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 task but computationally challenging. Numerical evidence suggests that the infinite-swapping algorithm (isa) is a promising method. The isa can be seen as an improvement of the replica methods. We rigorously analyze the ergodic properties of the isa in the low temperature regime, deducing ab Eyring-Kramers formula for the spectral gap (or Poincaré constant) 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 overdamped Langevin dynamics. As a corollary, we derive a deviation 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 overdamped 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 [RC05, GSC$^+$13]. 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.

A lot of effort has 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 independent copies of the underlying dynamics. One copy evolves at the desired low temperature $\tau_1 > 0$ and the other copy with a higher temperature $1 \gg \tau_2 \gg \tau_1$. At random times the positions of both particles are swapped. This approach has the advantage that the particle at a low temperature

Date: January 17, 2020.

correctly samples the low temperature Gibbs measure whereas the particle at a high temperature can explore the full state space discovering the relevant states of the system.

Replica exchange methods and parallel tempering have been applied successfully in many different situations and they seem to accelerate sampling in low-temperature situations quite well. To the best of our knowledge, almost all evaluations of the performance of those methods are empirical and numerical. In an attempt to study the sampling performance of parallel tempering via large deviations, it was discovered that the large deviation rate function is a monotone function of the swapping rate (see [DLPD12]). It means that sampling can only improve as the swapping rate increases. This led to the discovery of the infinite swapping algorithm/process (isa), which can be interpreted as the limit of parallel tempering when swapping the particles infinitely fast (see [DLPD12], or Section 2.1 for details). Formally, given the underlying energy landscape $H : \mathbb{R}^n \to \mathbb{R}$, the isa is defined as the evolution of two particles X_t^1 and X_t^2 varying between two different temperatures $0 < \tau_1 \ll \tau_2$, given by the stochastic differential equations (SDEs):

$$\begin{cases} dX_t^1 &= -\nabla H(X_t^1)\,dt + \sqrt{2\tau_1 \rho(X_t^1, X_t^2) + 2\tau_2 \rho(X_t^2, X_t^1)}\,dB_t^1\,, \\ dX_t^2 &= -\nabla H(X_t^2)\,dt + \sqrt{2\tau_2 \rho(X_t^1, X_t^2) + 2\tau_1 \rho(X_t^2, X_t^1)}\,dB_t^2\,, \end{cases} \tag{1.1}$$

with

$$\rho(x_1, x_2) := \frac{\pi(x_1, x_2)}{\pi(x_1, x_2) + \pi(x_2, x_1)} \quad \text{and} \quad \pi(x_1, x_2) := \frac{1}{Z} \exp\left(-\frac{H(x_1)}{\tau_1} - \frac{H(x_2)}{\tau_2} \right),$$

where Z is the normalizing constant making π a probability measure. Numerical and heuristic studies [DDN17] indicate that there is an exponential gain when using the isa for sampling instead of the classical overdamped Langevin dynamics. However, no rigorous result has been established so far.

In this article we take the analysis of [DDN17] to the next level. We carry out the first rigorous study of the ergodic properties of the isa at low temperatures. Under standard non-degeneracy assumptions, we deduce the low-temperature asymptotics for the Poincaré constant and a good estimate for the log-Sobolev constant of the isa (see Theorem 2.6 and Theorem 2.7 below). In the context of metastability, those type of formulas are also known as Eyring-Kramers formulas. Comparing our results to the Eyring-Kramers formula for the overdamped Langevin dynamics (see e.g. [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 the result of Theorem 2.6 is optimal.

To the best of our knowledge, this is the first time that an Eyring-Kramers formula was derived for inhomogeneous diffusions. The reason is that usually, if the diffusion coefficient σ is inhomogeneous, the stationary and ergodic distribution μ is unknown. But for the isa (1.1), the ergodic distribution μ is explicitly known. It is given by $\mu(x_1, x_2) = \frac{1}{2}\left(\pi(x_1, x_2) + \pi(x_2, x_1)\right)$. This makes a rigorous analysis of (1.1) feasible.

For the proof of Theorem 2.6 and Theorem 2.7, we follow the transportation approach of [MS14]. There are several other methods which could be used to deduce the Eyring-Kramers formula for the Poincaré constant. For example, one could consider to adapt the potential theoretic approach (see [BEGK04, BGK05]) or the approach using semiclassical analysis (see [HKN04, HN05, HN06]). However, it seems that only the approach of [MS14] is robust enough to deduce good estimates for the log-Sobolev constant. This 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 overdamped Langevin dynamics. It is well known that the Poincaré and 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 deviation inequalities (see [CG08, WY08] and references therein). Hence, our main results yield a precise quantitative control on the convergence of the time average to the ensemble average, quantifying the ergodic theorem. As a consequence, we conclude that sampling at low temperatures using isa is exponentially faster than using the overdamped Langevin dynamics.

In the second application, we study simulated annealing for the isa and compare it to simulated annealing for the overdamped 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 and n is large. 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. It is a process that aims to increase the size of the crystals by a process involving 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 and parallel tempering have been successfully applied to 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 overdamped Langevin dynamics. Therefore, one expects that the isa also outperforms the overdamped Langevin dynamics for simulated annealing. In this article, we show that this is indeed the case. Unfortunately, from our theoretic study it is unclear if the isa could compete in practice 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 further directions from this article. We plan to extend the study of the isa to the underdamped Langevin dynamics. One could also extend the isa to Lévy flights and apply it to simulated annealing to get 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

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

2.1. Infinite-swapping as the weak limit of parallel tempering.

Before describing parallel tempering, let us consider a simpler situation: a single diffusion on an energy landscape given by a sufficiently smooth, non-convex Hamiltonian function $H : \mathbb{R}^n \to \mathbb{R}$ at a single temperature $\tau > 0$, given by the SDE

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

where B_t is a standard Brownian motion on $\mathbb{R}^n$. The generator of the diffusion is

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

The associated Dirichlet form is

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

and the Fisher information is

$$\mathcal{I}_{\nu^\tau}(f^2) := 2\mathcal{E}_{\nu^\tau}(f).$$

Under some growth assumptions on H (e.g. those of [MS14, Section 1.2]), this process (known as the overdamped Langevin dynamics) has an invariant measure given by

$$\nu^\tau(x) := \frac{1}{Z^\tau} \exp\left(-\frac{H(x)}{\tau}\right) \tag{2.1}$$

where Z^τ is the normalization constant. Due to the non-convexity of H, this process 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.
- In the long run, the process stays near a local minimum for an 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))

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

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

$$\mathrm{Ent}_{\nu^\tau}(f^2) := \int f^2 \ln \frac{f^2}{\int f^2 d\nu^\tau} d\nu^\tau \le \frac{1}{\alpha}\mathcal{I}_{\nu^\tau}(f)$$

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

In the present work, we extend these results to a non-homogeneous diffusion, the "infinite swapping process". It arises from parallel tempering, 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 $\mathbb{R}^n \times \mathbb{R}^n$

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

Let us identify the symbol $\sigma = +, -$ with the identity and swap permutation on $\{1, 2\}$, respectively. Then π^σ is the invariant measure of the following 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 a standard Brownian motion in $\mathbb{R}^n \times \mathbb{R}^n$. Its generator is

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

and the associated Dirichlet form is

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

The idea of parallel tempering is to swap between the positions of X_1 and X_2. At 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 the 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. 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 infinite swapping process, whose dynamics is governed by the SDE:

$$\begin{cases} dX_1 = -\nabla H(X_1)\, dt + \sqrt{2a_1(X_1, X_2)}\, dB_1 \,, \\ dX_2 = -\nabla H(X_2)\, dt + \sqrt{2a_2(X_1, X_2)}\, dB_2 \,, \end{cases} \tag{2.2}$$

where the diffusion coefficients a_1, a_2 are given by

$$a_1 := \tau_1 \rho^+ + \tau_2 \rho^- \quad \text{and} \quad a_2 := \tau_2 \rho^+ + \tau_1 \rho^-,$$

$$\text{where } \rho^+ := \frac{\pi^+}{\pi^+ + \pi^-} \quad \text{and} \quad \rho^- := \frac{\pi^-}{\pi^+ + \pi^-}.$$

