \ge P_{\gamma_1^{\frac{\tau_1}{\varepsilon}}}(|X| \le r) \cdot \inf_{|x|\le r} f_1 \ge \left(1 - n e^{-\frac{r^2}{2n}\frac{\varepsilon}{\tau_1}}\right) \cdot \inf_{[0,r]} h. \tag{3.32}$$

In the following $R_n > 0$ is the number such that $\exp\left(-\frac{R_n^2}{2n}\right) = \frac{1}{2}$.

Case 1: $n \ge 3$. We choose h to be a compactly supported smooth function such that $h = 1$ on $[0, R_n]$, decreases to 0 on $[R_n, 2R_n]$ and is 0 outside $[0, 2R_n]$. Then

$$\tau_2 \, \mathbb{E}_{\gamma_1^{\tau_2}} |\nabla f_\varepsilon|^2 \overset{(3.30)}{\lesssim} (\varepsilon/\tau_2)^{(n-2)/2}, \ \mathbb{E}_{\gamma_1^{\tau_2}} f_\varepsilon \overset{(3.31)}{\lesssim} (\varepsilon/\tau_2)^{n/2}, \ \mathbb{E}_{\gamma_1^{\tau_1}} f_\varepsilon \overset{(3.32)}{\ge} \frac{1}{2},$$

where the implicit constants only depend on the dimension n and the function h. Since $h' = 0$ on $[0, R_n]$

$$\tau_1 \, \mathbb{E}_{\nu_1^{\tau_1}} |\nabla f_\varepsilon|^2 \overset{(3.29)}{\le} \frac{\tau_1}{\varepsilon} \|h'\|_{L^\infty}^2 P_{\nu_1^{\frac{\tau_1}{\varepsilon}}}(|X| \ge R_n) \le \frac{\tau_1}{\varepsilon} \|h'\|_{L^\infty}^2 C_H e^{-c_H \frac{\varepsilon}{\tau_1}} \lesssim_m (\tau_1/\varepsilon)^m,$$

for every positive integer m, where the constants $c_H, C_H > 0$ only depend on the Hamiltonian H. The second inequality is a consequence of Assumption 2.2 (see

[MS14, Lemma 3.13]). Now, for any $0 < \eta < \frac{1}{2}$, set $\varepsilon = \tau_1^{1-\eta}\tau_2^{\eta}$, and choose m large enough so that $\eta m \geq (1-\eta)(n-2)/2$, we obtain

$$\mathcal{E}_{\bar{\mu}}(f) \overset{(3.28)}{\lesssim_\eta} (\tau_1/\tau_2)^{(1-\eta)(n-2)/2}, \ \mathrm{Var}_{\bar{\mu}}(f) \overset{(3.27)}{\gtrsim_\eta} (\tau_2/\tau_1)^{(1-\eta)(n-2)/2}\mathcal{E}_{\bar{\mu}}(f),$$

if τ_2, τ_1/τ_2 are both small enough.

Case 2: $n = 2$. Let h be the function given by

$$h(r) = \begin{cases} 1 & \text{for } 0 \leq r \leq r_0, \\ 2(1-r^\alpha) & \text{for } r_0 \leq r \leq 1, \\ 0 & \text{for } r \geq 1, \end{cases}$$

for parameters $0 < \alpha < 1, 0 < r_0 < 1$ satisfying $r_0^\alpha = \frac{1}{2}$, to be specified later. Then h is absolutely continuous, $h' = 0$ on $[0, r_0]$, and by direct computation

$$\|f_1\|_{L^1} \leq \pi\alpha, \quad \|\nabla f_1\|_{L^\infty}^2 = \alpha^2 r_0^{-2}, \quad \|\nabla f_1\|_{L^2}^2 = 3\pi\alpha.$$

