



# Quantum Concentration Inequalities

Giacomo De Palma and Cambyse Rouzé

**Abstract.** We establish Transportation Cost Inequalities (TCIs) with respect to the quantum Wasserstein distance by introducing quantum extensions of well-known classical methods: First, we generalize the Dobrushin uniqueness condition to prove that Gibbs states of 1D commuting Hamiltonians satisfy a TCI at any positive temperature and provide conditions under which this first result can be extended to non-commuting Hamiltonians. Next, using a non-commutative version of Ollivier’s coarse Ricci curvature, we prove that high temperature Gibbs states of commuting Hamiltonians on arbitrary hypergraphs  $H = (V, E)$  satisfy a TCI with constant scaling as  $O(|V|)$ . Third, we argue that the temperature range for which the TCI holds can be enlarged by relating it to recently established modified logarithmic Sobolev inequalities. Fourth, we prove that the inequality still holds for fixed points of arbitrary reversible local quantum Markov semigroups on regular lattices, albeit with slightly worsened constants, under a seemingly weaker condition of local indistinguishability of the fixed points. Finally, we use our framework to prove Gaussian concentration bounds for the distribution of eigenvalues of quasi-local observables and argue the usefulness of the TCI in proving the equivalence of the canonical and microcanonical ensembles and an exponential improvement over the weak Eigenstate Thermalization Hypothesis.

## 1. Introduction

Given a random variable  $X$  of law  $\mu$  taking values on a metric space  $(\Omega, d_\Omega)$  and a function  $f : \Omega \rightarrow \mathbb{R}$ , a concentration of measure inequality quantifies the probability that the random variable  $f(X)$  deviates from its mean or its median. Since the early age of the theory, concentration inequalities have seen many new methods, refinements and exciting applications to various areas of mathematics [1–3]. Among the different classes of concentration inequalities, Gaussian concentration is arguably the most standard one: the measure  $\mu$  is said to be *sub-Gaussian* if there exist constants  $K, \kappa > 0$  such that, for all  $A \subseteq \Omega$  with  $\mu(A) \geq 1/2$  we have for any  $r \geq 0$

$$\mu(\{x \in \Omega : d_\Omega(x, A) < r\}) \geq 1 - Ke^{-\kappa r^2}. \tag{1}$$

In her seminal work [4], Marton made the beautiful observation that the above behavior can be obtained as a consequence of a *transportation cost* inequality: if there exists  $c > 0$  such that, for any probability measure  $\nu \ll \mu$ ,

$$W_1(\mu, \nu) \leq \sqrt{cS(\nu\|\mu)}, \tag{TC(c)}$$

then (1) holds with constants  $\kappa = \frac{1}{c}$  and  $K = 1$  for all  $r > \sqrt{c \ln 2}$ . Here,  $S(\nu\|\mu)$  refers to the relative entropy between the measures  $\nu$  and  $\mu$ , whereas the quantity  $W_1(\mu, \nu)$  in (TC(c)) is the Wasserstein distance between the two measures  $\mu, \nu$ :

$$W_1(\mu, \nu) := \sup_{\|f\|_L \leq 1} |\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]|.$$

Later, Bobkov and Götze [5] proved that transportation cost inequalities are in fact equivalent to the property of sub-Gaussianity: more precisely, (TC(c)) holds if and only if for all Lipschitz functions  $f : \Omega \rightarrow \mathbb{R}$ ,

$$\mathbb{P}_\mu\left(|f(X) - \mathbb{E}_\mu[f(X)]| > t\right) \leq 2e^{-\frac{t^2}{c}}, \forall t \geq 0.$$

One of the main advantages of transportation cost inequalities is their tensorization property: assume that  $\mu$  satisfies TC(c), then  $\mu^{\otimes n}$  satisfies TC(nc) for all  $n \in \mathbb{N}$ , where the set  $\Omega^n$  is provided with the metric

$$d_n(x^n, y^n) := \sum_{i=1}^n d_\Omega(x_i, y_i).$$

Perhaps the simplest example of that sort is given by taking  $\Omega^n = [d]^n$  endowed with the Hamming distance  $d_H$ . In the case  $n = 1$ , the corresponding Wasserstein distance reduces to the total variation, and TC(1/2) holds for any measure  $\mu$ , since it simply reduces to Pinsker’s inequality. For  $n \geq 1$ ,  $\mu^{\otimes n}$  satisfies TC(n/2).

While the theory of concentration inequalities for i.i.d. random variables is by now well understood, things become more challenging when the random variables are allowed to depend on each other [3, 6]. One way to extend concentration bounds to weakly dependent random variables is to assume that their joint law  $\mu$  satisfies the so-called Dobrushin uniqueness condition [6]. Dobrushin’s uniqueness condition plays an important role in the study of Gibbs measures in the one-phase region; however, it often turns out to be a very strong requirement on the measure  $\mu$ . More recently, Marton gave an attempt at extending the i.i.d. theory beyond the mere Gibbs setting [7]. Her main result consists in a logarithmic Sobolev inequality for a generic measure  $\mu$ —well known to imply transportation cost inequalities—under the so-called Dobrushin–Shlosman mixing condition [8], the latter condition being weaker than Dobrushin’s uniqueness condition. As mentioned in [9], such paths to establish Gaussian concentration suffer from the difficulty of deriving explicit constants. Moreover, the result of [7] also relies on the crucial assumption that the measure  $\mu$  has full support.

Recently, concentration inequalities have attracted much attention in the communities of random matrix theory, quantum information theory and operator algebras [10–18]. In [14], a quantum Wasserstein distance of order 1 (or quantum  $W_1$  distance) was defined on the set of the quantum states of  $n$  qudits with the property that it strictly reduces to the classical Wasserstein distance on  $[d]^n$  for states that are diagonal in the computational basis. This quantum generalization of the Wasserstein distance is based on the notion of neighboring states. Two quantum states of  $n$  qudits are neighboring if they differ only in one qudit, *i.e.*, if they coincide after that qudit is discarded. The quantum  $W_1$  distance is then that induced by the maximum norm that assigns distance at most one to every couple of neighboring states [14, Definition 4]. Such norm is called quantum  $W_1$  norm and is denoted with  $\|\cdot\|_{W_1}$ . The quantum  $W_1$  norm proposed in Ref. [14] admits a dual formulation in terms of a quantum generalization of the Lipschitz constant. Denoting with  $\mathcal{O}_n$  the set of the observables of  $n$  qudits, the Lipschitz constant of the observable  $H \in \mathcal{O}_n$  is defined as [14, Sect. V]

$$\begin{aligned} \|H\|_L &:= 2 \max_{i \in [n]} \min \left\{ \left\| H - H^{(i)} \right\|_{\infty} : H^{(i)} \right. \\ &\quad \left. \in \mathcal{O}_n \text{ does not act on the } i\text{-th qudit} \right\}. \end{aligned} \quad (2)$$

Then, the quantum  $W_1$  distance between the states  $\rho$  and  $\omega$  can also be expressed as [14, Sect. V]

$$\|\rho - \omega\|_{W_1} = \max \{ \text{Tr}[(\rho - \omega)H] : H \in \mathcal{O}_n, \|H\|_L \leq 1 \}. \quad (3)$$

Moreover, in [14] it was showed that  $\text{TC}(n/2)$  holds for any tensor product  $\omega = \omega_1 \otimes \cdots \otimes \omega_n$  of quantum states, hence extending Marton's original inequality with the exact same constant: for any state  $\rho$  of  $n$  qudits,

$$\|\rho - \omega_1 \otimes \cdots \otimes \omega_n\|_{W_1} \leq \sqrt{\frac{n}{2} S(\rho \|\omega_1 \otimes \cdots \otimes \omega_n)},$$

where  $S(\rho \|\omega) := \text{Tr}[\rho(\ln \rho - \ln \omega)]$  denotes Umegaki's relative entropy between the states  $\rho$  and  $\omega$ .

**Main results:** In this paper, we prove that any of the following conditions implies a transportation cost inequality:

- (i) A non-commutative Dobrushin uniqueness condition (Sect. 3);
- (ii) A generalization of Ollivier's coarse Ricci curvature bound (Sect. 4);
- (iii) A modified logarithmic Sobolev inequality condition (Sect. 5);
- (iv) A condition of local indistinguishability of the state (Sect. 6).

Each of these methods comes with its strengths and weaknesses:

- (i) The non-commutative Dobrushin uniqueness condition implies a nontrivial TCI at any temperature (see Remark 1), but the scaling of the constant  $c$  with the number of subsystems is optimal only in one dimension (see Remark 2).

- (ii) The coarse Ricci curvature bound provides TC inequalities for essentially any geometry, but it is only valid above a threshold temperature that depends on the locality of the Hamiltonian. Furthermore, such threshold temperature is in practice strictly larger than the true critical temperature (see Proposition 9).
- (iii) Quantum modified logarithmic Sobolev inequalities are typically more difficult to prove than their classical counterparts and are currently only proven to hold in specific cases. However, for one-dimensional systems, a recently derived modified logarithmic Sobolev inequality [19, 20] provides us with TC (up to polylogarithmic overhead) at any positive temperature.
- (iv) The condition of local indistinguishability of the state for regular lattices. Although the condition can be checked for classical systems, we do not yet have a way to prove it in the quantum setting.

We conclude the article with two natural applications of our bounds. First, we derive Gaussian concentration bounds for a large class of Lipschitz observables whenever the state  $\omega$  is that of a commuting Hamiltonian at large enough temperature (Sect. 7). Second, we argue on the use of the transportation cost inequality in proving the equivalence between the microcanonical and the canonical ensembles and an exponential improvement over the weak Eigenstate Thermalization Hypothesis (Sect. 8).

## 2. Notations and Basic Definitions

Given a finite set  $V$ , we denote by  $\mathcal{H}_V = \bigotimes_{v \in V} \mathcal{H}_v$  the Hilbert space of  $n = |V|$  qudits (*i.e.*,  $\mathcal{H}_v \equiv \mathbb{C}^d$  for all  $v \in V$ ) and by  $\mathcal{B}_V$  the algebra of linear operators on  $\mathcal{H}_V$ .  $\mathcal{O}_V$  corresponds to the space of self-adjoint linear operators on  $\mathcal{H}_V$ , whereas  $\mathcal{O}_V^T \subset \mathcal{O}_V$  is the subspace of traceless self-adjoint linear operators.  $\mathcal{O}_V^+$  denotes the cone of positive semidefinite linear operators on  $\mathcal{H}_V$  and  $\mathcal{S}_V \subset \mathcal{O}_V^+$  denotes the set of quantum states. We denote by  $\mathcal{P}_V$  the set of probability measures on  $[d]^V$ . For any subset  $A \subseteq V$ , we use the standard notations  $\mathcal{O}_A, \mathcal{S}_A \dots$  for the corresponding objects defined on subsystem  $A$ . Given a state  $\rho \in \mathcal{S}_V$ , we denote by  $\rho_A$  its marginal onto the subsystem  $A$ . For any  $X \in \mathcal{O}_V$ , we denote by  $\|X\|_1$  its trace norm. The identity on  $\mathcal{O}_v, v \in V$ , is denoted by  $\mathbb{I}_v$ .

Given two states  $\rho, \omega \in \mathcal{S}_V$  such that  $\text{supp}(\rho) \subseteq \text{supp}(\omega)$ , their quantum relative entropy is defined as [21–23]

$$S(\rho\|\omega) = \text{Tr}[\rho(\ln \rho - \ln \omega)]. \tag{4}$$

Whenever  $\rho = \rho_{AB}$  is a bipartite state and  $\omega = \rho_A \otimes \rho_B$ , their relative entropy reduces to the mutual information

$$I(A; B)_\rho := S(\rho_{AB}\|\rho_A \otimes \rho_B). \tag{5}$$

In the next sections, we also utilize the measured relative entropy [24–27]

$$S_{\mathbb{M}}(\rho\|\omega) := \sup_{(\mathcal{X}, M)} S(P_{\rho, M}\|P_{\sigma, M}), \tag{6}$$

where the supremum above is over all positive operator valued measures  $M$  that map the input quantum state to a probability distribution on a finite set  $\mathcal{X}$  with probability mass function given by  $P_{\rho, M}(x) = \text{Tr} \rho M(x)$ .

In this paper, we study inequalities relating the  $W_1$  distance between two states to their relative entropy. More precisely, for a fixed state  $\omega \in \mathcal{S}_V$ , we are interested in upper bounding the best constant  $C(\omega) > 0$  such that, for all  $\rho \in \mathcal{S}_V$  with  $\text{supp}(\rho) \subseteq \text{supp}(\omega)$ ,

$$\|\rho - \omega\|_{W_1} \leq \sqrt{C(\omega) S(\rho \|\omega)}. \tag{7}$$

In general, given a constant  $c \geq C(\omega)$ , we refer to the above inequality for  $C(\omega)$  replaced by  $c$  as a *transportation cost inequality*, denoted by  $\text{TC}(c)$ . As mentioned in the introduction, the following holds [14, Theorem 2]:

**Proposition 1.** *For any product state  $\omega \in \mathcal{S}_V$ ,*

$$C(\omega) \leq \frac{|V|}{2}. \tag{8}$$

In the next sections, we aim at recovering the linear dependence of the constant  $C(\omega)$  on the size  $n = |V|$  of the system under various measures of independence.

We will need the following properties of the quantum  $W_1$  distance:

**Proposition 2** [14, Proposition 2]. *The quantum  $W_1$  distance coincides with the trace distance for quantum states that differ in only one site, i.e., for any  $X \in \mathcal{O}_V^T$  such that  $\text{Tr}_v X = 0$  for some  $v \in V$  we have*

$$\|X\|_{W_1} = \frac{1}{2} \|X\|_1. \tag{9}$$

**Proposition 3** [14, Proposition 5]. *The quantum  $W_1$  distance between two quantum states that differ only in the region  $A \subseteq V$  is at most  $2|A|$  times their trace distance, i.e., for any  $X \in \mathcal{O}_V^T$  such that  $\text{Tr}_A X = 0$  we have*

$$\|X\|_{W_1} \leq |A| \|X\|_1. \tag{10}$$

**Proposition 4** (Tensorization [14, Proposition 4]). *The quantum  $W_1$  distance is additive with respect to the tensor product, i.e., let  $A, B$  be disjoint subsets of  $V$ . Then, for any  $\rho_A, \sigma_A \in \mathcal{S}_A$  and any  $\rho_B, \sigma_B \in \mathcal{S}_B$  we have*

$$\|\rho_A \otimes \rho_B - \sigma_A \otimes \sigma_B\|_{W_1} = \|\rho_A - \sigma_A\|_{W_1} + \|\rho_B - \sigma_B\|_{W_1}. \tag{11}$$

**Proposition 5** [14, Proposition 13]. *Let  $\Phi : \mathcal{O}_V \rightarrow \mathcal{O}_V$  be a quantum channel. For any  $v \in V$ , let  $A_v \subseteq V$  be the light-cone of the site  $v$ , i.e., the minimum subset of  $V$  such that  $\text{Tr}_{A_v} \Phi(X) = 0$  for any  $X \in \mathcal{O}_V$  such that  $\text{Tr}_v X = 0$ . Then,  $\Phi$  can expand the quantum  $W_1$  distance by at most twice the size of the largest light-cone, i.e., for any  $X \in \mathcal{O}_V^T$  we have*

$$\|\Phi(X)\|_{W_1} \leq 2 \max_{v \in V} |A_v| \|X\|_{W_1}. \tag{12}$$

**Proposition 6** [14, Proposition 15]. *For any  $H \in \mathcal{O}_V$  and any  $v \in V$ , we have*

$$\left\| H - \mathbb{I}_v \otimes \frac{1}{d} \text{Tr}_v H \right\|_{\infty} \leq \|H\|_L. \tag{13}$$

**Proposition 7** [14, Corollary 1]. *For any  $\rho, \sigma \in \mathcal{S}_V$ ,*

$$\|\rho - \sigma\|_{W_1} \geq \frac{1}{2} \sum_{v \in V} \|\rho_v - \sigma_v\|_1, \tag{14}$$

*and equality holds whenever both  $\rho$  and  $\sigma$  are product states.*

**Theorem 1** ( $W_1$  continuity of the entropy [14, Theorem 1]). *For any  $\rho, \sigma \in \mathcal{S}_V$ , we have*

$$|S(\rho) - S(\sigma)| \leq g(\|\rho - \sigma\|_{W_1}) + \|\rho - \sigma\|_{W_1} \ln(d^2 |V|), \tag{15}$$

*where for any  $t \geq 0$*

$$g(t) = (t + 1) \ln(t + 1) - t \ln t. \tag{16}$$

### 3. Dobrushin Uniqueness Condition

In this section, we consider a spin chain and prove the transportation cost inequality under a quantum generalization of Dobrushin’s uniqueness condition [6]. Such condition is formulated in terms of the conditional probability distributions of the state of a subset of  $V$  conditioned on the state of a second disjoint subset of  $V$ . Therefore, formulating a quantum version of Dobrushin’s uniqueness condition requires a quantum counterpart of the conditional probability distribution. In the classical setting, given two random variables  $X$  and  $Y$  taking values in finite sets and with joint probability distribution  $\omega_{XY}$ , the conditional probability distribution  $\omega_{Y|X}$  of  $Y$  given  $X$  with probability mass function

$$\omega_{Y|X=x}(y) = \frac{\omega_{XY}(x, y)}{\omega_X(x)} \tag{17}$$

represents the knowledge that we have on  $Y$  when we know only the value of  $X$ . We can associate with such conditional distribution the stochastic map  $\Phi_{X \rightarrow XY}$  that has as input a probability distribution  $p_X$  for  $X$  and as output the joint probability distribution of  $XY$  with probability mass function given by

$$\Phi_{X \rightarrow XY}(p_X)(x, y) = \omega_{Y|X=x}(y) p_X(x) = \frac{\omega_{XY}(x, y)}{\omega_X(x)} p_X(x). \tag{18}$$