The invariant measure of this process is the symmetric measure

$$\mu := \tfrac{1}{2}(\pi^+ + \pi^-). \tag{2.3}$$

The generator of this process is

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

The associated Dirichlet form is

$$\mathcal{E}_\mu(f) := \int (-\mathcal{L} f) f \, d\mu = \frac{1}{2} \mathcal{E}_{\pi^+}(f) + \frac{1}{2} \mathcal{E}_{\pi^-}(f) = \int \sum_{k=1}^{2} a_k |\nabla_{x_k} f|^2 d\mu$$

and the Fisher information is

$$\mathcal{I}_\mu(f^2) := 2\mathcal{E}_\mu(f). \tag{2.4}$$

2.2. **Growth and non-degeneracy assumptions.** In this article, we use the same assumptions on the potential 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 non-degenerate on the set of critical points. That is, for some $1 \le C_H < \infty$ holds*

$$\forall x \in \mathcal{S} := \big\{ x \in \mathbb{R}^n : \nabla H = 0 \big\} : \frac{|\xi|}{C_H} \le \big| \nabla^2 H(x) \xi \big| \le C_H |\xi|. \tag{2.5}$$

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*

$$\begin{aligned} \liminf_{|x| \to \infty} |\nabla H| \quad &\ge \quad C_H, \\ \liminf_{|x| \to \infty} \big(|\nabla H|^2 - \Delta H \big) \quad &\ge \quad -K_H. \end{aligned}$$

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 $K_H \geq 0$ holds*

$$\begin{aligned} \liminf_{|x|\to\infty} \frac{|\nabla H(x)|^2 - \Delta H(x)}{|x|^2} &\geq C_H, \\ \inf_x \nabla^2 H(x) &\geq -K_H. \end{aligned}$$

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

- The condition (2.6) and $H(x) \geq 0$ 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 (2.5) together with the growth condition (2.6) 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 $\mathcal{M} = \{m_1, \ldots, m_N\}$.
- Together with the rest of Assumption 2.2, the Lyapunov-type condition (2.7) 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 potential 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 an indication that LSI is stronger than PI.

To keep the presentation clear, we also make some non-degeneracy assumptions on the potential H. The *saddle height* $\widehat{H}(m_i, m_j)$ between two local minima m_i, m_j is defined by

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

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

$$H(m_N) \geq H(m_{N-1}) \geq \cdots \geq H(m_2) \geq \delta \quad \text{and} \quad H(m_1) = 0. \tag{2.8}$$

(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.

2.3. **The Eyring-Kramers formula.** 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 usual overdamped Langevin dynamic. The lower temperature cannot be arbitrarily small and there is an effective restriction on the ratio between the two temperatures τ_1 and τ_2. We comment on this observation in Subsection 2.4.

Theorem 2.6 (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 infinite swapping process defined by (2.3). Suppose that the potential H satisfies Assumptions 2.2 and 2.5. Then the Gibbs measure μ satisfies the Poincaré inequality*

$$\mathrm{Var}_\mu(f) \leq \frac{1}{\rho}\mathcal{E}_\mu(f),$$

with the constant ρ satisfying

$$\frac{1}{\rho} \leq \frac{1}{\sqrt{|\det \nabla^2 H(m_p)|}} \frac{2\pi\sqrt{|\det \nabla^2 H(s_{p1})|}}{|\lambda^-(s_{p1})|} \times \exp\left(\frac{H(s_{p1}) - H(m_p)}{\tau_2}\right)\left(1 + O(\sqrt{\tau_2}\,|\ln \tau_2|^{\frac{3}{2}})\right) + O(1)\Phi_n\left(\frac{\tau_2}{\tau_1}\right). \tag{2.9}$$

Here $\lambda^-(s_{p1})$ is the negative eigenvalue of the Hessian $\nabla^2 H(s_{p1})$ at the communicating saddle point s_{p1}, and $\Phi_n : [1,\infty) \to [0,\infty)$ is the function

$$\Phi_n(x) = \begin{cases} 1 & \text{for } n = 1 \\ 1 + \ln x & \text{for } n = 2 \\ 1 + x^{(n-2)/2} & \text{for } n \geq 3. \end{cases} \tag{2.10}$$

Theorem 2.7 (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 infinite swapping process defined by (2.3). Suppose that the potential H satisfies Assumptions 2.3 and 2.5. Then the Gibbs measure μ satisfies the log-Sobolev inequality*

$$\mathrm{Ent}_\mu(f) := \int f \ln f \, d\mu - \int f \, d\mu \ln \int f \, d\mu \leq \frac{1}{\alpha}\mathcal{I}_\mu(f), \tag{2.11}$$

with

$$\frac{2}{\alpha} \leq 2N^2\left(\frac{H(m_p)}{\tau_1} + \frac{H(m_p)}{\tau_2}\right)\frac{1}{\sqrt{|\det \nabla^2 H(m_p)|}} \frac{2\pi\sqrt{|\det \nabla^2 H(s_{p1})|}}{|\lambda^-(s_{p1})|} \times \exp\left(\frac{H(s_{p1}) - H(m_p)}{\tau_2}\right)\left(1 + O(\sqrt{\tau_2}\,|\ln \tau_2|^{\frac{3}{2}})\right) + O(\tau_1^{-1})\Phi_n\left(\frac{\tau_2}{\tau_1}\right). \tag{2.12}$$

Here, N is the number of local minima of H, $\lambda^-(s_{p1})$ is the negative eigenvalue of the Hessian $\nabla^2 H(s_{p1})$ at the communicating saddle point s_{p1}, and Φ_n is the function defined in (2.10).

Remark 2.8. *If we can ensure that τ_1 is not too low compared to τ_2, e.g. imposing a condition like*

$$\tau_1 \geq e^{-o\left(\frac{1}{\tau_2}\right)}$$

then the error terms involving $\Phi_n\left(\frac{\tau_2}{\tau_1}\right)$ in (2.9) and (2.12) become negligible, as can be seen from the form of the function Φ_n. (In fact, this restriction can be entirely dropped in dimension $n = 1$, and relaxed to $\tau_1 \geq e^{-e^{o\left(\frac{1}{\tau_2}\right)}}$ in dimension $n = 2$.) Then in this regime of temperatures τ_1, τ_2, the estimates (2.9) and (2.12) for the isa essentially reduce to the corresponding Eyring-Kramers formulas for the overdamped Langevin dynamics ***at the higher temperature*** *τ_2, given in* [MS14, Corollary 2.18]*. For the Poincaré constant this is true to the exact pre-factor, and for the LSI constant this is true to the leading exponential order. Because we choose $\tau_2 \geq K\tau_1$, this means that the effective energy barrier $H(m_p) - H(m_1)$ is reduced by a factor of $K > 1$.*

More precisely, the estimate we give for the LSI constant differs from the one in [MS14, Corollary 2.18] *(with temperature set to be τ_2) by two additional factors: first, we have $H(m_p)/\tau_1$ instead of $H(m_p)/\tau_2$ in the pre-factor, which amounts to an additional factor of τ_2/τ_1; and second, we also have a combinatorial factor on the order of N^2. Below, we show the change from $H(m_p)/\tau_2$ to $H(m_p)/\tau_1$ is necessary in a generic one-dimensional case. However, presently we do not know whether the combinatorial factor is necessary. It would be interesting to study whether this factor of N^2 can be removed from the LSI constant.*

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 formula of Φ_n that describes this dependence is close to being optimal.

Proposition 2.9. *If $\tau_2, \tau_1/\tau_2$ are sufficiently small, then for every $\eta > 0$, there exists a constant $C_\eta > 0$ such that*

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

2.5. **Optimality of the Eyring-Kramers formula in dimension one.** For the overdamped Langevin dynamics, the corresponding Eyring-Kramers formula for Poincaré inequality has been shown to be optimal. For the isa, the Poincaré constant of (2.9) is optimal in a generic one-dimensional case. This gives a strong indication of optimality in higher dimensions.

Proposition 2.10. *Assume that $\tau_2 \geq K\tau_1$ for some constant $K > 1$. 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*

$$\inf_{f \in H^1(\mu)} \frac{\mathcal{E}_\mu(f)}{\mathrm{Var}_\mu(f)} \leq \rho$$

where ρ is given by the asymptotic formula (2.9)*.*

For the overdamped 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 (2.12) to be optimal. However, up to some combinatorial factor in N, it has the asymptotic behavior for a generic one-dimensional case.