We choose $\varepsilon = \tau_2$ and $r_0^2\frac{\tau_2}{\tau_1} = R_2^2$ (which is possible once τ_1/τ_2 is small enough). Then:

$$\mathbb{E}_{\gamma_1^{\tau_2}} f_\varepsilon \overset{(3.31)}{\leq} \frac{1}{2\pi}\frac{\varepsilon}{\tau_2}\|f_1\|_{L^1} \leq \frac{\alpha}{2}, \ \mathbb{E}_{\gamma_1^{\tau_1}} f_\varepsilon \overset{(3.32)}{\geq} \frac{1}{2},$$

$$\tau_1 \mathbb{E}_{\nu_1^{\tau_1}} |\nabla f_\varepsilon|^2 \overset{(3.29)}{\leq} \frac{\tau_1}{\varepsilon}\|\nabla f_1\|_{L^\infty}^2 \leq \frac{\alpha^2}{R_2^2}, \ \tau_2 \mathbb{E}_{\gamma_1^{\tau_2}} |\nabla f_\varepsilon|^2 \overset{(3.30)}{\leq} \frac{1}{2\pi}\|\nabla f_1\|_{L^2}^2 = \frac{3\alpha}{2}.$$

Since $r_0^\alpha = \frac{1}{2}$, $\frac{1}{\alpha} = \frac{1}{2\ln 2}\ln\left(\frac{\tau_2}{\tau_1 R_2^2}\right)$. Thus

$$\mathcal{E}_{\bar{\mu}}(f) \overset{(3.28)}{\lesssim} \frac{\alpha^2}{R_2^2} + \frac{3\alpha}{2}, \quad \mathrm{Var}_{\bar{\mu}}(f) \overset{(3.27)}{\gtrsim} \frac{1}{\alpha}\mathcal{E}_{\bar{\mu}}(f) \gtrsim \ln\left(\frac{\tau_2}{\tau_1}\right)\mathcal{E}_{\bar{\mu}}(f),$$

if $\tau_2, \tau_1/\tau_2$ are both small enough.

3.6. **Proof of Proposition 2.13 and Proposition 2.14.** It suffices to consider test functions of the form $f(x, y) = g(x)g(y)$. This is equivalent to replacing μ by $\pi = \nu^{\tau_1} \otimes \nu^{\tau_2}$. In this case, $\mathrm{Var}_\mu(f), \mathrm{Ent}_\mu(f^2), \mathcal{E}_\mu(f), \mathcal{I}_\mu(f)$ reduce to

$$\mathrm{Var}_\pi(f) = \mathbb{E}_{\nu^{\tau_1}} g^2\, \mathbb{E}_{\nu^{\tau_2}} g^2 - (\mathbb{E}_{\nu^{\tau_1}} g)^2(\mathbb{E}_{\nu^{\tau_2}} g)^2,$$
$$\mathrm{Ent}_\pi(f) = \mathbb{E}_{\nu^{\tau_1}} g^2\, \mathrm{Ent}_{\nu^{\tau_2}} g^2 + \mathbb{E}_{\nu^{\tau_2}} g^2\, \mathrm{Ent}_{\nu^{\tau_1}} g^2,$$
$$\frac{1}{2}\mathcal{I}_\pi(f^2) = \mathcal{E}_\pi(f) = \tau_1\, \mathbb{E}_{\nu^{\tau_1}}(g')^2\, \mathbb{E}_{\nu^{\tau_2}} g^2 + \tau_2\, \mathbb{E}_{\nu^{\tau_1}} g^2\, \mathbb{E}_{\nu^{\tau_2}}(g')^2.$$

We represent ν^{τ_i} for $i = 1, 2$ as the mixture

$$\nu^{\tau_i} = Z_1^{\tau_i}\nu_1^{\tau_i} + Z_2^{\tau_i}\nu_2^{\tau_i} \quad \text{where } \nu_1^{\tau_i} := \nu^{\tau_i}|_{\Omega_1}, \nu_2^{\tau_i} := \nu^{\tau_i}|_{\Omega_2},$$

where $\Omega_1 := (-\infty, s), \Omega_2 := (s, \infty)$. Denote

$$Z_1^{\tau_i} := \nu^{\tau_i}(\Omega_1) \approx_{\tau_2} 1, \quad Z_2^{\tau_i} = \nu^{\tau_i}(\Omega_2) \approx_{\tau_2} \frac{\sqrt{H''(m_1)}}{\sqrt{H''(m_2)}} e^{-H(m_2)/\tau_i}.$$