In the quantum setting, we consider a bipartite quantum system  $AB$  and a joint quantum state  $\omega_{AB}$  of  $AB$ . The quantum counterpart of the stochastic map (18) is called quantum recovery map [28, 29], and its action on a quantum state  $\rho_A$  of  $A$  is

$$\Phi_{A \rightarrow AB}(\rho_A) = \int_{\mathbb{R}} \omega_{AB}^{\frac{1-it}{2}} \omega_A^{\frac{it-1}{2}} \rho_A \omega_A^{-\frac{1+it}{2}} \omega_{AB}^{\frac{1+it}{2}} d\mu_0(t), \tag{19}$$

where  $\mu_0$  is the probability distribution on  $\mathbb{R}$  with density

$$d\mu_0(t) = \frac{\pi dt}{2(\cosh(\pi t) + 1)}. \tag{20}$$

We stress that (19) reduces to (18) whenever  $\rho_A, \omega_A$  and  $\omega_{AB}$  commute. If  $A$  is in the state  $\omega_A$ , the recovery map  $\Phi_{A \rightarrow AB}$  recovers the joint state  $\omega_{AB}$ , i.e.,  $\Phi_{A \rightarrow AB}(\omega_A) = \omega_{AB}$ . The relevance of the recovery map comes from the recoverability theorem [29], which states that  $\Phi_{A \rightarrow AB}$  can recover a generic joint state  $\rho_{AB}$  from its marginal  $\rho_A$  if removing the subsystem  $B$  does not significantly decrease the relative entropy between  $\rho$  and  $\omega$ . More precisely, for any quantum state  $\rho_{AB}$  of  $AB$  we have

$$S(\rho_{AB} \parallel \omega_{AB}) - S(\rho_A \parallel \omega_A) \geq S_M(\rho_{AB} \parallel \Phi_{A \rightarrow AB}(\rho_A)). \tag{21}$$

We consider the setting where  $V$  is partitioned as

$$V = A_1 \sqcup \dots \sqcup A_m. \tag{22}$$

For any  $i \in [m]$ , we denote with  $A_1^i$  the union  $A_1 \sqcup \dots \sqcup A_i$ . The recoverability theorem implies the following Lemma 1, which we will employ several times:

**Lemma 1.** *For any  $\rho, \omega \in \mathcal{S}_V$ , we have*

$$S(\rho \parallel \omega) \geq \frac{1}{2m} \left( \sum_{i=1}^m \left\| \rho_{A_1^i} - \Phi_{A_1^{i-1} \rightarrow A_1^i}(\rho_{A_1^i}) \right\|_1 \right)^2, \tag{23}$$

$$\left( \frac{1}{2m} \sum_{i=1}^m \left\| \rho_{A_1^i} - \Phi_{A_1^{i-1} \rightarrow A_1^i}(\rho_{A_1^i}) \right\|_1 \right)^2 \leq 1 - \exp\left(-\frac{S(\rho \parallel \omega)}{m}\right), \tag{24}$$

where  $\Phi_{A_1^{i-1} \rightarrow A_1^i}$  are the recovery maps associated with  $\omega$ .

*Proof.* Eq. (19) and Pinsker’s inequality imply for any  $i \in [m]$

$$\begin{aligned} S(\rho_{A_1^i} \parallel \omega_{A_1^i}) - S(\rho_{A_1^{i-1}} \parallel \omega_{A_1^{i-1}}) &\geq S_M\left(\rho_{A_1^i} \parallel \Phi_{A_1^{i-1} \rightarrow A_1^i}(\rho_{A_1^{i-1}})\right) \\ &\geq \frac{1}{2} \left\| \rho_{A_1^i} - \Phi_{A_1^{i-1} \rightarrow A_1^i}(\rho_{A_1^{i-1}}) \right\|_1^2. \end{aligned} \tag{25}$$

Summing (25) over  $i$  and using the convexity of the square function yields

$$\begin{aligned} S(\rho \parallel \omega) &\geq \frac{1}{2} \sum_{i=1}^m \left\| \rho_{A_1^i} - \Phi_{A_1^{i-1} \rightarrow A_1^i}(\rho_{A_1^{i-1}}) \right\|_1^2 \\ &\geq \frac{1}{2m} \left( \sum_{i=1}^m \left\| \rho_{A_1^i} - \Phi_{A_1^{i-1} \rightarrow A_1^i}(\rho_{A_1^i}) \right\|_1 \right)^2. \end{aligned} \tag{26}$$

The claim (23) follows.

With an analogous proof, applying the improved Pinsker’s inequality

$$\frac{1}{4} \|\sigma - \tau\|_1^2 \leq 1 - e^{-S_M(\sigma \parallel \tau)} \tag{27}$$

and the convexity of the function  $t \mapsto -\ln(1 - t^2)$ , we get

$$S(\rho \parallel \omega) \geq -m \ln \left( 1 - \left( \frac{1}{2m} \sum_{i=1}^m \left\| \rho_{A_1^i} - \Phi_{A_1^{i-1} \rightarrow A_1^i}(\rho_{A_1^i}) \right\|_1 \right)^2 \right). \tag{28}$$

The claim (24) follows. □

The following property of the recovery map will be fundamental:

**Lemma 2.** *Let  $\omega_{ABC}$  be a joint state of the tripartite quantum system  $ABC$ . Let us assume that  $\omega_{ABC}$  is Markovian, i.e.,*

$$I(A; C|B)_\omega = 0. \tag{29}$$

*Then, the recovery map  $\Phi_{AB \rightarrow ABC}$  associated with  $\omega_{ABC}$  does not act on the subsystem  $A$ .*

*Proof.* From the characterization of the states that saturate the strong subadditivity [30], the Hilbert space  $\mathcal{H}_B$  of  $B$  has a decomposition

$$\mathcal{H}_B = \bigoplus_{i=1}^k \mathcal{H}_{B_i^L} \otimes \mathcal{H}_{B_i^R}, \tag{30}$$

where the Hilbert spaces  $\{\mathcal{H}_{B_i^L}\}_{i=1}^k$  and  $\{\mathcal{H}_{B_i^R}\}_{i=1}^k$  are pairwise orthogonal, and  $\omega_{ABC}$  can be expressed as

$$\omega_{ABC} = \bigoplus_{i=1}^k p_i \omega_{AB_i^L}^{(i)} \otimes \omega_{B_i^R C}^{(i)}, \tag{31}$$

where  $p$  is a probability distribution on  $[k]$ , and each  $\omega_{AB_i^L}^{(i)}$  or  $\omega_{B_i^R C}^{(i)}$  is a quantum state with support in the corresponding  $\mathcal{H}_A \otimes \mathcal{H}_{B_i^L}$  or  $\mathcal{H}_{B_i^R} \otimes \mathcal{H}_C$ . We have for any quantum state  $\rho_{AB}$  of  $AB$

$$\Phi_{AB \rightarrow ABC}(\rho_{AB}) = \int_{\mathbb{R}} \omega_{ABC}^{\frac{1-it}{2}} \omega_{AB}^{\frac{it-1}{2}} \rho_{AB} \omega_{AB}^{-\frac{1+it}{2}} \omega_{ABC}^{\frac{1+it}{2}} d\mu_0(t). \tag{32}$$

We have for any  $z \in \mathbb{C}$  that

$$\omega_{ABC}^z \omega_{AB}^{-z} = \bigoplus_{i=1}^k \left( \omega_{B_i^R C}^{(i)} \right)^z \left( \omega_{B_i^L}^{(i)} \right)^{-z} \tag{33}$$

does not act on  $A$ , and the claim follows choosing  $z = (1 - it) / 2$ . □

### 3.1. Markovian Case

In this subsection, we assume that  $\omega \in \mathcal{S}_V$  is a one-dimensional quantum Markov state. More precisely, let  $\{A_1, \dots, A_m\}$  be a partition of  $V$  and let  $K = \max(|A_1|, \dots, |A_m|)$ . Then, we assume that

$$I(A_i; A_1^{i-2} | A_{i-1})_\omega = 0 \tag{34}$$

for any  $i = 3, \dots, m$ . For any  $i \in [m]$ , let  $\Phi_i$  be the recovery map (19) associated with  $\omega_{A_1^i}$  that recovers  $A_i$  from  $A_1^{i-1}$ . From Lemma 2,  $\Phi_i$  acts only on  $A_{i-1}$ , i.e., it is a map  $\Phi_i : \mathcal{O}_{A_{i-1}} \rightarrow \mathcal{O}_{A_{i-1} A_i}$ . We also define

$$\tilde{\Phi}_i = \text{Tr}_{A_{i-1}} \circ \Phi_i : \mathcal{O}_{A_{i-1}} \rightarrow \mathcal{O}_{A_i}. \tag{35}$$

We can now state the main result of this section:



**Theorem 2.** *Let us assume that for any  $i \in [m]$ ,  $\tilde{\Phi}_i$  is a contraction with respect to the trace norm for all the couples of quantum states of  $A_1^{i-1}$  that differ only on the subsystem  $A_{i-1}$ , i.e., that coincide after discarding  $A_{i-1}$ . More precisely, we assume that there exists  $0 \leq \eta < 1$  such that for any  $i \in [m]$  and any  $X \in \mathcal{O}_{A_1^{i-1}}^T$  with  $\text{Tr}_{A_{i-1}} X = 0$  we have*

$$\left\| \tilde{\Phi}_i(X) \right\|_1 \leq \eta \|X\|_1. \tag{36}$$

Then, we have

$$C(\omega) \leq 2m K^2 \left( \frac{1}{1-\eta} + 1 \right)^2. \tag{37}$$

Furthermore, for any  $\rho \in \mathcal{S}_V$  we have

$$\|\rho - \omega\|_{W_1} \leq K \left( \frac{1}{1-\eta} + 1 \right) 2m \sqrt{1 - e^{-\frac{S(\rho\|\omega)}{m}}}. \tag{38}$$

*Proof.* Let  $\rho \in \mathcal{S}_V$ . On the one hand, we have from Lemma 3

$$\begin{aligned} \|\rho - \omega\|_{W_1} &= \left\| \sum_{i=1}^m (\Phi_m \circ \dots \circ \Phi_{i+1})(\rho_{A_1^i} - \Phi_i(\rho_{A_1^{i-1}})) \right\|_{W_1} \\ &\leq \sum_{i=1}^m \left\| (\Phi_m \circ \dots \circ \Phi_{i+1})(\rho_{A_1^i} - \Phi_i(\rho_{A_1^{i-1}})) \right\|_{W_1} \\ &\leq K \left( \frac{1}{1-\eta} + 1 \right) \sum_{i=1}^m \left\| \rho_{A_1^i} - \Phi_i(\rho_{A_1^{i-1}}) \right\|_1. \end{aligned} \tag{39}$$

On the other hand, we have from (23) of Lemma 1

$$S(\rho\|\omega) \geq \frac{1}{2m} \left( \sum_{i=1}^m \left\| \rho_{A_1^i} - \Phi_i(\rho_{A_1^{i-1}}) \right\|_1 \right)^2, \tag{40}$$

and the claim (37) follows. The claim (38) follows by employing (24) in place of (23).  $\square$

*Remark 1.* Condition (36) holds for some  $\eta < 1$  iff  $\tilde{\Phi}_i$  strictly decreases the trace distance between any two quantum states that differ only in the subsystem  $A_i$  on which  $\tilde{\Phi}_i$  acts, i.e., that coincide after discarding  $A_i$ . We expect this condition to hold for any strictly positive temperature.

*Remark 2.* An example of quantum state satisfying (34) is a Gibbs state of a nearest-neighbor Hamiltonian on the  $D$ -dimensional cubic lattice  $\Lambda = [L]^D$ , where  $x, y \in \Lambda$  are neighbors iff  $\|x - y\|_1 = 1$ . We can then choose  $m = L + 1$  and

$$A_i = \{x \in \Lambda : x_1 = i\}, \quad i = 0, \dots, L, \tag{41}$$

with

$$K = (L + 1)^{D-1}, \tag{42}$$

and get from Theorem 2

$$C(\omega) \leq 2(L+1)^{2D-1} \left( \frac{1}{1-\eta} + 1 \right)^2 = 2|V|^{\frac{2D-1}{D}} \left( \frac{1}{1-\eta} + 1 \right)^2. \tag{43}$$

We stress that, assuming that  $\eta$  remains bounded away from 1, we get  $C(\omega) = O(|V|)$  iff  $D = 1$ , i.e., for one-dimensional systems.

*Remark 3.* We can choose

$$\begin{aligned} \eta &= \max \left\{ \left\| \tilde{\Phi}_i(X) - \omega_{A_i} \otimes \text{Tr}_{A_{i-1}} X \right\|_1 : i \in [m], X \in \mathcal{O}_{A_{i-1}}, \|X\|_1 = 1 \right\} \\ &\leq \max_{i \in [m]} \left\| \tilde{\Phi}_i - \omega_{A_i} \otimes \text{Tr}_{A_{i-1}} \right\|_{\diamond}, \end{aligned} \tag{44}$$

where  $\omega_{A_i} \otimes \text{Tr}_{A_{i-1}} : \mathcal{O}_{A_{i-1}} \rightarrow \mathcal{O}_{A_i}$  is the quantum channel that replaces the input quantum state with  $\omega_{A_i}$  and

$$\|\Phi\|_{\diamond} = \sup \left\{ \left\| (\Phi \otimes \mathbb{I}_{\mathcal{B}(\mathcal{H})})(X) \right\|_1 : X \in \mathcal{B}(\mathcal{H}^{\otimes 2}), \|X\|_1 = 1 \right\} \tag{45}$$

denotes the diamond norm of the linear map  $\Phi$  on  $\mathcal{B}(\mathcal{H})$ .

**Proposition 8.** *Let  $\omega \in S_V$  satisfy (34), and assume*

$$a = \max_{i \in [m-1]} S_{\infty}(\omega_{A_i} \otimes \omega_{A_{i+1}} \| \omega_{A_i A_{i+1}}) < \frac{1}{2}, \tag{46}$$

where

$$S_{\infty}(\rho \| \sigma) = \ln \inf \{ \lambda \in \mathbb{R} : \rho \leq \lambda \sigma \} \tag{47}$$

denotes the quantum max-divergence [31] between the quantum states  $\rho$  and  $\sigma$ . Then, we can choose in (36)

$$\eta = \sqrt{2a}. \tag{48}$$

*Proof.* From Remark 1, we can choose

$$\begin{aligned} \eta &= \max_{i \in [m]} \max_{|\psi_i\rangle} \left\| \tilde{\Phi}_i(|\psi_i\rangle\langle\psi_i|) - \omega_{A_i} \otimes \text{Tr}_{A_{i-1}} |\psi_i\rangle\langle\psi_i| \right\|_1 \\ &\leq \max_{i \in [m]} \max_{|\psi_i\rangle} \left\| \Phi_i(|\psi_i\rangle\langle\psi_i|) - \omega_{A_i} \otimes |\psi_i\rangle\langle\psi_i| \right\|_1, \end{aligned} \tag{49}$$

where each  $|\psi_i\rangle$  is a unit vector in  $\mathcal{H}_{A_1^{i-1}}$ . We have from Pinsker’s inequality

$$\left\| \Phi_i(|\psi_i\rangle\langle\psi_i|) - |\psi_i\rangle\langle\psi_i| \otimes \omega_{A_i} \right\|_1 \leq \sqrt{2 S_{\mathbb{M}}(|\psi_i\rangle\langle\psi_i| \otimes \omega_{A_i} \| \Phi_i(|\psi_i\rangle\langle\psi_i|))}. \tag{50}$$

(46) implies

$$\ln \omega_{A_{i-1} A_i} \geq \ln \omega_{A_{i-1}} + \ln \omega_{A_i} - a. \tag{51}$$

From the characterization of the states that saturate the strong subadditivity [30], we get

$$\ln \omega_{A_1^{i-1}} + \ln \omega_{A_{i-1} A_i} = \ln \omega_{A_{i-1}} + \ln \omega_{A_i}, \tag{52}$$

therefore, (51) can be rewritten as

$$\ln \omega_{A_1^i} \geq \ln \omega_{A_1^{i-1}} + \ln \omega_{A_i} - a. \tag{53}$$

Choosing in (25)  $\rho_{A_i} = |\psi_i\rangle\langle\psi_i| \otimes \omega_{A_i}$  we get with the help of (53)

$$\begin{aligned} & S_{\mathbb{M}}(|\psi_i\rangle\langle\psi_i| \otimes \omega_{A_i} \| \Phi_i(|\psi_i\rangle\langle\psi_i|)) \\ & \leq \langle\psi_i| \left( \ln \omega_{A_i^{i-1}} - \text{Tr}_{A_i} \left[ \omega_{A_i} \ln \omega_{A_i} \right] \right) |\psi_i\rangle - S(A_i)_\omega \leq a, \end{aligned} \tag{54}$$

and the claim follows. □

*Remark 4.* Condition (36) is reminiscent of the so-called Dobrushin uniqueness condition (see [6, Theorem 4]).