Proposition 2.11. *Assume that $\tau_2 \geq K\tau_1$ for some constant $K > 1$. 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*

$$\inf_{f \in H^1(\mu)} \frac{\mathcal{I}_\mu(f^2)}{\operatorname{Ent}_\mu(f^2)} \lesssim_N \alpha$$

where α is given by the asymptotic formulas (2.12).

2.6. **Application to sampling.** It is well known that estimates on the Poincaré and log-Sobolev constant yield estimates of the rate of convergence to equilibrium of the underlying process. Applying to the isa, we obtain the following direct consequence of Theorem 2.6 and Theorem 2.7. We refer to [Sch12, Theorem 1.7] for a proof in the setting of the overdamped Langevin dynamics. The argument directly carries over to the isa.

Corollary 2.12. *Let f_t be the relative density of the infinite swapping process (2.2) at time t. Under the same assumptions as in Theorem 2.6 it holds that*

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

where ρ satisfies the estimate (2.9). Under the same assumptions as in Theorem (2.7) it holds that

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

where α satisfies the estimate (2.12).

Another well-known consequence is that the Poincaré or log-Sobolev constant allows one to quantify the ergodic theorem i.e. to estimate speed of convergence of the time average to the ensemble mean. See [CG06, Proposition 1.2.] and [Wu00, Corollary 4] for a proof in the setting of the overdamped Langevin dynamics. The same argument carries over to the isa.

Corollary 2.13. *Let ν denote the initial law of the isa X_t. Under the same assumptions as in Theorem 2.6 it holds that for all functions $f : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ such that $\sup |f| = 1$, all $0 < R \leq 1$ and all $t > 0$*

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

where ρ satisfies the estimate (2.9).
Under the same assumptions as in Theorem 2.7 it holds that for all functions $f \in L^1(\mu)$ and all $R, t > 0$

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t f(X_s)ds - \int f d\mu \geq R\right) \leq \left\|\frac{d\nu}{d\mu}\right\|_{L^2} \exp\bigl(-t\alpha H^*(R)\bigr),$$

where α satisfies the estimate (2.12) and

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

One consequence of Corollary 2.13 is that the isa has an exponential gain compared to the overdamped Langevin dynamics for sampling (see also Remark 2.8). See [DLPD12] for the details on the use of the isa to sample from the Gibbs measure $\frac{1}{Z}\exp\big(-\frac{H}{\tau_1}\big)$ at temperature τ_1.

2.7. **Application to simulated annealing.** In this section, we apply the log-Sobolev inequality of Theorem 2.7 to the simulated annealing of the isa.

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 and lives in a high-dimensional space. Let us explain the main idea of the stochastic version of simulated annealing. One considers a stochastic process on H that is 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 cooling is slow enough. Hence, another goal is to find the best stochastic process with the fastest possible cooling schedule that still allows one to find the global minimum.

Simulated annealing for the overdamped Langevin dynamics was studied in [GH86, Mic92]. As we will see below, the cooling schedule has to be logarithmically slow. This implies long computation times in order to find the global minimum. There are many ways to improve this behavior. Luckily, one has the freedom to choose the underlying stochastic process which is 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 can only improve when particles are swapped faster, which makes the isa a natural candidate for simulated annealing.

In [Mic92] it was shown that for the overdamped Langevin dynamics the fastest successful cooling schedule is characterized by the Eyring-Kramers formula for the log-Sobolev constant. However, at that time no estimates on the associated log-Sobolev constant for 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 for the underdamped Langevin dynamics, showing that the Langevin dynamics is at least as good as the overdamped Langevin dynamics for simulated annealing. The main result of [HKS89] (see also [Mon18]) is stated as follows.

Theorem 2.14 ([HKS89, Mic92]). *Let X_t be given by the classical overdamped Langevin dynamics*

$$dX_t = -\nabla H(X_t)\,dt + \sqrt{2\tau(t)}\,dB_t. \tag{2.13}$$

Let $E_* := H(s_{1p}) - H(m_p)$ denote the critical depth of the potential H. Then:

If $\tau(t) \geq \frac{E}{\ln t}$ for t large enough with $E > E_*$, then for all $\delta > 0$

$$\mathbb{P}\big(H(X_t) \leq H(m_1) + \delta\big) \underset{t\to\infty}{\to} 1.$$

If $\tau(t) \leq \frac{E}{\ln t}$ for t large enough with $0 < E < E_*$, then for δ small enough

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

In this section we study simulated annealing for the infinite swapping dynamics given by the following SDE

$$\begin{cases} dX_1 = -\nabla H(X_1)\,dt + \sqrt{2\,\tau_1(t)\,\rho(X_1,X_2) + 2\,\tau_2(t)\,\rho(X_2,X_1)}\,dB_1\,, \\ dX_2 = -\nabla H(X_2)\,dt + \sqrt{2\,\tau_2(t)\,\rho(X_1,X_2) + 2\,\tau_1(t)\,\rho(X_2,X_1)}\,dB_2\,. \end{cases} \tag{2.14}$$

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 2.6 and Theorem 2.7, we showed that the infinite swapping dynamics mixes faster than the overdamped Langevin dynamics. Choosing $\tau_2 = K\tau_1$, the effective critical depth of the potential H is $\frac{E_*}{K}$ compared to E_* for the classical overdamped Langevin dynamics given by (2.13). This indicates that the infinite swapping dynamics could outperform the overdamped Langevin dynamics for simulated annealing. The main result of this section shows that this is true.

Theorem 2.15. *Assume that the potential H satisfies Assumptions 2.3 and 2.5. Let $E_* := H(s_{p1}) - H(m_p)$ be the critical depth of the potential H. For $K > 1$ and $E > \frac{E_*}{K}$, let*

$$\tau_1(t) = \frac{E}{\ln(2+t)} \quad \textit{and} \quad \tau_2(t) = \frac{KE}{\ln(2+t)}. \tag{2.15}$$

Let X_1, X_2 be given by (2.14) 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 $p \geq 1$, there exists a constant C_p such that

$$\int \big(H(x_1) + H(x_2)\big)^p dm(x_1,x_2) \leq C_p. \tag{2.16}$$

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

$$\mathbb{P}(\min\{H(X_1(t)), H(X_2(t))\} > \delta) \lesssim \left(\frac{1}{1+t}\right)^{\min\left(\frac{\delta}{E}, \frac{1}{2} - \frac{E_*}{2KE}\right) - \varepsilon}. \tag{2.17}$$

3. Proofs

3.1. Proof of Theorem 2.6 and Theorem 2.7.

Our overall approach follows that of [MS14], which was via a decomposition of the state space $\mathbb{R}^n$ into an "admissible partition" of metastable regions $\{\Omega_i\}_{i=1}^N$ for the Gibbs measure ν^τ defined in (2.1), as described 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 $i \in [N]$,

$$\nu^\tau(\Omega_i)Z^\tau := \int_{\Omega_i} \exp\left(-\frac{H(x)}{\tau}\right)dx = \frac{(2\pi\tau)^{n/2}}{\sqrt{\det \nabla^2 H(m_i)}} \exp\left(-\frac{H(m_i)}{\tau}\right)(1 + O(\sqrt{\tau}|\ln\tau|^{3/2})). \tag{3.1}$$

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,

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

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 $\mathbb{R}^n$

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

$$Z_i^\tau := \nu^\tau(\Omega_i) = \frac{\sqrt{\det \nabla^2 H(m_1)}}{\sqrt{\det \nabla^2 H(m_i)}} \exp\left(-\frac{H(m_i)}{\tau}\right)(1 + O(\sqrt{\tau}|\ln\tau|^{3/2})).$$

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

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

where $Z_{ij}^+ := Z_i^{\tau_1} Z_j^{\tau_2}, Z_{ij}^- := Z_i^{\tau_2} Z_j^{\tau_1}$ 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*

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

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

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

where the second line is summing over unordered pairs $(i,j) \neq (k,l)$ 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*

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

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

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

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

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

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

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

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

$$\operatorname{Var}_{\pi_{ij}^\sigma}(f) \overset{(3.20)}{\le} O(1)\, \mathbb{E}_{\pi_{ij}^\sigma}(\tau_{\sigma(1)} |\nabla_{x_1} f|^2 + \tau_{\sigma(2)} |\nabla_{x_2} f|^2).$$

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

$$\mathrm{Ent}_{\pi_{ij}^{\sigma}}(f^2) \overset{(3.21)}{\leq} O(1)\, \mathbb{E}_{\pi_{ij}^{\sigma}}(|\nabla_{x_1} f|^2 + |\nabla_{x_2} f|^2).$$

We defer the details of the proof of Lemmas 3.5 and 3.6 to Section 3. 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]. It follows that

$$\begin{aligned} Z_{ij}^{\sigma}\, \mathrm{Var}_{\pi_{ij}^{\sigma}}(f) &\leq O(1)\mathcal{E}_{\pi^{\sigma}}(f)[\Omega_i \times \Omega_j], \\ Z_{ij}^{\sigma}\, \mathrm{Ent}_{\pi_{ij}^{\sigma}}(f) &\leq O(\tau_1^{-1})\mathcal{E}_{\pi^{\sigma}}(f)[\Omega_i \times \Omega_j]. \end{aligned} \tag{3.11}$$

Here and below, 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.

