A Berry-Esseen Theorem for Quantum Lattice Systems and the Equivalence of Statistical Mechanical Ensembles

Fernando G.S.L. Brandão*
Department of Computer Science, University College London

Marcus Cramer†
Department of Physics, University of Ulm, Germany

Madalin Guta‡
Department of Mathematics, University of Nottingham

We prove a version of the Berry-Esseen theorem for quantum lattice systems. Given a local quantum Hamiltonian on N particles and a quantum state with a finite correlation length, it states that measurement of the energy in the state follows a normal distribution, up to error scaling as $O(N^{-1/2}\,\mathrm{polylog}(N))$. This is optimal up to a poly-logarithmic factor.

We then give an application of the theorem to the problem of showing the equivalence of the canonical and microcanonical ensembles for quantum lattice systems. For any model with short ranged interactions and any temperature for which the system has a finite correlation length, we prove that the canonical state of N particles has its local reduced density matrices of $\tilde{O}(\sqrt{N})$ particles equal to the reduced density matrices of the microcanonical state of the same mean energy. This result is established by combining the Berry-Esseen theorem for quantum lattice systems with techniques from quantum information theory.

I. QUANTUM BERRY-ESSEEN THEOREM FOR QUANTUM LATTICE SYSTEMS

In this paper we prove a variant of the Berry-Esseen Theorem in the setting of quantum lattice systems. Before stating the result let us introduce the setting and the notation we consider.

We let $\Lambda = \{1, \ldots, n\}^d$ be a finite collection of *vertices* or *lattice sites* in d dimensions with $N = n^d$ particles [1]. We consider k-local Hamiltonians, acting on the Hilbert space $\mathcal{H} = \otimes_{i \in \mathcal{X}} \mathcal{H}_i$, given by

$$H = \sum_{i \in \Lambda} H_i, \tag{1}$$

where we assume that the H_i are Hermitian, $\|H_i\| \leq 1$, and local in the sense that H_i acts only on sites j with $d(i,j) \leq k$ (for the Manhattan metric $d(.,.)$ in the lattice).

We say a state $\rho \in \mathcal{D}(\mathcal{H})$ (the set of density matrices in $\mathcal{H}$) has correlation length ξ if for every two regions X, Y,

$$\begin{aligned} \mathrm{cor}(X,Y)_\rho \;&:=\; \max_{M \in X, N \in Y} \frac{|\mathrm{tr}((M \otimes N)(\rho_{XY} - \rho_X \otimes \rho_Y))|}{\|M\|\|N\|} \\ &\leq\; 2^{-\mathrm{dist}(X,Y)/\xi}, \end{aligned} \tag{2}$$

*Electronic address: f.brandao@ucl.ac.uk
†Electronic address: marcus.cramer@uni-ulm.de
‡Electronic address: madalin.guta@nottingham.ac.uk

[1] The result of this section can be generalized to more general geometries; see Section IV

where

$$\mathrm{dist}(X,Y) := \min_{x\in X, y\in Y} \mathrm{dist}(x,y). \tag{3}$$

Write the spectral decomposition of H as

$$H = \sum_k E_k |E_k\rangle\langle E_k| \tag{4}$$

with $E_0 \leq E_1 \leq \ldots \leq E_{2^N}$.

Define the function

$$f(\xi,k,s) := \left(1+\frac{\xi}{k}\right)^{2d} k^{d/2}\left(\max\left\{\frac{1}{s},\frac{1}{s^3}\right\} + \mathrm{e}^{-1/\xi}\xi^D \max\left\{\frac{1}{s^2}, s^3\right\}\right) \tag{5}$$

Our main result is the following:

Theorem 1. *Let H be a k-local Hamiltonian in $\Lambda = [n]^d$ with $N = n^d$ particles and ρ a state with correlation length $\xi > 0$. Let*

$$\mu = tr(\rho H), \quad \sigma = tr(\rho(H-\mu)^2)^{1/2}, \quad s = \frac{\sigma}{N^{1/2}k^{d/2}}. \tag{6}$$

Then

$$\sup_y |F(y) - G(y)| \leq C f(\xi,k,s)\frac{\log^{2d}(N)}{\sqrt{N}} \tag{7}$$

where $C > 0$ depends only on the dimension of the lattice,

$$F(y) := \sum_{k:E_k\leq y} \langle E_k|\rho|E_k\rangle, \tag{8}$$

and

$$G(y) := \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{y} \mathrm{d}z\, \mathrm{e}^{-\frac{(z-\mu)^2}{2\sigma^2}}. \tag{9}$$

is the Gaussian cumulative distribution with mean μ and variance σ^2.

Comparison with Previous Work: Quantum central limit theorems, so the fact that $\sup_y |F(y) - G(y)|$ goes to zero as N goes to infinity, have been proven in [4, 5]. To the best of our knowledge, the rate of convergence has, in the present context, not been considered so far. Of course, the classical Berry-Esseen bound for sums of random variables has a long tradition starting in the 1940-ties with [1, 6]; see [7] and references therein for more recent work on the subject.

Proof Idea: We closely follow the classical proof for sums of random variables in Refs [2] and [3]. Our main innovation is to introduce auxiliary operators that will allow us to, despite non-commutativity, arrive at the same scaling. We also give a method for bounding the error when approximating products of matrix exponentials in terms of their Taylor expansion, which may be of independent interest.

Theorem 1, in addition to its own value in quantum statistics, has an interesting application to the equivalence of canonical and microcanonical ensembles for quantum many-body systems, which we now consider.

II. EQUIVALENCE OF STATISTICAL MECHANICS ENSEMBLES

In statistical mechanics there are two main ensembles (at zero chemical potential) that can be used to compute equilibrium properties of large systems: the microcanonical and the canonical ensemble. Roughly, the first describes the physics of a system that is isolated and has total fixed energy. The second describes the physics of a system that is at thermal equilibrium with a large environment at fixed temperature T. It turns out that in many cases the two ensembles give the same predictions for very large systems. There is a long sequence of studies aiming at elucidating under what conditions the two ensembles can be used interchangeably (see e.g. [15–20] and discussion below).

In textbooks the canonical ensemble is commonly introduced by considering the microcanonical ensemble of the system and a large environment and restricting to observables of the system only. Then under the assumption that the interactions of the system and environment are very weak, the canonical ensemble emerges. However in many situations the assumption of weak coupling is not justified. It is therefore an important question to find more general conditions for the equivalence of the two ensembles. In this section we give such condition: we show that short ranged interactions and a finite correlation length already lead to the equivalence of ensembles for every sufficiently large finite volume.

Let $\rho_T := e^{-H/T}/Z(T)$ be the canonical state at temperature T (also known as Gibbs state or thermal state) and $Z(T) := \text{tr}(e^{-H/T})$ the partition function (we set Boltzmann constant to unit). In the canonical ensemble at temperature T, averages are computed using ρ_T.

Define the energy density at temperature T as

$$u(T) := \frac{1}{N}\text{tr}(H\rho_T), \tag{10}$$

the entropy density at temperature T as

$$s(T) := \frac{1}{N}S(\rho_T), \tag{11}$$

and the specific heat capacity at temperature T as

$$c(T) := \left.\frac{du(T')}{dT'}\right|_{T'=T} = \frac{1}{NT^2}\left(\text{tr}(H^2\rho_T) - \text{tr}(H\rho_T)^2\right). \tag{12}$$

Note that if ρ_T has a correlation length ξ, then $c(T) \leq O(\xi)$.

Let

$$M_{e,\delta} := \{k : |E_k - eN| \leq \delta\sqrt{N}\}, \tag{13}$$

and define the microcanonical state of mean energy e and energy spread $\delta\sqrt{N}$ by

$$\tau_{e,\delta} := \frac{1}{|M_{e,\delta}|}\sum_{k\in M_{e,\delta}} |E_k\rangle\langle E_k|. \tag{14}$$

In the microcanonical ensemble averages are computed using $\tau_{e,\delta}$.

We can now state the main result of this section. It shows that for general quantum many-body systems at non-critical temperatures (meaning that the canonical, Gibbs, state has a finite correlation length), the canonical ensemble gives essentially the same predictions as the microcanonical ensemble, for every observable that acts on roughly $\sqrt{N}$ particles, with N the total number of particles of the system.

Given a region R, we denote by $\text{tr}_{\Lambda\setminus R}$ the partial trace over the complement of R in Λ.

Theorem 2. *Let $\mathcal{C}_l$ be the set of all hypercubes contained in $\Lambda := \{1, \ldots, n\}^d$ of volume l^d. Let H be a k-local Hamiltonian on Λ. Suppose that T is such that the Gibbs state ρ_T has correlation length ξ. Let $N := n^d$ and*

$$E := \left\{ e : |eN - tr(H\rho_T)| \leq \frac{1}{4} Tc(T)^{\frac{1}{2}} \sqrt{N} \right\}. \tag{15}$$

Set $\nu := 4\pi f(\xi, k, c(T)^{\frac{1}{2}} k^{-d/2}) Tc(T)^{\frac{1}{2}}$. Let δ be any number such that $\frac{\nu \log^{2d}(N)}{\sqrt{N}} \leq \delta \leq \frac{1}{8} Tc(T)^{\frac{1}{2}}$. Then for every triple (ε, l, e) such that $\varepsilon > 0$,

$$l \leq \left(\frac{T\varepsilon^2}{12\nu} \frac{N}{\log^{2d}(N)} \right)^{\frac{1}{2d}}, \tag{16}$$

and $e \in E$:

$$\mathop{\mathbb{E}}_{C \in \mathcal{C}_l} \left\| tr_{\Lambda \setminus C}(\tau_{e,\delta}) - tr_{\Lambda \setminus C}(\rho_T) \right\|_1 \leq \varepsilon, \tag{17}$$

where the expectation is taken uniformly over $\mathcal{C}_l$.

We note that the condition of a finite correlation length is necessary. Indeed as shown in [8], the two ensembles differ in the Ising model close to criticality, when the correlation length diverges, for regions of size $O(\log(N))$. It is an open question if a similar result can be obtained for critical systems and small enough regions, assuming that correlations decay algebraically; in our proof it is important that the correlations decay exponentially.

Any system is expected to have a finite correlation length whenever it is away from a critical point. One dimensional systems always have a finite correlation length at any temperature [9], while in any dimensions there is a critical temperature (depending only on the geometry of the lattice) above which every system has a finite correlation length [10] [2].

We also note we do not need to take the average over regions $R \in \mathcal{R}_l$ if we assume the Hamiltonian is translation invariant, as in this case σ_R is the same for all $R \in \mathcal{R}_l$.

Beyond Microcanonical States: How crucial is the use of the microcanonical ensemble? It turns out that Theorem 2 can be generalized to a much larger class of states. In a nutshell all that is required is that the state is concentrated around a given energy and has sufficiently large entropy. To state the condition precisely, define the max-relative entropy of two states π and σ [12]:

$$S_{\max}(\pi || \sigma) := \{\min \lambda : \rho \leq 2^{\lambda} \sigma\}, \tag{18}$$

and its smooth version

$$S^{\varepsilon}_{\max}(\pi || \sigma) := \min_{\tilde{\pi} \in B_{\varepsilon}(\pi)} D_{\max}(\tilde{\pi} || \sigma), \tag{19}$$

with $B_{\varepsilon}(\pi) := \{\tilde{\pi} : \|\pi - \tilde{\pi}\|_1 \leq \varepsilon\}$.

Then we have

[2] Ref. [10] proves a finite correlation length in the high temperature phase for a different correlation function than the one used in Eq. (2). But using [11] one can show that the two notions are equivalent in the high temperature regime.

Corollary 3. *Let $\mathcal{C}_l$ be the set of all hypercubes contained in $\Lambda := \{1, \ldots, n\}^d$ of volume l^d. Let H be a k-local Hamiltonian on Λ. Suppose that T is such that the Gibbs state ρ_T has correlation length ξ. Let $N := n^d$ and*

$$E := \left\{ e : |eN - tr(H\rho_T)| \leq \frac{1}{4} Tc(T)^{\frac{1}{2}} \sqrt{N} \right\}. \tag{20}$$

Set $\nu := 4\pi f(\xi, k, c(T)^{\frac{1}{2}} k^{-d/2}) Tc(T)^{\frac{1}{2}}$. Let δ be any number such that $\frac{\nu \log^{2d}(N)}{\sqrt{N}} \leq \delta \leq \frac{1}{8} Tc(T)^{\frac{1}{2}}$. Consider a triple (ε, l, e) such that $1/4 \geq \varepsilon > 0$,

$$l \leq \left(\frac{T\varepsilon^2}{12\nu} \frac{N}{\log^{2d}(N)} \right)^{\frac{1}{2d}}, \tag{21}$$

and $e \in E$. Then

1. *For every π such that $S^{\varepsilon}_{\max}(\pi || \tau_{e,\delta}) \leq n^{\frac{1}{2d}}$,*

$$\mathop{\mathbb{E}}_{C \in \mathcal{C}_l} \left\| tr_{\Lambda \setminus C}(\pi) - tr_{\Lambda \setminus C}(\rho_T) \right\|_1 \leq 2\varepsilon, \tag{22}$$

2. *For states $|\psi\rangle$ drawn from the Haar measure in span$\{|E_k\rangle : k \in M_{e,\delta}\}$, with probability $1 - 2^{-O(\varepsilon^{-2}|M_{e,\delta}|)}$,*

$$\mathop{\mathbb{E}}_{C \in \mathcal{C}_l} \left\| tr_{\Lambda \setminus C}(|\psi\rangle\langle\psi|) - tr_{\Lambda \setminus C}(\rho_T) \right\|_1 \leq 2\varepsilon, \tag{23}$$

The second part of the corollary is a direct consequence of Theorem 2 and the result of [13] that a generic state in a energy subspace is equal to the microcanonical state in small regions.

Comparison with Previous Work: The problem of equivalence of ensembles was considered already in the foundational work of Boltzmann and Gibbs. See [14] for a historical perspective. An intuitive explanation for the equivalence in non-critical temperatures is the following: Whenever there is a finite correlation length, the specific heat capacity is a constant. Then as the energy is extensive, we find that it is concentrated around its mean value. However this argument is too simplistic. Indeed it is easy to see that for large enough n, $\tau_{e,\delta}$ and ρ_T (with $e = u(T)$) are almost orthogonal. Therefore any meaningful argument for equivalence of ensembles must go beyond the distribution of energies and somehow restrict the kind of observables considered (e.g. observables acting in small regions).

The most fruitful direction so far has been to consider systems in the thermodynamical limit. In this regime one can prove the equivalence of ensembles on the level of thermodynamical functions [15, 17–19] (showing that the thermodynamical limits of the entropy density of the microcanonical ensemble is the Legendre transform of the limit of the free energy density) and also on the level of states, as we do here, both for classical [16] and, only very recently, also for quantum systems [20]. However the price of considering the thermodynamical limit – instead of the physically relevant regime of very large but finite sizes – is that no finite bounds can be obtained on the size of the regions on which the states are close.

In this respect Theorem 2 goes beyond the earlier work in several aspects:

- It covers the general case of non translation-invariant models.

- It is based on the assumption of a finite correlation length, which is simpler than the assumption of a unique phase region employed in [18–20].
- It gives explicit finite size bounds; for quite big regions of order $\tilde{O}(\sqrt{N})$ the two ensembles already look the same.
- It shows that the equivalence holds true even for microcanonical states with very small energy spread, of order $O(\log^d(N))$ and substantially smaller than the value $O(N^{1/2})$ that could have been expected.
- It shows that that any two microcanonical states $\tau_{e,\delta}$ and $\tau_{e',\delta'}$ are locally equivalent whenever $|e-e'| \leq O(\sqrt{N})$ and $O(\log^d(N)) \leq \delta+\delta' \leq \tilde{O}(\sqrt{N})$ (when $\rho_{T(e)}$ has a finite correlation length).
- It covers more general states than the microcanonical, showing that the important condition is that the state is concentrated around a fixed energy and has sufficiently large entropy.