Proof of Proposition 2.13 (**Optimality of PI in 1d**): Imposing $\mathbb{E}_{\nu^{\tau_1}}\, g = 0$, we get

$$\frac{\mathcal{E}_\pi(f)}{\mathrm{Var}_\pi(f)} = \tau_1 \frac{\mathbb{E}_{\nu^{\tau_1}}(g')^2}{\mathbb{E}_{\nu^{\tau_1}}\, g^2} + \tau_2 \frac{\mathbb{E}_{\nu^{\tau_2}}(g')^2}{\mathbb{E}_{\nu^{\tau_2}}\, g^2}.$$

We make the following ansatz for g:

$$g(x) = \begin{cases} g(m_1) & \text{for } x \le s-\delta, \\ g(m_1) + \frac{g(m_2)-g(m_1)}{\sqrt{2\pi\sigma\tau_2}} \cdot \kappa \int_{s-\delta}^x e^{-(y-s)^2/(2\sigma\tau_2)} dy & \text{for } s-\delta < x < s+\delta, \\ g(m_2) & \text{for } x > s+\delta, \end{cases}$$

where σ is a positive constant to be specified later, $\delta = \sqrt{2r_0\tau_2|\ln\tau_2|}$ for some positive constant r_0 to be chosen later, and κ is chosen so that g is continuous at $s+\delta$. (This is the same kind of ansatz used in [MS14, Section 2.4].) Then $\kappa = 1 + O(\tau_2^{-r_0/\sigma}) \approx 1$ once r_0 is large enough. Fix such a choice of r_0. For τ_2 small enough, δ is small enough so that

$$\mathbb{E}_{\nu^{\tau_i}}\, g \approx_{\tau_2} g(m_1) Z_1^{\tau_i} + g(m_2) Z_2^{\tau_i}.$$

This motivates the choice

$$g(m_1) \approx_{\tau_2} -1, g(m_2) \approx_{\tau_2} 1/Z_2^{\tau_1}$$

such that $\mathbb{E}_{\nu^{\tau_1}}\, g = 0$. Then

$$\begin{aligned} \mathbb{E}_{\nu^{\tau_2}}\, g^2 &\approx_{\tau_2} Z_1^{\tau_2} g(m_1)^2 + Z_2^{\tau_2} g(m_2)^2 \approx_{\tau_2} g(m_2)^2 Z_2^{\tau_2}, \\ \mathbb{E}_{\nu^{\tau_1}}\, g^2 &\approx_{\tau_2} Z_1^{\tau_1} g(m_1)^2 + Z_2^{\tau_1} g(m_2)^2 \approx_{\tau_2} g(m_2)^2 Z_2^{\tau_1}. \end{aligned}$$

Finally, we compute the Dirichlet forms. By Taylor expansion of H around s,

$$\begin{aligned} \mathbb{E}_{\nu^{\tau_2}}(g')^2 &\approx_{\tau_2} \frac{g(m_2)^2}{2\pi\sigma\tau_2} \frac{1}{Z^{\tau_2}} \int_{B_\delta(s)} e^{-(x-s)^2/(\sigma\tau_2) - H(x)/\tau_2} dx \\ &\approx_{\tau_2} \frac{g(m_2)^2}{2\pi\sigma\tau_2} \frac{\sqrt{H''(m_1)}}{\sqrt{2\pi\tau_2}} e^{-H(s)/\tau_2} \int_{B_\delta(s)} e^{-(x-s)^2/(2\tau_2)(2/\sigma + H''(s))} dx \\ &\approx_{\tau_2} g(m_2)^2 \frac{\sqrt{H''(m_1)}}{2\pi\tau_2} e^{-H(s)/\tau_2} \sqrt{|H''(s)|}, \end{aligned}$$

where we set $\sigma = 1/|H''(s)| = -1/H''(s)$. This implies

$$\tau_2 \frac{\mathbb{E}_{\nu^{\tau_2}}(g')^2}{\mathbb{E}_{\nu^{\tau_2}}\, g^2} \approx_{\tau_2} \frac{\sqrt{H''(m_2)|H''(s)|}}{2\pi} e^{(H(m_2)-H(s))/\tau_2} \approx_{\tau_2} \rho.$$