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. However, in order to ensure we only get exponential dependence on $1/\tau_2$ rather than $1/\tau_1$ in the Eyring-Kramers formula, we can 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 $\pi_{ij}^{+}, \pi_{ik}^{+}$ and for $\pi_{ji}^{-}, \pi_{ki}^{-}$)**.**

$$Z_{ik}^{+}(\mathbb{E}_{\pi_{ij}^{+}} f - \mathbb{E}_{\pi_{ik}^{+}} f)^2 \lessapprox C_{kj}^{\tau_2} \cdot \mathcal{E}_{\pi^{+}}(f)[\Omega_i \times \mathbb{R}^n], \tag{3.12}$$

$$Z_{ki}^{-}(\mathbb{E}_{\pi_{ji}^{-}} f - \mathbb{E}_{\pi_{ki}^{-}} f)^2 \lessapprox C_{kj}^{\tau_2} \cdot \mathcal{E}_{\pi^{-}}(f)[\mathbb{R}^n \times \Omega_i], \tag{3.13}$$

where

$$C_{kj}^{\tau_2} := \frac{1}{\sqrt{\det \nabla^2 H(m_k)}} \frac{2\pi\sqrt{\det \nabla^2 H(s_{kj})}}{|\lambda^{-}(s_{kj})|} \exp\left(\frac{H(s_{kj}) - H(m_k)}{\tau_2}\right).$$

Here and below, $\approx$ (resp. $\lessapprox$) means equality (resp. less than or equal) up to a multiplicative factor of $1 + O(\sqrt{\tau_2}|\ln \tau_2|^{3/2})$.

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

$$\begin{aligned} Z_{ik}^{+}(\mathbb{E}_{\pi_{ij}^{+}} f - \mathbb{E}_{\pi_{ik}^{+}} f)^2 &\leq Z_i^{\tau_1} Z_k^{\tau_2}\, \mathbb{E}_{\nu_i^{\tau_1}}^{x_1} (\mathbb{E}_{\nu_j^{\tau_2}}^{x_2} f - \mathbb{E}_{\nu_k^{\tau_2}}^{x_2} f)^2 \\ &\lessapprox Z_i^{\tau_1}\, \mathbb{E}_{\nu_i^{\tau_1}}^{x_1}\, C_{kj}^{\tau_2} \int \tau_2 |\nabla_{x_2} f|^2 d\nu^{\tau_2}(x_2) \\ &\leq C_{kj}^{\tau_2} \cdot \mathcal{E}_{\pi^{+}}(f)[\Omega_i \times \mathbb{R}^n]. \end{aligned}$$

The second estimate is completely analogous. □

To deal with the other mean-differences in (3.5) and (3.6), 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 come at a cost that is polynomial in 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}^-)**.** *In the same setting as Lemma 3.17,*

$$(\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\pi_{ij}^-} f)^2 \leq \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(\tau_2)(\mathbb{E}_{\pi_{ij}^+} |\nabla_{x_2} f|^2 + \mathbb{E}_{\pi_{ij}^-} |\nabla_{x_1} f|^2) + \omega(\tau_2) \sum_{\sigma \in \{+,-\}} \mathbb{E}_{\pi_{ij}^\sigma}(\tau_{\sigma(1)}|\nabla_{x_1} f|^2 + \tau_{\sigma(2)}|\nabla_{x_2} f|^2).$$

for any smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, where $\Phi_n : [1,\infty) \to [0,\infty)$ is the function

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

and $\omega(\tau_2) := O(\sqrt{\tau_2}|\ln \tau_2|^{3/2})$.

We defer the proof of this lemma to the next two sections. It follows that

$$\min(Z_{ij}^+, Z_{ij}^-)(\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\pi_{ij}^-} f)^2 \leq \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f)[\Omega_i \times \Omega_j]. \tag{3.14}$$

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 $\sigma \in \{+,-\}$*

$$Z_{ip}^+ Z_{11}^\sigma (\mathbb{E}_{\pi_{ip}^+}(f) - \mathbb{E}_{\pi_{11}^\sigma}(f))^2 \lessapprox C_{p1}^{\tau_2} \cdot \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n] + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f),$$
$$Z_{pj}^- Z_{11}^\sigma (\mathbb{E}_{\pi_{pj}^-}(f) - \mathbb{E}_{\pi_{11}^\sigma}(f))^2 \lessapprox C_{p1}^{\tau_2} \cdot \mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_j] + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f).$$

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

$$Z_{ij}^{\sigma_1} Z_{kl}^{\sigma_2} (\mathbb{E}_{\pi_{ij}^{\sigma_1}}(f) - \mathbb{E}_{\pi_{kl}^{\sigma_2}}(f))^2 \leq \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f).$$

Finally, for any other $\{(i,j)^{\sigma_1},(k,l)^{\sigma_2}\}$*, the term* $Z_{ij}^{\sigma_1}Z_{kl}^{\sigma_2}(\mathbb{E}_{\pi_{ij}^{\sigma_1}}(f)-\mathbb{E}_{\pi_{kl}^{\sigma_2}}(f))^2$ *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. Let Γ be the graph whose vertices are labelled $\cdot_{ij}^{\sigma}$ and have three kinds of edges:

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

We decompose the mean-difference between any two measures $\pi_{ij}^{+},\pi_{kl}^{-}$ as a sum of mean-differences of the types in (3.12), (3.13), and (3.14), corresponding to a sequence of "moves" on the graph Γ. Given any sequence of moves $v_0\to v_1\to\cdots\to v_m$ on graph Γ, we have

$$\begin{aligned} Z_{v_0}Z_{v_m}(\mathbb{E}_{\pi_{v_0}}f-\mathbb{E}_{\pi_{v_m}}f)^2 &= Z_{v_0}Z_{v_m}\left(\sum_{t=1}^{m}\sqrt{\omega_t}\frac{1}{\sqrt{\omega_t}}(\mathbb{E}_{\pi_{v_{t-1}}}f-\mathbb{E}_{\pi_{v_t}}f)\right)^2 \\ &= \sum_{t=1}^{m}\frac{1}{\omega_t}Z_{v_0}Z_{v_m}(\mathbb{E}_{\pi_{v_{t-1}}}f-\mathbb{E}_{\pi_{v_t}}f)^2 \end{aligned} \tag{3.15}$$

for any $\omega_t>0,\sum_{t=1}^{m}\omega_t=1$. 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: $\cdot_{ij}^{-}\to\cdot_{1j}^{-}\to\cdot_{1j}^{+}\to\cdot_{11}^{+}\to\cdot_{1l}^{+}\to\cdot_{1l}^{-}\to\cdot_{kl}^{-}$; and
- 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 (3.15) for a Type I sequence. For the 1st move,

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

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,

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

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,

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

and for the 2nd move in this sequence,

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

Thus, for this sequence, we can assign $\omega_1 = 1 - Z_p^{\tau_2} \approx 1, \omega_2 = Z_p^{\tau_2}$, 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*