It is an interesting open question how small δ can be taken. We note that the eigenstate thermalization hypothesis (ETH) states that even for $\delta = 0$, i.e. for a single eigenstate, one should have thermal expectation values for regions small enough [21]. However, while believed to hold true for several systems, there are known counterexamples to ETH – for instance systems with many-body localization.

As we outline next, the proof of Theorem 2 is based on our quantum Berry-Essen theorem and ideas of quantum information theory. Thus the result gives a new application of quantum information to statistical mechanics, complementing recent studies in this direction (e.g. [13, 22, 23]).

Proof Idea: The proof of Theorem 2 is very different from the approach followed in [18–20], being based on a combination of Theorem 1 and arguments from quantum information theory. We note that the close-to-optimal error of $\tilde{O}(N^{-\frac{1}{2}})$ in Eq. (7) is crucial; if it were instead $\Omega(N^{-1/2+\nu})$ for any $\nu > 0$ our proof would fail.

A quick summary of the argument is the following: First using Theorem 1 we show in Lemma 4 that

$$S_{\max}(\tau_{e,\delta}||\rho_T) \leq O(\log^d(N)). \tag{24}$$

We then consider a partition of the lattice $[n]^d$ into regions $A_1, ..., A_m$ and R, where each A_i is a d-dimensional cube of size l^d separated from each other by a distance of $2\xi l^d$ and R is composed of the remaining sites of the lattice (see Fig 1). Denote $\rho_{A_1...A_m} := \text{tr}_R(\rho_T)$ and $\tau_{A_1...A_m} := \text{tr}_R(\tau_{e,\delta})$. Then by the data processing inequality for $S_{\max}$ and Eq. (24):

$$S_{\max}(\tau_{A_1...A_m}||\rho_{A_1...A_m}) \leq O(\log^d(N)). \tag{25}$$

From the fact that ρ_T has a finite correlation length we show in Lemma 6 that $\|\rho_{A_1...A_m} - \rho_{A_1} \otimes \ldots \otimes \rho_{A_m}\|_1 \leq m2^{-l^d}$. Then in Lemma 5 we show that this bound together with Eq. (25) implies that

$$S^{\varepsilon}_{\max}(\tau_{A_1...A_m}||\rho_{A_1} \otimes \ldots \otimes \rho_{A_m}) \leq O(\log^d(N)). \tag{26}$$

for $\varepsilon = O(e^{\nu \log^2(d)/T})m^{1/2}2^{-l^d/2}$.

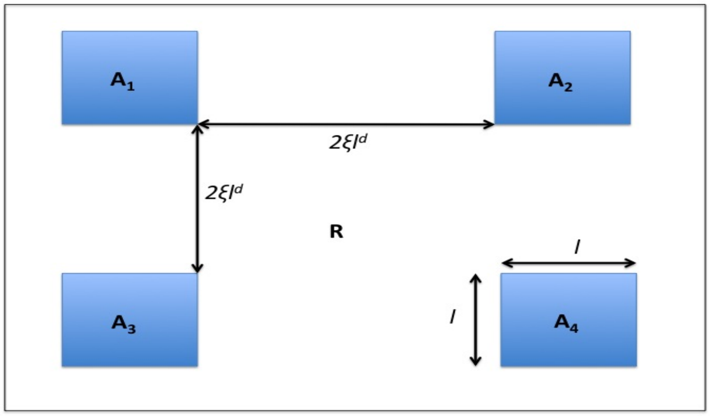

FIG. 1: Depiction of regions $A_1, \ldots, A_m$, each of linear size l and distance $2\xi l^d$ from each other, and their complement R.

The final part of the proof is to argue that the equation above implies that for most i, τ_{A_i} is close to ρ_{A_i}. We do so by making use of basic properties of the quantum entropy and relative entropy, in particular the subadditivity of entropy and Pinsker's inequality. Indeed with

$$S(\rho||\sigma) := \mathrm{tr}(\rho(\log\rho - \log\sigma)) \tag{27}$$

the relative entropy and

$$S^{\varepsilon}(\rho||\sigma) := \min_{\tilde{\rho}\in B_{\varepsilon}(\rho)} S(\tilde{\rho}||\sigma), \tag{28}$$

we find that Eq. (26) also holds for $S^{\varepsilon}(\tau_{A_1\ldots A_m}||\rho_{A_1}\otimes\ldots\otimes\rho_{A_m})$. Then by subadditivity of the von Neumann entropy we can show that

$$\mathop{\mathbb{E}}_{i} S^{\varepsilon}(\tau_{A_i}||\rho_{A_i}) \leq O\left(\frac{\log^d(N)}{m}\right). \tag{29}$$

The result then follows from Pinsker's inequality, which gives

$$2\ln(2)\sqrt{S^{\varepsilon}(\tau_{A_i}||\rho_{A_i})} \geq \|\tau_{A_i} - \rho_{A_i}\|_1 - \varepsilon. \tag{30}$$

III. EQUIVALENCE OF STATISTICAL ENSEMBLES: PROOF OF THEOREM 2

We begin with the following lemma:

Lemma 4. *Let H be a k-local Hamiltonian in $\Lambda = [n]^d$, with $N = n^d$ particles. Let $T > 0$ be such that ρ_T has a finite correlation length. Let $\nu := 4\pi f(\xi, k, c(T)^{1/2}k^{-d/2})c(T)^{1/2}T$, $\delta = \frac{\nu\log^{2d}(N)}{N^{1/2}}$, and e be such that $|e - u(T)| \leq \frac{c(T)^{\frac{1}{2}}T}{4\sqrt{N}}$. Then*

$$S_{\max}(\tau_{e,\delta}||\rho_T) \leq \frac{2\nu}{T}\log^d(N) + \frac{1}{2}\log(N). \tag{31}$$

Proof. Define

$$Z(T,e,\delta) := \sum_{k\in M_{e,\delta}} e^{-E_k/T}. \tag{32}$$

Using Theorem 1,

$$\begin{aligned} \frac{Z(T,e,\delta)}{Z(T)} &= \text{tr}\left(\rho_T \sum_{k\in M_{e,\delta}} |E_k\rangle\langle E_k|\right) \\ &= F(eN+\delta\sqrt{N}) - F(eN-\delta\sqrt{N}) \\ &\geq G(eN+\delta\sqrt{N}) - G(eN-\delta\sqrt{N}) - 2f(\xi,k,s)\frac{\log^{2d}(N)}{\sqrt{N}} \end{aligned} \tag{33}$$

with $F(y)$ given by Eq. (8), with $\rho = \rho_T$, and $f(\xi,k,s)$ given by Eq. (5). Since $\sigma^2 = \text{tr}(H^2\rho_T) - \text{tr}(H\rho_T)^2 = c(T)T^2N$, we get $s = c(T)^{1/2}Tk^{-d/2}$.

We have

$$\begin{aligned} G(eN+\delta\sqrt{N}) - G(eN-\delta\sqrt{N}) &= \text{erf}\left(\frac{eN+\delta\sqrt{N}-u(T)N}{\sigma}\right) - \text{erf}\left(\frac{eN-\delta\sqrt{N}-u(T)N}{\sigma}\right) \\ &= \text{erf}\left(\frac{1}{4} + \frac{\delta}{C(T)^{\frac{1}{2}}T}\right) - \text{erf}\left(\frac{1}{4} - \frac{\delta}{C(T)^{\frac{1}{2}}T}\right) \\ &=: M. \end{aligned} \tag{34}$$

where

$$\text{erf}(x) := \frac{1}{\sqrt{\pi}}\int_0^x e^{-t^2/2}dt. \tag{35}$$

Using the Maclaurin series $\text{erf}(z) = 2\pi^{-1/2}\left(z - \frac{1}{3}z^3 + \frac{1}{10}z^5 - \frac{1}{42}z^7 + \ldots\right)$,

$$M \geq 3f(\xi,k,s)\frac{\log^{2d}(N)}{\sqrt{N}}, \tag{36}$$

and from Eq. (33)

$$\frac{Z(T,e,\delta)}{Z(T)} \geq 2f(\xi,k,s)\frac{\log^{2d}(N)}{\sqrt{N}}. \tag{37}$$

Since for every $k \in M_{en,\delta}$,

$$\frac{1}{|M_{e,\delta}|} \leq e^{2\nu\log^d(N)/T}\frac{e^{-E_k/T}}{Z(T,e,\delta)}. \tag{38}$$

we find

$$\begin{aligned} \tau_{e,\delta} &\leq e^{2\nu\log^d(N)/T}\frac{1}{Z(T,e,\delta)}\sum_{k\in M_{e,\delta}} e^{-E_k/T}|E_k\rangle\langle E_k| \\ &\leq N^{1/2}e^{2\nu\log^d(N)/T}\frac{1}{Z(T)}\sum_{k\in M_{e,\delta}} e^{-E_k/T}|E_k\rangle\langle E_k| \\ &\leq \sqrt{N}e^{2\nu\log^d(N)/T}\rho_T, \end{aligned} \tag{39}$$

from which Eq. (31) follows. □

Lemma 5. *Let H be a local Hamiltonian in $\Lambda = [n]^d$, with $N = n^d$ particles. Let $T > 0$ be such that ρ_T has correlation length ξ. Let $\nu > 0$, $\delta = \nu \frac{\log^d(N)}{N^{1/2}}$, and e be such that $|e - u(T)| \leq \frac{c(T)^{1/2} T}{4N^{1/2}}$. Consider a partition of the lattice $[n]^d$ into regions $A_1, ..., A_m$ and R, where each A_i is a d-dimensional cube of size l^d separated from each other by a distance of $2\xi l^d$ and R is composed of the remaining sites of the lattice (see Fig 1). Let $\rho_{A_1 \ldots A_m} := tr_R(\rho_T)$ and $\tau_{A_1 \ldots A_m} := tr_R(\tau_{e,\delta})$. Then*

$$S^{\kappa}_{\max}(\tau_{A_1 \ldots A_m} || \rho_{A_1} \otimes \ldots \otimes \rho_{A_m}) \leq \frac{2\nu}{T(1-\kappa)} \log^d(N) + \frac{1}{2(1-\kappa)} \log(N) \tag{40}$$

with

$$\kappa := 8 N^{1/4} e^{\nu \log^d(N)/T} m^{1/2} 2^{-l^d} 2^{l^d/2}. \tag{41}$$

Proof. By Lemma 4 and the monotonicity of $S_{\max}$ under partial trace,

$$S_{\max}(\tau_{A_1 \ldots A_m} || \rho_{A_1} \otimes \ldots \otimes \rho_{A_m}) \leq \frac{2\nu}{T} \log^d(N) + \frac{1}{2} \log(N). \tag{42}$$

By Lemma 6,

$$\left\| \mathrm{tr}_R(\rho_T) - \mathrm{tr}_{\Lambda \backslash A_1}(\rho_T) \otimes \ldots \otimes \mathrm{tr}_{\Lambda \backslash A_m}(\rho_T) \right\|_1 \leq m 2^{-2l^d} 2^{l^d}. \tag{43}$$

The statement then follow from Lemma 8.

□

We are now in position to prove Theorem 2. The idea is to combine Lemma 5 with basic properties of the quantum relative entropy. Given two quantum states ρ and σ, the relative entropy is defined as

$$S(\rho || \sigma) := \mathrm{tr}(\rho(\log(\rho) - \log(\sigma))). \tag{44}$$

It satisfies the following properties

- (Pinsker's inequality)

$$S(\rho || \sigma) \geq \frac{1}{2 \ln(2)} \|\rho - \sigma\|_1^2. \tag{45}$$

- (Relation with $S_{\max}$ [12])

$$S(\rho || \sigma) \leq S_{\max}(\rho || \sigma). \tag{46}$$

- (super-additivity)

$$S(\rho_{A_1 \ldots A_m} || \sigma_{A_1} \otimes \ldots \otimes \sigma_{A_m}) \geq \sum_{k=1}^{m} S(\rho_{A_i} || \sigma_{A_i}) \tag{47}$$

The third property is a easy consequence of subadditivity of entropy. Indeed:

$$\begin{aligned} S(\rho_{A_1 \ldots A_m} || \sigma_{A_1} \otimes \ldots \otimes \sigma_{A_m}) &= -S(\rho_{A_1 \ldots A_m}) - \mathrm{tr}(\rho_{A_1 \otimes \ldots \otimes A_m} \log(\sigma_{A_1} \otimes \ldots \otimes \sigma_{A_m})) \\ &\geq -\sum_{k=1}^{m} S(\rho_{A_i}) - \mathrm{tr}(\rho_{A_1 \ldots A_m} \log(\sigma_{A_1} \otimes \ldots \otimes \sigma_{A_m})) \\ &= \sum_{k=1}^{m} S(\rho_{A_i} || \sigma_{A_i}). \end{aligned} \tag{48}$$

Theorem 2 (restatement). *Let $\mathcal{C}_l$ be the set of all hypercubes contained in $\Lambda := \{1,\ldots,n\}^d$ of volume l^d. Let H be a k-local Hamiltonian on Λ. Suppose that T is such that the Gibbs state ρ_T has correlation length ξ. Let $N := n^d$ and*

$$E := \left\{ e : |eN - tr(H\rho_T)| \leq \frac{1}{4} Tc(T)^{\frac{1}{2}} \sqrt{N} \right\}. \tag{49}$$

Set $\nu := 4\pi f(\xi, k, c(T)^{\frac{1}{2}} k^{-d/2}) Tc(T)^{\frac{1}{2}}$. Let δ be any number such that $\frac{\nu \log^{2d}(N)}{\sqrt{N}} \leq \delta \leq \frac{1}{8} Tc(T)^{\frac{1}{2}}$. Then for every triple (ε, l, e) such that $\varepsilon > 0$,

$$l \leq \left(\frac{T\varepsilon^2}{12\nu} \frac{N}{\log^{2d}(N)} \right)^{\frac{1}{2d}}, \tag{50}$$

and $e \in E$:

$$\mathop{\mathbb{E}}_{C \in \mathcal{C}_l} \left\| tr_{\Lambda \setminus C}(\tau_{e,\delta}) - tr_{\Lambda \setminus C}(\rho_T) \right\|_1 \leq \varepsilon, \tag{51}$$

where the expectation is taken uniformly over $\mathcal{C}_l$.

Proof. We prove the statement of the theorem with $\delta = \frac{\nu \log^{2d}(N)}{N^{1/2}}$. The case of larger δ will then follow by a simple averaging argument.

As in Lemma 5, consider a partition of the lattice $[n]^d$ into regions $A_1, \ldots, A_m$ and R, where each A_i is a d-dimensional cube of size l^d separated from each other by a distance of $2\xi l^d$ and R is composed of the remaining sites of the lattice (see Fig 1).