### 3.2. Non-Markovian States

Here, we prove an alternative version of Theorem 2 where the Markov condition (34) is replaced by exponential decay of correlations.

**Theorem 3.** *Let  $V = [n]$  be a one-dimensional lattice, and let  $\omega \in \mathcal{S}_V$ . For any  $i \in [n]$ , let  $\Phi_i$  be the recovery map associated with  $\omega_{1\dots i}$  that recovers the site  $i$  from the sites  $1 \dots i - 1$ . We assume that  $\omega$  has exponentially decaying correlations, in the sense that there exist  $C \geq 0$  and  $0 \leq \eta < 1$  such that for any  $i \in [n]$ , any  $k = 0, \dots, \max(i, n - i)$  and any  $\tau \in \mathcal{S}_{1\dots i}$ ,*

$$\| \text{Tr}_{i-k+1\dots i+k} (\Phi_n \circ \dots \circ \Phi_{i+1})(\tau_{1\dots i}) - \tau_{1\dots i-k} \otimes \omega_{i+k+1\dots n} \|_1 \leq C \eta^k. \tag{55}$$

*We also assume that for any  $i \in [n]$ , any  $k = 0, \dots, i - 1$  and any  $\tau \in \mathcal{S}_{1\dots i-1}$*

$$\| \text{Tr}_{i-k\dots i} \Phi_i(\tau_{1\dots i-1}) - \tau_{1\dots i-k-1} \|_1 \leq C \eta^k. \tag{56}$$

*Then,*

$$C(\omega) \leq 8n \left( 2 + \frac{C+1}{1-\eta} - \frac{\ln(C^2 n)}{2 \ln \eta} \right)^2. \tag{57}$$

*Proof.* From (23) of Lemma 1, we have for any  $\rho \in \mathcal{S}_V$

$$S(\rho \| \omega) \geq \frac{1}{2n} \left( \sum_{i=1}^n \| \rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1}) \|_1 \right)^2. \tag{58}$$

We have

$$\begin{aligned} \| \rho - \omega \|_{W_1} &= \left\| \sum_{i=1}^n (\Phi_n \circ \dots \circ \Phi_{i+1})(\rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1})) \right\|_{W_1} \\ &\leq \sum_{i=1}^n \| (\Phi_n \circ \dots \circ \Phi_{i+1})(\rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1})) \|_{W_1}. \end{aligned} \tag{59}$$

For any  $i \in [n]$ , we have from Lemma 5

$$\begin{aligned} & \| (\Phi_n \circ \dots \circ \Phi_{i+1})(\rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1})) \|_{W_1} \\ & \leq 2 \sum_{k=0}^{\max\{i, n-i\}} \| \text{Tr}_{i-k+1\dots i+k} (\Phi_n \circ \dots \circ \Phi_{i+1})(\rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1})) \|_1 \\ & \leq 2 \sum_{k=0}^{\max\{i, n-i\}} (\| \rho_{1\dots i-k} - \text{Tr}_{i-k+1\dots i} \Phi_i(\rho_{1\dots i-1}) \|_1 \\ & \quad + C \eta^k \| \rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1}) \|_1). \end{aligned} \tag{60}$$

We have for any  $k_0 \in \{0, \dots, i - 1\}$

$$\begin{aligned}
 & \sum_{k=0}^{i-1} \|\rho_{1\dots i-k} - \text{Tr}_{i-k+1\dots i} \Phi_i(\rho_{1\dots i-1})\|_1 \\
 & \leq \sum_{k=0}^{k_0} \|\rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1})\|_1 + \sum_{k=k_0+1}^{i-1} \|\rho_{1\dots i-k} - \omega_{1\dots i-k} \\
 & \quad - \text{Tr}_{i-k+1\dots i} \Phi_i(\rho_{1\dots i-k} - \omega_{1\dots i-k})\|_1 \\
 & \leq (k_0 + 1) \|\rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1})\|_1 + C \sum_{k=k_0+1}^{i-1} \eta^{k-1} \|\rho_{1\dots i-k} - \omega_{1\dots i-k}\|_1 \\
 & \leq (k_0 + 1) \|\rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1})\|_1 + \frac{C \eta^{k_0}}{1 - \eta} \|\rho - \omega\|_1, \tag{61}
 \end{aligned}$$

therefore

$$\begin{aligned}
 & \|(\Phi_n \circ \dots \circ \Phi_{i+1})(\rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1}))\|_{W_1} \\
 & \leq 2 \left( \left( k_0 + 1 + \frac{C}{1 - \eta} \right) \|\rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1})\|_1 + \frac{C \eta^{k_0}}{1 - \eta} \|\rho - \omega\|_1 \right), \tag{62}
 \end{aligned}$$

and

$$\begin{aligned}
 \|\rho - \omega\|_{W_1} & \leq 2 \left( \left( k_0 + 1 + \frac{C}{1 - \eta} \right) \sum_{i=1}^n \|\rho_{1\dots i} - \Phi_i(\rho_{1\dots i-1})\|_1 + \frac{n C \eta^{k_0}}{1 - \eta} \|\rho - \omega\|_1 \right) \\
 & \leq 2 \left( k_0 + 1 + C \frac{1 + \eta^{k_0} \sqrt{n}}{1 - \eta} \right) \sqrt{2 n S(\rho \|\omega)}, \tag{63}
 \end{aligned}$$

and the claim follows choosing

$$k_0 = \left\lceil -\frac{\ln(C^2 n)}{2 \ln \eta} \right\rceil. \tag{64}$$

□

### 3.3. Auxiliary Lemmas

**Lemma 3.** *Under the same hypotheses of Theorem 2, for any  $i \in [m]$  and any  $X \in \mathcal{O}_{A_i}^T$  such that*

$$\text{Tr}_{A_{i-1} A_i} X = 0 \tag{65}$$

we have

$$\|(\Phi_m \circ \dots \circ \Phi_{i+1})(X)\|_{W_1} \leq K \left( \frac{1}{1 - \eta} + 1 \right) \|X\|_1. \tag{66}$$

*Proof.* We have from Lemma 4 and from the contractivity of the trace distance

$$\begin{aligned}
 \|(\Phi_m \circ \dots \circ \Phi_{i+1})(X)\|_{W_1} & \leq |A_{i-1}| \|(\Phi_m \circ \dots \circ \Phi_{i+1})(X)\|_1 \\
 & \quad + \|(\Phi_m \circ \dots \circ \Phi_{i+1})(\text{Tr}_{A_{i-1}} X)\|_{W_1} \\
 & \leq K \|X\|_1 + \|(\Phi_m \circ \dots \circ \Phi_{i+1})(\text{Tr}_{A_{i-1}} X)\|_{W_1}. \tag{67}
 \end{aligned}$$

We have

$$\begin{aligned}
 & \|(\Phi_m \circ \dots \circ \Phi_{i+1})(\text{Tr}_{A_{i-1}} X)\|_{W_1} \\
 & \leq |A_i| \|(\Phi_m \circ \dots \circ \Phi_{i+1})(\text{Tr}_{A_{i-1}} X)\|_1 \\
 & \quad + \|\text{Tr}_{A_i}(\Phi_m \circ \dots \circ \Phi_{i+1})(\text{Tr}_{A_{i-1}} X)\|_{W_1} \\
 & \leq K \|\text{Tr}_{A_{i-1}} X\|_1 + \|(\Phi_m \circ \dots \circ \Phi_{i+2} \circ \tilde{\Phi}_{i+1})(\text{Tr}_{A_{i-1}} X)\|_{W_1}. \tag{68}
 \end{aligned}$$

Iterating the procedure, we get

$$\begin{aligned}
 & \|(\Phi_m \circ \dots \circ \Phi_{i+1})(\text{Tr}_{A_{i-1}} X)\|_{W_1} \\
 & \leq K \left( \|\text{Tr}_{A_{i-1}} X\|_1 + \|\tilde{\Phi}_{i+1}(\text{Tr}_{A_{i-1}} X)\|_1 \right. \\
 & \quad \left. + \dots + \|(\tilde{\Phi}_m \circ \dots \circ \tilde{\Phi}_{i+1})(\text{Tr}_{A_{i-1}} X)\|_1 \right) \\
 & \quad + \|\text{Tr}_{A_m}(\tilde{\Phi}_m \circ \dots \circ \tilde{\Phi}_{i+1})(\text{Tr}_{A_{i-1}} X)\|_{W_1} \\
 & \leq K(1 + \eta + \dots + \eta^{m-i}) \|\text{Tr}_{A_{i-1}} X\|_1 + \|\text{Tr}_{A_{i-1}A_i} X\|_1 \\
 & \leq \frac{K}{1 - \eta} \|\text{Tr}_{A_{i-1}} X\|_1 \leq \frac{K}{1 - \eta} \|X\|_1, \tag{69}
 \end{aligned}$$

where the last two inequalities follow from (36) and (65), respectively. The claim follows. □

**Lemma 4.** For any  $X \in \mathcal{O}_V^T$  and any  $A \subseteq V$ ,

$$\|X\|_{W_1} \leq |A| \|X\|_1 + \|\text{Tr}_A X\|_{W_1}. \tag{70}$$

*Proof.* Without loss of generality, we can assume that  $V = [n]$  and  $A = [k]$  for some  $k \in [n]$ . We have

$$\begin{aligned}
 \|X\|_{W_1} & \leq \left\| X - \frac{\mathbb{I}}{d} \otimes \text{Tr}_1 X \right\|_{W_1} + \left\| \frac{\mathbb{I}}{d} \otimes \text{Tr}_1 X \right\|_{W_1} \\
 & = \frac{1}{2} \left\| X - \frac{\mathbb{I}}{d} \otimes \text{Tr}_1 X \right\|_1 + \|\text{Tr}_1 X\|_{W_1} \\
 & \leq \|X\|_1 + \|\text{Tr}_1 X\|_{W_1}, \tag{71}
 \end{aligned}$$

where the equality follows from Propositions 2 and 4 and the last inequality follows from the triangle inequality for the trace norm and its contractivity with respect to partial traces. By induction, we get

$$\begin{aligned}
 \|X\|_{W_1} & \leq (\|X\|_1 + \dots + \|\text{Tr}_{1\dots k-1} X\|_1) + \|\text{Tr}_{1\dots k} X\|_{W_1} \\
 & \leq k \|X\|_1 + \|\text{Tr}_{1\dots k} X\|_{W_1}, \tag{72}
 \end{aligned}$$

and the claim follows. □

**Lemma 5.** Let  $V = [n]$ . Then, for any  $X \in \mathcal{O}_V^T$ ,

$$\|X\|_{W_1} \leq \|X\|_1 + \|\text{Tr}_1 X\|_1 + \dots + \|\text{Tr}_{1\dots n-1} X\|_1. \tag{73}$$

*Proof.* Follows from Lemma 4. □

### 4. Curvature Bound

In the seminal paper [32], Ollivier introduced a generalization of the notion of curvature to generic, possibly discrete, metric spaces. In his framework, the curvature of a metric space  $(\Omega, d)$  endowed with a classical stochastic map  $P$  acting on the probability measures on  $\Omega$  is defined as the following contraction property of the Wasserstein distance  $W_1$ : for any two probability measures  $\mu_1, \mu_2$ ,

$$W_1(P(\mu_1), P(\mu_2)) \leq (1 - \kappa) W_1(\mu_1, \mu_2). \tag{74}$$

The constant  $\kappa > 0$  is called the coarse Ricci curvature of the triple  $(\Omega, d, P)$ . In particular, it is easy to verify that the existence of a positive coarse Ricci curvature induces the uniqueness of the invariant measure  $\nu$  for the Markov kernel  $P$ . Moreover, it was recently proven in [33] that Ollivier’s coarse Ricci curvature provides an upper bound on the transportation cost inequality for the measure  $\nu$ , hence recovering the results from the smooth Riemannian setting.

Here, inspired by the works of [32] and [33], we prove that a contraction of the Lipschitz constant under a certain quantum channel constructed from the Petz recovery maps of the Gibbs state  $\omega$  can be used to conclude that  $\omega$  satisfies a transportation cost inequality. In particular, we do not need to assume that the underlying graph is  $\mathbb{Z}$ , in contrast with Sect. 3. Let  $G = (V, E)$  be a hypergraph with  $n = |V|$ , and let  $H := \sum_{A \in E} h_A$  be a Hamiltonian whose local terms  $h_A$  pairwise commute and are supported on the hyperedges  $A \in E$ . For a given site  $v \in V$ , we recall the composition of the partial trace  $\text{Tr}_v$  on  $v$  with the rotated Petz recovery map of  $v$ :

$$\Psi_v(\rho) = \Phi_v \circ \text{Tr}_v(\rho) = \int_{\mathbb{R}} \omega^{\frac{1-it}{2}} \omega_{v^c}^{\frac{-1+it}{2}} (\rho_{v^c} \otimes I_v) \omega_{v^c}^{\frac{-1-it}{2}} \omega^{\frac{1+it}{2}} d\mu_0(t) \tag{75}$$

for the probability density  $\mu_0(t) := \frac{\pi}{2} (\cosh(\pi t) + 1)^{-1}$ . Note that since we assumed  $\omega$  to be the Gibbs state of a commuting Hamiltonian, the map  $\Psi_v$  acts non-trivially on the neighborhood of  $v$

$$N_v := \bigcup \{A \in E : v \in A\}. \tag{76}$$

We also introduce the quantum channel

$$\Psi = \frac{1}{n} \sum_{v \in V} \Psi_v. \tag{77}$$

We assume that  $\Psi$  is a contraction with respect to the  $W_1$  norm, *i.e.*, that

$$\|\Psi\|_{W_1 \rightarrow W_1} = \max_{\Delta \in \mathcal{O}_V^T} \frac{\|\Psi(\Delta)\|_{W_1}}{\|\Delta\|_{W_1}} \leq 1 - \frac{\kappa}{n} \tag{78}$$

for some  $\kappa > 0$ , in analogy with (74). This contraction property was already derived in Ollivier’s original article [32] as a generalization of Dobrushin’s uniqueness condition. Here, we first prove that this condition implies the transportation cost inequality for the Gibbs state  $\omega \equiv \omega_\beta := e^{-\beta H} / \text{Tr} e^{-\beta H}$ :

**Theorem 4.** *With the conditions of the previous paragraph, we have*

$$C(\omega_\beta) \leq 2n \frac{N^2}{(1 - e^{-\kappa})^2}, \tag{79}$$

where  $N := \max_{v \in V} |N_v|$ .

*Proof.* We have for any state  $\rho \in \mathcal{S}_V$

$$\|\rho - \omega_\beta\|_{W_1} \leq \sum_{i=1}^n \|\Psi^{i-1}(\rho) - \Psi^i(\rho)\|_{W_1} + \|\Psi^n(\rho) - \omega_\beta\|_{W_1}. \tag{80}$$

The last term can be controlled by  $\|\rho - \omega_\beta\|_{W_1}$  thanks to the contraction (78):

$$\|\Psi^n(\rho) - \omega_\beta\|_{W_1} \leq \left(1 - \frac{\kappa}{n}\right)^n \|\rho - \omega_\beta\|_{W_1} \leq e^{-\kappa} \|\rho - \omega_\beta\|_{W_1}. \tag{81}$$

On the other hand, the sum on the right-hand side of (80) can be controlled as follows:

$$\begin{aligned} \sum_{i=1}^n \|\Psi^{i-1}(\rho) - \Psi^i(\rho)\|_{W_1} &\leq \frac{1}{n} \sum_{i=1}^n \sum_{v \in V} \|\Psi_v(\Psi^{i-1}(\rho)) - \Psi^{i-1}(\rho)\|_{W_1} \\ &\leq \frac{N}{n} \sum_{i=1}^n \sum_{v \in V} \|\Psi_v(\Psi^{i-1}(\rho)) - \Psi^{i-1}(\rho)\|_1, \end{aligned} \tag{82}$$

where the last inequality follows by Proposition 3. Proceeding as in the proof of Lemma 1, by the joint use of Pinsker’s inequality with the recoverability bound followed by the data processing inequality we can further bound the trace distances above so that

$$\begin{aligned} \sum_{i=1}^n \|\Psi^{i-1}(\rho) - \Psi^i(\rho)\|_{W_1} &\leq \frac{N}{n} \sum_{i=1}^n \sum_{v \in V} \sqrt{2 S_{\mathbb{M}}(\Psi^{i-1}(\rho) \|\Psi_v(\Psi^{i-1}(\rho)))} \\ &\leq N \sqrt{2 \sum_{i=1}^n \sum_{v \in V} S_{\mathbb{M}}(\Psi^{i-1}(\rho) \|\Psi_v(\Psi^{i-1}(\rho)))} \\ &\leq N \sqrt{2 \sum_{i=1}^n \sum_{v \in V} (S(\Psi^{i-1}(\rho) \|\omega_\beta) - S(\text{Tr}_v \Psi^{i-1}(\rho) \|\text{Tr}_v \omega_\beta))} \\ &\leq N \sqrt{2 \sum_{i=1}^n \sum_{v \in V} (S(\Psi^{i-1}(\rho) \|\omega_\beta) - S(\Psi_v(\Psi^{i-1}(\rho)) \|\omega_\beta))} \\ &\stackrel{(1)}{\leq} N \sqrt{2n \sum_{i=1}^n (S(\Psi^{i-1}(\rho) \|\omega_\beta) - S(\Psi(\Psi^{i-1}(\rho)) \|\omega_\beta))} \\ &\leq N \sqrt{2n S(\rho \|\omega_\beta)}. \end{aligned} \tag{83}$$

Inequality (1) above uses the concavity of the entropy, so that for any state  $\rho$

$$\begin{aligned}
 \frac{1}{n} \sum_{v \in V} (S(\rho \| \omega_\beta) - S(\Psi_v(\rho) \| \omega_\beta)) &= S(\rho \| \omega_\beta) + \frac{1}{n} \sum_{v \in V} S(\Psi_v(\rho)) \\
 &\quad + \frac{1}{n} \sum_{v \in V} \text{Tr} [\Psi_v(\rho) \ln \omega_\beta] \\
 &\leq S(\rho \| \omega_\beta) + S(\Psi(\rho)) + \text{Tr} [\Psi(\rho) \ln \omega_\beta] \\
 &= S(\rho \| \omega_\beta) - S(\Psi(\rho) \| \omega_\beta). \tag{84}
 \end{aligned}$$

Plugging (81) and (83) onto (80), the result follows. □

It remains to prove that (78) is satisfied at high enough temperature.