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

$$\frac{Z_{ip}^+ Z_{kl}^\sigma}{\Lambda(Z_{ip}^+, Z_{kl}^\sigma)} (\mathbb{E}_{\pi_{ip}^+}(f) - \mathbb{E}_{\pi_{kl}^\sigma}(f))^2 \lessapprox \frac{1}{\Lambda\left(\frac{Z_{ip}^+}{Z_{kl}^\sigma}, 1\right)} \left(C_{p1}^{\tau_2} \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n] + \Phi_n \left(\frac{\tau_2}{\tau_1}\right) O(1) \mathcal{E}_\mu(f) \right),$$

$$\frac{Z_{pi}^- Z_{kl}^\sigma}{\Lambda(Z_{pi}^-, Z_{kl}^\sigma)} (\mathbb{E}_{\pi_{pi}^-}(f) - \mathbb{E}_{\pi_{kl}^\sigma}(f))^2 \lessapprox \frac{1}{\Lambda\left(\frac{Z_{pi}^-}{Z_{kl}^\sigma}, 1\right)} \left(C_{p1}^{\tau_2} \mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_i] + \Phi_n \left(\frac{\tau_2}{\tau_1}\right) O(1) \mathcal{E}_\mu(f) \right).$$

Finally, for any other $\{(i,j)^{\sigma_1}, (k,l)^{\sigma_2}\}$, the term $\frac{Z_{ij}^{\sigma_1} Z_{kl}^{\sigma_2}}{\Lambda(Z_{ij}^{\sigma_1}, Z_{kl}^{\sigma_2})} (\mathbb{E}_{\pi_{ij}^{\sigma_1}}(f) - \mathbb{E}_{\pi_{kl}^{\sigma_2}}(f))^2$ 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

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

for $b \lessapprox 1, a \ll 1$. 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 2.6. Combining Lemma 3.3, (3.11) and Lemma 3.9, we get

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

as desired.

□

Proof of Theorem 2.7. Combining Lemma 3.4, (3.11), (3.11) and Lemma 3.10, we get

$$\begin{aligned}
\mathrm{Ent}_\mu(f) &\lessapprox \frac{1}{2}\sum_{(i,j)} O(\tau_1^{-1})\mathcal{E}_{\pi^+}(f)[\Omega_i \times \Omega_j] + \frac{1}{2}\sum_{(i,j)} O(\tau_1^{-1})\mathcal{E}_{\pi^-}(f)[\Omega_i \times \Omega_j] \\
&\quad + \frac{1}{2}\sum_{(i,j)} 2N^2 O(\tau_1^{-1})\mathcal{E}_{\pi^+}(f)[\Omega_i \times \Omega_j] + \frac{1}{2}\sum_{(i,j)} 2N^2 O(\tau_1^{-1})\mathcal{E}_{\pi^-}(f)[\Omega_i \times \Omega_j] \\
&\quad + \frac{1}{2}\sum_{i\leq p}\left(\sum_{(k,l)^\sigma}\frac{1}{\Lambda\left(\frac{Z^+_{ip}}{Z^\sigma_{kl}},1\right)}\right)\left(C^{\tau_2}_{p1}\cdot\mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n] + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f)\right) \\
&\quad + \frac{1}{2}\sum_{i\leq p}\left(\sum_{(k,l)^\sigma}\frac{1}{\Lambda\left(\frac{Z^-_{pi}}{Z^\sigma_{kl}},1\right)}\right)\left(C^{\tau_2}_{p1}\cdot\mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_j] + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f)\right) \\
&\leq 2N^2\left(O(\tau_1^{-1}) + H(m_p)(\tau_1^{-1} + \tau_2^{-1})C^{\tau_2}_{p1} + O(\tau_1^{-1})\Phi_n\left(\frac{\tau_2}{\tau_1}\right)\right)\mathcal{E}_\mu(f),
\end{aligned}$$

as desired. □

3.2. **Proof of Theorem 2.15.** With the help of Theorem 2.7, i.e. the low-temperature asymptotics of the log-Sobolev constant, the proof of Theorem 2.15 follows the arguments in [Mic92, Mon18].

For each $t > 0$, let μ_t be the probability measure given in (2.3) at temperatures $\tau_1 = \tau_1(t), \tau_2 = \tau_2(t)$ as defined in (2.15), 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 making π_t a probability measure. 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 constant C such that*

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

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

$$\begin{aligned}\mathbb{P}(\min\{H(X_1(t)), H(X_2(t))\} > \delta) &\le \mathbb{P}(\min\{H(\tilde{X}_1(t)), H(\tilde{X}_2(t))\} > \delta) + d_{TV}(\mu_t, m_t) \\ &\le C(2+t)^{-\frac{\delta-\varepsilon}{E}} + \sqrt{2\,\mathrm{Ent}(m_t|\mu_t)}, \end{aligned} \tag{3.16}$$

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.

$$\frac{d}{dt}\mathrm{Ent}(m_t|\mu_t) \le -2\mathcal{I}_{\mu_t}\left(\frac{m_t}{\mu_t}\right) + \frac{d}{dt}\left(\frac{1}{\tau_1(t)} + \frac{1}{\tau_2(t)}\right)\mathbb{E}[H(X_1(t)) + H(X_2(t))]. \tag{3.17}$$

Proof. First note that

$$\begin{aligned}\frac{d}{dt}\mathrm{Ent}(m_t|\mu_t) &= \int \frac{dm_t}{dt}\ln\left(\frac{m_t}{\mu_t}\right) dx + \int m_t \frac{d}{dt}\ln\left(\frac{m_t}{\mu_t}\right) dx \\ &= \int \frac{dm_t}{dt}\ln\left(\frac{m_t}{\mu_t}\right) dx + \int \frac{dm_t}{dt} dx - \int \frac{m_t}{\mu_t}\frac{d\mu_t}{dt} dx \\ &= \int \frac{dm_t}{dt}\ln\left(\frac{m_t}{\mu_t}\right) dx - \int \frac{d\ln(\mu_t)}{dt} dm_t. \end{aligned} \tag{3.18}$$

We consider the first term in (3.18). 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

$$\begin{aligned}\int \frac{dm_t}{dt}\ln\left(\frac{m_t}{\mu_t}\right) dx &= -\int \left(a_1\left|\nabla_{x_1}\left(\frac{m_t}{\mu_t}\right)\right|^2 + a_2\left|\nabla_{x_2}\left(\frac{m_t}{\mu_t}\right)\right|^2\right)\frac{\mu_t}{m_t} d\mu_t \\ &= -2\mathcal{I}_{\mu_t}\left(\frac{m_t}{\mu_t}\right), \end{aligned} \tag{3.19}$$

where $\mathcal{I}_{\mu_t}$ is the Fisher information defined in (2.4) for $\mu = \mu_t$. Next we consider the second term in (3.18). Using that $\min H = 0$ and that $\tau_1(t), \tau_2(t)$ are decreasing, direct calculation yields

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

Integrating this against dm_t and combining it with (3.19) yields (3.17). □

The second term on the right hand side of (3.17) are controlled via the following lemma.

Lemma 3.13. *For any $\varepsilon > 0$, there exists a constant C such that*

$$\mathbb{E}\big[H(X_1(t)) + H(X_2(t))\big] \leq C(1+t)^{\varepsilon}.$$

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 in (2.16) and growth assumptions on the potential H in Assumption 2.3.

Lemma 3.14. *For any $\varepsilon > 0$, there exists C such that*

$$\operatorname{Ent}(m_t|\mu_t) \leq C\left(\frac{1}{t}\right)^{1-\frac{E^*}{KE}-\varepsilon}.$$

Proof. Using the log-Sobolev inequality in Theorem 2.7, the estimate (3.17) becomes

$$\frac{d}{dt}\operatorname{Ent}(m_t|\mu_t) \leq -2\alpha_t \operatorname{Ent}(m_t|\mu_t) + \frac{2}{E}(2+t)^{-1}\,\mathbb{E}[H(X_1(t)) + H(X_2(t))],$$

where α_t is the LSI constant in (2.11) for $\mu = \mu_t$. From (2.12) we see that for any $\varepsilon > 0$, there exists $t_0 > 0$ and $C_1 > 0$ such that for $t > t_0$,

$$2\alpha_t \geq C_1(2+t)^{-\frac{E_*}{KE}-\varepsilon}.$$

Together with Lemma 3.13, we get that for $t > t_0$,

$$\frac{d}{dt}\operatorname{Ent}(m_t|\mu_t) \leq -C_1(1+t)^{-\frac{E_*}{E}-\varepsilon}\operatorname{Ent}(m_t|\mu_t) + C_2(1+t)^{-1+\varepsilon}.$$