By Lemma 5 and Eq. (46) there is a $\pi_{A_1 \ldots A_l}$ s.t. $\|\pi_{A_1\ldots A_l} - \tau_{A_1 \ldots A_l}\|_1 \leq \kappa$

$$S(\pi_{A_1 \ldots A_l} || \rho_{A_1} \otimes \ldots \otimes \rho_{A_l}) \leq \frac{2\nu}{T(1-\kappa)} \log^d(N) + \frac{1}{2(1-\nu)} \log(N) \tag{52}$$

By Eq. (47)

$$\sum_{k=1}^{m} S(\pi_{A_k} || \rho_{A_k}) \leq S(\pi_{A_1 \ldots A_l} || \rho_{A_1} \otimes \ldots \otimes \rho_{A_l}) \leq 3\nu \log^d(N)/T. \tag{53}$$

Noting that $m = N/(l^d + (2\xi)^d l^{d^2})$,

$$\mathop{\mathbb{E}}_{k} S(\pi_{A_k} || \rho_{A_k}) \leq \frac{3\nu}{T} \log^d(N) \frac{l^{d^2}}{N}. \tag{54}$$

Using Pinsker's inequality (Eq. 45) and the convexity of $x \mapsto x^2$,

$$\mathop{\mathbb{E}}_{k} \|\pi_{A_k} - \rho_{A_k}\|_1 \leq \sqrt{\frac{3\nu}{T} \log^d(N) \frac{l^{d^2}}{N}}. \tag{55}$$

Finally, by Eq. (40),

$$\mathop{\mathbb{E}}_{k} \|\tau_{A_k} - \rho_{A_k}\|_1 \leq \sqrt{\frac{3\nu}{T} \log^d(N) \frac{l^{d^2}}{N}} + 2N^{1/2} e^{2c \log^d(N)/T} m 2^{-l^d}. \tag{56}$$

Choosing l as in Eq. (50) we find that the L.H.S. of the equation above is smaller than ε. □

The following auxiliary lemmas were used in the proofs above.

Lemma 6. *Let $\pi_{A_1,\ldots,A_m} \in \mathcal{D}((\mathbb{C}^D)^{\otimes m})$ be such that for every $j \in [l]$,*

$$cor(A_1, \ldots, A_{j-1} : A_j) \leq \varepsilon. \tag{57}$$

Then

$$\|\pi_{A_1,\ldots,A_m} - \pi_{A_1} \otimes \ldots \otimes \pi_{A_m}\|_1 \leq mD^2\varepsilon. \tag{58}$$

Proof. By Lemma 7, for every $j \in [l]$,

$$\left\|\pi_{A_1,\ldots,A_{j-1},A_j} - \pi_{A_1,\ldots,A_{j-1}} \otimes \pi_{A_j}\right\|_1 \leq D^2 \text{cor}(A_1, \ldots, A_{j-1} : A_j) \leq D^2\varepsilon. \tag{59}$$

Then by a telescoping sum and triangle inequality,

$$\begin{aligned} \|\pi_{A_1,\ldots,A_m} - \pi_{A_1} \otimes \ldots \otimes \pi_{A_m}\|_1 &= \left\|\sum_{j=1}^{m} (L_j - L_{j-1})\right\| \\ &\leq \sum_{j=1}^{m} \|\pi_{A_1,\ldots,A_j} - \pi_{A_1,\ldots,A_{j-1}} \otimes \pi_{A_j}\|_1 \leq mD^2\varepsilon, \end{aligned} \tag{60}$$

with

$$L_j := \pi_{A_1,\ldots,A_j} \otimes \pi_{A_{j+1}} \otimes \ldots \otimes \pi_{A_m}. \tag{61}$$

□

Lemma 7. *[Lemma 20 of [24]] For every $L \in \mathcal{B}(\mathbb{C}^r \otimes \mathbb{C}^R)$ with $r \leq R$,*

$$\|L\|_1 \leq r^2 \max_{\|X\|,\|Y\| \leq 1} |tr((X \otimes Y)L)|. \tag{62}$$

The second Lemma was first proven by Datta and Renner in [25], in a different formulation, and appeared in a form equivalent to the one bellow as Lemma C.5 of [26].

Lemma 8. *Let $\rho, \sigma \in \mathcal{D}(\mathbb{H})$ be such that*

$$D_{\max}(\rho||\sigma) \leq \lambda. \tag{63}$$

Let $\tilde{\sigma} \in B_\varepsilon(\sigma)$. Then

$$D_{\max}^{2^\lambda 4\sqrt{\varepsilon}}(\rho||\tilde{\sigma}) \leq \frac{\lambda}{1 - 42^\lambda\sqrt{\varepsilon}} \tag{64}$$

Proof. The statement follows from Lemma C.5 of [26] with $Y = 2^\lambda\sigma$ and $\Delta := 2^\lambda|\sigma - \tilde{\sigma}|$. □

IV. QUANTUM BERRY-ESSEEN: PROOF OF THEOREM 1

To facilitate the exposition we will use a slightly different notation in this section from the rest of the paper.

We let $\mathcal{X}$ a finite collection of *vertices* or *lattice sites* equipped with a metric $d : \mathcal{X} \times \mathcal{X} \to \mathbb{N}$ and consider operators acting on the Hilbert space $\mathcal{H} = \otimes_{i\in\mathcal{X}}\mathcal{H}_i$ of the form

$$\hat{\mathcal{X}} = \sum_{i\in\mathcal{X}} \hat{X}_i, \tag{65}$$

where we assume that the $\hat{X}_i$ are hermitian, bounded, $\|\hat{X}_i\| \leq x/2$, and local in the sense that there is a $R \in \mathbb{N}$ such that $\hat{X}_i$ acts only on sites j with $d(i,j) \leq R$.

We will need the notion of dimension of $\mathcal{X}$: We call D the dimension of $\mathcal{X}$ if it is the smallest $D > 0$ such that there is a constant $c_D > 0$ such that for every $l > 0$

$$\left|\left\{i \in \mathcal{X} \,\middle|\, d(i,j) = l\right\}\right| = \sum_{\substack{i\in\mathcal{X}\\ d(i,j)=l}} 1 \leq c_D l^{D-1}. \tag{66}$$

We will assume that $D \geq 1$. For subsets $\mathcal{A}, \mathcal{B} \subset \mathcal{X}$, we denote the distance between them as

$$d(\mathcal{A},\mathcal{B}) = \min_{\substack{i\in\mathcal{A}\\ j\in\mathcal{B}}} d(i,j). \tag{67}$$

We let $\hat{\varrho}$ a state acting on $\mathcal{H}$, write

$$\langle \cdot \rangle = \mathrm{tr}[\hat{\varrho}\,\cdot\,] \tag{68}$$

and denote

$$\mu = \langle \hat{\mathcal{X}} \rangle, \quad \sigma = \left\langle (\hat{\mathcal{X}} - \mu)^2 \right\rangle^{1/2}, \quad s = \frac{\sigma}{x|\mathcal{X}|^{1/2} R^{D/2}}, \tag{69}$$

assuming that $s > 0$. We assume that the state $\hat{\varrho}$ is such that there is $l_0 \in \mathbb{N}$, $l_0 > 0$, and a non-increasing $f : \{l_0 + 1, l_0 + 1, \dots\} \to \mathbb{R}_{\geq 0}$ such that for all operators $\hat{A} = \hat{A} \otimes \mathrm{id}_{\mathcal{X}\setminus\mathcal{A}}$ and $\hat{B} = \hat{B} \otimes \mathrm{id}_{\mathcal{X}\setminus\mathcal{B}}$ with $d(\mathcal{A},\mathcal{B}) > l_0$ we have

$$\left|\langle \hat{A}\hat{B}\rangle - \langle \hat{A}\rangle\langle \hat{B}\rangle\right| \leq \|\hat{A}\|\|\hat{B}\| f(d(\mathcal{A},\mathcal{B})). \tag{70}$$

We write

$$c[f] = \left(1 + \sum_{l=1}^{\infty} f(l + l_0) l^{D-1}\right)^{1/2}. \tag{71}$$

Write

$$\hat{\mathcal{X}} = \sum_n x_n |n\rangle\langle n| \tag{72}$$

and consider

$$F(y) = \sum_{n: x_n \leq y} \langle n|\hat{\varrho}|n\rangle. \tag{73}$$

Define

$$\Delta = \sup_y |F(y) - G(y)| \tag{74}$$

for the Gaussian cumulative distribution

$$G(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{y} \mathrm{d}z \, \mathrm{e}^{-\frac{(z-\mu)^2}{2\sigma^2}}. \tag{75}$$

We prove

Theorem 9. *Let $\hat{\varrho}$ such that $f(l) = \mathrm{e}^{-l/\xi}$ for some $\xi > 0$ and let $\log(|\mathcal{X}|) > \max\{1, (\frac{l_0}{4R})^{5/3}\}$. Then*

$$\Delta \le C \left(\max\left\{1, \frac{1}{s^2}\right\} + c[f]s + \mathrm{e}^{-1/\xi}\xi^D \max\left\{\frac{1}{s}, s^2\right\} \right) R^{D/2} \frac{1}{s|\mathcal{X}|^{1/2}} \left(\frac{\xi}{R}\log(|\mathcal{X}|) + \frac{\log(|\mathcal{X}|)}{\log(\log(|\mathcal{X}|))} \right)^{2D}, \tag{76}$$

where $C > 0$ depends only on the dimension of the lattice.

Proof. We first note that

$$\Delta = \sup_y |F(\sigma y + \mu) - G(\sigma y + \mu)| = \sup_y \Big| \sum_{n : \frac{x_n - \mu}{\sigma} \le y} \langle n|\hat{\varrho}|n\rangle - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \mathrm{d}z \, \mathrm{e}^{-\frac{z^2}{2}} \Big|. \tag{77}$$

We will employ Esseen's integral bound [1]

$$\Delta \le \frac{c_1}{T} + \int_0^T \mathrm{d}t \, \frac{|\varphi(t) - \mathrm{e}^{-t^2/2}|}{|t|}, \tag{78}$$

where $T > 0$, otherwise arbitrary, c_1 is an absolute constant, and φ is the Fourier-Stieltjes transform of $F(\sigma y + \mu)$, i.e., the characteristic function

$$\varphi(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}ty} \, \mathrm{d}F(\sigma y + \mu) = \sum_n \langle n|\hat{\varrho}|n\rangle \mathrm{e}^{\mathrm{i}t(\frac{x_n}{\sigma} - \mu)} = \left\langle \mathrm{e}^{\mathrm{i}t(\hat{\mathcal{X}}/\sigma - \mu)} \right\rangle =: \left\langle \mathrm{e}^{\mathrm{i}t\hat{\mathcal{Y}}} \right\rangle, \tag{79}$$

where

$$\hat{\mathcal{Y}} = \frac{1}{\sigma} \sum_{i \in \mathcal{X}} (\hat{X}_i - \langle \hat{X}_i \rangle) =: \frac{1}{\sigma} \sum_{i \in \mathcal{X}} \hat{Y}_i, \quad \langle \hat{\mathcal{Y}} \rangle = 0, \quad \left\langle \left(\sum_{i \in \mathcal{X}} \hat{Y}_i \right) \right\rangle = \sigma^2, \quad \|\hat{Y}_i\| \le x. \tag{80}$$

To not overburden notation, we redefine $\mathcal{X} := \mathcal{Y}$ and $\hat{X}_i := \hat{Y}_i$.

In order to apply Esseen's bound, the goal is to bound $|\varphi(t) - \mathrm{e}^{-t^2/2}|$, which we do along the lines of [2, 3], setting up a differential equation for $\varphi(t)$ and bounding its derivative. The main difference to [2, 3] will be to introduce auxiliary operators (operators $\hat{R}_n(t)$ and $\hat{S}_n(t)$ below) in order to be able to bound the non-commuting terms.

Throughout, we denote the support of an operator $\hat{O}(t) = \hat{O}(t) \otimes \mathrm{id}_{\mathcal{X} \setminus \mathcal{S}_{\hat{O}(t)}}$ by $\mathcal{S}_{\hat{O}(t)}$ and its n-th derivative by $\hat{O}^{(n)}(t)$. We let $t \ge 0$. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi(t) = \mathrm{i}\langle \hat{\mathcal{X}} \mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t} \rangle = \frac{\mathrm{i}}{\sigma} \sum_{j \in \mathcal{X}} \langle \hat{X}_j \mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t} \rangle. \tag{81}$$

We now fix $j \in \mathcal{X}$ and $0 < h \in \mathbb{N}$. For $n \in \mathbb{N}$, $n > 0$, we let

$$\begin{aligned} \hat{Z}_n &= \frac{1}{\sigma} \sum_{\substack{i\in\mathcal{X} \\ d(i,j)\leq 2Rhn}} \hat{X}_i, \quad \hat{z}_n = \hat{\mathcal{X}} - \hat{Z}_n = \frac{1}{\sigma} \sum_{\substack{i\in\mathcal{X} \\ d(i,j)> 2Rhn}} \hat{X}_i, \\ \hat{\xi}_n(t) &= \mathrm{e}^{\mathrm{i}(\hat{z}_{n-1}-\hat{z}_n)t}\hat{R}_n(t) - \mathrm{id}, \quad \hat{\Xi}_n(t) = \hat{\xi}_1(t)\cdots\hat{\xi}_n(t), \\ \hat{\eta}_n(t) &= \hat{S}_n(t)\mathrm{e}^{-\mathrm{i}\hat{Z}_n t} - \mathrm{id}, \end{aligned} \tag{82}$$

and we write $\hat{Z}_0 = \hat{X}_j/\sigma$, $\hat{z}_0 = \hat{\mathcal{X}}$, $\hat{\Xi}_0 = \mathrm{id}$. Here, $\hat{R}_n(t)$ and $\hat{S}_n(t)$ are arbitrary operators and we will choose them in Section IV A. One finds (see Appendix A for details and compare Refs. [2, 3])

$$\langle \hat{X}_j \mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\rangle = \Big(\mathrm{i}\langle \hat{X}_j\mathcal{X}\rangle t + g(t)\Big)\varphi(t) + h(t), \tag{83}$$

where $g(t) = g_1(t) + g_2(t) + g_3(t)$, $h(t) = h_1(t) + h_2(t) + h_3(t)$,

$$\begin{aligned} g_1(t) &= -\mathrm{i}\big(\langle \hat{X}_j\hat{z}_1\rangle - \langle\hat{X}_j\rangle\langle\hat{z}_1\rangle\big)t, \\ g_2(t) &= \big\langle \hat{X}_j\hat{\xi}_1(t)\big\rangle + \mathrm{i}\langle\hat{X}_j\hat{z}_1\rangle t - \mathrm{i}\langle\hat{X}_j\mathcal{X}\rangle t, \\ g_3(t) &= \big\langle \hat{X}_j\hat{\xi}_1(t)\big\rangle\langle\hat{\eta}_2(t)\rangle + \sum_{n=3}^{k}\big\langle\hat{X}_j\hat{\Xi}_{n-1}(t)\big\rangle\langle(\hat{\eta}_n(t)+\mathrm{id})\rangle, \\ h_1(t) &= \sum_{n=1}^{k}\Big(\langle\hat{X}_j\hat{\Xi}_{n-1}(t)\mathrm{e}^{\mathrm{i}\hat{z}_n t}\rangle - \langle\hat{X}_j\hat{\Xi}_{n-1}(t)\rangle\langle\mathrm{e}^{\mathrm{i}\hat{z}_n t}\rangle\Big), \\ h_2(t) &= \sum_{n=2}^{k}\big\langle\hat{X}_j\hat{\Xi}_{n-1}(t)\big\rangle\big\langle\big(\hat{\eta}_n(t) - \langle\hat{\eta}_n(t)\rangle\big)\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\big\rangle, \\ h_3(t) &= \langle\hat{X}_j\hat{\Xi}_k(t)\mathrm{e}^{\mathrm{i}\hat{z}_k t}\rangle + \sum_{n=0}^{k-1}\langle\hat{X}_j\hat{\Xi}_n(t)\hat{r}_n(t)\mathrm{e}^{\mathrm{i}\hat{z}_{n+1}t}\rangle + \sum_{n=2}^{k}\big\langle\hat{X}_j\hat{\Xi}_{n-1}(t)\big\rangle\langle\hat{s}_n(t)\rangle, \\ \hat{r}_n(t) &= \mathrm{e}^{\mathrm{i}(\hat{z}_n-\hat{z}_{n+1})t}\Big(\mathrm{e}^{-\mathrm{i}(\hat{z}_n-\hat{z}_{n+1})t}\mathrm{e}^{\mathrm{i}\hat{z}_n t}\mathrm{e}^{-\mathrm{i}\hat{z}_{n+1}t} - \hat{R}_{n+1}(t)\Big) =: \mathrm{e}^{\mathrm{i}(\hat{z}_n-\hat{z}_{n+1})}\Big(\hat{Z}_{R,n+1}(t) - \hat{R}_{n+1}(t)\Big), \\ \hat{s}_n(t) &= \Big(\mathrm{e}^{-\mathrm{i}(-\hat{\mathcal{X}}+\hat{Z}_n)t}\mathrm{e}^{-\mathrm{i}\hat{\mathcal{X}}t}\mathrm{e}^{\mathrm{i}\hat{Z}_n t} - \hat{S}_n(t)\Big)\mathrm{e}^{-\mathrm{i}\hat{Z}_n t}\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t} =: \Big(\hat{Z}_{S,n}(t) - \hat{S}_n(t)\Big)\mathrm{e}^{-\mathrm{i}\hat{Z}_n t}\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}. \end{aligned} \tag{84}$$

In the following section, we make the operators $\hat{R}_n(t)$ and $\hat{S}_n(t)$ explicit and derive some of their properties and give bounds on $|g_2(t) + g_3(t)|$ and $|h_3(t)|$. In Sections IV B-IV D, we derive bounds to the remaining functions in Eq. (84). Equipped with these bounds, we come back to Eqs. (81) and (83) in Section IV E to finish the proof.