It remains to show the other term is asymptotically negligible:

$$\begin{aligned} \mathbb{E}_{\nu_{\tau_1}}(g')^2 &\lessapprox_{\tau_2} \frac{g(m_2)^2}{2\pi\sigma\tau_2} \frac{1}{Z_{\tau_1}} \int_{B_\delta(s)} e^{-(x-s)^2/(\sigma\tau_2)} dx \cdot \sup_{x\in B_\delta(s)} e^{-H(x)/\tau_1} \\ &\lessapprox_{\tau_2} \frac{g(m_2)^2}{2\pi} \frac{\sqrt{H''(m_1)|H''(s)|}}{\sqrt{2\tau_1\tau_2}} e^{-(1-\eta)H(s)/\tau_1}, \end{aligned}$$

where $\eta = O(\delta^2)$. Since $\tau_2 > K\tau_1$ for a constant $K > 1$, choosing δ sufficiently small, this implies $\tau_1 \frac{\mathbb{E}_{\nu^{\tau_1}}(g')^2}{\mathbb{E}_{\nu^{\tau_1}} g^2}$ is asymptotically negligible compared to ρ.

Proof of Proposition 2.14 (**Optimality of LSI in 1d up to constant factor**): In the same set-up as above, imposing $\mathbb{E}_{\nu^{\tau_1}} g^2 = 1$, we get

$$\frac{1}{2}\frac{\mathcal{I}_\pi(f^2)}{\mathrm{Ent}_\pi(f)} \le \tau_1 \frac{\mathbb{E}_{\nu^{\tau_1}}(g')^2}{\mathrm{Ent}_{\nu^{\tau_1}} g^2} + \tau_2 \frac{\mathbb{E}_{\nu^{\tau_2}}(g')^2}{\mathrm{Ent}_{\nu^{\tau_1}} g^2\, \mathbb{E}_{\nu^{\tau_2}} g^2}.$$

We use the same form of ansatz as before with

$$g(m_1)^2 \approx_{\tau_2} \frac{Z_2^{\tau_1}}{Z_1^{\tau_1}} \approx_{\tau_2} \frac{\sqrt{H''(m_1)}}{\sqrt{H''(m_2)}} e^{-H(m_2)/\tau_1}, \quad g(m_2)^2 = \frac{1}{g(m_1)^2}$$

such that $\mathbb{E}_{\nu^{\tau_1}} g^2 = 1$. Then

$$\mathbb{E}_{\nu^{\tau_2}} g^2 \approx_{\tau_2} Z_1^{\tau_2} g(m_1)^2 + Z_2^{\tau_2} g(m_2)^2 \approx_{\tau_2} Z_2^{\tau_2} g(m_2)^2,$$

$$\mathrm{Ent}_{\nu^{\tau_1}} g^2 \approx_{\tau_2} Z_1^{\tau_1} g(m_1)^2 \ln g(m_1)^2 + Z_2^{\tau_1} g(m_2)^2 \ln g(m_2)^2 \approx_{\tau_2} \ln g(m_2)^2 \approx_{\tau_2} \frac{H(m_2)}{\tau_1},$$

and the same computation as before shows

$$\mathbb{E}_{\nu^{\tau_1}}(g')^2 \lessapprox_{\tau_2} g(m_2)^2 \frac{\sqrt{H''(m_1)|H''(s)|}}{2\pi\sqrt{2\tau_1\tau_2}} e^{-(1-\eta)H(s)/\tau_1},$$

$$\mathbb{E}_{\nu^{\tau_2}}(g')^2 \approx_{\tau_2} g(m_2)^2 \frac{\sqrt{H''(m_1)|H''(s)|}}{2\pi\tau_2} e^{-H(s)/\tau_2},$$

where $\eta = O(\delta^2)$. This implies

$$\tau_2 \frac{\mathbb{E}_{\nu^{\tau_2}}(g')^2}{\mathrm{Ent}_{\nu^{\tau_1}} g^2\, \mathbb{E}_{\nu^{\tau_2}} g^2} \approx_{\tau_2} \tau_1 \frac{\sqrt{H''(m_2)|H''(s)|}}{2\pi H(m_2)} e^{(H(m_2)-H(s))/\tau_2} \lesssim \alpha,$$

and that $\tau_1 \frac{\mathbb{E}_{\nu^{\tau_1}}(g')^2}{\mathrm{Ent}_{\nu^{\tau_1}} g^2}$ is asymptotically negligible compared to α.