**Proposition 9.** *There exists an inverse temperature  $\beta_c > 0$  such that for all  $\beta < \beta_c$ , (78) holds for some constant  $\kappa(\beta) > 0$ . In particular, whenever  $N > 1$ , one can choose*

$$\beta_c = (5N \max_{A \in E} \|h_A\|_\infty)^{-1} W\left(\frac{1}{16d^3}\right), \tag{85}$$

where  $W$  denotes the Lambert function and is defined as the inverse of  $x \mapsto xe^x$ .

*Proof.* We have

$$\|\Psi\|_{W_1 \rightarrow W_1} = \max_{\Delta \in \mathcal{O}_V^T} \frac{\|\Psi(\Delta)\|_{W_1}}{\|\Delta\|_{W_1}}. \tag{86}$$

Any  $\Delta \in \mathcal{O}_V^T$  can be expressed as [14, Sect. III]

$$\Delta = \sum_{v \in V} \Delta_v \tag{87}$$

such that for any  $v \in V$ ,  $\Delta_v \in \mathcal{O}_V^T$  satisfies  $\text{Tr}_v \Delta_v = 0$  and

$$\|\Delta\|_{W_1} = \sum_{v \in V} \|\Delta_v\|_{W_1} = \frac{1}{2} \sum_{v \in V} \|\Delta_v\|_1. \tag{88}$$

Therefore, we have

$$\|\Psi\|_{W_1 \rightarrow W_1} = \max_{v \in V} \max \{ \|\Psi(\Delta_v)\|_{W_1} : \Delta_v \in \mathcal{O}_V^T, \text{Tr}_v \Delta_v = 0, \|\Delta_v\|_1 = 2 \}. \tag{89}$$

We have

$$\begin{aligned}
 \|\Psi(\Delta_v)\|_{W_1} &\leq \left\| \Psi(\Delta_v) - \frac{\mathbb{I}_v}{d} \otimes \text{Tr}_v \Psi(\Delta_v) \right\|_{W_1} + \left\| \frac{\mathbb{I}_v}{d} \otimes \text{Tr}_v \Psi(\Delta_v) \right\|_{W_1} \\
 &= \frac{1}{2} \left\| \Psi(\Delta_v) - \frac{\mathbb{I}_v}{d} \otimes \text{Tr}_v \Psi(\Delta_v) \right\|_1 + \|\text{Tr}_v \Psi(\Delta_v)\|_{W_1} \\
 &\leq \frac{1}{2} \|\Psi(\Delta_v)\|_1 + \frac{1}{2} \|\text{Tr}_v \Psi(\Delta_v)\|_1 + \|\text{Tr}_v \Psi(\Delta_v)\|_{W_1}, \tag{90}
 \end{aligned}$$



where the equality follows from Proposition 2 and 4. Since  $\text{Tr}_v \Delta_v = 0$ , we have  $\Psi_v(\Delta_v) = 0$ , and

$$\frac{1}{2} \|\Psi(\Delta_v)\|_1 \leq \frac{1}{2n} \sum_{w \in V \setminus v} \|\Psi_w(\Delta_v)\|_1 \leq 1 - \frac{1}{n}. \tag{91}$$

For any  $w \in V \setminus N_v$ , we have

$$\text{Tr}_v \Psi_w(\Delta_v) = \Psi_w(\text{Tr}_v \Delta_v) = 0. \tag{92}$$

Then,

$$\text{Tr}_v \Psi(\Delta_v) = \frac{1}{n} \sum_{w \in N_v \setminus v} \text{Tr}_v \Psi_w(\Delta_v). \tag{93}$$

We have for any  $w \in N_v \setminus v$ , recalling that  $v \in N_w$ ,

$$\text{Tr}_{N_w \setminus v} \text{Tr}_v \Psi_w(\Delta_v) = \text{Tr}_{N_w} \Psi_w(\Delta_v) = \text{Tr}_{N_w} \Delta_v = 0, \tag{94}$$

therefore,

$$\|\text{Tr}_v \Psi_w(\Delta_v)\|_{W_1} \leq (N - 1) \|\text{Tr}_v \Psi_w(\Delta_v)\|_1, \tag{95}$$

and

$$\|\text{Tr}_v \Psi(\Delta_v)\|_{W_1} \leq \frac{N - 1}{n} \sum_{w \in N_v \setminus v} \|\text{Tr}_v \Psi_w(\Delta_v)\|_1. \tag{96}$$

Moreover,

$$\|\text{Tr}_v \Psi(\Delta_v)\|_1 \leq \frac{1}{n} \sum_{w \in N_v \setminus v} \|\text{Tr}_v \Psi_w(\Delta_v)\|_1. \tag{97}$$

Putting together (90), (91), (96) and (97), we get

$$\begin{aligned} \|\Psi(\Delta_v)\|_{W_1} &\leq 1 - \frac{1}{n} + \frac{N - \frac{1}{2}}{n} \sum_{w \in N_v \setminus v} \|\text{Tr}_v \Psi_w(\Delta_v)\|_1 \\ &\leq 1 - \frac{1}{n} + \frac{N - \frac{1}{2}}{n} \sum_{w \in N_v \setminus v} (\|\omega_w \otimes \text{Tr}_{v \setminus w} \Delta_v\|_1 + 2 \|\Psi_w - \omega_w \otimes \text{Tr}_w\|_\diamond) \\ &= 1 - \frac{1}{n} + \frac{2N - 1}{n} \sum_{w \in N_v \setminus v} \|\Psi_w - \omega_w \otimes \text{Tr}_w\|_\diamond, \end{aligned} \tag{98}$$

where  $\omega_w \otimes \text{Tr}_w$  is the quantum channel that replaces with  $\omega_w$  the state of the site  $w$ . We then have

$$\|\Psi\|_{W_1 \rightarrow W_1} \leq 1 - \frac{1}{n} + \frac{2N - 1}{n} \sum_{w \in N_v \setminus v} \|\Psi_w - \omega_w \otimes \text{Tr}_w\|_\diamond. \tag{99}$$

We have

$$\begin{aligned} &\|\Psi_w - \omega_w \otimes \text{Tr}_w\|_\diamond \\ &\leq \int_{\mathbb{R}} \left\| \omega^{\frac{1-it}{2}} \omega_w^{\frac{it-1}{2}} (\mathbb{I}_w \otimes \text{Tr}_w[\cdot]) \omega_w^{-\frac{it+1}{2}} \omega^{\frac{1+it}{2}} - \omega_w \otimes \text{Tr}_w[\cdot] \right\|_\diamond d\mu_0(t). \end{aligned} \tag{100}$$

Since the Hamiltonian terms  $h_A$  commute we have that, given  $H_v := \sum_{A \ni v} h_A$ ,

$$\omega^{\frac{1-it}{2}} \omega_{v,c}^{\frac{it-1}{2}} = e^{-\beta \frac{1-it}{2} H_v} \left( \text{Tr}_v [e^{-\beta H_v}] \right)^{\frac{it-1}{2}}. \tag{101}$$

Now,

$$\begin{aligned} \left\| \omega^{\frac{1-it}{2}} \omega_{v,c}^{\frac{it-1}{2}} - d^{\frac{it-1}{2}} \mathbb{I} \right\|_{\infty} &\leq \left\| e^{-\beta \frac{1-it}{2} H_v} - \mathbb{I} \right\|_{\infty} \left\| \left( \text{Tr}_v [e^{-\beta H_v}] \right)^{\frac{it-1}{2}} \right\|_{\infty} \\ &\quad + \left\| \left( \text{Tr}_v [e^{-\beta H_v}] \right)^{\frac{it-1}{2}} - d^{\frac{it-1}{2}} \mathbb{I} \right\|_{\infty} \\ &\stackrel{(1)}{\leq} \beta \frac{\sqrt{1+t^2}}{2} \|H_v\|_{\infty} e^{\beta \frac{\sqrt{1+t^2}}{2} \|H_v\|_{\infty}} d^{-\frac{1}{2}} e^{\frac{\beta}{2} \|H_v\|_{\infty}} \\ &\quad + \frac{\sqrt{1+t^2}}{2} M^{1+\frac{\sqrt{1+t^2}}{2}} d \left\| d^{-1} \text{Tr}_v [e^{-\beta H_v}] - \mathbb{I} \right\|_{\infty} \\ &\leq \beta \sqrt{1+t^2} \|H_v\|_{\infty} d^{1+\frac{\sqrt{1+t^2}}{2}} e^{\beta \|H_v\|_{\infty}} \left( 2 + \frac{\sqrt{1+t^2}}{2} \right) \\ &\equiv f_v(\beta, t). \end{aligned} \tag{102}$$

Inequality (1) above follows from the operator convexity of  $x \mapsto x^{-\frac{1}{2}}$  as well as Lemma 6, where  $M := \max\{\|\text{Tr}_v [e^{-\beta H_v}]\|_{\infty}, \|\text{Tr}_v [e^{-\beta H_v}]^{-1}\|_{\infty}, d\} \leq d e^{\beta \|H_v\|_{\infty}}$ . Moreover,

$$e^{-2\beta \|H_v\|_{\infty}} d^{-1} \mathbb{I} \leq \omega_v \leq e^{2\beta \|H_v\|_{\infty}} d^{-1} \mathbb{I} \quad \Rightarrow \quad \|\omega_v - d^{-1} \mathbb{I}\|_1 \leq 2\beta \|H_v\|_{\infty} e^{2\beta \|H_v\|_{\infty}}. \tag{103}$$

Therefore,

$$\begin{aligned} &\left\| \omega^{\frac{1-it}{2}} \omega_{v,c}^{\frac{it-1}{2}} (\mathbb{I}_v \otimes \text{Tr}_v [\cdot]) \omega_{v,c}^{-\frac{it+1}{2}} \omega^{\frac{1+it}{2}} - \omega_v \otimes \text{Tr}_v [\cdot] \right\|_{\diamond} \\ &\leq d^{\frac{1}{2}} (e^{\beta \|H_v\|_{\infty}} + 1) f_v(\beta, t) + \|d^{-1} \mathbb{I} - \omega_v\|_{\infty} \\ &\leq d^{\frac{1}{2}} (e^{\beta \|H_v\|_{\infty}} + 1) f_v(\beta, t) + 2\beta \|H_v\|_{\infty} e^{2\beta \|H_v\|_{\infty}}, \end{aligned} \tag{104}$$

and the integrand in (100) tends to zero pointwise for  $\beta \rightarrow 0$ . On the other hand, we have for any  $t \in \mathbb{R}$

$$\begin{aligned} &\left\| \omega^{\frac{1-it}{2}} \omega_{v,c}^{\frac{it-1}{2}} (\mathbb{I}_v \otimes \text{Tr}_v [\cdot]) \omega_{v,c}^{-\frac{it+1}{2}} \omega^{\frac{1+it}{2}} - \omega_v \otimes \text{Tr}_v [\cdot] \right\|_{\diamond} \\ &\leq \left\| \omega^{\frac{1-it}{2}} \omega_{v,c}^{\frac{it-1}{2}} (\mathbb{I}_v \otimes \text{Tr}_v [\cdot]) \omega_{v,c}^{-\frac{it+1}{2}} \omega^{\frac{1+it}{2}} \right\|_{\diamond} + \|\omega_v \otimes \text{Tr}_v [\cdot]\|_{\diamond} \\ &\leq 2, \end{aligned} \tag{105}$$

therefore the integrand in (100) is uniformly bounded. Then, we get for all  $t \in \mathbb{R}_+$  that

$$\|\Psi_v - \omega_v \otimes \text{Tr}_v\|_{\diamond} \leq d^{\frac{1}{2}} (e^{\beta \|H_v\|_{\infty}} + 1) f_v(\beta, t) + 2\beta \|H_v\|_{\infty} e^{2\beta \|H_v\|_{\infty}} + 2\mu_0([-t, t]^c) \tag{106}$$

Therefore, for any  $0 < \kappa < 1$  there exists  $\beta(\kappa) > 0$  such that condition (78) is satisfied for all  $0 \leq \beta \leq \beta(\kappa)$ . More precisely, in view of (106) and (98), it is

sufficient that

$$4\beta \sqrt{1+t^2} C d^{\frac{3+\sqrt{1+t^2}}{2}} e^{\beta C(3+\frac{\sqrt{1+t^2}}{2})} + 2\mu_0([-t, t]^c) \leq \frac{N-1}{2N-1} \tag{107}$$

where  $C := \sup_v \|H_v\|_\infty$ . Moreover, it is clear that  $\mu_0([-t, t]^c) \leq 2e^{-\pi t}$ . The result follows after choosing  $t$  so that the exponentially decaying term  $4e^{-\pi t}$  counts for at most half the upper bound and solving (107) for  $\beta_c$ , up to some numerical simplifications.  $\square$

*Remark 5.* The lower bound (85) can be compared to that in the classical setting [32, Example 17] (see also [34]): there, the author showed that for a Hamiltonian of the form  $U(S) := -\sum_{x \sim y \in G} S(x)S(y) - H \sum_x S(x)$ , where  $S(x) \in \{-1, 1\}$  denotes the spin configuration at the site  $x$  of a graph  $G$ , i.e.,  $d = 2$ ,

$$\beta_c \geq \frac{1}{2} \ln \left( \frac{N+1}{N-1} \right) \sim_{N \rightarrow \infty} \frac{1}{N},$$

which shows asymptotic optimality of our result, up to numerical multiplicative constants. For comparison, the exact value of  $\beta_c$  for the Ising model on the regular infinite tree with degree  $N$  is known to be equal to  $\frac{1}{2} \ln \left( \frac{N}{N-2} \right)$ .

### 4.1. Auxiliary Lemma

**Lemma 6.** *For any positive, definite matrices  $A, B$  and all  $z \in \mathbb{C}$ ,*

$$\|A^z - B^z\|_\infty \leq |z| \max\{\|A\|_\infty, \|A^{-1}\|_\infty, \|B\|_\infty, \|B^{-1}\|_\infty\}^{1+|\operatorname{Re}(z)|} \|A - B\|_\infty, \tag{108}$$

*Proof.* It suffices to use a linear interpolation between  $A$  and  $B$ :  $A(s) := sA + (1-s)B$ . We have

$$\begin{aligned} A^z - B^z &= \int_0^1 \frac{d}{ds} A(s)^z ds \\ &= z \iint_{[0,1]^2} A(s)^{zu} \frac{d}{ds} \ln(A(s)) A^{z(1-u)} dsdu \\ &= z \iint_{[0,1]^2} \int_0^\infty A(s)^{zu} (A(s) + v)^{-1} (A - B)(A(s) + v)^{-1} \\ &\quad A(s)^{z(1-u)} dv du ds. \end{aligned} \tag{109}$$

Then,

$$\begin{aligned} \|A^z - B^z\|_\infty &\leq |z| \int_0^1 \int_0^\infty \|A(s)^{\operatorname{Re}(z)}\|_\infty \|(A(s) + v)^{-1}\|_\infty^2 \|A - B\|_\infty dv ds \\ &\leq |z| \|A - B\|_\infty M(z) \int_0^1 \|A(s)^{-1}\|_\infty ds \\ &\leq |z| M(z) \cdot M' \|A - B\|_\infty. \end{aligned} \tag{110}$$

by the operator convexity of  $x \mapsto x^{-1}$  where  $M(z) := \max_{s \in [0,1]} \|(sA + (1-s)B)^{\operatorname{Re}(z)}\|_\infty$  and  $M' := \max\{\|A^{-1}\|_\infty, \|B^{-1}\|_\infty\}$ . The result follows by operator convexity of the inverse function and further simple estimates.  $\square$

### 5. Modified Logarithmic Sobolev Inequalities

In this section, we pursue a different approach to prove transportation cost inequalities for  $W_1$ , namely through the existence of a non-commutative entropic inequality known as the *modified logarithmic Sobolev inequality* [35,36]. In order to introduce our main result, we need a variation of the Lipschitz constant that was introduced in [13]. This definition departs from a noncommutative differential structure, which we define below (see [37]):

**Definition 1.** (*Differential structure*) A set of operators  $L_k \in \mathcal{O}_V$  and constants  $\omega_k \in \mathbb{R}$  define a differential structure  $\{L_k, \omega_k\}_{k \in \mathcal{K}}$  for a full rank state  $\omega \in \mathcal{S}_V$  if

- 1  $\{L_k\}_{k \in \mathcal{K}} = \{L_k^\dagger\}_{k \in \mathcal{K}}$ ;
- 2  $\{L_k\}_{k \in \mathcal{K}}$  consists of eigenvectors of the modular operator  $\Delta_\omega(X) := \omega X \omega^{-1}$  with

$$\Delta_\omega(L_k) = e^{-\omega_k} L_k. \tag{111}$$

Such a differential structure can be used to provide the set of matrices with a Lipschitz constant that is tailored to  $\omega$ , see, e.g., [13,37] for more on this. In order to distinguish that constant from  $\|\cdot\|_L$ , we refer to it as the differential Lipschitz constant and denote it by  $\|\|\nabla X\|\|$ . It is defined as:

$$\|\|\nabla X\|\| := \left( \sum_{k \in \mathcal{K}} (e^{-\omega_k/2} + e^{\omega_k/2}) \|\partial_k X\|_\infty^2 \right)^{1/2}, \tag{112}$$

where  $\partial_k X \equiv [L_k, X]$ . For ease of notations, we will denote the differential structure by the couple  $(\nabla, \omega)$ . The notion of a differential structure is also intimately connected to that of the generator of a quantum dynamical semigroup converging to  $\omega$  [37], and properties of that semigroup immediately translate to properties of the metric. This is because the differential structure can be used to define an operator that behaves in an analogous way to the Laplacian on a smooth manifold, which in turn induces a heat semigroup. We refer to [13,37] for more details on this connection and interpretation.