□

A. Operators $\hat{R}_n(t)$ and $\hat{S}_n(t)$

We let $M \in \mathbb{N}$, $M \geq 2$, and put

$$\begin{aligned} \hat{R}_n(t) &= \mathrm{id} + \sum_{m=2}^{M}\hat{Z}_{R,n}^{(m)}(0)\frac{t^m}{m!}, \\ \hat{S}_n(t) &= \mathrm{id} + \sum_{m=2}^{M}\hat{Z}_{S,n}^{(m)}(0)\frac{t^m}{m!}. \end{aligned} \tag{85}$$

For such operators, we have the following Lemma, which we prove in Appendix C.

Lemma 10. *Let $M \in \mathbb{N}$, $M \geq 2$. Let $\mathcal{A}, \mathcal{B} \subset \mathcal{X}$ and $\hat{\mathcal{A}} = \sum_{i\in\mathcal{A}} \hat{Y}_i$, $\hat{\mathcal{B}} = \sum_{i\in\mathcal{B}} \hat{Y}_i$ with $\hat{Y}_i$ local (supported on $\{k \in \mathcal{X} \mid d(i,k) \leq R\}$) and bounded, $\|\hat{Y}_i\| \leq y$. Denote the support of $[\hat{\mathcal{A}}, \hat{\mathcal{B}}]$ by $\mathcal{S}$. Let*

$$\hat{Z}(t) = \mathrm{e}^{-\mathrm{i}t(\hat{\mathcal{A}}+\hat{\mathcal{B}})}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{A}}}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}}, \quad \hat{Z}_M(t) = \mathrm{id} + \sum_{m=2}^{M} \hat{Z}^{(m)}(0)\frac{t^m}{m!}. \tag{86}$$

Then $\mathcal{S}_{\hat{Z}_M(t)} \subset \{i \in \mathcal{X} \mid d(i,\mathcal{S}) \leq 2(M-2)R\}$. Further, let $\mathcal{C} \subset \mathcal{X} \times \mathcal{X}$ such that $[\hat{\mathcal{A}}, \hat{\mathcal{B}}] = \sum_{(i,j)\in\mathcal{C}}[\hat{Y}_i, \hat{Y}_j]$ and let β such that $\sum_{(i,j)\in\mathcal{C}} \|[\hat{Y}_i, \hat{Y}_j]\| \leq \beta$. If $\tau := 2t\max\{2yc_D(2R)^D, \sqrt{\beta/2}\} < 1/2$ then

$$\|\hat{Z}(t) - \hat{Z}_M(t)\| \leq 2\tau^{M+1} \quad \textit{and} \quad \|\hat{Z}_M^{(m)}(t)\| \leq 2m!(2\tau/t)^m. \tag{87}$$

We set out to apply the lemma to $\hat{Z}_{R,n}(t)$ and $\hat{Z}_{S,n}(t)$, which we defined in Eq. (84) and which are of the form (86) with $\hat{Y}_i = \hat{X}_i/\sigma$ (such that $y = x/\sigma$). To this end, we compute the corresponding commutators: We have $[\hat{z}_0, -\hat{z}_1] = [\hat{z}_1, \hat{\mathcal{X}}] = [\hat{z}_1, \hat{Z}_1]$, $[-\hat{\mathcal{X}}, \hat{Z}_n] = [\hat{Z}_n, \hat{z}_n]$, and for $n > 1$

$$[\hat{z}_{n-1}, -\hat{z}_n] = [\hat{z}_n, \hat{\mathcal{X}}] - [\hat{z}_n, \hat{Z}_{n-1}] = [\hat{z}_n, \hat{Z}_n] - [\hat{z}_n, \hat{Z}_{n-1}] = [\hat{z}_n, \hat{Z}_n], \tag{88}$$

where the last inequality holds as $h \geq 1$.[3] Now, for $n \geq 1$, we have

$$[\hat{z}_n, \hat{Z}_n] = \frac{1}{\sigma^2} \sum_{\substack{i,k\in\mathcal{X} \\ d(i,j)>2Rhn \\ d(k,j)\leq 2Rhn}} [\hat{X}_i, \hat{X}_k] = \frac{1}{\sigma^2} \sum_{\substack{i,k\in\mathcal{X} \\ d(i,j)>2Rhn \\ 2Rhn-2R<d(k,j)\leq 2Rhn \\ d(i,k)\leq 2R}} [\hat{X}_i, \hat{X}_k], \tag{89}$$

i.e.,

$$\begin{aligned} \|[\hat{z}_n, \hat{Z}_n]\| &\leq \frac{1}{\sigma^2} \sum_{\substack{i,k\in\mathcal{X} \\ d(i,j)>2Rhn \\ 2Rhn-2R<d(k,j)\leq 2Rhn \\ d(i,k)\leq 2R}} \|[\hat{X}_i, \hat{X}_k]\| \leq \frac{2x^2}{\sigma^2} \sum_{\substack{i,k\in\mathcal{X} \\ 2Rhn-2R<d(k,j)\leq 2Rhn \\ d(i,k)\leq 2R}} 1 \\ &\leq \frac{2x^2 c_D(2R)^D}{\sigma^2} \sum_{\substack{k\in\mathcal{X} \\ 2Rhn-2R<d(k,j)\leq 2Rhn}} 1 \leq \frac{2x^2 c_D^2(2R)^D}{\sigma^2} \sum_{l=2Rhn-2R+1}^{2Rhn} l^{D-1} \\ &\leq 2\left(\frac{2xc_D(2R)^D}{\sigma}\right)^2 (hn)^{D-1} =: \beta_n, \end{aligned} \tag{90}$$

for which we have $\sqrt{\beta_n/2} = \max\{2xc_D(2R)^D/\sigma, \sqrt{\beta_n/2}\}$, i.e., applying the lemma, we have for

$$\tau_n = 2t\sqrt{\beta_n/2} = 4t\frac{xc_D(2R)^D}{\sigma}(hn)^{\frac{D-1}{2}} \leq 1/2 \tag{91}$$

that

$$\|\hat{Z}_{R,n}(t) - \hat{R}_n(t)\|, \|\hat{Z}_{S,n}(t) - \hat{S}_n(t)\| \leq 2\tau_n^{M+1} \quad \text{and} \quad \|\hat{R}_n^{(m)}(t)\|, \|\hat{S}_n^{(m)}(t)\| \leq 2m!\big(4\sqrt{\beta_n/2}\big)^m. \tag{92}$$

[3] $\hat{z}_n$ is the sum of all $\hat{X}_i/\sigma$ with $d(i,j) > 2Rhn$ and $\hat{Z}_{n-1}$ is the sum of all $\hat{X}_i/\sigma$ with $d(i,j) \leq 2Rh(n-1)$, i.e., the distance between the support of $\hat{z}_n$ and the support of $\hat{Z}_{n-1}$ is at least $2Rh - 2R$.

Further, Eq. (89) implies that the support of $[\hat{z}_n, \hat{Z}_n]$ is a subset of

$$\begin{aligned}\mathcal{S} &\subset \bigcup_{\substack{i,k\in\mathcal{X}\\ d(i,j)>2Rhn\\ 2Rhn-2R<d(k,j)\leq 2Rhn\\ d(i,k)\leq 2R}} \{l\in\mathcal{X}\,|\,d(i,l)\leq R\}\cup\{m\in\mathcal{X}\,|\,d(m,k)\leq R\}\\ &\subset \bigcup_{\substack{i,k\in\mathcal{X}\\ d(i,j)>2Rhn\\ 2Rhn-2R<d(k,j)\leq 2Rhn\\ d(i,k)\leq 2R}} \{l\in\mathcal{X}\,|\,2Rhn-R<d(l,j)\leq 3R+2Rhn\}\cup\{m\in\mathcal{X}\,|\,2Rhn-3R<d(m,j)\leq 2Rhn+R\\ &\subset \{l\in\mathcal{X}\,|\,2Rhn-3R<d(l,j)\leq 3R+2Rhn\}.\end{aligned} \tag{93}$$

We now derive a few implications of Eqs. (92) and (93), which we formulate as a lemma.

Lemma 11. *Let $n, M\in\mathbb{N}$, $n\geq 1$, $M\geq 2$. The support of $\mathcal{S}_{\hat{\Xi}_n(t)}$ and the support of $\mathcal{S}_{\hat{\eta}^j_n(t)}$ fulfil*

$$\mathcal{S}_{\hat{\Xi}_n(t)}, \mathcal{S}_{\hat{\eta}^j_n(t)} \subset \left\{i\in\mathcal{X}\,\middle|\, d(i,j)\leq 2R(hn+M)-R\right\}. \tag{94}$$

Let $\tau_n = 2t\sqrt{\beta_n/2}\leq 1/2$. Then

$$\begin{aligned}\|\hat{\Xi}_n(t)\| &\leq \left(\tfrac{9}{2}h^{\frac{D+1}{2}}\tau_1\right)^n (n!)^{D-1},\\ \|\hat{\eta}_2(t)\| &\leq 5\cdot 2^{D-1}h^{\frac{D+1}{2}}\tau_1.\end{aligned} \tag{95}$$

Let $\frac{9}{2}\tau_k h^{\frac{D+1}{2}}k^{\frac{D-1}{2}}\leq 1/2$. Then

$$\begin{aligned}|h_3(t)| &\leq x\left(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\right)^2\left(2^{D+1}\frac{1}{2^k}+\frac{24}{81}\left(\tfrac{1}{9}h^{-\frac{D+1}{2}}k^{-\frac{D-1}{2}}\right)^{M-1}h^{-(D+1)}\right),\\ |g_2(t)+g_3(t)| &\leq x\left(\frac{145}{324}+\frac{14}{9}2^D+4\cdot 6^{D-1}\right)\left(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\right)^2.\end{aligned} \tag{96}$$

Proof. Direct application of Lemma 10 to Eq. (93) yields[4]

$$\mathcal{S}_{\hat{R}_n(t)}, \mathcal{S}_{\hat{R}_n(t)} \subset \left\{i\in\mathcal{X}\,\middle|\, 2R(hn-M)+R<d(i,j)\leq 2R(hn+M)-R\right\}, \tag{97}$$

which implies that for $n\geq 1$

$$\begin{aligned}\mathcal{S}_{\hat{\Xi}_n(t)} &\subset \bigcup_{m=1}^{n}\mathcal{S}_{\hat{z}_{m-1}-\hat{z}_m}\cup\mathcal{S}_{\hat{R}_m(t)}\\ &\subset \bigcup_{m=1}^{n}\left\{i\in\mathcal{X}\,\middle|\,2Rh(m-1)-R<d(i,j)\leq 2Rhm+R\right\}\cup\left\{i\in\mathcal{X}\,\middle|\,2R(hm-M)+R<d(i,j)\leq 2R(hm+\right.\\ &\subset \left\{i\in\mathcal{X}\,\middle|\,d(i,j)\leq 2R(hn+M)-R\right\}\end{aligned} \tag{98}$$

[4] We have that $2Rhn-3R<d(s,j)\leq 3R+2Rhn$ for all s in $\mathcal{S}$. From the lemma we have that $\mathcal{S}_{\hat{R}_n(t)}, \mathcal{S}_{\hat{R}_n(t)}\subset \left\{i\in\mathcal{X}\,\middle|\,d(i,\mathcal{S})\leq 2(M-2)R\right\}$. Thus, $2Rhn-3R<d(s,j)\leq d(i,j)+d(i,s)$ and $d(i,j)\leq d(i,s)+d(s,j)\leq d(i,s)+3R+2Rhn\leq 2(M-2)R+3R+2Rhn$, i.e., with the choice $d(i,\mathcal{S})=d(i,s)$, we have $2Rhn-3R-2(M-2)R<d(i,j)\leq 2(M-2)R+3R+2Rhn$.

and

$$\mathcal{S}_{\hat{\eta}^j_n(t)} \subset \mathcal{S}_{\hat{Z}_n} \cup \mathcal{S}_{\hat{S}_n(t)} \subset \left\{i \in \mathcal{X} \,\middle|\, d(i,j) \leq 2Rhn\right\} \cup \left\{i \in \mathcal{X} \,\middle|\, 2R(hn-M)+R < d(i,j) \leq 2R(hn+M)-R\right\}$$
$$\subset \left\{i \in \mathcal{X} \,\middle|\, d(i,j) \leq 2R(hn+M)-R\right\}. \tag{99}$$

As $\hat{R}_1(0) = \mathrm{id}$ and $\hat{R}_1^{(1)}(0) = 0$, we have

$$\begin{aligned}
|g_2(t)| &\leq x\left\|\mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)t}\hat{R}_1(t) - \mathrm{id} - \mathrm{i}(\mathcal{X}-\hat{z}_1)t\right\| \\
&= x\left\|\int_0^t \mathrm{d}s\,\left(\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)\mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)s}\hat{R}_1(s) + \mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)s}\hat{R}_1^{(1)}(s) - \mathrm{i}(\mathcal{X}-\hat{z}_1)\right)\right\| \\
&= x\left\|\int_0^t \mathrm{d}s\,\int_0^s \mathrm{d}u\,\left(-(\hat{\mathcal{X}}-\hat{z}_1)^2\mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)u}\hat{R}_1(u) + 2\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)\mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)u}\hat{R}_1^{(1)}(u) + \mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)u}\hat{R}_1^{(2)}(u)\right)\right\| \\
&\leq x\int_0^t \mathrm{d}s\,\int_0^s \mathrm{d}u\,\left(\|\hat{Z}_1\|^2\|\hat{R}_1(u)\| + 2\|\hat{Z}_1\|\|\hat{R}_1^{(1)}(u)\| + \|\hat{R}_1^{(2)}(u)\|\right) \\
&\leq x\int_0^t \mathrm{d}s\,\int_0^s \mathrm{d}u\,\left(\left(\frac{1}{2}\sqrt{\beta_1/2}h^{\frac{D+1}{2}}\right)^2\|\hat{R}_1(u)\| + \sqrt{\beta_1/2}h^{\frac{D+1}{2}}\|\hat{R}_1^{(1)}(u)\| + \|\hat{R}_1^{(2)}(u)\|\right),
\end{aligned} \tag{100}$$

i.e., Eq. (92) implies that for $2t\sqrt{\beta_1/2} = 4t\frac{xc_D(2R)^D}{\sigma}h^{\frac{D-1}{2}} \leq 1/2$, we have

$$|g_2(t)| \leq xt^2\frac{\beta_1}{2}\left(\frac{1}{4}h^{D+1} + 4h^{\frac{D+1}{2}} + 32\right) \leq \frac{145}{324}x\left(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\right)^2. \tag{101}$$