Acknowledgment

The authors want to thank Max Fathi and Paul Bressloff for the fruitful discussions and the referee for fruitful feedback and additional references. Georg Menz and André Schlichting want to thank the University of Bonn for financial support via the CRC 1060 *The Mathematics of Emergent Effects* of the University of Bonn that is funded through the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). André Schlichting is also funded by the DFG under Germany's Excellence Strategy EXC 2044–390685587, Mathematics Münster: Dynamics–Geometry–Structure. Wenpin Tang gratefully acknowledges financial support through an NSF grant DMS-2113779 and a start-up grant at Columbia University.

References

[And80] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(4):2384–2393, 1980.

[BEGK04] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)*, 6(4):399–424, 2004.

[Ber13] N. Berglund. Kramers' law: validity, derivations and generalisations. *Markov Process. Related Fields*, 19(3):459–490, 2013.

[BGK05] A. Bovier, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc. (JEMS)*, 7(1):69–99, 2005.

[BR16] F. Bouchet and J. Reygner. Generalisation of the Eyring-Kramers transition rate formula to irreversible diffusion processes. *Ann. Henri Poincaré*, 17(12):3499–3532, 2016.

[CCD+19] Y. Chen, J. Chen, J. Dong, J. Peng, and Z. Wang. Accelerating nonconvex learning via replica exchange Langevin diffusion. In *International Conference on Learning Representations (ICLR)*, 2019.

[CG08] P. Cattiaux and A. Guillin. Deviation bounds for additive functionals of Markov processes. *ESAIM Probab. Stat.*, 12:12–29, 2008.

[CS11] J. D. Chodera and M. R. Shirts. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *J. Chem. Phys.*, 135(19):194110, 2011.

[Dal17] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):651–676, 2017.

[DCWY19] R. Dwivedi, Y. Chen, M. J. Wainwright, and B. Yu. Log-concave sampling: Metropolis-Hastings algorithms are fast. *J. Mach. Learn. Res.*, 20(183):1–42, 2019.

[DDN18] J. Doll, P. Dupuis, and P. Nyquist. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms. *Appl. Math. Optim.*, 78(1):103–144, 2018.

[DLPD12] P. Dupuis, Y. Liu, N. Plattner, and J. D. Doll. On the infinite swapping limit for parallel tempering. *Multiscale Model. Simul.*, 10(3):986–1022, 2012.

[DM17] A. Durmus and E. Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587, 2017.

[DT21] J. Dong and X. T. Tong. Replica exchange for non-convex optimization. *J. Mach. Learn. Res.*, 22(173):1–59, 2021.

[Gao08] Y. Q. Gao. An integrate-over-temperature approach for enhanced sampling. *The Journal of Chemical Physics*, 128(6):064105, February 2008.

[GCS+14] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. CRC Press, Boca Raton, FL, third edition, 2014.

[GH86] S. Geman and C.-R. Hwang. Diffusions for global optimization. *SIAM Journal on Control and Optimization*, 24(5):1031–1043, 1986.

[GLPN16] G. D. Gesù, T. Lelièvre, D. L. Peutrec, and B. Nectoux. Jump Markov models and transition state theory: the quasi-stationary distribution approach. *Faraday Discuss.*, 195:469–495, 2016.

[GLPN19] G. D. Gesù, T. Lelièvre, D. L. Peutrec, and B. Nectoux. Sharp Asymptotics of the First Exit Point Density. *Annals of PDE*, 5(1), 3 2019.

[HHS11] F. Hérau, M. Hitrik, and J. Sjöstrand. Tunnel effect and symmetries for Kramers-Fokker-Planck type operators. *J. Inst. Math. Jussieu*, 10(3):567–634, 2011.

[HKN04] B. Helffer, M. Klein, and F. Nier. Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach. *Mat. Contemp.*, 26:41–85, 2004.

[HKS89] R. A. Holley, S. Kusuoka, and D. W. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, 83(2):333–347, 1989.

[HN05] B. Helffer and F. Nier. *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten Laplacians*, volume 1862 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2005.

[HN06] B. Helffer and F. Nier. Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach: the case with boundary. *Mem. Soc. Math. Fr., Nouv. Ser.*, (105):vi+89, 2006.

[HNR20] H. Hult, P. Nyquist, and C. Ringqvist. Infinite swapping algorithm for training restricted Boltzmann machines. In *Monte Carlo and quasi-Monte Carlo methods*, volume 324, pages 285–307. Springer, Cham, 2020.