When the state  $\omega$  is a quantum Gibbs state corresponding to a local, commuting Hamiltonian associated with a uniformly bounded interaction defined on a lattice  $V \subset \mathbb{Z}^D$ , the differential structure  $(\nabla, \omega)$  can be chosen as local. This means that the operators  $L_k \equiv L_{i,\alpha}$  are indexed by a site  $i \in V$  and an index  $\alpha$  of a set  $\Gamma$  whose cardinality only depends on the local dimension  $d$  and the locality  $\kappa$  of  $\omega$ . Moreover, we assume that the operators  $L_{i,\alpha}$  are supported on a neighborhood  $\mathcal{N}_i$  of site  $i$  of diameter  $r \equiv r(\kappa)$  and the corresponding constants  $\omega_{i,\alpha}$  are uniformly bounded:  $\sup_{i \in \mathbb{Z}^D} \max_{\alpha \in \Gamma} |\omega_{i,\alpha}| \equiv \Omega < \infty$ . The definition in Eq. (112) yields a metric on states by duality:

$$W_\nabla(\rho, \omega) := \sup_{X=X^\dagger, \|\|\nabla X\|\| \leq 1} |\text{Tr}(X(\rho - \omega))|.$$

**Proposition 10.** *Given the Gibbs state  $\omega$  of a local commuting Hamiltonian  $H$  on  $V \subset \mathbb{Z}^D$  with  $|V| = n$  and associated local differential structure  $(\nabla, \omega)$ ,*

the following bound holds for all  $\rho \in \mathcal{S}_V$ :

$$\|\rho - \omega\|_{W_1} \leq C \sqrt{n} W_{\nabla}(\rho, \omega),$$

for some constant  $C$  independent of  $n$ .

*Proof.* By duality, it is equivalent to prove that for all  $H \in \mathcal{O}_V$

$$\|\nabla H\| \leq C \sqrt{n} \|H\|_L.$$

First, we have

$$\|\nabla H\| = \left( \sum_{i \in V} \sum_{\alpha \in \Gamma} (e^{-\omega_{i,\alpha}/2} + e^{\omega_{i,\alpha}/2}) \|[L_{i,\alpha}, H]\|_{\infty}^2 \right)^{\frac{1}{2}} \tag{113}$$

$$\leq \sqrt{n|\Gamma|} \sqrt{2 e^{\Omega/2}} \max_{i \in V} \max_{\alpha \in \Gamma} \|[L_{i,\alpha}, H]\|_{\infty}. \tag{114}$$

Now, since for each pair  $(i, \alpha)$ ,  $L_{i,\alpha}$  is supported on a neighborhood  $\mathcal{N}_i$  of site  $i \in V$ ,

$$\|[L_{i,\alpha}, H]\|_{\infty} = \|[L_{i,\alpha}, H - \frac{\mathbb{I}_{\mathcal{N}_i}}{d^{|\mathcal{N}_i|}} \otimes \text{Tr}_{\mathcal{N}_i}(H)]\|_{\infty} \tag{115}$$

$$\leq 2 \|L_{i,\alpha}\|_{\infty} \|H - \frac{\mathbb{I}_{\mathcal{N}_i}}{d^{|\mathcal{N}_i|}} \otimes \text{Tr}_{\mathcal{N}_i}(H)\|_{\infty}. \tag{116}$$

Next, by a telescopic sum argument, we can further control the last infinity norm on the right hand side above as follows: given an arbitrary ordering of the region  $\mathcal{N}_i$ ,

$$\begin{aligned} & \|H - \frac{\mathbb{I}_{\mathcal{N}_i}}{d^{|\mathcal{N}_i|}} \otimes \text{Tr}_{\mathcal{N}_i}(H)\|_{\infty} \\ & \leq \sum_{j=1}^{|\mathcal{N}_i|} \left\| \left( \frac{\mathbb{I}_{1\dots j-1}}{d^{j-1}} \otimes \text{Tr}_{1\dots j-1} - \frac{\mathbb{I}_{1\dots j}}{d^j} \otimes \text{Tr}_{1\dots j} \right) (H) \right\|_{\infty} \end{aligned} \tag{117}$$

$$\stackrel{(2)}{\leq} |\mathcal{N}_i| \max_{j \in \mathcal{N}_i} \|H - \frac{\mathbb{I}_j}{d} \otimes \text{Tr}_j(H)\|_{\infty}, \tag{118}$$

where (1) follows from the triangle inequality, whereas (2) follows from the fact that the maps  $\frac{\mathbb{I}_{1\dots j-1}}{d^{j-1}} \otimes \text{Tr}_{1\dots j-1}$  are completely positive and unital, and therefore contract the operator norm. All in all, we have derived the following bound on the differential Lipschitz constant of  $H$ :

$$\|\nabla H\| \leq \sqrt{n|\Gamma|} 2\sqrt{2 e^{\Omega/2}} \max_{i \in V} \max_{\alpha \in \Gamma} \|L_{i,\alpha}\|_{\infty} |\mathcal{N}_i| \max_{j \in \mathcal{N}_i} \|H - \frac{\mathbb{I}_j}{d} \otimes \text{Tr}_j(H)\|_{\infty} \tag{119}$$

$$\stackrel{(3)}{\leq} \frac{d^2 - 1}{d^2} \sqrt{n|\Gamma|} 2\sqrt{2 e^{\Omega/2}} \max_{i \in V} \max_{\alpha \in \Gamma} \|L_{i,\alpha}\|_{\infty} |\mathcal{N}_i| \|H\|_L \tag{120}$$

$$\equiv C \sqrt{n} \max_{i \in V} \|H - \frac{\mathbb{I}_j}{d} \otimes \text{Tr}_j(H)\|_{\infty}, \tag{121}$$

for some constant  $C$  independent of  $n$ , and where (3) follows from Proposition 6. □

The advantage of  $W_1$  as compared to  $W_{\nabla}$  is that it does not depend on the state  $\omega$ . On the other hand, the bound derived in Proposition 10 can be used in conjunction with recently proved transportation cost inequalities for  $W_{\nabla}$  through the proof of the existence of a modified logarithmic Sobolev inequality in order to get analogous inequalities for  $W_1$  (see [38] for more details):

**Theorem 5.** *Let  $\omega$  be the Gibbs state of a local commuting Hamiltonian  $H$  at inverse temperature  $\beta$  on  $V \subset \mathbb{Z}^D$ . Then, there exists a critical inverse temperature  $\beta_c$  such that  $C(\omega_\beta) \leq Cn$  for some constant  $C$  independent of  $n = |V|$  whenever  $\beta < \beta_c$  if any of the two conditions below is satisfied:*

- (i)  $H$  is classical;
- (ii)  $H$  is a nearest neighbor Hamiltonian.

Moreover, we can drop the assumption of 2-locality in the 1D case, where  $\beta_c = 0$  at the cost of getting a slightly worsened constant  $C(\omega_\beta) \leq Cn \text{polylog}(n)$ , so that we recover the result of Theorem 2.

*Proof.* In [19,20,38], the existence of local differential structures associated with  $\omega$  that satisfy the so-called modified logarithmic Sobolev inequality was proved under the conditions of the theorem. Moreover, the modified logarithmic Sobolev inequality implies the transportation cost inequality for the differential Wasserstein distance [13]: there exists a constant  $C'$  independent of  $n$  such that

$$W_{\nabla}(\rho, \omega) \leq \sqrt{C' S(\rho \parallel \omega)} \quad (122)$$

for all state  $\rho \in \mathcal{S}_V$ . This fact in conjunction with Proposition 10 allows us to conclude.  $\square$

## 6. Local Indistinguishability

In this section, we provide a transportation cost inequality under a condition of local indistinguishability [39–41]. In the classical setting, this condition constitutes a weakening of Dobrushin Shlosman’s mixing condition [8] recently considered by Marton [7]. Moreover, as opposed to the latter, our technique has the benefit of not requiring the local specifications of the state to be uniformly lower bounded by a positive number, at the cost of getting a slightly worsened constant.

### 6.1. Transportation Cost from Local Indistinguishability

We start by proving our general result in the quantum setting. Here, we assume that the  $n = (2m + 1)^D$  qudits are arranged on a  $D$ -dimensional regular lattice  $V := [-m, m]^D$ . Before we state our main result, we need to introduce the notion of a non-commutative conditional expectation.

**Definition 2.** (*Conditional expectations*) Let  $\mathcal{N} \subseteq \mathcal{B}_V$  be a von Neumann sub-algebra<sup>1</sup> of  $\mathcal{B}_V$ . A conditional expectation onto  $\mathcal{N}$  is a completely positive unital map  $E_{\mathcal{N}}^{\dagger} : \mathcal{B}_V \rightarrow \mathcal{N}$  satisfying

- (i) for all  $X \in \mathcal{N}$ ,  $E_{\mathcal{N}}^{\dagger}(X) = X$ ;
- (ii) for all  $a, b \in \mathcal{N}$ ,  $X \in \mathcal{B}_V$ ,  $E_{\mathcal{N}}^{\dagger}(aXb) = aE_{\mathcal{N}}(X)b$ .

We denote by  $E_{\mathcal{N}}$  its adjoint map with respect to the trace inner product, i.e.,

$$\text{Tr}(E_{\mathcal{N}}(X)Y) = \text{Tr}(XE_{\mathcal{N}}^{\dagger}(Y)).$$

As a simple example, we consider a full-rank state  $\sigma \in \mathcal{S}_V$  and let  $(e^{t\mathcal{L}})_{t \geq 0}$  be a quantum Markov semigroup. Under the following detailed balance condition, the limit  $\lim_{t \rightarrow \infty} e^{t\mathcal{L}^{\dagger}} = E_{\mathcal{N}}^{\dagger}$  is a conditional expectation onto the algebra  $\mathcal{N}$  of fixed points of the semigroup:

$$\forall X, Y \in \mathcal{B}_V, \quad \text{Tr}(\sigma X^{\dagger} \mathcal{L}^{\dagger}(Y)) = \text{Tr}(\sigma \mathcal{L}^{\dagger}(X)^{\dagger} Y).$$

Next, for a state  $\rho$ , the relative entropy with respect to  $\mathcal{N}$  is defined as follows

$$S(\rho \| \mathcal{N}) := S(\rho \| E_{\mathcal{N}}(\rho)) = \inf_{E_{\mathcal{N}}(\sigma) = \rho} S(\rho \| \sigma),$$

where the infimum is always attained by  $\sigma = E_{\mathcal{N}}(\rho)$ . Indeed, for any  $\sigma$  satisfying  $E_{\mathcal{N}}(\sigma) = \rho$ , we have the following chain rule (see [42, Lemma 3.4])

$$S(\rho \| \sigma) = S(\rho \| E_{\mathcal{N}}(\rho)) + S(E_{\mathcal{N}}(\rho) \| \sigma). \tag{123}$$

Hence, the infimum is attained if and only if  $S(E_{\mathcal{N}}(\rho) \| \sigma) = 0$ .

**Definition 3.** (*Local indistinguishability*) Let  $\{\mathcal{N}_C\}_{C \subseteq V}$  be a set of subalgebras of  $\mathcal{B}_V$  such that  $\mathcal{B}_{C^c} \subset \mathcal{N}_C$  and  $\mathbf{E} := \{E_C\}_{C \subseteq V}$  be a set of compatible conditional expectations  $E_C^{\dagger} : \mathcal{B}_V \rightarrow \mathcal{N}_C$  acting non-trivially on region  $C$ , i.e., they satisfy the property that for any  $C \subseteq C'$ ,  $E_C \circ E_{C'} = E_{C'} \circ E_C = E_{C'}$ . Then, we say that  $\mathbf{E}$  satisfies local indistinguishability if there exists a fast decaying function  $\varphi : \mathbb{N} \rightarrow \mathbb{R}$  independent of  $V$  such that for every regions  $XYZ \subset V$  with  $\text{dist}(X, Z) \geq \ell$ , and for all states  $\rho \in \mathcal{S}_V$ ,

$$\|E_{YZ} \circ (E_{XYZ} - E_{XY})(\rho)\|_1 \leq |XYZ| \varphi(\ell),$$

For instance, take a product state  $\omega \in \mathcal{S}_V$  and for each region  $C \subseteq V$ , denote  $E_C(\rho) = \text{Tr}_C(\rho) \otimes \omega_C$ . One can easily verify that the maps  $E_C$  are conditional expectations and satisfy the local indistinguishability condition with  $\varphi = 0$ . We are now ready to state and prove the main theorem of this section. For a strictly decreasing function  $\varphi : \mathbb{N} \rightarrow \mathbb{R}_+$  and a positive real number  $a > 0$ , we denote by  $\varphi^{-1}(a) := \min\{\ell \in \mathbb{N} : \varphi(\ell) \leq a\}$ .

**Theorem 6.** *Let  $\mathbf{E}$  be a set of compatible conditional expectations satisfying local indistinguishability with fast decaying function  $\varphi$ . Then, for all hypercubes  $V_0 \subset V$  and all  $\rho \in \mathcal{S}_V$ ,*

$$\|\rho - E_{V_0}(\rho)\|_{W_1} \leq 2\sqrt{20} (7\varphi^{-1}(|V_0|^{-3/2}))^D \sqrt{|V_0| S(\rho \| E_{V_0}(\rho))}, \tag{124}$$

<sup>1</sup>We recall that a finite dimensional von Neumann algebra is a matrix algebra that is close under taking the adjoint.

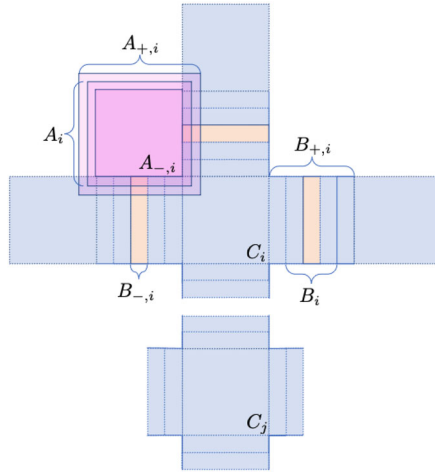


FIGURE 1. Geometry of the lattice in the proof of Theorem 6 with tiling by regions  $A_+ := \cup_i A_{+,i}$ ,  $B_+ := \cup_i B_{+,i}$  and  $C := \cup C_i$

for some fixed constant  $c$  of order 1. In particular, whenever  $E_V(\rho) = \omega_V \in \mathcal{S}_V$  for all states  $\rho$ , and assuming the exponential clustering function  $\varphi(\ell) := \kappa e^{-\ell/\xi}$ , the state  $\omega_V$  satisfies  $\text{TC}(c)$  with  $c = O(n \text{ polylog}(n))$ .