Further, for $2t\sqrt{\beta_n/2} \leq 1/2$, Eq. (92) implies

$$\begin{aligned}
\|\hat{\xi}_n(t)\| = \left\|\mathrm{e}^{\mathrm{i}(\hat{z}_{n-1}-\hat{z}_n)t}\hat{R}_n(t) - \mathrm{id}\right\| &\leq \int_0^t \mathrm{d}s\,\left(\|\hat{z}_{n-1}-\hat{z}_n\|\|\hat{R}_n(s)\| + \|\hat{R}_n^{(1)}(s)\|\right) \\
&\leq 2\|\hat{z}_{n-1}-\hat{z}_n\|t + 8\sqrt{\beta_n/2}t,
\end{aligned} \tag{102}$$

where

$$\begin{aligned}
\|\hat{z}_{n-1}-\hat{z}_n\| &= \begin{cases} \|\hat{Z}_1\| & \text{if } n=1, \\ \|\hat{Z}_{n-1}-\hat{Z}_n\| & \text{if } n>1, \end{cases} \\
&\leq \frac{x}{\sigma} \times \begin{cases} c_D(2Rh)^D & \text{if } n=1, \\ \sum_{i\in\mathcal{X}:2Rh(n-1)<d(i,j)\leq 2Rhn} 1 & \text{if } n>1, \end{cases} \\
&\leq \frac{x}{\sigma} \times \begin{cases} c_D(2Rh)^D & \text{if } n=1, \\ c_D(2Rhn)^{D-1}2Rh & \text{if } n>1, \end{cases} \\
&= \frac{1}{2}\frac{2xc_D(2R)^D}{\sigma}(hn)^{D-1}h = \frac{1}{2}\sqrt{\beta_n/2}(hn)^{\frac{D-1}{2}}h
\end{aligned} \tag{103}$$

i.e.,

$$\|\hat{\xi}_n(t)\| \leq \sqrt{\beta_n/2}(hn)^{\frac{D-1}{2}}ht + 8\sqrt{\beta_n/2}t = \sqrt{\beta_1/2}n^{\frac{D-1}{2}}t\left((hn)^{\frac{D-1}{2}}h + 8\right) \leq 9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}n^{D-1}t \tag{104}$$

and thus

$$\|\hat{\Xi}_n(t)\| \leq \left(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\right)^n(n!)^{D-1}. \tag{105}$$

Now,

$$\begin{aligned}|h_3(t)| &\le x\|\hat{\Xi}_k(t)\| + x\sum_{n=0}^{k-1}\|\hat{\Xi}_n(t)\|\|\hat{r}_n(t)\| + x\sum_{n=2}^{k}\|\hat{\Xi}_{n-1}(t)\|\|\hat{s}_n(t)\| \\ &\le x\|\hat{\Xi}_k(t)\| + x\sum_{n=1}^{k}\|\hat{\Xi}_{n-1}(t)\|\|\hat{Z}_{R,n}(t)-\hat{R}_n(t)\| + x\sum_{n=2}^{k}\|\hat{\Xi}_{n-1}(t)\|\|\hat{Z}_{S,n}(t)-\hat{S}_n(t)\|,\end{aligned} \tag{106}$$

i.e., Eq. (92) implies that for $2t\sqrt{\beta_n/2}\le 1/2$, we have

$$\begin{aligned}|h_3(t)| &\le x\|\hat{\Xi}_k(t)\| + 2x\sum_{n=1}^{k}\|\hat{\Xi}_{n-1}(t)\|\tau_n^{M+1} + 2x\sum_{n=2}^{k}\|\hat{\Xi}_{n-1}(t)\|\tau_n^{M+1} \\ &= x\|\hat{\Xi}_k(t)\| + 2x\tau_1^{M+1} + 4x\sum_{n=2}^{k}\|\hat{\Xi}_{n-1}(t)\|\tau_n^{M+1} \\ &\le x\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)^k (k!)^{D-1} + 2x\tau_1^{M+1} + 4x\sum_{n=1}^{k-1}\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)^n (n!)^{D-1}\tau_{n+1}^{M+1} \\ &= 2x\tau_1^{M+1} + x\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)\left(\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)^{k-1}(k!)^{D-1} + 4\sum_{n=0}^{k-2}\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)^n((n+1)!)^{D-1}\tau_{n+2}^{M+1}\right) \\ &\le 2x\tau_1^{M+1} + x\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)\left(2^{D-1}\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}k^{D-1}\big)^{k-2} + 4\sum_{n=0}^{k-2}\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}k^{D-1}\big)^n\tau_{n+2}^{M+1}\right),\end{aligned} \tag{107}$$

for which we have (we recall that $\tau_k = 4t\frac{xc_D(2R)^D}{\sigma}(hk)^{\frac{D-1}{2}} = \tau_1 k^{\frac{D-1}{2}}$)

$$\begin{aligned}|h_3(t)| &\le 2x\tau_1^{M+1} + x\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)\left(2^{D-1}\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)\frac{1}{2^{k-2}} + 8\tau_k^{M+1}\right) \\ &= x\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)^2\left(2^{D+1}\frac{1}{2^k} + \frac{8}{81}(\tau_k k^{-\frac{D-1}{2}})^{M-1}h^{-(D+1)} + \frac{16}{9}\tau_k^M h^{-\frac{D+1}{2}}k^{\frac{D-1}{2}}\right) \\ &\le x\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)^2\left(2^{D+1}\frac{1}{2^k} + \frac{24}{81}\tau_k^{M-1}h^{-(D+1)}\right) \\ &\le x\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)^2\left(2^{D+1}\frac{1}{2^k} + \frac{24}{81}\big(\tfrac{1}{9}h^{-\frac{D+1}{2}}k^{-\frac{D-1}{2}}\big)^{M-1}h^{-(D+1)}\right)\end{aligned} \tag{108}$$

if $\frac{9}{2}\tau_k h^{\frac{D+1}{2}}k^{\frac{D-1}{2}} \le 1/2$.

Eq. (92) further implies that for $2t\sqrt{\beta_n/2}\le 1/2$

$$\begin{aligned}\|\hat{\eta}_2(t)\| &= \|\hat{S}_2(t)\mathrm{e}^{-\mathrm{i}\hat{Z}_2 t} - \mathrm{id}\| \le \int_0^t \mathrm{d}s\,\big(\|\hat{S}_2^{(1)}(s)\| + \|\hat{S}_2(s)\|\|\hat{Z}_2\|\big) \\ &\le \big(8\sqrt{\beta_2/2} + \frac{2c_D x(4Rh)^D}{\sigma}\big)t \\ &= \big(8 + (2h)^{\frac{D+1}{2}}\big)\sqrt{\beta_2/2}\,t \le 5(2h)^{\frac{D+1}{2}}\sqrt{\beta_2/2}\,t,\end{aligned} \tag{109}$$

i.e.,

$$\begin{aligned}
|g_2(t)+g_3(t)| &\le |g_2(t)| + x\|\hat{\xi}_1(t)\|\|\hat{\eta}_2(t)\| + x\sum_{n=3}^{k}\|\hat{\Xi}_{n-1}(t)\|\|\hat{S}_n(t)\| \\
&\le |g_2(t)| + \frac{45}{81}2^D x\big(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\big)^2 + 2x\sum_{n=2}^{k-1}\big(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\big)^n(n!)^{D-1} \\
&\le x\left(\frac{145}{324}+\frac{45}{81}2^D\right)\big(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\big)^2 + 2x\big(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\big)^2\sum_{n=0}^{k-3}\big(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\big)^n((n+2)!)^{D-1} \\
&= x\left(\frac{145}{324}+\frac{14}{9}2^D\right)\big(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\big)^2 + 2x\big(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\big)^2\sum_{n=1}^{k-3}\big(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\big)^n((n+2)!)^{D-1} \\
&\le x\left(\frac{145}{324}+\frac{14}{9}2^D\right)\big(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\big)^2 + 2\cdot 6^{D-1}x\big(9\sqrt{\beta_1/2}h^{\frac{D+1}{2}}t\big)^2\sum_{n=1}^{k-3}\big(9\sqrt{\beta_n/2}h^{\frac{D+1}{2}}n^{\frac{D-1}{2}}t\big)^n,
\end{aligned} \tag{110}$$

for which we have

$$|g_2(t)+g_3(t)| \le x\left(\frac{145}{324}+\frac{14}{9}2^D+4\cdot 6^{D-1}\right)\left(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\right)^2 \tag{111}$$

if $\frac{9}{2}\tau_k h^{\frac{D+1}{2}}k^{\frac{D-1}{2}} \le 1/2$.

□

B. Bound on $|g_1(t)|$

We have

$$|g_1(t)| \le \frac{t}{\sigma}\sum_{\substack{i\in\mathcal{X}\\ d(i,j)>2Rh}}\big|\langle\hat{X}_j\hat{X}_i\rangle - \langle\hat{X}_j\rangle\langle\hat{X}_i\rangle\big|, \tag{112}$$

where for $l\in\mathcal{S}_{\hat{X}_j}$ we have $d(l,j)\le R$ and for $k\in\mathcal{S}_{\hat{X}_i}$ we have $d(k,i)\le R$, i.e., $2Rh<d(i,j)\le d(i,k)+d(k,l)+d(l,j)\le 2R+d(k,l)$, i.e., $d(\mathcal{S}_{\hat{X}_j},\mathcal{S}_{\hat{X}_i})\ge d(i,j)-2R>2R(h-1)$. Hence, for $2R(h-1)\ge l_0$

$$|g_1(t)| \le \frac{x^2t}{\sigma}\sum_{\substack{i\in\mathcal{X}\\ d(i,j)>2Rh}} f(d(i,j)-2R) \le \frac{c_D x^2 t}{\sigma}\sum_{l=2Rh+1}^{\infty} f(l-2R)l^{D-1}. \tag{113}$$

C. Bound on $|h_1(t)|$

We have

$$|h_1(t)| \le \big|\langle\hat{X}_j e^{i\hat{z}_1 t}\rangle - \langle\hat{X}_j\rangle\langle e^{i\hat{z}_1 t}\rangle\big| + \sum_{n=2}^{k}\big|\langle\hat{X}_j\hat{\Xi}_{n-1}(t)e^{i\hat{z}_n t}\rangle - \langle\hat{X}_j\hat{\Xi}_{n-1}(t)\rangle\langle e^{i\hat{z}_n t}\rangle\big|, \tag{114}$$

where, as $\langle \hat{X}_j \rangle = 0$,

$$\begin{aligned} \big|\langle \hat{X}_j \mathrm{e}^{\mathrm{i}\hat{z}_1 t}\rangle - \langle \hat{X}_j\rangle\langle \mathrm{e}^{\mathrm{i}\hat{z}_1 t}\rangle\big| &\leq \int_0^t \mathrm{d}s\, \big|\langle \hat{X}_j \hat{z}_1 \mathrm{e}^{\mathrm{i}\hat{z}_1 t}\rangle - \langle \hat{X}_j\rangle\langle \hat{z}_1 \mathrm{e}^{\mathrm{i}\hat{z}_1 t}\rangle\big| \\ &\leq \frac{1}{\sigma} \sum_{\substack{i\in\mathcal{X} \\ d(i,j)>2Rh}} \int_0^t \mathrm{d}s\, \big|\langle \hat{X}_j \hat{X}_i \mathrm{e}^{\mathrm{i}\hat{z}_1 t}\rangle - \langle \hat{X}_j\rangle\langle \hat{X}_i \mathrm{e}^{\mathrm{i}\hat{z}_1 t}\rangle\big|, \end{aligned} \tag{115}$$

i.e., analogous to Section IV B, we have for $2R(h-1) \geq l_0$ that

$$\big|\langle \hat{X}_j \mathrm{e}^{\mathrm{i}\hat{z}_1 t}\rangle - \langle \hat{X}_j\rangle\langle \mathrm{e}^{\mathrm{i}\hat{z}_1 t}\rangle\big| \leq \frac{c_D x^2 t}{\sigma} \sum_{l=2Rh+1}^{\infty} f(l-2R) l^{D-1}. \tag{116}$$

Now,

$$\mathcal{S}_{\hat{z}_n} \subset \big\{ i \in \mathcal{X} \,\big|\, d(i,j) > 2Rhn - R \big\}, \tag{117}$$

which, together with Lemma 11, implies that $d(\mathcal{S}_{\hat{z}_n}, \mathcal{S}_{\hat{\Xi}_{n-1}(t)}) > 2R(h-M)$. Hence, for $2R(h-M) \geq l_0$, we have

$$|h_1(t)| \leq \frac{c_D x^2 t}{\sigma} \sum_{l=2Rh+1}^{\infty} f(l-2R) l^{D-1} + x f(2R(h-M)+1) \sum_{n=1}^{k-1} \|\hat{\Xi}_n(t)\|, \tag{118}$$

where, again due to Lemma 11, for $\tau_n \leq 1/2$

$$\begin{aligned} \sum_{n=1}^{k-1} \|\hat{\Xi}_n(t)\| &\leq \sum_{n=1}^{k-1} \big(\tfrac{9}{2} h^{\frac{D+1}{2}} \tau_1\big)^n (n!)^{D-1} \\ &\leq \big(\tfrac{9}{2} h^{\frac{D+1}{2}} \tau_1\big) \sum_{n=0}^{k-2} \big(\tfrac{9}{2} h^{\frac{D+1}{2}} k^{D-1} \tau_1\big)^n, \end{aligned} \tag{119}$$

i.e., for $\frac{9}{2} h^{\frac{D+1}{2}} k^{\frac{D-1}{2}} \tau_k \leq 1/2$

$$|h_1(t)| \leq \frac{c_D x^2 t}{\sigma} \sum_{l=2Rh+1}^{\infty} f(l-2R) l^{D-1} + 2x f(2R(h-M)+1) \big(\tfrac{9}{2} h^{\frac{D+1}{2}} \tau_1\big). \tag{120}$$