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 $t > t_0$,

$$\frac{d}{dt}Q(t) \leq -C_1(1+t)^{-\frac{E_*}{KE}-\varepsilon}Q(t),$$

$$Q(t) \le Q(t_0) e^{-C_1 \int_{t_0}^{t} (1+t)^{-\frac{E_*}{KE}+\varepsilon}},$$
$$\mathrm{Ent}(m_t|\mu_t) \le \frac{2C_2}{C_1}(1+t)^{-1+\frac{E_*}{KE}+2\varepsilon} + \mathrm{Ent}(m_{t_0}|\mu_{t_0}) e^{-\frac{C_1}{\nu}((1+t)^{\nu}-(1+t_0)^{\nu})},$$

where $\nu = 1 - \frac{E_*}{KE} - \varepsilon > 0$, and the conclusion follows. □

Combining (3.16) and Lemma 3.14, we get that for any $\delta > 0, \varepsilon > 0$, there exists a constant C such that

$$\mathbb{P}\Big(\min\{H(X_1(t)), H(X_2(t))\} > \delta\Big) \le C\left(\left(\frac{1}{1+t}\right)^{\frac{(\delta-\varepsilon)}{E}} + \left(\frac{1}{1+t}\right)^{\frac{1-\frac{E^*}{KE}-\varepsilon}{2}}\right),$$

which implies (2.17).

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. It may be verified by basic properties of variance and entropy.

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 $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*

$$\mathrm{Var}_\pi(f) = \mathbb{E}_{\nu_j}^{x_2}\big(\mathrm{Var}_{\nu_i}^{x_1}(f)\big) + \mathrm{Var}_{\nu_j}^{x_2}\big(\mathbb{E}_{\nu_i}^{x_1}(f)\big) \le \mathbb{E}_{\nu_j}^{x_2}\big(\mathrm{Var}_{\nu_i}^{x_1}(f)\big) + \mathbb{E}_{\nu_i}^{x_1}\big(\mathrm{Var}_{\nu_j}^{x_2}(f)\big) \tag{3.20}$$

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

$$\mathrm{Ent}_\pi(g) = \mathbb{E}_{\nu_j}^{x_2}\big(\mathrm{Ent}_{\nu_i}^{x_1}(g)\big) + \mathrm{Ent}_{\nu_j}^{x_2}\big(\mathbb{E}_{\nu_i}^{x_1}(g)\big) \le \mathbb{E}_{\nu_j}^{x_2}\big(\mathrm{Ent}_{\nu_i}^{x_1}(g)\big) + \mathbb{E}_{\nu_i}^{x_1}\big(\mathrm{Ent}_{\nu_j}^{x_2}(g)\big) \tag{3.21}$$

Definition 3.16 (Local PI and LSI for ν_i^τ)**.** *We say the local Gibbs measure ν_i^τ 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 we denote PI(ρ). We say ν_i^τ 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 we denote 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 $\nu_i^{\tau_1}, \nu_i^{\tau_2}$ **defined on 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 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*

$$\frac{L_\tau W_\tau}{W_\tau} \leq -\lambda + b\mathbb{1}_{U_i} \quad \forall x \in \Omega_i. \tag{3.22}$$

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

$$\int_{\Omega_i} (-L_\tau W_\tau) g d\nu_i^\tau = \int_{\Omega_i} \nabla g \cdot \nabla W_\tau d\nu_i^\tau. \tag{3.23}$$

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-Emery 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 (3.22), the Hamiltonian H was replace by a perturbed one H_τ such that $\|H - H_\tau\|_\infty = O(\tau)$. In order to satisfy the condition (3.23), 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 $\tau \le \varepsilon$, we choose the Lyapunov function to be $W_\tau = \exp \frac{H_\varepsilon}{2\tau}$. Then (3.23) is satisfied by [MS14, Theorem B.1] and

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

where the last inequality holds as long as $|\nabla V_\varepsilon| \le |\nabla H|$. Then once (3.22) 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 $\mathcal{M} = \{m_1, m_2, \ldots, m_N\}$ 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 that* *(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 (3.23)).

Then ν_i^τ satisfies LSI(α) with

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

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 (cf. [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, $R > 0$, consider the ellipsoid*

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

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

Lemma 3.25 (Approximation by truncated Gaussian)**.** *For $\tau \le \tau_2$, 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*

$$\frac{d\gamma_i^\tau}{d\nu_i^\tau}(x) = 1 + \omega(\tau_2), \tag{3.24}$$

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

$$(\mathbb{E}_{\nu_i^\tau} f - \mathbb{E}_{\gamma_i^\tau} f)^2 \le \mathrm{Var}_{\nu_i^\tau}\left(\frac{d\gamma_i^\tau}{d\nu_i^\tau}\right)\mathrm{Var}_{\nu_i^\tau}(f) \le \omega(\tau_2)\tau\,\mathbb{E}_{\nu_i^\tau}|\nabla f|^2.$$

where $\omega(\tau_2) := O(\sqrt{\tau_2}|\ln\tau_2|^{3/2})$.

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 $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*

$$\left(\mathbb{E}_{\pi^\sigma_{ij}} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}} \otimes \gamma_j^{\tau_{\sigma(2)}}} f\right)^2 \le \omega(\tau_2)\, \mathbb{E}_{\pi^\sigma_{ij}} \left(\tau_{\sigma(1)} |\nabla_{x_1} f|^2 + \tau_{\sigma(2)} |\nabla_{x_2} f|^2\right).$$

where $\omega(\tau_2) := O(\sqrt{\tau_2} |\ln \tau_2|^{3/2})$.

Proof. This follows from the previous lemma by writing

$$\begin{aligned}\mathbb{E}_{\pi^\sigma_{ij}} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}} \otimes \gamma_j^{\tau_{\sigma(2)}}} f = &\left(\mathbb{E}_{\nu_i^{\tau_{\sigma(1)}} \otimes \nu_j^{\tau_{\sigma(2)}}} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}} \otimes \nu_j^{\tau_{\sigma(2)}}} f\right) \\ &+ \left(\mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}} \otimes \nu_j^{\tau_{\sigma(2)}}} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}} \otimes \gamma_j^{\tau_{\sigma(2)}}} f\right).\end{aligned}$$

□

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 (2.10), 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

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

where, for some $0 < \kappa \le R$ to be specified later,

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

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

Estimate for I_2: By Cauchy-Schwarz,

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

$$\leq \sqrt{\tau_2} \int_{S^{n-1}} d\omega \int_{\kappa\sqrt{\tau_2}}^{R\sqrt{\tau_2}} |\nabla f(s\omega)|^2 \left(\int_{\frac{s}{\sqrt{\tau_2}}}^{R} \frac{e^{-\frac{r^2}{2}}}{Z_R} r^n dr \right) ds.$$

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

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

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

$$I_2 \leq C_n(1+\kappa^{-(n-1)})\tau_2 \, \mathbb{E}_{\gamma_i^{\tau_2}} |\nabla f|^2.$$

Estimate for I_1: By Cauchy-Schwarz

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

where C_n is a constant only depending on n. The conclusion now follows if we choose $\kappa = R$ when $R < 1$ and $\kappa = 1$ when $R \geq 1$. □

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

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

Proof. This follows from the previous lemma and (3.24) by writing

$$\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_2} \otimes \gamma_j^{\tau_1}} f = (\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\gamma_1}} f) + (\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_1}} f - \mathbb{E}_{\gamma_i^{\tau_2} \otimes \gamma_j^{\tau_1}} f).$$

□

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*

$$\begin{aligned} (\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\pi_{ij}^-} f)^2 &= (\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\nu_i^{\tau_1} \otimes \nu_j^{\tau_1}} f + \mathbb{E}_{\nu_i^{\tau_1} \otimes \nu_j^{\tau_1}} f - \mathbb{E}_{\pi_{ij}^-} f)^2 \\ &\leq 2\, \mathbb{E}^{x_1}_{\nu_i^{\tau_1}} (\mathbb{E}^{x_2}_{\nu_j^{\tau_2}} f - \mathbb{E}^{x_2}_{\nu_j^{\tau_1}} f)^2 + 2\, \mathbb{E}^{x_2}_{\nu_j^{\tau_1}} (\mathbb{E}^{x_1}_{\nu_i^{\tau_1}} f - \mathbb{E}^{x_1}_{\nu_i^{\tau_2}} f)^2. \end{aligned}$$