[KAJ94] C. Koulamas, S. R. Antony, and R. Jaen. A survey of simulated annealing applications to operations research problems. *Omega*, 22(1):41–56, 1994.

[KGV83] S. Kirkpatrick, J. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[KZ09] S. Kannan and M. Zacharias. Simulated annealing coupled replica exchange molecular dynamics–an efficient conformational sampling method. *J. Struct. Biol.*, 166(3):288–294, 2009.

[LLPN19] T. Lelièvre, D. Le Peutrec, and B. Nectoux. Exit event from a metastable state and Eyring-Kramers law for the overdamped Langevin dynamics. In *Stochastic dynamics out of equilibrium*, volume 282, pages 331–363. Springer, Cham, 2019.

[LPA$^+$09] Y. Li, V. A. Protopopescu, N. Arnold, X. Zhang, and A. Gorin. Hybrid parallel tempering and simulated annealing method. *Appl. Math. Comput.*, 212(1):216–228, 2009.

[LVE13] J. Lu and E. Vanden-Eijnden. Infinite swapping replica exchange molecular dynamics leads to a simple simulation patch using mixture potentials. *The Journal of Chemical Physics*, 138(8):084105, February 2013.

[LVE17] J. Lu and E. Vanden-Eijnden. Methodological and computational aspects of parallel tempering methods in the infinite swapping limit, 2017.

[Mic92] L. Miclo. Recuit simulé sur $\mathbf{R}^n$. Étude de l'évolution de l'énergie libre. *Ann. Inst. H. Poincaré Probab. Statist.*, 28(2):235–266, 1992.

[MLLVE19] A. Martinsson, J. Lu, B. Leimkuhler, and E. Vanden-Eijnden. The simulated tempering method in the infinite switch limit with adaptive weight learning. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(1):013207, January 2019.

[Mon18] P. Monmarché. Hypocoercivity in metastable settings and kinetic simulated annealing. *Probab. Theory Related Fields*, 172(3-4):1215–1248, 2018.

[MS14] G. Menz and A. Schlichting. Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *Ann. Probab.*, 42(5):1809–1884, 2014.

[Nar99] K. Nara. Simulated annealing applications. In *Modern Optimisation Techniques in Power Systems*, pages 15–38. Springer, Dordrecht, 1999.

[Pav07] I. Pavlyukevich. Lévy flights, non-local search and simulated annealing. *J. Comput. Phys.*, 226(2):1830–1844, 2007.

[RC04] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, second edition, 2004.

[RT96] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[Sch12] A. Schlichting. *The Eyring-Kramers formula for Poincaré and logarithmic Sobolev inequalities*. PhD thesis, Universität Leipzig, 2012. Available at http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-97965.

[TZ21] W. Tang and X. Y. Zhou. Simulated annealing from continuum to discretization: a convergence analysis via the Eyring–Kramers law. 2021. arXiv:2102.02339.

[Č85] V. Černý. Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. *J. Optim. Theory Appl.*, 45(1):41–51, 1985.

[vLA87] P. J. M. van Laarhoven and E. H. L. Aarts. *Simulated annealing: theory and applications*, volume 37. D. Reidel Publishing Co., Dordrecht, 1987.

[Wu00] L. Wu. A deviation inequality for non-reversible Markov processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 36(4):435–445, 2000.

[WY08] L. Wu and N. Yao. Large deviation principles for Markov processes via Φ-Sobolev inequalities. *Electron. Commun. Probab.*, 13:10–23, 2008.

[YD09] X.-S. Yang and S. Deb. Cuckoo Search via Lévy flights. In *2009 World Congress on Nature Biologically Inspired Computing (NaBIC)*, pages 210–214, 2009.

[YLLG18] Z. You, L. Li, J. Lu, and H. Ge. Integrated tempering enhanced sampling method as the infinite switching limit of simulated tempering. *The Journal of Chemical Physics*, 149(8):084114, August 2018.

Department of Mathematics, UCLA.
Email address: `menz@math.ucla.edu`

Institut for Analysis and Numerics, WWU Münster.
Email address: `a.schlichting@uni-muenster.de`

Department of Industrial Engineering and Operations Research, Columbia University.
Email address: `wt2319@columbia.edu`

Department of Mathematics, UCLA.
Email address: `timwu@ucla.edu`