D. Bound on $|h_2(t)|$

We recall that $\hat{\Xi}_{n-1}(t)$ and $\hat{\eta}_n(t)$ depend on j, which we denote by a superscript. Employing the Cauchy-Schwarz inequality, we find

$$\begin{aligned} \Big|\sum_{j\in\mathcal{X}} \big\langle \hat{X}_j \hat{\Xi}^j_{n-1}(t)\big\rangle \big\langle \big(\hat{\eta}^j_n(t) - \langle \hat{\eta}^j_n(t)\rangle\big) \mathrm{e}^{\mathrm{i}\hat{\mathcal{X}} t} \big\rangle\Big|^2 &= \Big|\Big\langle \sum_{j\in\mathcal{X}} \big\langle \hat{X}_j \hat{\Xi}^j_{n-1}(t)\big\rangle \big(\hat{\eta}^j_n(t) - \langle \hat{\eta}^j_n(t)\rangle\big) \mathrm{e}^{\mathrm{i}\hat{\mathcal{X}} t} \Big\rangle\Big|^2 \\ &\leq \Big\langle \Big(\sum_{j\in\mathcal{X}} \big\langle \hat{X}_j \hat{\Xi}^j_{n-1}(t)\big\rangle \big(\hat{\eta}^j_n(t) - \langle \hat{\eta}^j_n(t)\rangle\big)\Big) \Big(\sum_{j\in\mathcal{X}} \big\langle \hat{X}_j \hat{\Xi}^j_{n-1}(t)\big\rangle \big(\hat{\eta}^j_n(t) - \langle \hat{\eta}^j_n(t) \\ &\leq x^2 \sum_{i,j\in\mathcal{X}} \|\hat{\Xi}^i_{n-1}(t)\| \|\hat{\Xi}^j_{n-1}(t)\| \big(\langle \hat{\eta}^i_n(t) (\hat{\eta}^j_n(t))^\dagger\rangle - \langle \hat{\eta}^i_n(t)\rangle \langle (\hat{\eta}^j_n(t))^\dagger\rangle\big). \end{aligned} \tag{121}$$

Lemma 11 implies $d(\mathcal{S}_{\hat{\eta}^j_n(t)}, \mathcal{S}_{\hat{\eta}^i_n(t)}) \geq d(i,j) - 4R(hn+M) + 2R =: d(i,j) - r_n$, and therefore (we use $|\langle \hat{\eta}^i_n(t)(\hat{\eta}^j_n(t))^\dagger\rangle - \langle \hat{\eta}^i_n(t)\rangle\langle(\hat{\eta}^j_n(t))^\dagger\rangle| \leq \|\hat{\eta}^i_n(t)\|\|\hat{\eta}^j_n(t)\|$)

$$\begin{aligned}
\Big|\sum_{j\in\mathcal{X}} \big\langle \hat{X}_j \hat{\Xi}^j_{n-1}(t)\big\rangle\big\langle\big(\hat{\eta}^j_n(t) - \langle\hat{\eta}^j_n(t)\rangle\big)\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\big\rangle\Big|^2 \leq{}& x^2 \sum_{\substack{i,j\in\mathcal{X}\\ d(i,j)\leq r_n+l_0}} \|\hat{\Xi}^i_{n-1}(t)\|\|\hat{\Xi}^j_{n-1}(t)\|\|\hat{\eta}^i_n(t)\|\|\hat{\eta}^j_n(t)\| \\
&+ x^2 \sum_{\substack{i,j\in\mathcal{X}\\ d(i,j)> r_n+l_0}} \|\hat{\Xi}^i_{n-1}(t)\|\|\hat{\Xi}^j_{n-1}(t)\|\|\hat{\eta}^i_n(t)\|\|\hat{\eta}^j_n(t)\| f(d(i,j)-r_n) \\
\leq{}& c_D x^2 |\mathcal{X}| \Big((r_n+l_0)^D + \sum_{l=r_n+l_0+1}^{\infty} f(l-r_n) l^{D-1}\Big) \\
&\times \max_{i,j\in\mathcal{X}} \|\hat{\Xi}^i_{n-1}(t)\|\|\hat{\Xi}^j_{n-1}(t)\|\|\hat{\eta}^i_n(t)\|\|\hat{\eta}^j_n(t)\| \\
=:{}& (c_{r_n}[f])^2 \max_{i,j\in\mathcal{X}} \|\hat{\Xi}^i_{n-1}(t)\|\|\hat{\Xi}^j_{n-1}(t)\|\|\hat{\eta}^i_n(t)\|\|\hat{\eta}^j_n(t)\|,
\end{aligned} \tag{122}$$

where for $\tau_n \leq 1/2$, again using Lemma 11,

$$\begin{aligned}
\sqrt{\|\hat{\Xi}^i_{n-1}(t)\|\|\hat{\Xi}^j_{n-1}(t)\|\|\hat{\eta}^i_n(t)\|\|\hat{\eta}^j_n(t)\|} &\leq \big(\tfrac{9}{2} h^{\frac{D+1}{2}} \tau_1\big)^{n-1} ((n-1)!)^{D-1} \sqrt{\|\hat{\eta}^i_n(t)\|\|\hat{\eta}^j_n(t)\|} \\
&\leq \big(\tfrac{9}{2} h^{\frac{D+1}{2}} \tau_1\big)^{n-1} ((n-1)!)^{D-1} \times \begin{cases} 5\cdot 2^{D-1} h^{\frac{D+1}{2}}\tau_1 & \text{for } n=2, \\ 3 & \text{for } n>2, \end{cases}
\end{aligned} \tag{123}$$

i.e.,

$$\begin{aligned}
\Big|\sum_{j\in\mathcal{X}} h^j_2(t)\Big| &\leq 45\cdot 2^{D-2} c_{r_2}[f] h^{D+1}\tau_1^2 + 3\sum_{n=2}^{k-1} c_{r_{n+1}}[f] \big(\tfrac{9}{2} h^{\frac{D+1}{2}}\tau_1\big)^n (n!)^{D-1} \\
&= 12\cdot 2^{D-3} \left(\frac{10}{27} c_{r_2}[f] + \frac{1}{2^{D-1}} \sum_{n=0}^{k-3} c_{r_{n+3}}[f] \big(\tfrac{9}{2} h^{\frac{D+1}{2}}\tau_1\big)^n ((n+2)!)^{D-1}\right) \big(\tfrac{9}{2} h^{\frac{D+1}{2}}\tau_1\big)^2 \\
&\leq 12\cdot 2^{D-3} \left(\frac{10}{27} c_{r_2}[f] + \sum_{n=0}^{k-3} c_{r_{n+3}}[f] \big(\tfrac{9}{2}\tau_k h^{\frac{D+1}{2}} k^{\frac{D-1}{2}}\big)^n\right) \big(\tfrac{9}{2} h^{\frac{D+1}{2}}\tau_1\big)^2,
\end{aligned} \tag{124}$$

where for $2R(h-M) \geq l_0$ and $n \geq 2$

$$\begin{aligned}
(c_{r_n}[f])^2 &= c_D x^2 |\mathcal{X}| \Big((r_n+l_0)^D + \sum_{l=1}^{\infty} f(l+l_0)(l+r_n+l_0)^{D-1}\Big) \\
&\leq c_D x^2 |\mathcal{X}| (r_n+l_0)^D \Big(1 + \sum_{l=1}^{\infty} f(l+l_0)(l+1)^{D-1}\Big) \\
&\leq c_D x^2 |\mathcal{X}| (6R)^D (hn)^D \Big(1 + \sum_{l=1}^{\infty} f(l+l_0)(l+1)^{D-1}\Big),
\end{aligned} \tag{125}$$

i.e.,

$$\Big|\sum_{j\in\mathcal{X}} h_2^j(t)\Big| \le 12\cdot 2^{D-3} c_D^{1/2} x (6R)^{D/2}\Big(1+\sum_{l=1}^{\infty} f(l+l_0)(l+1)^{D-1}\Big)^{1/2} \\ \times \left(\frac{10}{27}2^{D/2} + \sum_{n=0}^{k-3}(n+3)^{D/2}\big(\tfrac{9}{2}\tau_k h^{\frac{D+1}{2}} k^{\frac{D-1}{2}}\big)^n\right)\big(\tfrac{9}{2}h^{\frac{D+1}{2}}\tau_1\big)^2 |\mathcal{X}|^{1/2} h^{D/2}, \tag{126}$$

which we combine with the bound on $|h_3(t)|$ in Lemma 11 to find that for $\frac{9}{2}\tau_k h^{\frac{D+1}{2}} k^{\frac{D-1}{2}} \le 1/2$

$$\Big|\sum_{j\in\mathcal{X}} h_2^j(t)\Big| + \Big|\sum_{j\in\mathcal{X}} h_3^j(t)\Big| \le x\Big[C_D[f]\left(\frac{10}{27}2^{D/2}+\sum_{n=0}^{k-3}\frac{(n+3)^{D/2}}{2^n}\right)\frac{h^{D/2}}{|\mathcal{X}|^{1/2}} \\ + \left(2^{D+1}\frac{1}{2^k} + \frac{24}{81}\big(\tfrac{1}{9}h^{-\frac{D+1}{2}}k^{-\frac{D-1}{2}}\big)^{M-1}h^{-(D+1)}\right)\Big]|\mathcal{X}|\big(\tfrac{9}{2}\tau_1 h^{\frac{D+1}{2}}\big)^2 \\ = \Big[C_D[f]\left(\frac{10}{27}2^{D/2}+\sum_{n=0}^{k-3}\frac{(n+3)^{D/2}}{2^n}\right)\frac{h^{D/2}}{|\mathcal{X}|^{1/2}} \\ + \left(2^{D+1}\frac{1}{2^k} + \frac{24}{81}\frac{1}{\big(9h^{\frac{D+1}{2}}k^{\frac{D-1}{2}}\big)^{M-1}h^{D+1}}\right)\Big]\frac{|\mathcal{X}|^{3/2}x^3}{\sigma^2}\big(18c_D 2^D\big)^2 R^{2D}\frac{h^{2D}}{|\mathcal{X}|^{1/2}}t^2, \tag{127}$$

where

$$C_D[f] = 3\cdot 2^{D-1}c_D^{1/2}(6R)^{D/2}\sqrt{1+\sum_{l=1}^{\infty} f(l+l_0)(l+1)^{D-1}}. \tag{128}$$

E. Final steps

We denote by $C>0$ a constant, not always the same, that depends only on D. Further, we write $S=\frac{\sigma}{x|\mathcal{X}|^{1/2}}$. Combining Eqs. (81) and (83) and the bounds obtained in the previous sections, we have

$$\varphi^{(1)}(t) = \big(-t+g(t)\big)\varphi(t) + h(t), \quad \varphi(0)=1, \tag{129}$$

where, letting $k,h,M\in\mathbb{N}$, $k,M\ge 2$, $2R(h-M)\ge l_0$, and

$$t \le \frac{CS}{R^D}\frac{|\mathcal{X}|^{1/2}}{h^D k^{D-1}} =: T_1 \tag{130}$$

the functions g and h are bounded by

$$|g(t)| \le c_1 t + c_2 t^2, \\ |h(t)| \le c_3 t + c_4 t^2, \tag{131}$$

where

$$
\begin{aligned}
c_1 &= \frac{C}{S^2}\sum_{l=2Rh+1}^{\infty} f(l-2R)l^{D-1},\\
c_3 &= c_1 + \frac{CR^D}{S^2} f(2R(h-M)+1)h^D =: c_1 + c_5,\\
c_2 &= \frac{CR^{2D}}{S^3}\frac{h^{2D}}{|\mathcal{X}|^{1/2}},\\
c_4 &= CC[f]\frac{S^3}{R^{2D}}\frac{1}{h^{3D/2}}c_2^2 + \Big(\frac{1}{2^k} + \frac{1}{\big(9h^{\frac{D+1}{2}}k^{\frac{D-1}{2}}\big)^{M-1}h^{D+1}}\Big)c_2 =: CC[f]\frac{S^3}{R^{2D}}\frac{1}{h^{3D/2}}c_2^2 + c_6 \le CC[f]\frac{S^3}{R^{2D}}c_2^2 + c_6,\\
C[f] &= R^{D/2}\Big(1+\sum_{l=1}^{\infty} f(l+l_0)(l+1)^{D-1}\Big)^{1/2}.
\end{aligned}
\tag{132}
$$

Now let $c_1 < 1/2$ and

$$
t \le \frac{1}{4c_2} = \frac{CS^3}{R^{2D}}\frac{|\mathcal{X}|^{1/2}}{h^{2D}} =: T_2. \tag{133}
$$

Then the solution to Eq. (129) with functions g and h bounded as above fulfils (see Appendix B)

$$
|\varphi(t) - \mathrm{e}^{-t^2/2}| \le \left(\frac{c_1}{2} + \frac{c_2}{3}t\right) t^2\mathrm{e}^{-t^2/6} + 4(1-\mathrm{e}^{-\frac{t^2}{4}})(c_3 + c_4 t), \tag{134}
$$

i.e., for

$$
1 < T := \min\{T_1, T_2\} \tag{135}
$$

we have, using Esseen's inequality [1],

$$
\begin{aligned}
\Delta &\le C\frac{1}{T} + C\int_0^T \mathrm{d}t\,\frac{|\varphi(t)-\mathrm{e}^{-t^2/2}|}{t}\\
&\le C\frac{1}{\min\{T_1,T_2\}} + C\int_0^\infty \mathrm{d}t\ (c_1+c_2 t)\,t\mathrm{e}^{-t^2/6} + Cc_3\left(\int_0^1 \mathrm{d}t\,\frac{1-\mathrm{e}^{-\frac{t^2}{4}}}{t} + \int_1^T \mathrm{d}t\,\frac{1}{t}\right) + Cc_4 T\\
&\le C\max\left\{\frac{1}{T_1},\frac{1}{T_2}\right\} + C\,(c_1+c_2) + Cc_3\,(1+\log(T)) + C\frac{C[f]S^3}{R^{2D}}c_2^2 T + Cc_6 T\\
&\le C\left(\max\left\{\frac{1}{c_2 T_1},1\right\} + \frac{C[f]S^3}{R^{2D}}\right)c_2 + Cc_3\max\{1,\log(\min\{T_1,T_2\})\} + C\frac{c_6}{c_2},
\end{aligned}
\tag{136}
$$

which also holds for $c_1 \ge 1/2$ and $0 < T \le 1$ as, trivially, $\Delta \le 2$.