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

$$(\mathbb{E}_{\nu_k^{\tau_2}} g - \mathbb{E}_{\nu_k^{\tau_1}} g)^2 = \left(\mathbb{E}_{\nu_k^{\tau_2}} g - \mathbb{E}_{\nu_k^{\tau_2}} g \frac{d\nu_k^{\tau_1}}{d\nu_k^{\tau_2}} \right)^2 = \operatorname{Cov}_{\nu_k^{\tau_2}} \left(g, \frac{d\nu_k^{\tau_1}}{d\nu_k^{\tau_2}} \right)^2$$

$$\leq \operatorname{Var}_{\nu_k^{\tau_2}}(g)\operatorname{Var}_{\nu_k^{\tau_2}}\left(\frac{d\nu_k^{\tau_1}}{d\nu_k^{\tau_2}}\right) \leq O(\tau_2)\,\mathbb{E}_{\nu_k^{\tau_2}}|\nabla g|^2\,\mathbb{E}_{\nu_k^{\tau_1}}\left(\frac{d\nu_k^{\tau_1}}{d\nu_k^{\tau_2}}\right).$$

Finally, using the partition size given in (3.1) we have a uniform estimate on the relative density

$$\frac{d\nu_k^{\tau_1}}{d\nu_k^{\tau_2}} = \frac{\nu_k^{\tau_2}(\Omega_k)}{\nu_k^{\tau_1}(\Omega_k)}e^{-H(x)(\tau_1^{-1}-\tau_2^{-1})} \leq \frac{\nu_k^{\tau_2}(\Omega_k)}{\nu_k^{\tau_1}(\Omega_k)} \leq \left(\frac{\tau_2}{\tau_1}\right)^{\frac{n}{2}}(1+O(\sqrt{\tau_2}|\ln\tau_2|^{3/2})).$$

Applying the preceding with $k=j, g(\cdot)=f(x_1,\cdot)$ and $k=i, g(\cdot)=f(\cdot,x_2)$, respectively, we obtain the estimate

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

In comparison to Lemma 3.8, the dependence in the ratio τ_2/τ_1 appearing in front of the Dirichlet form is about one order worse in exponent.

3.5. **Proof of Proposition 2.9.** It suffices to consider test functions of the form $f(x,y)=f(x)$. This is equivalent to replace μ by its first marginal, which is $\bar\mu = \frac{1}{2}(\nu^{\tau_1}+\nu^{\tau_2})$. In this case, $\operatorname{Var}_\mu(f)$ and $\mathcal{E}_\mu(f)$ reduces to

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

We further restrict f to $C_c(\Omega_i)$ with $i=1$. By (3.1) and (2.8), $\nu^\tau(\Omega_i)\approx 1$ once τ is small enough. Then

$$\begin{aligned}
\operatorname{Var}_{\bar\mu}(f) &\gtrsim (\mathbb{E}_{\nu^{\tau_1}} f)^2 - 4(\mathbb{E}_{\nu^{\tau_2}} f)^2 \gtrsim (\mathbb{E}_{\nu_i^{\tau_1}} f)^2 - 5(\mathbb{E}_{\nu_1^{\tau_2}} f)^2,\\
\mathcal{E}_{\bar\mu}(f) &\lesssim \tau_1\,\mathbb{E}_{\nu_i^{\tau_1}}|\nabla f|^2 + \tau_2\,\mathbb{E}_{\nu_i^{\tau_2}}|\nabla f|^2).
\end{aligned}$$

By change of variables, we may assume $m_i=0, \Sigma_i=(\nabla^2 H(m_i))^{-1}=\mathrm{Id}$. We consider a test function of the form

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

where $h\geq 0$ is a compactly supported, absolutely continuous function and $\tau_1\leq\varepsilon\leq\tau_2$ 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 $\varepsilon\leq\tau_2$, f_ε is supported in the support of $\gamma_i^{\tau_2}$. By Lemma 3.25,

$$\operatorname{Var}_{\bar\mu}(f) \gtrsim (\mathbb{E}_{\gamma_i^{\tau_1}} f_\varepsilon)^2 - 6(\mathbb{E}_{\gamma_i^{\tau_2}} f_\varepsilon)^2, \tag{3.25}$$

$$\mathcal{E}_{\bar\mu}(f) \lesssim \tau_1\,\mathbb{E}_{\nu_i^{\tau_1}}|\nabla f_\varepsilon| + \tau_2\,\mathbb{E}_{\gamma_i^{\tau_2}}|\nabla f_\varepsilon|^2, \tag{3.26}$$

if τ_2 is small enough. We have:

$$\tau_1\,\mathbb{E}_{\nu_i^{\tau_1}}|\nabla f_\varepsilon|^2 = \frac{\tau_1}{\varepsilon}\,\mathbb{E}_{\nu_i^{\frac{\tau_1}{\varepsilon}}}|\nabla f_1|^2, \tag{3.27}$$

$$\tau_2\,\mathbb{E}_{\gamma_i^{\tau_2}}|\nabla f_\varepsilon|^2 = \frac{\tau_2}{\varepsilon}\,\mathbb{E}_{\gamma_i^{\frac{\tau_2}{\varepsilon}}}|\nabla f_1|^2 \leq \frac{1}{\sqrt{2\pi}^n}(\varepsilon/\tau_2)^{(n-2)/2}\|\nabla f_1\|_{L^2}^2, \tag{3.28}$$

$$\mathbb{E}_{\gamma_i^{\tau_2}} f_\varepsilon = \mathbb{E}_{\gamma_i^{\frac{\tau_2}{\varepsilon}}} f_1 \le \frac{1}{\sqrt{2\pi}^n} (\varepsilon/\tau_2)^{n/2} \|f_1\|_{L^1}, \tag{3.29}$$

and for any $r \ge 0$,

$$\mathbb{E}_{\gamma_i^{\tau_1}} f_\varepsilon = \mathbb{E}_{\gamma_i^{\frac{\tau_1}{\varepsilon}}} f_1 \ge (\inf_{|x|\le r} f_1) P_{\gamma_i^{\frac{\tau_1}{\varepsilon}}}(|X| \le r) \ge (\inf_{[0,r]} h) \left(1 - n \exp\left(-\frac{r^2}{2n}\frac{\varepsilon}{\tau_1}\right)\right) \tag{3.30}$$

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

$$\tau_2 \, \mathbb{E}_{\gamma_i^{\tau_2}} |\nabla f_\varepsilon|^2 \overset{(3.28)}{\lesssim} (\varepsilon/\tau_2)^{(n-2)/2}, \ \mathbb{E}_{\gamma_i^{\tau_2}} f_\varepsilon \overset{(3.29)}{\lesssim} (\varepsilon/\tau_2)^{n/2}, \ \mathbb{E}_{\gamma_i^{\tau_1}} f_\varepsilon \overset{(3.30)}{\ge} \frac{1}{2},$$

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

$$\tau_1 \, \mathbb{E}_{\nu_i^{\tau_1}} |\nabla f_\varepsilon|^2 \overset{(3.27)}{\le} \frac{\tau_1}{\varepsilon} \|g'\|_{L^\infty}^2 P_{\nu_i^{\frac{\tau_1}{\varepsilon}}}(|X| \ge R_n) \le \frac{\tau_1}{\varepsilon} \|g'\|_{L^\infty}^2 C_H e^{-c_H \frac{\varepsilon}{\tau_1}} \lesssim_m (\tau_1/\varepsilon)^{\eta m},$$

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 $\eta m \ge (1-\eta)(n-2)/2$, we obtain

$$\mathcal{E}_{\bar\mu}(f) \overset{(3.26)}{\lesssim_\eta} (\tau_1/\tau_2)^{(1-\eta)(n-2)/2}, \ \mathrm{Var}_{\bar\mu}(f) \overset{(3.25)}{\gtrsim_\eta} (\tau_1/\tau_2)^{(1-\eta)(n-2)/2} \mathcal{E}_{\bar\mu}(f),$$

if τ_2, τ_1/τ_2 are both small enough.