*Proof.* For sake of clarity, we provide the proof for  $D = 2$  only, although the general case follows similarly. First, we partition the hypercube  $V_0$  into regions  $A$ ,  $B_+$  and  $C_+$  in the same way as done in [41] (see also Fig. 1). Then, by triangle inequality

$$\begin{aligned}
 & \|\rho - E_{V_0}(\rho)\|_{W_1} \\
 & \leq \|\rho - E_{A_+} E_{B_- C}(\rho)\|_{W_1} + \|E_{A_+} E_{B_- C}(\rho) - E_{V_0}(\rho)\|_{W_1} \\
 & \leq \|(\text{id} - E_{A_+} E_{B_+} E_C)(\rho)\|_{W_1} + \|E_{A_+}(E_{B_+} E_C - E_{B_- C})(\rho)\|_{W_1} \\
 & \quad + \|(E_{A_+} E_{B_- C} - E_{V_0})(\rho)\|_{W_1} \\
 & \leq (I) + 2|A_+^{\max}| (II) + (III), \tag{125}
 \end{aligned}$$

where

$$\begin{aligned}
 (I) & := \|(\text{id} - E_{A_+} E_{B_+} E_C)(\rho)\|_{W_1} \\
 (II) & := \|(E_{B_+} E_C - E_{B_- C})(\rho)\|_{W_1} \\
 (III) & := \|(E_{A_+} E_{B_- C} - E_{V_0})(\rho)\|_{W_1}
 \end{aligned}$$

and where the last bound in (125) follows from Proposition 5 with  $|A_+^{\max}| := \max_i |A_{+,i}|$ . Now, we control each of the norms on the right-hand side of (125) separately. First, we denote by  $E_{C^{(0)}} := \text{id}$ ,  $C^{(i)} := \cup_{j \leq i} C_j$  given an arbitrary ordering of the connected subregions in  $C$ , and similarly for the other regions



$A_+$  and  $B_+$ . Then,

$$(I) \leq \|(\text{id} - E_C)(\rho)\|_{W_1} + \|(E_C - E_{B_+} E_C)(\rho)\|_{W_1} + \|(E_{B_+} E_C - E_{A_+} E_{B_+} E_C)(\rho)\|_{W_1} \tag{126}$$

$$\begin{aligned} &\leq |C^{\max}| \sum_{i \in \mathcal{I}_C} \|(E_{C^{(i-1)}} - E_{C^{(i)}})(\rho)\|_1 \\ &\quad + |B_+^{\max}| \sum_{i \in \mathcal{I}_{B_+}} \|(E_{B_+^{(i-1)}} - E_{B_+^{(i)}})(E_C(\rho))\|_1 \\ &\quad + |A_+^{\max}| \sum_{i \in \mathcal{I}_{A_+}} \|(E_{A_+^{(i-1)}} - E_{A_+^{(i)}})E_{B_+} E_C(\rho)\|_1 \end{aligned} \tag{127}$$

$$\begin{aligned} &\leq |V_0^{\max}| \left\{ \sqrt{2|\mathcal{I}_C|} \sqrt{\sum_{i \in \mathcal{I}_C} S(E_{C^{(i-1)}}(\rho) \| E_{C^{(i)}}(\rho))} \right. \\ &\quad + \sqrt{2|\mathcal{I}_{B_+}|} \sqrt{\sum_{i \in \mathcal{I}_{B_+}} S(E_{B_+^{(i-1)}}(E_C(\rho)) \| E_{B_+^{(i)}}(E_C(\rho)))} \\ &\quad \left. + \sqrt{2|\mathcal{I}_{A_+}|} \sqrt{\sum_{i \in \mathcal{I}_{A_+}} S(E_{A_+^{(i-1)}} E_{B_+} E_C(\rho) \| E_{A_+^{(i)}} E_{B_+} E_C(\rho))} \right\} \end{aligned} \tag{128}$$

$$\begin{aligned} &\leq |V_0^{\max}| \sqrt{2|\mathcal{I}_{\max}|} \left\{ \sqrt{S(\rho \| E_C(\rho))} + \sqrt{S(E_C(\rho) \| E_{B_+} E_C(\rho))} \right. \\ &\quad \left. + \sqrt{S(E_{B_+} E_C(\rho) \| E_{A_+} E_{B_+} E_C(\rho))} \right\} \end{aligned} \tag{129}$$

with

$$\begin{aligned} |C^{\max}| &:= \max_i |C_i| \\ |V_0^{\max}| &:= \max\{|A_+^{\max}|, |B_+^{\max}|, |C^{\max}|\} \\ |\mathcal{I}_{\max}| &:= \max\{|\mathcal{I}_{A_+}|, |\mathcal{I}_{B_+}|, |\mathcal{I}_C|\} \end{aligned}$$

where  $|\mathcal{I}_C|$  denotes the number of connected components in  $C$ , and similarly for the other sets. Above, (126) follows by the triangle inequality, (127) by the triangle inequality and Proposition 3, and (128) by Pinsker’s inequality as well as Jensen’s inequality for  $x \mapsto x^{\frac{1}{2}}$ . (129) follows from the chain rule in (123), and the fact that the regions  $C_i$ , resp.  $A_{+,i}$ , resp.  $B_{+,i}$ , do not overlap, so that for instance  $E_{A_{+,i}} E_{A_{+,j}} = E_{A_{+,j}} E_{A_{+,i}} = E_{A_{+,j} \cup A_{+,i}}$ . Next, we control the second norm on the right-hand side of (125): using Proposition 3, we have

$$\begin{aligned} \text{(II)} &\leq 2|B_+| \|(E_{B_+} E_C - E_{B_- C})(\rho)\|_1 \\ &= 2|B_+| \|(E_{B_+} E_C - E_{B_- C})(E_C - E_{B_- C})(\rho)\|_1 \end{aligned} \tag{130}$$

$$\begin{aligned} &\leq 2|B_+| \|E_{B_+} E_C - E_{B_- C}\|_{1 \rightarrow 1} \|(E_C - E_{B_- C})(\rho)\|_1 \\ &\leq 2|B_+| \max_{\rho' \in \mathcal{S}_V} \|E_B(E_C - E_{B_- C})(\rho')\|_1 \|E_C(\rho) - E_{B_- C}(\rho)\|_1 \end{aligned} \tag{131}$$

$$\leq 2|B_+| |B_- C| \varphi(\ell) \|E_C(\rho) - E_{B_- C}(\rho)\|_1 \tag{132}$$

$$\leq 2|B_+||B_-C|\varphi(\ell)\sqrt{2S(E_C(\rho)\|E_{B_-C}(\rho))}. \tag{133}$$

Above, (130) follows from the compatibility of the conditional expectations, and (131) from  $E_{B_+} \circ E_B = E_{B_+}$  and the monotonicity of the trace-distance under such CPTP map. (132) follows from the condition of local indistinguishability when taking  $X = C \setminus B$ ,  $Y = B \setminus B_-$  and  $Z = B_-$ , and assuming that  $\text{dist}(X, Z) \geq \ell$ . Finally, (133) follows from an application of Pinsker’s inequality. Similarly, we find

$$\text{(III)} \leq 2|A_+||V_0|\varphi(\ell)\sqrt{2S(E_{B_-C}(\rho)\|E_{V_0}(\rho))} \tag{134}$$

Then, by inserting (129), (133) and (134) into (125), we have

$$\begin{aligned} &\|\rho - E_{V_0}(\rho)\|_{W_1} \\ &\leq 2\sqrt{2}\max\{|V_0^{\max}|\sqrt{|\mathcal{I}_{\max}|}, |A_+||V_0|\varphi(\ell), |B_+||B_-C|\varphi(\ell)\} \\ &\quad \left\{ \sqrt{S(\rho\|E_C(\rho))} + \sqrt{S(E_C(\rho)\|E_{B_+}E_C(\rho))} + \sqrt{S(E_{B_+}E_C(\rho)\|E_{A_+}E_{B_+}E_C(\rho))} \right. \\ &\quad \left. + \sqrt{S(E_{B_-C}(\rho)\|E_{V_0}(\rho))} + \sqrt{S(E_C(\rho)\|E_{B_-C}(\rho))} \right\} \\ &\leq 2\sqrt{10}\max\{|V_0^{\max}|\sqrt{|\mathcal{I}_{\max}|}, |A_+||V_0|\varphi(\ell), |B_+||B_-C|\varphi(\ell)\} \\ &\quad \left( S(\rho\|E_C(\rho)) + S(E_C(\rho)\|E_{B_+}E_C(\rho)) + S(E_{B_+}E_C(\rho)\|E_{A_+}E_{B_+}E_C(\rho)) \right. \\ &\quad \left. + S(E_C(\rho)\|E_{B_-C}(\rho)) + S(E_{B_-C}(\rho)\|E_{V_0}(\rho)) \right)^{\frac{1}{2}} \end{aligned} \tag{135}$$

$$\leq 2\sqrt{20}\max\{|V_0^{\max}|\sqrt{|\mathcal{I}_{\max}|}, |A_+||V_0|\varphi(\ell), |B_+||B_-C|\varphi(\ell)\}S(\rho\|E_{V_0}(\rho))^{\frac{1}{2}} \tag{136}$$

where (135) is another directly application of Jensen’s inequality for  $x \mapsto x^{\frac{1}{2}}$ , whereas (136) follows from two uses of the chain rule (123) after adding the positive term  $S(E_{A_+}E_{B_+}E_C(\rho)\|E_{V_0}(\rho))$  to the square root and a final use of the data processing inequality. The result then follows after choosing the length  $\ell := \varphi^{-1}(|V_0|^{-3/2})$  so that

$$\max\{|A_+||V_0|, |B_+||B_-C|\}\varphi(\ell) \leq |V_0|^2\varphi(\ell) \leq |V_0^{\max}|\sqrt{|\mathcal{I}_{\max}|}.$$

With this choice, and estimating  $|V_0^{\max}| \leq (7\ell)^D$  the bound found in (136) can be further controlled by

$$\|\rho - E_{V_0}(\rho)\|_{W_1} \leq 2\sqrt{20}(7\ell)^D\sqrt{|V_0|S(\rho\|E_{V_0}(\rho))} \tag{137}$$

$$\leq 2\sqrt{20}(7\varphi^{-1}(|V_0|^{-3/2}))^D\sqrt{|V_0|S(\rho\|E_{V_0}(\rho))}. \tag{138}$$

□

### 6.2. Classical Case

In this section, we restrict our analysis to classical conditional expectations and probability measures. In this setting, it is easy to see that the property of local indistinguishability is implied by the following condition. Here, with a slight abuse of notations, we will use the same symbol for a probability measure  $\mu$  on the Borel sets of  $[d]^V$  and its corresponding probability mass function.

**Definition 4** (*Local indistinguishability, classical case.*) Let  $\mu$  be a probability measure on  $[d]^V$ , and  $\{\mu_C\}_{C \subseteq V}$  be a set of compatible conditional probability measures  $\mu_C(\cdot|x_{C^c})$  acting on the sets  $[d]^C$ , i.e., they satisfy the property that for any  $C \subseteq C'$ ,  $\mathbb{E}_{\mu_C} \circ \mathbb{E}_{\mu_{C'}} = \mathbb{E}_{\mu_{C'}} \circ \mathbb{E}_{\mu_C} = \mathbb{E}_{\mu_{C'}}$ . Then, we say that the measure  $\mu$  satisfies local indistinguishability if there exists a fast decaying function  $\varphi : \mathbb{N} \rightarrow \mathbb{R}_+$  such that for every regions  $V' = XYZ \subset V$  such that  $\text{dist}(i, j) \geq \ell$  for any  $i \in X$  and  $j \in Z$ ,

$$\max_{x_X \in [d]^X} \max_{x_{V' \setminus C} \in [d]^{V' \setminus C}} \sum_{y_{V'}} |\mu_{Y|XV' \setminus C}(y_Y|x_X x_{V' \setminus C}) \mu_{XY|ZV' \setminus C}(y_{XY}|y_Z x_{V' \setminus C}) - \mu_{V'|V' \setminus C}(y_{V'}|x_{V' \setminus C})| \leq |V'| \varphi(\ell),$$

where  $\partial Z$  denotes the boundary of  $Z$ .

**Corollary 1.** *Let  $\mu$  be a probability measure on  $[d]^V$  satisfying local indistinguishability with fast decaying function  $\varphi$ . Then, for all  $\nu \ll \mu$ ,*

$$W_1(\nu, \mu) \leq 2\sqrt{20} (\tau \varphi^{-1}(n^{-3/2}))^D \sqrt{n S(\nu|\mu)}. \tag{139}$$

*Equivalently, the measure  $\mu$  satisfies the following sub-Gaussian tail: for any function  $f$  such that  $\|f\|_L \leq 1$ ,*

$$\mathbb{P}_\mu \left( |f(X) - \mathbb{E}_\mu[f(X)]| > t \right) \leq 2 \exp \left( -\frac{t^2}{80n(\tau \varphi^{-1}(n^{-3/2}))^{2D}} \right). \tag{140}$$

### 7. Gaussian Concentration

As mentioned before, the classical transportation cost inequalities for a measure  $\mu$  are equivalent to the sub-Gaussian bounds on the tail probability of any Lipschitz function  $f$  of a random variable  $X$  drawn according to  $\mu$ . One way to see this is by using the variational formulation of the relative entropy in order to bound the Laplace transform of  $f(X)$ . In the non-commutative setting, this leads to the following characterization of the transportation cost constant  $C(\omega)$ :

**Proposition 11.** *For any  $\omega \in \mathcal{S}_V$ ,*

$$C(\omega) = 4 \sup_{K \in \mathcal{O}_V} \frac{\ln \text{Tr} \exp(K + \ln \omega) - \text{Tr}[\omega K]}{\|K\|_L^2}, \tag{141}$$

*and the sup can be restricted to  $K \in \mathcal{O}_V$  such that  $\text{Tr}[\omega K] = 0$ .*

*Proof.* Let  $\tilde{C}(\omega)$  be the right-hand side of (141). On the one hand, let  $K \in \mathcal{O}_V$  satisfy  $\text{Tr}[\omega K] = 0$ , and let

$$\rho = \frac{\exp(K + \ln \omega)}{\text{Tr} \exp(K + \ln \omega)} \in \mathcal{S}_V. \tag{142}$$

We have

$$\begin{aligned} \ln \text{Tr} \exp(K + \ln \omega) &= \text{Tr}[\rho K] - S(\rho|\omega) \leq \|\rho - \omega\|_{W_1} \|K\|_L \\ &\quad - \frac{\|\rho - \omega\|_{W_1}^2}{C(\omega)} \leq \frac{C(\omega) \|K\|_L^2}{4}, \end{aligned} \tag{143}$$

therefore  $\tilde{C}(\omega) \leq C(\omega)$ .

On the other hand, let  $\rho \in \mathcal{S}_V$ , and let  $K \in \mathcal{O}_V$  such that

$$\|K\|_L = \frac{2\|\rho - \omega\|_{W_1}}{\tilde{C}(\omega)}, \quad \text{Tr}[\omega K] = 0, \quad \text{Tr}[\rho K] = \frac{2\|\rho - \omega\|_{W_1}^2}{\tilde{C}(\omega)}. \tag{144}$$

We have

$$S(\rho|\omega) \geq \text{Tr}[\rho K] - \ln \text{Tr} \exp(K + \ln \omega) \geq \frac{\|\rho - \omega\|_{W_1}^2}{\tilde{C}(\omega)}, \tag{145}$$

where the last inequality follows from the definition of  $\tilde{C}(\omega)$ , therefore  $C(\omega) \leq \tilde{C}(\omega)$ , and the claim follows.  $\square$

In the tracial setting [10], and more generally whenever  $[K, \omega] = 0$  the quantity  $\text{Tr} \exp(K + \ln \omega)$  can be interpreted as the Laplace transform of  $K$  in the state  $\omega$ , and therefore the equivalence between Gaussian concentration and the transportation cost inequality holds. However, this is no longer true when  $K$  and  $\omega$  do not commute, and the following bound can turn out to be strictly stronger to the transportation cost inequality as a consequence of the Golden–Thompson inequality: for any  $K \in \mathcal{O}_V$  such that  $\text{Tr}[\omega K] = 0$ ,

$$\ln \text{Tr}[\omega e^K] \leq \frac{C'(\omega)}{4} \|K\|_L^2. \tag{146}$$

In other words,  $C'(\omega) \geq C(\omega)$ . Recently, bounds of the form of (146) were obtained for some subclasses of Lipschitz observables  $K$  (typically local observables) when the  $\omega$  is the Gibbs state of a (possibly non-commuting) quasi-local Hamiltonian  $H$  [43] using cluster expansion techniques. However, the existence of the Gaussian concentration inequality for general Lipschitz observables was left open.

Here instead, we pursue a different approach using our transportation cost inequality. In particular, we prove that (146) can be approximately recovered for Gibbs states of commuting Hamiltonians for a larger class of Lipschitz observables than those considered in [43]. For this, we adapt the result of [13, Theorem 8] which was written for  $W_{1,\nabla}$  to the case of  $W_1$ . In this section, we denote by  $X_R$ , respectively,  $X_I$  the real, respectively, imaginary parts of an operator  $X \in \mathcal{B}_V$ . Given an observable  $O \in \mathcal{O}_V$  with spectral decomposition  $O := \sum_\lambda \lambda P_\lambda$ , a state  $\omega \in \mathcal{S}_V$  and a real number  $r \in \mathbb{R}$ , we denote by

$$\mathbb{P}_\omega(O \geq r) := \sum_{\lambda \geq r} \text{Tr}(\omega P_\lambda) \tag{147}$$

the probability of getting an eigenvalue  $\lambda \geq r$  when measuring  $O$  on the state  $\omega$ .

**Theorem 7.** *Assume that the full-rank state  $\omega \in \mathcal{S}_V$  satisfies TC( $c$ ) for some  $c > 0$ . Then, for any observable  $O \in \mathcal{O}_V$ ,*

$$\mathbb{P}_\omega(\|O - \text{Tr}(\omega O)\| \geq r) \leq 2 \exp\left(-\frac{r^2}{4c \max\{\|(\omega^{-\frac{1}{2}} O \omega^{\frac{1}{2}})_R\|_L^2, \|(\omega^{-\frac{1}{2}} O \omega^{\frac{1}{2}})_I\|_L^2\}}\right). \tag{148}$$

Whenever  $[O, \omega] = 0$  the bound can be tightened into

$$\mathbb{P}_\omega(\|O - \text{Tr}(\omega O)\| \geq r) \leq 2 \exp\left(-\frac{r^2}{c\|O\|_L^2}\right). \tag{149}$$

Therefore, whenever  $\omega \in \mathcal{S}_V$  corresponds to the Gibbs state of a local commuting Hamiltonian on a hypergraph at inverse temperature  $\beta$ , the above bounds hold as long as  $0 < \beta < \beta_c$  where  $\beta_c$  is defined in (85).