We bound

$$
c_6 \le C\Big(\frac{1}{2^k} + \frac{1}{\big(9h^{\frac{D+1}{2}}k^{\frac{D-1}{2}}\big)^{M-1}h^{D+1}}\Big)c_2 \le C\Big(\frac{1}{2^k} + h^{-M}\Big)c_2. \tag{137}
$$

Now set $h = 2M$, $k = \lceil M\log_2(M)\rceil$, and let $\log(\log(|\mathcal{X}|)) > 0$ and $M \ge \frac{2\log(|\mathcal{X}|)}{\log(\log(|\mathcal{X}|))} \ge \max\{2, \frac{l_0}{2R}\}$. Then all the assumptions on k, h, M are fulfilled and we have $2 \le \alpha M\log_2(M) \le k \le M\log_2(M) + 1 \le 2M\log_2(M)$ and

$$
c_6 \le CM^{-M}c_2 \le Cc_2\exp\left(-\frac{2\log(|\mathcal{X}|)}{\log(\log(|\mathcal{X}|))}\log\left(\frac{2\log(|\mathcal{X}|)}{\log(\log(|\mathcal{X}|))}\right)\right) \le Cc_2|\mathcal{X}|^{-1}. \tag{138}
$$

and

$$c_2T_1 = \frac{CR^D}{S^2}\frac{M^D}{k^{D-1}} \geq \frac{CR^D}{S^2}\frac{M}{[\log_2(M)]^{D-1}} \geq \frac{CR^D}{S^2}. \tag{139}$$

Hence, writing $s = S/R^{D/2}$, $c[f] = C[f]/R^{D/2}$,

$$\begin{aligned}\Delta &\leq C\left(\max\left\{s^2,1\right\} + c[f]s^3\right)\frac{R^{D/2}}{s^3}\frac{M^{2D}}{|\mathcal{X}|^{1/2}} + C|\mathcal{X}|^{-1} + Cc_3\max\left\{1,\log\left(\frac{s^3}{R^{D/2}}\frac{|\mathcal{X}|^{1/2}}{M^{2D}}\right)\right\} \\ &\leq C\left(\max\left\{1,\frac{1}{s^2}\right\} + c[f]s\right)R^{D/2}\frac{M^{2D}}{s|\mathcal{X}|^{1/2}} + Cc_3s^2\max\left\{\frac{1}{s^2},s\right\}|\mathcal{X}|^{1/2}.\end{aligned} \tag{140}$$

1. *Exponential decay*

Now let $f(l) \leq \mathrm{e}^{-l/\xi}$ with $\xi > 0$. Then

$$\begin{aligned}Cs^2c_3 &= \frac{1}{R^D}\mathrm{e}^{2R/\xi}\sum_{l=1}^{\infty}\mathrm{e}^{-(l+4RM)/\xi}(l+4RM)^{D-1} + \mathrm{e}^{-(2RM+1)/\xi}M^D \\ &= \frac{1}{R^D}\mathrm{e}^{2R/\xi}\xi^{D-1}\sum_{l=1}^{\infty}\int_{l-1}^{l}\mathrm{d}x\,\mathrm{e}^{-(l+4RM)/\xi}[(l+4RM)/\xi]^{D-1} + \mathrm{e}^{-(2RM+1)/\xi}M^D,\end{aligned} \tag{141}$$

i.e., for $D \leq 1 + 4RM/\xi$,

$$\begin{aligned}Cs^2c_3 &\leq \frac{1}{R^D}\mathrm{e}^{2R/\xi}\xi^D\int_{4RM/\xi}^{\infty}\mathrm{d}x\,\mathrm{e}^{-x}x^{D-1} + \mathrm{e}^{-(2RM+1)/\xi}M^D \\ &\leq \frac{1}{R^D}\mathrm{e}^{2R/\xi}\xi^D\mathrm{e}^{-3RM/\xi}\int_0^{\infty}\mathrm{d}x\,\mathrm{e}^{-x/4}x^{D-1} + \frac{\xi^DD^D}{R^D}\mathrm{e}^{-(RM+1)/\xi}\left(\mathrm{e}^{-\frac{RM}{D\xi}}\frac{RM}{D\xi}\right)^D \\ &\leq \left(\int_0^{\infty}\mathrm{d}x\,\mathrm{e}^{-x/4}x^{D-1} + D^D\right)\mathrm{e}^{-1/\xi}\frac{\xi^D}{R^D}\mathrm{e}^{-RM/\xi}.\end{aligned} \tag{142}$$

Now set $M = \lceil D\xi\log(|\mathcal{X}|)/R + \frac{2\log(|\mathcal{X}|)}{\log(\log(|\mathcal{X}|))}\rceil$ (and recall that we assumed $\log(\log(|\mathcal{X}|)) > 0$ and $\frac{2\log(|\mathcal{X}|)}{\log(\log(|\mathcal{X}|))} \geq \max\{2, \frac{l_0}{2R}\}$, both of which are implied by $\log(|\mathcal{X}|) > \max\{1, (\frac{l_0}{4R})^{5/3}\}$). Then $4RM/\xi + 1 \geq 4D\log(|\mathcal{X}|) > D$, i.e., the above assumption is fulfilled and we have

$$\begin{aligned}\Delta &\leq C\left(\max\left\{1,\frac{1}{s^2}\right\} + c[f]s + \mathrm{e}^{-1/\xi}\xi^D\max\left\{\frac{1}{s},s^2\right\}\right)R^{D/2}\frac{M^{2D}}{s|\mathcal{X}|^{1/2}} \\ &= C\left(\max\left\{1,\frac{1}{s^2}\right\} + c[f]s + \mathrm{e}^{-1/\xi}\xi^D\max\left\{\frac{1}{s},s^2\right\}\right)R^{D/2}\frac{1}{s|\mathcal{X}|^{1/2}}\left(\frac{\xi}{R}\log(|\mathcal{X}|) + \frac{\log(|\mathcal{X}|)}{\log(\log(|\mathcal{X}|))}\right)^{2D}.\end{aligned} \tag{143}$$

---

[1] C.G. Esseen, Acta Math. **77**, 1 (1945).

[2] J. Sunklodas, *Rate of convergence in the central limit theorem for random variables with strong mixing*, Lithuanian Mathematical Journal **24**, 182 (1984).

[3] A.N. Tikhomirov, *On the Convergence Rate in the Central Limit Theorem for Weakly Dependent Random Variables*, Theory Probab. Appl. **25**, 790 (1980).

[4] D. Goderis and P. Vets. Central limit theorem for mixing quantum systems and the CCR-algebra of fluctuations. Commun. Math. Phys. 122, 249 (1989).
[5] M. Hartmann, G. Mahler, and O. Hess, Gaussian quantum fluctuations in interacting many particle systems, Lett. Math. Phys. 68, 103 (2004).
[6] A.C. Berry, The accuracy of the Gaussian approximation to the sum of independent variates, Trans. Amer. Math. Soc. 49, 122 (1941).
[7] L.H. Y. Chen and Q.-M. Shao. Normal approximation under local dependence. Ann. Probab. 32, 198 (2004).
[8] M. Deserno. Microcanonical and canonical two-dimensional Ising model: An example.
[9] H. Araki. Gibbs states of a one dimensional quantum lattice. Comm. Math. Phys. **14**, 120 (1969).
[10] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, J. Eisert. Locality of temperature. Phys. Rev. X **4**, 031019 (2014).
[11] G. Bouch. Complex-time singularity and locality estimates for quantum lattice systems. arXiv:1011.1875.
[12] N. Datta. Min- and Max- Relative Entropies and a New Entanglement Monotone. IEEE Transactions on Information Theory 55, 2816 (2009).
[13] S. Popescu, A.J. Short, A. Winter. Entanglement and the foundations of statistical mechanics. Nature Physics **2**, 754 (2006).
[14] H. Touchette. Equivalence of statistical-mechanical ensembles: A collection of quotes and notes. 2006.
[15] J.L. Lebowitz and E.H. Lieb, Existence of thermodynamics for real matter with Coulomb forces. Phys. Rev. Lett. **22**, 631 (1969).
[16] H.-O. Georgii, The equivalence of ensembles for classical systems of particles. J. Stat. Phys. **80**, 1341 (1995).
[17] H. Touchette, The large deviation approach to statistical mechanics. Phys. Rep. 478, 1 (2009).
[18] R. Lima. Equivalence of ensembles in quantum lattice systems. Annales de l'I. H. P. 15 (1), 6168 (1971).
[19] R. Lima. Equivalence of Ensembles in Quantum Lattice Systems: States. Commun. Math. Phys. 24, 180192 (1972).
[20] M.P. Müller, E. Adlam, Ll. Masanes, and Nathan Wiebe. Thermalization and canonical typicality in translation-invariant quantum lattice systems. arXiv:1312.7420.
[21] M. Srednicki. Chaos and Quantum Thermalization. Physical Review E **50**, 888 (1994).
[22] N. Linden, S. Popescu, A.J. Short, A. Winter. Quantum mechanical evolution towards thermal equilibrium. Phys. Rev. E **79**, 061103 (2009).
[23] A. Riera, C. Gogolin, J. Eisert. Thermalization in Nature and on a Quantum Computer. Phys. Rev. Lett. **108**, 080402 (2012).
[24] F.G.S.L. Brandão and M. Horodecki. Exponential Decay of Correlations Implies Area Law. To appear in CMP. arXiv:1206.2947.
[25] N. Datta and R. Renner. Smooth Renyi Entropies and the Quantum Information Spectrum. IEEE Transactions on Information Theory 55, 2807 (2009).
[26] F.G.S.L. Brandão and M.B. Plenio. A Generalization of Quantum Steins Lemma. Commun. Math. Phys. **295**, 791 (2010).

Appendix A: The first step

Omitting the argument, we write $\hat{\Xi}_n = \hat{\Xi}_n(t)$, $\hat{\xi}_n = \hat{\xi}_n(t)$, $\hat{\eta}_n = \hat{\eta}_n(t)$, $\hat{S}_n = \hat{S}_n(t)$, and $\hat{R}_n = \hat{R}_n(t)$ throughout this section. For $k \in \mathbb{N}$, $k \geq 2$, we have

$$\begin{aligned} \mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t} &= (\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\mathrm{e}^{-\mathrm{i}\hat{z}_1 t} - \mathrm{id})\mathrm{e}^{\mathrm{i}\hat{z}_1 t} + \mathrm{e}^{\mathrm{i}\hat{z}_1 t} \\ &= (\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\mathrm{e}^{-\mathrm{i}\hat{z}_1 t} - \mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)t}\hat{R}_1)\mathrm{e}^{\mathrm{i}\hat{z}_1 t} + (\mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)t}\hat{R}_1 - \mathrm{id})\mathrm{e}^{\mathrm{i}\hat{z}_1 t} + \mathrm{e}^{\mathrm{i}\hat{z}_1 t} \\ &= \mathrm{e}^{\mathrm{i}\hat{z}_1 t} + \sum_{n=2}^{k} \hat{\Xi}_{n-1}\mathrm{e}^{\mathrm{i}\hat{z}_n t} + (\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\mathrm{e}^{-\mathrm{i}\hat{z}_1 t} - \mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{z}_1)t}\hat{R}_1)\mathrm{e}^{\mathrm{i}\hat{z}_1 t} + \hat{\xi}_1\mathrm{e}^{\mathrm{i}\hat{z}_1 t} - \sum_{n=2}^{k} \hat{\Xi}_{n-1}\mathrm{e}^{\mathrm{i}\hat{z}_n t}. \end{aligned} \tag{A1}$$

For $n \in \mathbb{N}$, $1 \leq n \leq k-1$, denote

$$\hat{Y}_n = \hat{\Xi}_n e^{i\hat{z}_n t} - \sum_{m=n+1}^{k} \hat{\Xi}_{m-1} e^{i\hat{z}_m t}. \tag{A2}$$

Then

$$e^{i\hat{\mathcal{X}}t} = e^{i\hat{z}_1 t} + \sum_{n=2}^{k} \hat{\Xi}_{n-1} e^{i\hat{z}_n t} + (e^{i\hat{\mathcal{X}}t} e^{-i\hat{z}_1 t} - e^{i(\hat{\mathcal{X}}-\hat{z}_1)t} \hat{R}_1) e^{i\hat{z}_1 t} + \hat{Y}_1. \tag{A3}$$

We have

$$\begin{aligned}
\hat{Y}_n &= \hat{\Xi}_n e^{i\hat{z}_n t} - \hat{\Xi}_n e^{i\hat{z}_{n+1} t} - \sum_{m=n+2}^{k} \hat{\Xi}_{m-1} e^{i\hat{z}_m t} \\
&= \hat{Y}_{n+1} + \hat{\Xi}_n e^{i\hat{z}_n t} - \hat{\Xi}_n e^{i\hat{z}_{n+1} t} - \hat{\Xi}_{n+1} e^{i\hat{z}_{n+1} t} \\
&= \hat{Y}_{n+1} + \hat{\Xi}_n \left(e^{i\hat{z}_n t} e^{-i\hat{z}_{n+1} t} - e^{i(\hat{z}_n - \hat{z}_{n+1})} \hat{R}_{n+1} \right) e^{i\hat{z}_{n+1} t} \\
&=: \hat{Y}_{n+1} + \hat{\Delta}_n
\end{aligned} \tag{A4}$$

for $1 \leq n \leq k-2$ and, letting $\hat{\Delta}_{k-1} = \hat{\Xi}_{k-1} \left(e^{i\hat{z}_{k-1} t} e^{-i\hat{z}_k t} - e^{i(\hat{z}_{k-1} - \hat{z}_k)} \hat{R}_k \right) e^{i\hat{z}_k t}$,

$$\begin{aligned}
\hat{Y}_{k-1} &= \hat{\Xi}_{k-1} e^{i\hat{z}_{k-1} t} - \hat{\Xi}_{k-1} e^{i\hat{z}_k t} \\
&= \hat{\Xi}_{k-1} \left(e^{i\hat{z}_{k-1} t} e^{-i\hat{z}_k t} - \mathrm{id} \right) e^{i\hat{z}_k t} - \hat{\Delta}_{k-1} + \hat{\Delta}_{k-1} \\
&= \hat{\Xi}_{k-1} \left(e^{i(\hat{z}_{k-1} + \hat{z}_k)} \hat{R}_k - \mathrm{id} \right) e^{i\hat{z}_k t} + \hat{\Delta}_{k-1} \\
&= \hat{\Xi}_k e^{i\hat{z}_k t} + \hat{\Delta}_{k-1}.
\end{aligned} \tag{A5}$$

Hence,

$$\hat{Y}_1 = \hat{\Xi}_k e^{i\hat{z}_k t} + \sum_{n=1}^{k-1} \hat{\Delta}_n \tag{A6}$$

and therefore, writing $\hat{\Delta}_0 = (e^{i\hat{\mathcal{X}}t} e^{-i\hat{z}_1 t} - e^{i(\hat{\mathcal{X}}-\hat{z}_1)t} \hat{R}_1) e^{i\hat{z}_1 t}$,

$$e^{i\hat{\mathcal{X}}t} = e^{i\hat{z}_1 t} + \sum_{n=2}^{k} \hat{\Xi}_{n-1} e^{i\hat{z}_n t} + \hat{\Xi}_k e^{i\hat{z}_k t} + \sum_{n=0}^{k-1} \hat{\Delta}_n. \tag{A7}$$

Now, for $n \geq 1$

$$\langle e^{i\hat{z}_n t} \rangle = \langle e^{i(\hat{\mathcal{X}} - \hat{Z}_n)t} \rangle = \langle e^{i(\hat{\mathcal{X}} - \hat{Z}_n)t} \rangle - \langle \hat{S}_n e^{-i\hat{Z}_n t} e^{i\hat{\mathcal{X}}t} \rangle + \langle \hat{S}_n e^{-i\hat{Z}_n t} e^{i\hat{\mathcal{X}}t} \rangle, \tag{A8}$$

where

$$\begin{aligned}
\langle \hat{S}_n e^{-i\hat{Z}_n t} e^{i\hat{\mathcal{X}}t} \rangle &= \langle (\hat{\eta}_n + \mathrm{id}) e^{i\hat{\mathcal{X}}t} \rangle = \langle (\hat{\eta}_n + \mathrm{id}) e^{i\hat{\mathcal{X}}t} \rangle - \langle (\hat{\eta}_n + \mathrm{id}) \rangle \varphi(t) + \langle (\hat{\eta}_n + \mathrm{id}) \rangle \varphi(t) \\
&= \langle (\hat{\eta}_n - \langle \hat{\eta}_n \rangle) e^{i\hat{\mathcal{X}}t} \rangle + \langle (\hat{\eta}_n + \mathrm{id}) \rangle \varphi(t).
\end{aligned} \tag{A9}$$