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

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

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

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

We choose $\varepsilon = \tau_2$ and $r_0^2 \frac{\tau_2}{\tau_1} = R_2^2$ (which is possible once $\tau_1/\tau_2 \le 1/R_2^2$). Then:

$$\mathbb{E}_{\gamma_i^{\tau_2}} f_\varepsilon \overset{(3.29)}{\le} \frac{1}{2\pi}\frac{\varepsilon}{\tau_2}\|f_1\|_{L^1} \le \frac{\alpha}{2}, \ \mathbb{E}_{\gamma_i^{\tau_1}} f_\varepsilon \overset{(3.30)}{\ge} \frac{1}{2},$$

$$\tau_1 \, \mathbb{E}_{\nu_i^{\tau_1}} |\nabla f_\varepsilon|^2 \overset{(3.27)}{\le} \frac{\tau_1}{\varepsilon}\|\nabla f_1\|_{L^\infty}^2 \le \frac{\alpha^2}{R_2^2}, \ \tau_2 \, \mathbb{E}_{\gamma_i^{\tau_2}} |\nabla f_\varepsilon|^2 \overset{(3.28)}{\le} \frac{1}{2\pi}\|\nabla f_1\|_{L^2}^2 = \frac{3\alpha}{2}.$$

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

$$\mathcal{E}_{\bar\mu}(f) \overset{(3.26)}{\lesssim} \frac{\alpha^2}{2R_2^2} + \frac{3\alpha}{4}, \quad \mathrm{Var}_{\bar\mu}(f) \overset{(3.25)}{\gtrsim} \frac{1}{\alpha}\mathcal{E}_{\bar\mu}(f) \gtrsim \ln\left(\frac{\tau_2}{\tau_1}\right)\mathcal{E}_{\bar\mu}(f),$$

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

3.6. **Proof of Proposition 2.10 and Proposition 2.11.** It suffices to consider test functions of the form $f(x,y) = g(x)g(y)$. This is equivalent to replace 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} 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

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

Here and below, $\approx$ (resp. $\lessapprox$) means equality (resp. less than or equal) up to a multiplicative factor of $1 + O(\sqrt{\tau_2}|\ln\tau_2|^{3/2})$.

Proof of Proposition 2.10: 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 $\delta = \sqrt{2r_0\tau_2|\ln\tau_2|}$, 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 $\kappa = 1 + O(\tau_2^{-r_0/\sigma}) \approx 1$ for r_0 large enough. For δ sufficiently small,

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

This motivates the choice

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

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

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

$$\mathbb{E}_{\nu^{\tau_1}}\, g^2 \approx Z_1^{\tau_1} g(m_1)^2 + Z_2^{\tau_1} g(m_2)^2 \approx g(m_2)^2 Z_2^{\tau_1}.$$

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

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

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

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

It remains to show the other term is asymptotically negligible:

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

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 2.10: 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

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

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

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

and the same computation as before shows

$$\begin{aligned}
\mathbb{E}_{\nu^{\tau_1}}(g')^2 &\lessapprox g(m_2)^2\frac{\sqrt{H''(m_1)}}{2\pi\sqrt{2\tau_1\tau_2\sigma}}e^{-(1-\eta)H(s)/\tau_1}, \\
\mathbb{E}_{\nu^{\tau_2}}(g')^2 &\approx g(m_2)^2\frac{\sqrt{H''(m_1)|H''(s)|}}{2\pi\tau_2}e^{-H(s)/\tau_2},
\end{aligned}$$

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

$$\tau_2 \frac{\mathbb{E}_{\nu^{\tau_2}}(g')^2}{\operatorname{Ent}_{\nu^{\tau_2}} g^2} \approx \frac{1}{\ln(1/Z_2^{\tau_1})} \frac{\sqrt{H''(m_2)|H''(s)|}}{2\pi} e^{(H(m_2)-H(s))/\tau_2} \lesssim_N \alpha,$$

and that $\tau_1 \frac{\mathbb{E}_{\nu^{\tau_1}}(g')^2}{\operatorname{Ent}_{\nu^{\tau_2}} g^2 \, \mathbb{E}_{\nu^{\tau_1}} g^2}$ is asymptotically negligible.

Acknowledgment

This research has been partially supported by NSF grant DMS-1712632. The authors want to thank Max Fathi and Paul Bressloff for the fruitful discussions. Georg Menz and André Schlichting want to thank the University of Bonn for financial support via the CRC 1060 'The Mathematics of Emergent Effects'.

References

[And80] Hans C Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980.

[BEGK04] Anton Bovier, Michael Eckhoff, Véronique Gayrard, and Markus Klein. Metastability in reversible diffusion processes I: Sharp asymptotics for capacities and exit times. *Journal of the European Mathematical Society*, 6(4):399–424, 2004.

[BGK05] Anton Bovier, Véronique Gayrard, and Markus Klein. Metastability in reversible diffusion processes II: Precise asymptotics for small eigenvalues. *Journal of the European Mathematical Society*, 7(1):69–99, 2005.

[CG06] Patrick Cattiaux and Arnaud Guillin. Deviation bounds for additive functionals of Markov processes. *arXiv:math/0603021*, 2006.

[CG08] Patrick Cattiaux and Arnaud Guillin. Deviation bounds for additive functionals of Markov processes. *ESAIM: Probability and Statistics*, 12:12–29, 2008.

[CS11] John D Chodera and Michael R Shirts. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *The Journal of Chemical Physics*, 135(19):194110, 2011.

[DDN17] J. Doll, Paul Dupuis, and Pierre Nyquist. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms. *Applied Mathematics & Optimization*, pages 1–42, 2017.

[DLPD12] Paul Dupuis, Yufei Liu, Nuria Plattner, and J. Doll. On the infinite swapping limit for parallel tempering. *Multiscale Modeling & Simulation*, 10(3):986–1022, 2012.

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

[GSC+13] Andrew Gelman, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

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

[HKS89] Richard A Holley, Shigeo Kusuoka, and Daniel Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *Journal of functional analysis*, 83(2):333–347, 1989.

[HN05] Bernard Helffer and Francis 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] Bernard Helffer and Francis Nier. Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach: the case with boundary. *Mém. Soc. Math. Fr. (N.S.)*, (105):vi+89, 2006.

[KAJ94] C Koulamas, SR 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] Srinivasaraghavan Kannan and Martin Zacharias. Simulated annealing coupled replica exchange molecular dynamics–an efficient conformational sampling method. *Journal of structural biology*, 166(3):288–294, June 2009.

[LPA+09] Yaohang Li, Vladimir Protopopescu, Nikita Arnold, Xinyu Zhang, and Andrey Gorin. Hybrid parallel tempering and simulated annealing method. In *Applied Mathematics and Computation*, volume 212, pages 216–228. 06 2009.

[Mic92] Laurent Miclo. Recuit simulé sur $\mathbb{R}^n$. étude de l'évolution de l'énergie libre. *Annales de l'Institut Henri Poincaré*, 28(2):235–266, 1992.

[Mon18] Pierre Monmarché. Hypocoercivity in metastable settings and kinetic simulated annealing. *Probability Theory and Related Fields*, pages 1–34, 2018.

[MS14] Georg Menz and André Schlichting. Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *The Annals of Probability*, 42(5):1809–1884, 2014.

[Nar99] K. Nara. Simulated annealing applications. In YH. Song, editor, *Modern Optimisation Techniques in Power Systems. International Series on Microprocessor-Based and Intelligent Systems Engineering.*, volume 20. Springer, Dordrecht, 1999.

[Pav07] Ilya Pavlyukevich. Lévy flights, non-local search and simulated annealing. *Journal of Computational Physics*, 226(2):1830 – 1844, 2007.

[RC05] Christian P Robert and George Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, 2005.

[Sch12] André 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.

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

[vLA87] Peter J. M. van Laarhoven and Emile H. L. Aarts. *Simulated annealing*, pages 7–15. Springer Netherlands, Dordrecht, 1987.

[Wu00] Liming Wu. A deviation inequality for non-reversible Markov processes. *Annales de l'I.H.P. Probabilités et statistiques*, 36(4):435–445, 2000.

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

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

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

Institute for Applied Mathematics, University of Bonn.
Current address: Institute for Geometry and Practical Mathematics, RWTH Aachen University
Email address: schlichting@iam.uni-bonn.de schlichting@igpm.rwth-aachen.de

Department of Statistics, University of California, Berkeley.
Email address: wenpintang@stat.berkeley.edu

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