*Proof.* Given  $X \in \mathcal{B}_V$ , we denote by  $X := X_R + iX_I$  its decomposition onto real and imaginary parts. We also assume that  $\text{Tr}(\omega X) = 0$  and  $\|X_R\|_L, \|X_I\|_L \leq 1$ . By assumption, we have that for any  $\rho \in \mathcal{S}_V$

$$|\text{Tr}(\rho X)| \leq |\text{Tr}(\rho X_R)| + |\text{Tr}(\rho X_I)| \leq 2\|\rho - \omega\|_{W_1} \leq 2\sqrt{cS(\rho|\omega)}. \tag{150}$$

Then, since  $\inf_{\theta>0} \left(\frac{a}{\theta} + \frac{b\theta}{2}\right) = \sqrt{2ab}$  for any  $a, b \geq 0$ , we have that for all  $\theta > 0$ :

$$|\text{Tr}(\rho X)| \leq \sqrt{2} \frac{S(\rho|\omega)}{\theta} + \frac{c\theta}{\sqrt{2}} \iff \theta|\text{Tr}(\rho X)| - \frac{c}{\sqrt{2}}\theta^2 \leq \sqrt{2}S(\rho|\omega). \tag{151}$$

Next, we further upper bound the relative entropy in terms of the maximal divergence  $\widehat{S}(\rho|\omega) := \text{Tr}[\omega(\omega^{-\frac{1}{2}}\rho\omega^{-\frac{1}{2}}) \ln(\omega^{-\frac{1}{2}}\rho\omega^{-\frac{1}{2}})]$  [44]. Choosing  $\rho = \omega^{\frac{1}{2}}e^{\theta O}\omega^{\frac{1}{2}}/\text{Tr}(\omega e^{\theta O})$  for some observable  $O \in \mathcal{O}_V$ , we arrive at

$$\theta|\text{Tr}(\rho X)| - \frac{c}{\sqrt{2}}\theta^2 \leq \sqrt{2}\theta \frac{\text{Tr}(\omega e^{\theta O})}{\text{Tr}(\omega e^{\theta O})} - \sqrt{2} \ln(\text{Tr}(\omega e^{\theta O})). \tag{152}$$

Next, we choose  $X = \sqrt{2}\omega^{-\frac{1}{2}}O\omega^{\frac{1}{2}}$ , so that the previous inequality reduces to

$$\ln(\text{Tr}(\omega e^{\theta O})) \leq \frac{c}{2}\theta^2. \tag{153}$$

The above inequality can be interpreted as a bound on the log-Laplace transform of the non-commutative variable  $O$  in the state  $\omega$ . By a use of Markov's inequality followed by an optimization over the variable  $\theta > 0$ , we finally get

$$\mathbb{P}_\omega(\|O\| \geq r) \leq 2e^{-\frac{r^2}{2c}}. \tag{154}$$

The result follows after simple rescalings. The tightening in the case of an observable commuting with  $\omega$  can be found by following the same steps as the ones above.  $\square$

In general, there is no way to precisely relate the Lipschitz constants of the real and imaginary parts of  $\omega^{-\frac{1}{2}}O\omega^{\frac{1}{2}}$  to the Lipschitz constant of  $O$  when  $[O, \omega] \neq 0$ . In the next result, we, however, prove that the constants

have similar scalings in the case of a commuting Gibbs measure  $\omega$  of a local Hamiltonian.

**Lemma 7.** *Let  $O = \sum_{A \subset V} \lambda_A O_A \otimes \mathbb{I}_{A^c}$  be the decomposition of an observable  $O$  in  $\mathcal{O}_V$ , where for each subregion  $A$ ,  $O_A$  is exactly supported in  $A$  with  $\|O_A\|_\infty \leq 1$ , and  $\lambda_A \in \mathbb{R}$ . Let further  $\omega$  be the Gibbs state of a geometrically  $k$ -local, commuting Hamiltonian  $H_V := \sum_{B \subset V} h_B \otimes \mathbb{I}_{B^c}$  at inverse temperature  $\beta$ . Then,*

$$\|(\omega^{-\frac{1}{2}} O \omega^{\frac{1}{2}})_R\|_L, \|(\omega^{-\frac{1}{2}} O \omega^{\frac{1}{2}})_I\|_L \leq 4 \max_{i \in V} \sum_{\substack{A \subset V \\ A_{\partial k} \ni i}} |\lambda_A| \exp\left(\beta \sum_{B \cap A \neq \emptyset} \|h_B\|_\infty\right), \tag{155}$$

where  $A_{\partial k} := \{j \in V : \text{dist}(j, A) \leq k\}$  denotes the  $k$ -enlargement of  $A$ . In particular, whenever the state  $\omega$  satisfies TC( $c$ ) with  $c = O(n)$ , any local observable  $O$  gives rise to a sub-Gaussian random variable variance  $O(n)$  when measured in the state  $\omega$ .

*Proof.* We prove the bound for the real part of  $(\omega^{-\frac{1}{2}} O \omega^{\frac{1}{2}})_R$  since the proof for the imaginary part follows the exact same reasoning. First, by Proposition 6, since for any  $A \subset V$ ,  $(\sigma^{-\frac{1}{2}} O_A \sigma^{\frac{1}{2}})_R$  is supported in region  $A_{\partial k}$ , we have that

$$\|(\omega^{-\frac{1}{2}} O \omega^{\frac{1}{2}})_R\|_L \leq 2 \max_{i \in V} \left\| \sum_{\substack{A \subset V \\ A_{\partial k} \ni i}} \lambda_A \left[ (\omega^{-\frac{1}{2}} O_A \omega^{\frac{1}{2}})_R - \text{Tr}_i[(\omega^{-\frac{1}{2}} O_A \omega^{\frac{1}{2}})_R] \otimes \frac{\mathbb{I}_d}{d} \right] \right\|_\infty \tag{156}$$

$$\leq 4 \max_{i \in V} \sum_{\substack{A \subset V \\ A_{\partial k} \ni i}} |\lambda_A| \|\omega^{-\frac{1}{2}} O_A \omega^{\frac{1}{2}}\|_\infty \tag{157}$$

$$\leq 4 \max_{i \in V} \sum_{\substack{A \subset V \\ A_{\partial k} \ni i}} |\lambda_A| \|e^{\beta \sum_{B \cap A \neq \emptyset} h_B} O_A e^{-\beta \sum_{B \cap A \neq \emptyset} h_B}\|_\infty \tag{158}$$

$$\leq 4 \max_{i \in V} \sum_{\substack{A \subset V \\ A_{\partial k} \ni i}} |\lambda_A| e^{\beta \sum_{B \cap A \neq \emptyset} \|h_B\|_\infty}. \tag{159}$$

The result follows. □

### 7.1. Comparison to Previous Tail Bounds

Our main result can be compared to other recently derived concentration bounds for quantum Gibbs states: in [45, Corollary 5.4], the authors consider a product state  $\rho = \bigotimes_{v \in V} \rho_v \in \mathcal{S}_V$  as well as a Hamiltonian  $H = \sum_{A \in E_{k,m}} h_A$ , where the set  $E_{k,m}$  of subsets of  $V$  has the following properties: for any  $A \in E_{k,m}$ ,

- (i)  $|A| \leq k$ ;
- (ii)  $|\{A' \in E_{k,m} : A \cap A' \neq \emptyset\}| \leq m$ .

With these conditions, he was able to prove that

$$\mathbb{P}_\rho(|H - \text{Tr}(\rho H)| \geq r) \leq 2e^{-\frac{r^2}{4eN^3kn}},$$

where number  $N := \max_{v \in V} \sum_{A \in E_{k,m} | v \in A} 1$  is the number of local terms acting non-trivially on spin  $v$ . A similar bound was previously derived by Kuwahara [46, Theorem 7], under a notion of  $g$ -extensivity: a local Hamiltonian  $H$  is said to be  $g$ -extensive if for every spin  $v$ ,  $\sum_{A \in E_{k,m} | v \in A} \|h_A\|_\infty \leq g$ . Under this condition, he shows that

$$\mathbb{P}_\rho(|H - \text{Tr}(\rho H)| \geq r) = \mathcal{O}(1) e^{-\frac{r^2}{cn \log(\frac{r}{\sqrt{n}})}},$$

where  $c$  is a  $\mathcal{O}(1)$  constant which depends only on  $k$  and  $g$ . Although these results recover the Gaussian tails of our Theorem 7 (up to logarithmic overheads), they only work for tensor product states and a subclass of Lipschitz observables. In particular, the tails become trivial whenever the Hamiltonian is a sum of terms acting on non-intersecting regions  $A$  of arbitrary size. In contrast, our bound is still non-trivial for this class of observables, since their Lipschitz constant is still  $\mathcal{O}(1)$ .

More recently, Kuwahara and Saito derived new concentration bounds for Gibbs states of interacting Hamiltonians in order to study the problem of equivalence of quantum statistical ensembles [43, 47] (see Sect. 8): in [47] first, the authors consider a Gibbs state  $\omega$  of a local Hamiltonian on a  $D$ -dimensional regular lattice  $\mathbb{Z}^D$ . They further assume the following  $(r_0, \xi)$  clustering: for any operators  $O_A, O_B$  supported on the subsets  $A$  and  $B$ ,

$$|\text{Tr}(\omega O_A O_B) - \text{Tr}(\omega O_A)\text{Tr}(\omega O_B)| \leq \|O_A\|_\infty \|O_B\|_\infty e^{-\text{dist}(A,B)/\xi}, \quad (160)$$

whenever  $\text{dist}(A, B) \geq r_0$ . Under this condition, they were able to show in Equation (S.17) (see also [45, Theorem 4.2] for a similar bound)

$$\mathbb{P}_\omega(|O - \text{Tr}(\omega O)| \geq r) \leq \min \left\{ 1, (e + 3e\xi) \max \left( e^{-(r^2/(cn))^{\frac{1}{D+1}}}, e^{-\frac{r^2}{c'\ell^D n}} \right) \right\},$$

for some constants  $c, c'$  further depending on  $\xi$  and  $D$ , and where  $\ell$  denotes the locality of the observable  $O$ . Therefore, and although the clustering of correlations is known to hold at high enough temperature [48], the bound is suboptimal for two reasons: firstly, whenever  $r$  is small enough, the exponent has the worse scaling  $r^{2/(D+1)}$ . Secondly, the bounds badly dependence on the locality of  $O$ , and becomes trivial whenever  $O$  is a sum of highly non-local terms. This second limitation also holds for the Gaussian concentration bound found in [43, Corollary 1] for high-temperature Gibbs states of Hamiltonians with long-range interactions. In comparison with the works cited above, our bound always provides better dependence of the tail on the locality of the observable, albeit under the condition that the Hamiltonian is made of local commuting terms.

### 8. Equivalence of Statistical Mechanical Ensembles

The three main ensembles employed in quantum statistical mechanics to compute the equilibrium properties of quantum systems are the canonical ensemble, the microcanonical ensemble and the diagonal ensemble. The quantum state associated with the canonical ensemble is the Gibbs state, which describes

the physics of a system that is at thermal equilibrium with a large bath at a given temperature. The diagonal and microcanonical ensembles both describe the physics of an isolated quantum system, and the associated states are convex combinations of the eigenstates of the Hamiltonian. The microcanonical ensemble assumes a uniform probability distribution for the energy in a given energy shell. The diagonal ensemble includes all the states that are diagonal in the eigenbasis of the Hamiltonian, and in particular it includes the eigenstates themselves.

For many quantum systems, the microcanonical and canonical ensembles give the same expectation values for local observables if the corresponding states have approximately the same average energy. A lot of effort has been devoted to determining conditions under which the two ensembles are equivalent [49–52]. The most prominent among such conditions are short ranged interactions and a finite correlation length, but analytical proofs can be obtained only in the case of regular lattices [52]. The situation is more complex for the diagonal ensemble. The condition under which this ensemble is equivalent to the microcanonical and canonical ensembles is called Eigenstate Thermalization Hypothesis (ETH) [53–57], and states that the expectation values of local observables on the eigenstates of the Hamiltonian are a smooth function of the energy, *i.e.*, for any given local observable, *any* two eigenstates with approximately the same energy yield approximately the same expectation value. The ETH is an extremely strong condition on the Hamiltonian and several quantum systems, including all integrable systems, do not satisfy it. A weak version of the ETH has been formulated [47, 58], stating that for any given local observable, *most* eigenstates in an energy shell yield approximately the same expectation value, or, more precisely, that the fraction of eigenstates yielding expectation values far from the Gibbs state with the same average energy vanishes in the thermodynamical limit. The weak ETH implies the equivalence between the canonical and microcanonical ensembles, but is not sufficient to prove their equivalence with the diagonal ensemble. Under the hypothesis of finite correlation length in the Gibbs state, an analytical proof of the weak ETH is available only for regular lattices [47].

A connection between a transportation cost inequality and the ETH was made by one of the authors in the case of a regular lattice and a nearest neighbour Hamiltonian [38]. Here, we look at the general problem of the equivalence of the statistical mechanical ensembles and of the weak ETH from the perspective of optimal mass transport and show that such equivalence can be formulated as closeness of the respective states in the  $W_1$  distance. The closeness in the  $W_1$  distance implies closeness of the expectation values of all Lipschitz observables, which constitute a significantly larger class than local observables. Therefore, the perspective of optimal mass transport can significantly extend the previous results. Moreover, we will show that the equivalence of the ensembles is intimately linked to the constant of the transportation cost inequality for the Gibbs states.

As in the rest of the paper, we consider a quantum system made by  $n$  qudits located at the vertices of a graph with vertex set  $V$ . Let us assume



that a Gibbs state  $\omega \in \mathcal{S}_V$  satisfies the transportation cost inequality with a constant

$$C(\omega) \leq nC, \tag{161}$$

where  $C$  does not depend on  $n$ . This condition is satisfied under the hypotheses of Theorem 2, Theorem 4 or Theorem 5. We stress that, contrarily to the results of Refs. [47, 52], the condition does not require us to restrict to regular lattices, since Theorem 4 does not need this hypothesis. The following Proposition 12 implies that any state  $\rho \in \mathcal{S}_V$  is close in  $W_1$  distance to the Gibbs state  $\omega$  with the same average energy, provided that  $\rho$  and  $\omega$  have approximately the same entropy, *i.e.*,

$$S(\omega) - S(\rho) \ll n. \tag{162}$$

Moreover, under the same hypothesis, the average reduced states over one qudit of  $\rho$  and  $\omega$  are close in trace distance.

**Proposition 12.** *Let  $\omega \in \mathcal{S}_V$  be a Gibbs state for the Hamiltonian  $H \in \mathcal{O}_V$ . Then, any quantum state  $\rho \in \mathcal{S}_V$  with the same average energy as  $\omega$  satisfies*

$$\frac{1}{n} \|\rho - \omega\|_{W_1} \leq \sqrt{C \frac{S(\omega) - S(\rho)}{n}}. \tag{163}$$

Moreover, let  $\Lambda : \mathcal{O}_V \rightarrow \mathcal{O}(\mathbb{C}^d)$  be the quantum channel that computes the average marginal state over one qudit, *i.e.*, for any  $\rho \in \mathcal{S}_V$ ,

$$\Lambda(\rho) = \frac{1}{n} \sum_{v \in V} \rho_v. \tag{164}$$

Then,

$$\|\Lambda(\rho) - \Lambda(\omega)\|_1 \leq 2\sqrt{C \frac{S(\omega) - S(\rho)}{n}}. \tag{165}$$

*Proof.* We have from the transportation cost inequality

$$\|\rho - \omega\|_{W_1} \leq \sqrt{C(\omega) S(\rho|\omega)} = \sqrt{C(\omega) (S(\omega) - S(\rho))}, \tag{166}$$

where the last equality follows since  $\text{Tr} [\rho \ln \omega] = \text{Tr} [\omega \ln \omega]$ .

We have from Proposition 7

$$\|\Lambda(\rho) - \Lambda(\omega)\|_1 \leq \frac{1}{n} \sum_{v \in V} \|\rho_v - \omega_v\|_1 \leq \frac{2}{n} \|\rho - \omega\|_{W_1} \leq \frac{2}{n} \sqrt{C(\omega) (S(\omega) - S(\rho))}, \tag{167}$$

and the claim follows. □

Choosing  $\rho$  to be diagonal in the eigenbasis of the Hamiltonian, Proposition 12 implies that any convex combination of a sufficiently large number of eigenstates is close in  $W_1$  distance to the Gibbs state with the same average energy. Such number of eigenstates can even be an exponentially small fraction of the total number of eigenstates appearing in a microcanonical state, since the uniform superposition of a fraction  $e^{-n\epsilon}$  of the eigenstates decreases the

entropy by  $\epsilon n$ . Therefore, Proposition 12 constitutes an exponential improvement over the weak ETH.

A natural question is whether also the strong ETH can be captured by the  $W_1$  distance. Unfortunately the answer is negative. Indeed, proving the strong ETH via optimal mass transport would mean to prove that all the eigenstates of the Hamiltonian are close in  $W_1$  distance to the Gibbs states with the corresponding average energy. However, Theorem 1 implies that any state with low entropy, and in particular any pure state, is far from any state with large entropy, and in particular from a Gibbs state with temperature  $\Omega(1)$ . More precisely, for any two states  $\rho, \omega \in \mathcal{S}_V$ ,

$$\|\rho - \omega\|_{W_1} \geq \frac{S(\omega) - S(\rho) - \ln(n + 1) - 1}{\ln(d^2n)}. \tag{168}$$

Equation (168) also implies that any quantum state which is close in  $W_1$  distance to the Gibbs state with the same average energy must have approximately also the same entropy, and in this sense Proposition 12 is optimal.

### 8.1. Comparison with Previous Results

To make our result more easily comparable to the literature, let us introduce more formally the microcanonical ensemble: given the decomposition  $H = \sum_E EP(E)$ , we define the microcanonical ensemble state

$$\omega_{E,\Delta} := \frac{P(E, \Delta)}{\text{Tr}(P(E, \Delta))},$$

where  $P(E, \Delta)$  corresponds to the projection onto the subspace spanned by the eigenvectors whose energy belongs to the interval  $(E - \Delta, E]$ .

**Corollary 2.** *Assume the Gibbs state  $\omega$  satisfies  $C(\omega) \leq Cn$ . Then, for any Lipschitz observable  $O$ ,*

$$\frac{1}{n} |\text{Tr}(\omega O) - \text{Tr}(\omega_{E,\Delta} O)| \leq \|O\|_L o_{n \rightarrow \infty}(1).$$

*Proof.* In view of Proposition 12, it suffices to control the relative entropy between the microcanonical and canonical ensemble states. Then,

$$S(\omega_{E,\Delta} \|\omega) = \ln \frac{\text{Tr}(e^{-\beta H})}{\text{Tr}(P(E, \Delta))} + \beta \text{Tr} \left[ H \frac{P(E, \Delta)}{\text{Tr}(P(E, \Delta))} \right] \leq \beta E + \ln \left[ \frac{\text{Tr}(e^{-\beta H})}{\text{Tr}(P(E, \Delta))} \right]. \tag{169}$$

Next, we control the ratio  $\frac{\text{Tr}(e^{-\beta H})}{\text{Tr}(P(E, \Delta))}$ . For this, we use an argument which was already used in [47, Equation (S.56)]: First, we have found in (149) that

$$\mathbb{P}_\omega(\|H - \text{Tr}(\omega H)\| \geq r) \leq 2e^{-\frac{r^2}{Cn\|H\|_L^2}}.$$

Therefore, choosing the interval  $\tilde{\Delta} := (\text{Tr}(\omega H) - \sqrt{Cn \ln(4)}\|H\|_L, \text{Tr}(\omega H) + \sqrt{Cn \ln(4)}\|H\|_L]$ , we have

$$\text{Tr} \left[ \sum_{E \in \tilde{\Delta}} \frac{e^{-\beta E}}{\text{Tr}(e^{-\beta H})} P(E) \right] \geq \frac{1}{2}$$

Next, we define

$$\tilde{Z} := \text{Tr} \left[ \sum_{E \in \tilde{\Delta}} e^{-\beta E} P(E) \right] \geq \frac{\text{Tr}(e^{-\beta H})}{2} \tag{170}$$

Choosing a slightly extended interval  $\tilde{\Delta}' := (\text{Tr}(\omega H) - \sqrt{Cn \ln(4)} \|H\|_L - \Delta, \text{Tr}(\omega H) + \sqrt{Cn \ln(4)} \|H\|_L + \Delta]$ , we have

$$\tilde{Z} \leq \sum_{\nu \in \mathbb{Z}: \nu \Delta \in \tilde{\Delta}'} \text{Tr}(P(\nu \Delta, \Delta)) e^{-\beta \Delta(\nu-1)} \tag{171}$$

Now, for  $E := \text{argmax} e^{-\beta E} \text{Tr}(P(E, \Delta))$ , we have

$$\text{Tr}(\nu \Delta, \Delta) e^{-\beta \Delta(\nu-1)} \leq e^{\beta \Delta} e^{-\beta E} \text{Tr}(P(E, \Delta)).$$

Replacing in (171), we have that

$$\tilde{Z} \leq e^{\beta \Delta} (2 + 2\sqrt{Cn \ln(4)} \|H\|_L / \Delta) e^{-\beta E} \text{Tr}(P(E, \Delta))$$

Finally, using the lower bound (170), we have that

$$\ln \left[ \frac{\text{Tr}(e^{-\beta H})}{\text{Tr}(P(E, \Delta))} \right] \leq \beta(\Delta - E) + \ln(4 + 4\sqrt{Cn \ln(4)} \|H\|_L / \Delta)$$

Therefore, plugging this last bound into (169), we have found that

$$S(\omega_{E,\Delta} \| \omega) \leq \beta \Delta + \ln(4 + 4\sqrt{Cn \ln(4)} \|H\|_L / \Delta).$$

Therefore,  $S(\omega_{E,\Delta} \| \omega) = o(n)$  whenever  $\Delta = e^{-o(n)}$ , and the result follows. □

In [47, Theorem 2], it is showed that, under the  $(r_0, \xi)$ -clustering of correlations (160), for any observable  $O := \sum_v O_v$  where each  $O_v$  acts on spin  $v$  as well as other spins  $w$  with  $\text{dist}(v, w) \leq \ell$  and has  $\|O_v\|_\infty \leq 1$ ,

$$\frac{1}{n} \left| \text{Tr}(\omega_{E,\Delta} O) - \text{Tr}(\omega O) \right| \leq \frac{1}{\sqrt{n}} \max(c_1 B_1, c_2 B_2),$$

where  $B_1 := \log(\sqrt{n}/\Delta)^{\frac{d+1}{2}}$ ,  $B_2 := (\ell^D \log(\sqrt{n}/\Delta))^{\frac{1}{2}}$ , and the constants  $c_1$  and  $c_2$  depend on  $D, r_0, \xi$  and the locality  $k$  of  $H$ . Therefore, as long as the energy shell  $\Delta$  is chosen as  $\Delta \sim e^{-\mathcal{O}(n^{\frac{1}{D+1}})}$  the averages of the operator density  $\frac{O}{n}$  in the canonical and microcanonical ensemble states converge to the same number as  $n \rightarrow \infty$ . Similar bounds were also derived in [43, Corollary 3] for larger classes of non-local Hamiltonians and observables above some threshold temperature. Corollary 2 constitutes an improvement over these results in two senses: Firstly, it applies to a more general class of Lipschitz observables. Secondly, it allows for a smaller energy shell  $\Delta = e^{-o(n)}$ . However, the condition  $\text{TC}(Cn)$  is currently only known to hold for the smaller class of local commuting Hamiltonians.

## Acknowledgements

The research of CR has been supported by project QTraj (ANR-20-CE40-0024-01) of the French National Research Agency (ANR) and by a Junior Researcher START Fellowship from the MCQST.

**Funding Information** Open Access funding enabled and organized by Projekt DEAL.

**Open Access.** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- [1] Raginsky, M., Sason, I.: Concentration of measure inequalities in information theory, communications and coding. arXiv preprint [arXiv:1212.4663](https://arxiv.org/abs/1212.4663) (2012)
- [2] Boucheron, S., Lugosi, G., Massart, P.: Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford University Press, Oxford (2013)
- [3] Kontorovich, A., Raginsky, M.: Concentration of measure without independence: a unified approach via the martingale method. In: Carlen, E., Madiman, M., Werner, E. (eds.) Convexity and Concentration, pp. 183–210. Springer, New York, NY (2017)
- [4] Marton, K.: A simple proof of the blowing-up lemma (corresp.). IEEE Trans. Inf. Theory **32**(3), 445–446 (1986)
- [5] Bobkov, S.G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. **163**(1), 1–28 (1999)
- [6] Dobrushin, R.L.: Prescribing a system of random variables by conditional distributions. Theory Probab. Appl. **15**(3), 458–486 (1970)
- [7] Marton, K.: Logarithmic Sobolev inequalities in discrete product spaces. Comb. Probab. Comput. **28**(6), 919–935 (2019)
- [8] Dobrushin, R.L., Shlosman, S.B.: Completely analytical interactions: constructive description. J. Stat. Phys. **46**(5), 983–1014 (1987)
- [9] Külske, C.: Concentration inequalities for functions of Gibbs fields with application to diffraction and random Gibbs measures. Commun. Math. Phys. **239**(1), 29–51 (2003)

- [10] Junge, M., Zeng, Q.: Noncommutative martingale deviation and Poincaré type inequalities with applications. *Probab. Theory Relat. Fields* **161**(3–4), 449–507 (2014)
- [11] Tropp, J.A.: An introduction to matrix concentration inequalities. *Found. Trends® Mach. Learn.* **8**(1–2), 1–230 (2015)
- [12] Huang, D., Tropp, J.A.: From Poincaré inequalities to nonlinear matrix concentration. *Bernoulli* **27**(3), 1724–1744 (2021)
- [13] Rouzé, C., Datta, N.: Concentration of quantum states from quantum functional and transportation cost inequalities. *J. Math. Phys.* **60**(1), 012202 (2019)
- [14] De Palma, G., Marvian, M., Trevisan, D., Lloyd, S.: The quantum Wasserstein distance of order 1. *IEEE Trans. Inf. Theory* **67**(10), 6627–6643 (2021)
- [15] De Palma, G., Trevisan, D.: Quantum optimal transport with quantum channels. *Ann. Henri Poincaré* **22**, 3199–3234 (2021)
- [16] Golse, F., Mouhot, C., Paul, T.: On the mean field and classical limits of quantum mechanics. *Commun. Math. Phys.* **343**(1), 165–205 (2016)
- [17] Cole, S., Eckstein, M., Friedland, S., Życzkowski, K.: Quantum optimal transport. arXiv preprint [arXiv:2105.06922](https://arxiv.org/abs/2105.06922) (2021)
- [18] Carlen, E.A., Maas, J.: Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *J. Funct. Anal.* **273**(5), 1810–1869 (2017)
- [19] Bardet, I., Capel, Á., Gao, L., Lucia, A., Pérez-García, D., Rouzé, C.: Entropy decay for Davies semigroups of a one dimensional quantum lattice (2021)
- [20] Bardet, I., Capel, Á., Gao, L., Lucia, A., Pérez-García, D., Rouzé, C.: Rapid thermalization of spin chain commuting Hamiltonians (2021)
- [21] Nielsen, M.A., Chuang, I.: Quantum computation and quantum information (2002)
- [22] Wilde, M.M.: Quantum Information Theory. Cambridge University Press, Cambridge (2013)
- [23] Holevo, A.S.: Quantum Systems, Channels, Information: A Mathematical Introduction, vol. 16. Walter de Gruyter, Berlin (2012)
- [24] Donald, M.J.: On the relative entropy. *Commun. Math. Phys.* **105**(1), 13–34 (1986)
- [25] Petz, D.: Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Commun. Math. Phys.* **105**(1), 123–131 (1986)
- [26] Hiai, F., Petz, D.: The proper formula for relative entropy and its asymptotics in quantum probability. *Commun. Math. Phys.* **143**(1), 99–114 (1991)
- [27] Berta, M., Fawzi, O., Tomamichel, M.: On variational expressions for quantum relative entropies. *Lett. Math. Phys.* **107**(12), 2239–2265 (2017)
- [28] Junge, M., Renner, R., Sutter, D., Wilde, M.M., Winter, A.: Universal recovery maps and approximate sufficiency of quantum relative entropy. *Ann. Henri Poincaré* **19**, 2955–2978 (2018)
- [29] Sutter, D., Berta, M., Tomamichel, M.: Multivariate trace inequalities. *Commun. Math. Phys.* **352**(1), 37–58 (2016)
- [30] Hayden, P., Jozsa, R., Petz, D., Winter, A.: Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Commun. Math. Phys.* **246**(2), 359–374 (2004)

- [31] Tomamichel, M.: Quantum Information Processing with Finite Resources: Mathematical Foundations. SpringerBriefs in Mathematical Physics. Springer International Publishing, Berlin (2015)
- [32] Ollivier, Y.: Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256**(3), 810–864 (2009)
- [33] Eldan, R., Lee, J.R., Lehec, J.: Transport-entropy inequalities and curvature in discrete-space Markov chains. In: Loebl, M., Nešetřil, J., Thomas, R. (eds.) *A Journey Through Discrete Mathematics*, pp. 391–406. Springer, Cham (2017)
- [34] Griffiths, R.B.: Correlations in ising ferromagnets. III. *Commun. Math. Phys.* **6**(2), 121–127 (1967)
- [35] Kastoryano, M.J., Temme, K.: Quantum logarithmic Sobolev inequalities and rapid mixing. *J. Math. Phys.* **54**(5), 052202 (2013)
- [36] Gao, L., Rouzé, C.: Complete entropic inequalities for quantum Markov chains. arXiv preprint [arXiv:2102.04146](https://arxiv.org/abs/2102.04146) (2021)
- [37] Carlen, E.A., Maas, J.: Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems. *J. Stat. Phys.* **178**(2), 319–378 (2019)
- [38] Capel, Á., Rouzé, C., França, D.S.: The modified logarithmic Sobolev inequality for quantum spin systems: classical and commuting nearest neighbour interactions. arXiv preprint [arXiv:2009.11817](https://arxiv.org/abs/2009.11817) (2020)
- [39] Cubitt, T.S., Lucia, A., Michalakis, S., Perez-Garcia, D.: Stability of local quantum dissipative systems. *Commun. Math. Phys.* **337**, 1275–1315 (2015)
- [40] Kastoryano, M.J., Brandao, F.G.: Quantum Gibbs samplers: the commuting case. *Commun. Math. Phys.* **344**(3), 915–957 (2016)
- [41] Brandão, F.G.S.L., Kastoryano, M.J.: Finite correlation length implies efficient preparation of quantum thermal states. *Commun. Math. Phys.* **365**(1), 1–16 (2019)
- [42] Junge, M., LaRacunte, N., Rouzé, C.: Stability of logarithmic Sobolev inequalities under a noncommutative change of measure. arXiv preprint [arXiv:1911.08533](https://arxiv.org/abs/1911.08533) (2019)
- [43] Kuwahara, T., Saito, K.: Gaussian concentration bound and ensemble equivalence in generic quantum many-body systems including long-range interactions. *Ann. Phys.* **421**, 168278 (2020)
- [44] Matsumoto, K.: A new quantum version of f-divergence. In: Ozawa, M., Butterfield, J., Halvorson, H., Rédei, M., Kitajima, Y., Buscemi, F. (eds.) *Nagoya Winter Workshop: Reality and Measurement in Algebraic Quantum Theory*, pp. 229–273. Springer, Singapore (2015)
- [45] Anshu, A.: Concentration bounds for quantum states with finite correlation length on quantum spin lattice systems. *New J. Phys.* **18**(8), 083011 (2016)
- [46] Kuwahara, T.: Connecting the probability distributions of different operators and generalization of the Chernoff–Hoeffding inequality. *J. Stat. Mech.: Theory Exp.* **2016**(11), 113103 (2016)
- [47] Kuwahara, T., Saito, K.: Eigenstate thermalization from the clustering property of correlation. *Phys. Rev. Lett.* **124**(20), 200604 (2020)
- [48] Kliesch, M., Gogolin, C., Kastoryano, M.J., Riera, A., Eisert, J.: Locality of temperature. *Phys. Rev. X* **4**, 031019 (2014)

- [49] Lima, R.: Equivalence of ensembles in quantum lattice systems. *Annales de l'IHP Physique théorique* **15**(1), 61–68 (1971)
- [50] Lima, R., et al.: Equivalence of ensembles in quantum lattice systems: states. *Commun. Math. Phys.* **24**(3), 180–192 (1972)
- [51] Müller, M.P., Adlam, E., Masanes, L., Wiebe, N.: Thermalization and canonical typicality in translation-invariant quantum lattice systems. *Commun. Math. Phys.* **340**(2), 499–561 (2015)
- [52] Brandão, F.G.S.L., Cramer, M.: Equivalence of statistical mechanical ensembles for non-critical quantum systems. [arXiv:1502.03263](https://arxiv.org/abs/1502.03263) [cond-mat, physics:quant-ph], February 2015
- [53] Deutsch, J.M.: Quantum statistical mechanics in a closed system. *Phys. Rev. A* **43**(4), 2046 (1991)
- [54] Srednicki, M.: Chaos and quantum thermalization. *Phys. Rev. E* **50**(2), 888–901 (1994)
- [55] Gogolin, C., Eisert, J.: Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems. *Rep. Prog. Phys.* **79**(5), 056001 (2016)
- [56] Reimann, P.: Dynamical typicality approach to eigenstate thermalization. *Phys. Rev. Lett.* **120**(23), 230601 (2018)
- [57] De Palma, G., Serafini, A., Giovannetti, V., Cramer, M.: Necessity of eigenstate thermalization. *Phys. Rev. Lett.* **115**(22), 220401 (2015)
- [58] Biroli, G., Kollath, C., Läuchli, A.M.: Effect of rare fluctuations on the thermalization of isolated quantum systems. *Phys. Rev. Lett.* **105**(25), 250401 (2010)

Giacomo De Palma  
Scuola Normale Superiore  
56126 Pisa  
Italy  
e-mail: [giacomo.depalma@sns.it](mailto:giacomo.depalma@sns.it)

Cambyse Rouzé  
Technische Universität München  
85748 Garching  
Germany  
e-mail: [cambyse.rouze@tum.de](mailto:cambyse.rouze@tum.de)

Communicated by Claude-Alain Pillet.

Received: July 24, 2021.

Accepted: March 10, 2022.