Hence,

$$\sum_{n=2}^{k}\big\langle\hat{X}_j\hat{\Xi}_{n-1}\big\rangle\big\langle\big(\hat{\eta}_n-\langle\hat{\eta}_n\rangle\big)\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\big\rangle=\sum_{n=2}^{k}\big\langle\hat{X}_j\hat{\Xi}_{n-1}\big\rangle\langle\mathrm{e}^{\mathrm{i}\hat{z}_n t}\rangle-\sum_{n=2}^{k}\big\langle\hat{X}_j\hat{\Xi}_{n-1}\big\rangle\langle(\hat{\eta}_n+\mathrm{id})\rangle\varphi(t)$$
$$+\sum_{n=2}^{k}\big\langle\hat{X}_j\hat{\Xi}_{n-1}\big\rangle\left(\langle\hat{S}_n\mathrm{e}^{-\mathrm{i}\hat{Z}_n t}\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\rangle-\langle\mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{Z}_n)t}\rangle\right) \tag{A10}$$

and therefore, using that $\langle\hat{X}_j\rangle=0$,

$$\begin{aligned}\langle\hat{X}_j\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\rangle=\;&\mathrm{i}\langle\hat{X}_j\mathcal{X}\rangle\varphi(t)t-\mathrm{i}\left(\langle\hat{X}_j\hat{z}_1\rangle-\langle\hat{X}_j\rangle\langle\hat{z}_1\rangle\right)\varphi(t)t\\&+\left(\big\langle\hat{X}_j\hat{\xi}_1(t)\big\rangle+\mathrm{i}\langle\hat{X}_j\hat{z}_1\rangle t-\mathrm{i}\langle\hat{X}_j\mathcal{X}\rangle t\right)\varphi(t)\\&+\big\langle\hat{X}_j\hat{\xi}_1(t)\big\rangle\langle\hat{\eta}_2(t)\rangle\varphi(t)+\sum_{n=3}^{k}\big\langle\hat{X}_j\hat{\Xi}_{n-1}(t)\big\rangle\langle(\hat{\eta}_n(t)+\mathrm{id})\rangle\varphi(t)\\&+\sum_{n=1}^{k}\left(\langle\hat{X}_j\hat{\Xi}_{n-1}(t)\mathrm{e}^{\mathrm{i}\hat{z}_n t}\rangle-\langle\hat{X}_j\hat{\Xi}_{n-1}(t)\rangle\langle\mathrm{e}^{\mathrm{i}\hat{z}_n t}\rangle\right)\\&+\sum_{n=2}^{k}\big\langle\hat{X}_j\hat{\Xi}_{n-1}(t)\big\rangle\big\langle\big(\hat{\eta}_n(t)-\langle\hat{\eta}_n(t)\rangle\big)\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\big\rangle+\langle\hat{X}_j\hat{\Xi}_k(t)\mathrm{e}^{\mathrm{i}\hat{z}_k t}\rangle\\&+\sum_{n=0}^{k-1}\langle\hat{X}_j\hat{\Delta}_n(t)\rangle-\sum_{n=2}^{k}\big\langle\hat{X}_j\hat{\Xi}_{n-1}(t)\big\rangle\left(\langle\hat{S}_n(t)\mathrm{e}^{-\mathrm{i}\hat{Z}_n t}\mathrm{e}^{\mathrm{i}\hat{\mathcal{X}}t}\rangle-\langle\mathrm{e}^{\mathrm{i}(\hat{\mathcal{X}}-\hat{Z}_n)t}\rangle\right),\end{aligned} \tag{A11}$$

where for $n\in\mathbb{N}$, $0\le n\le k-1$,

$$\hat{\Delta}_n(t)=\hat{\Xi}_n(t)\left(\mathrm{e}^{\mathrm{i}\hat{z}_n t}\mathrm{e}^{-\mathrm{i}\hat{z}_{n+1}t}-\mathrm{e}^{\mathrm{i}(\hat{z}_n-\hat{z}_{n+1})}\hat{R}_{n+1}(t)\right)\mathrm{e}^{\mathrm{i}\hat{z}_{n+1}t}. \tag{A12}$$

Appendix B: The differential equation and the bound on its solution

The solution of

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t}(t)=-t\varphi(t)+g(t)\varphi(t)+h(t),\quad \varphi(0)=1, \tag{B1}$$

is given by

$$\varphi(t)=\mathrm{e}^{-a(t)}\Big(1+\int_0^t\mathrm{d}s\,h(s)\mathrm{e}^{a(s)}\Big),\quad a(t)=\int_0^t\mathrm{d}s\;(s-g(s))=t^2/2-\int_0^t\mathrm{d}s\,g(s) \tag{B2}$$

and fulfils

$$\begin{aligned}|\varphi(t)-\mathrm{e}^{-t^2/2}|&=\mathrm{e}^{-t^2/2}\Big|\mathrm{e}^{\int_0^t\mathrm{d}s\,g(s)}+\mathrm{e}^{\int_0^t\mathrm{d}s\,g(s)}\int_0^t\mathrm{d}s\,h(s)\mathrm{e}^{s^2/2-\int_0^s\mathrm{d}u\,g(u)}-1\Big|\\&\le\mathrm{e}^{-t^2/2}\big|\mathrm{e}^{\int_0^t\mathrm{d}s\,g(s)}-1\big|+\mathrm{e}^{-t^2/2}\Big|\mathrm{e}^{\int_0^t\mathrm{d}s\,g(s)}\int_0^t\mathrm{d}s\,h(s)\mathrm{e}^{s^2/2-\int_0^s\mathrm{d}u\,g(u)}\Big|\\&\le\mathrm{e}^{-t^2/2}\mathrm{e}^{\int_0^t\mathrm{d}s\,|g(s)|}\int_0^t\mathrm{d}s\,|g(s)|+\mathrm{e}^{-t^2/2}\int_0^t\mathrm{d}s\,|h(s)|\mathrm{e}^{s^2/2+\int_s^t\mathrm{d}u\,|g(u)|}.\end{aligned} \tag{B3}$$

If $|g(u)| \leq c_1 u + c_2 u^2$ with $c_{1,2} \geq 0$, we find

$$\int_0^t \mathrm{d}s\, |h(s)| \mathrm{e}^{s^2/2 + \int_s^t \mathrm{d}u\, |g(u)|} \leq \mathrm{e}^{c_1 t^2/2 + c_2 t^3/3} \int_0^t \mathrm{d}s\, |h(s)| \mathrm{e}^{(1-c_1)s^2/4 + (1-c_1)s^2/4 - c_2 s^3/3}, \tag{B4}$$

where, for $2c_2 s \leq 1 - c_1$, the function $(1-c_1)s^2/4 - c_2 s^3/3$ is non-decreasing in s, i.e., for $2c_2 t \leq 1 - c_1$ and $1 > c_1$

$$\begin{aligned}
\mathrm{e}^{-t^2/2} \int_0^t \mathrm{d}s\, |h(s)| \mathrm{e}^{s^2/2 + \int_s^t \mathrm{d}u\, |g(u)|} &\leq \mathrm{e}^{-(1-c_1)t^2/4} \int_0^t \mathrm{d}s\, |h(s)| \mathrm{e}^{(1-c_1)s^2/4} \\
&= \frac{2}{1-c_1} \mathrm{e}^{-(1-c_1)t^2/4} \int_0^t \mathrm{d}s\, \frac{|h(s)|}{s} \frac{\mathrm{d}}{\mathrm{d}s} \mathrm{e}^{(1-c_1)s^2/4} \\
&\leq \frac{2}{1-c_1} \mathrm{e}^{-(1-c_1)t^2/4} \int_0^t \mathrm{d}s\, \frac{\mathrm{d}}{\mathrm{d}s} \mathrm{e}^{(1-c_1)s^2/4} \max_{s\in[0,t]} \frac{|h(s)|}{s} \\
&= \frac{2}{1-c_1} (1 - \mathrm{e}^{-\frac{1-c_1}{4} t^2}) \max_{s\in[0,t]} \frac{|h(s)|}{s}.
\end{aligned} \tag{B5}$$

Hence, for $|g(t)| \leq c_1 t + c_2 t^2$ with $c_2 \geq 0$ and $0 \leq c_1 < 1/2$ and t such that $2c_2 t \leq 1 - c_1$, we have

$$\begin{aligned}
|\varphi(t) - \mathrm{e}^{-t^2/2}| &\leq \frac{c_1}{2} \mathrm{e}^{-(1-c_1)t^2/3} t^2 + \frac{c_2}{3} \mathrm{e}^{-(1-c_1)t^2/3} t^3 + \frac{2}{1-c_1} (1 - \mathrm{e}^{-\frac{1-c_1}{4} t^2}) \max_{s\in[0,t]} \frac{|h(s)|}{s} \\
&\leq \frac{c_1}{2} \mathrm{e}^{-t^2/6} t^2 + \frac{c_2}{3} \mathrm{e}^{-t^2/6} t^3 + 4(1 - \mathrm{e}^{-\frac{t^2}{4}}) \max_{s\in[0,t]} \frac{|h(s)|}{s}.
\end{aligned} \tag{B6}$$

Appendix C: Proof of Lemma 10

We start by bounding the support (Section C 2) and the operator norm (Section C 3) of $\hat{Z}^{(n)}(0)$. We do so by expressing it in terms of nested commutators in the following section. We complete the proof of Lemma 10 in Section C 4. For operators $\hat{A}$ and $\hat{B}$, we denote

$$[\hat{A}, \hat{B}]_n = [\hat{A}, [\hat{A}, \hat{B}]_{n-1}], \quad [\hat{A}, \hat{B}]_0 = \hat{B}. \tag{C1}$$

1. Nested commutator form

We have

$$\hat{Z}^{(1)}(t) = -\mathrm{i}\hat{Z}(t)\mathrm{e}^{-\mathrm{i}t\hat{\mathcal{B}}} \left(\mathrm{e}^{-\mathrm{i}t\hat{\mathcal{A}}}\hat{\mathcal{B}}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{A}}} - \hat{\mathcal{B}}\right) \mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}} =: -\mathrm{i}\hat{Z}(t)\mathrm{e}^{-\mathrm{i}t\hat{\mathcal{B}}}\hat{Y}(t)\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}} =: -\mathrm{i}\hat{Z}(t)\hat{X}(t), \tag{C2}$$

where

$$\hat{X}^{(1)}(t) = -\mathrm{i}\mathrm{e}^{-\mathrm{i}t\hat{\mathcal{B}}}[\hat{\mathcal{B}}, \hat{Y}(t)]\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}} + \mathrm{e}^{-\mathrm{i}t\hat{\mathcal{B}}}\hat{Y}^{(1)}(t)\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}} \tag{C3}$$

and, for $n \in \mathbb{N}$, by induction

$$\hat{X}^{(n)}(t) = \sum_{k=0}^{n} \binom{n}{k} (-\mathrm{i})^{n-k} \mathrm{e}^{-\mathrm{i}t\hat{\mathcal{B}}}[\hat{\mathcal{B}}, \hat{Y}^{(k)}(t)]_{n-k}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}}. \tag{C4}$$

Hence, employing the general product rule, we find for $m \in \mathbb{N}$

$$\begin{aligned}\hat{Z}^{(m+1)}(t) &= -\mathrm{i}\sum_{n=0}^{m}\binom{m}{n}\hat{Z}^{(n)}(t)\hat{X}^{(m-n)}(t)\\ &= -\mathrm{i}\sum_{n=0}^{m}\binom{m}{n}\hat{Z}^{(n)}(t)\sum_{k=0}^{m-n}\binom{m-n}{k}(-\mathrm{i})^{m-n-k}\mathrm{e}^{-\mathrm{i}t\hat{\mathcal{B}}}[\hat{\mathcal{B}},\hat{Y}^{(k)}(t)]_{m-n-k}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}}\\ &= -\mathrm{i}\sum_{n=0}^{m}\binom{m}{n}\hat{Z}^{(n)}(t)(-\mathrm{i})^{m-n}\mathrm{e}^{-\mathrm{i}t\hat{\mathcal{B}}}\big[\hat{\mathcal{B}},\big(\mathrm{e}^{-\mathrm{i}t\hat{\mathcal{A}}}\hat{\mathcal{B}}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{A}}}-\hat{\mathcal{B}}\big)\big]_{m-n}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}}\\ &\quad -\mathrm{i}\sum_{n=0}^{m-1}\binom{m}{n}\hat{Z}^{(n)}(t)\sum_{k=1}^{m-n}\binom{m-n}{k}(-\mathrm{i})^{m-n-k}\mathrm{e}^{-\mathrm{i}t\hat{\mathcal{B}}}[\hat{\mathcal{B}},-\mathrm{i}\mathrm{e}^{-\mathrm{i}t\hat{\mathcal{A}}}[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k\mathrm{e}^{\mathrm{i}t\hat{\mathcal{A}}}]_{m-n-k}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}},\end{aligned} \tag{C5}$$

i.e., $\hat{Z}(0)=\mathrm{id}$, $\hat{Z}^{(1)}(0)=0$, and for $m\in\mathbb{N}$, $m>0$,

$$\hat{Z}^{(m+1)}(0) = -\sum_{n=0}^{m-1}\binom{m}{n}\hat{Z}^{(n)}(0)\sum_{k=1}^{m-n}\binom{m-n}{k}(-\mathrm{i})^{m-n-k}[\hat{\mathcal{B}},[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k]_{m-n-k}. \tag{C6}$$

2. Support

We let $\mathcal{C}\subset\mathcal{X}$ and $\hat{\mathcal{C}}=\sum_{i\in\mathcal{C}}\hat{Y}_i$. Let $\hat{S}$ an operator and denote by $\mathcal{S}_n$ the support of $[\hat{\mathcal{C}},\hat{S}]_n$. Then

$$[\hat{\mathcal{C}},\hat{S}]_{n+1} = \sum_{i\in\mathcal{C}}\big[\hat{Y}_i,[\hat{\mathcal{C}},\hat{S}]_n\big] = \sum_{\substack{i\in\mathcal{C}\\ d(i,\mathcal{S}_n)\le R}}\big[\hat{Y}_i,[\hat{\mathcal{C}},\hat{S}]_n\big], \tag{C7}$$

i.e., by induction[5]

$$\mathcal{S}_n \subset \big\{i\in\mathcal{X}\,\big|\,d(i,\mathcal{S}_0)\le 2nR\big\}. \tag{C10}$$

We thus have

$$\begin{aligned}\mathcal{S}_{[\hat{\mathcal{B}},[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k]_{n-k}} &\subset \Big\{i\in\mathcal{X}\,\Big|\,d(i,\mathcal{S}_{[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k})\le 2(n-k)R\Big\},\\ \mathcal{S}_{[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k} = \mathcal{S}_{[\hat{\mathcal{A}},[\hat{\mathcal{A}},\hat{\mathcal{B}}]]_{k-1}} &\subset \big\{i\in\mathcal{X}\,\big|\,d(i,\mathcal{S})\le 2(k-1)R\big\},\end{aligned} \tag{C11}$$

---

[5] We have

$$\mathcal{S}_{n+1}\subset\mathcal{S}_n\bigcup_{\substack{i\in\mathcal{X}\\ d(i,\mathcal{S}_n)\le R}}\mathcal{S}_{\hat{X}_i}\subset\mathcal{S}_n\bigcup_{\substack{i\in\mathcal{X}\\ d(i,\mathcal{S}_n)\le R}}\big\{j\in\mathcal{X}\,\big|\,d(i,j)\le R\big\}. \tag{C8}$$

Now let (C10) hold. Then $d(i,\mathcal{S}_0)\le d(i,s)+d(s,\mathcal{S}_0)\le d(i,s)+2nR$ for all $s\in\mathcal{S}_n$. Hence, $d(i,j)\le R$ and $d(i,\mathcal{S}_n)\le R$ imply $d(j,\mathcal{S}_0)\le d(j,i)+d(i,\mathcal{S}_0)\le R+2nR+d(i,s)$, i.e., with the choice $d(i,\mathcal{S}_n)=d(i,s)$

$$\mathcal{S}_{n+1}\subset\mathcal{S}_n\bigcup_{\substack{i\in\mathcal{X}\\ d(i,\mathcal{S}_n)\le R}}\big\{j\in\mathcal{X}\,\big|\,d(j,\mathcal{S}_0)\le 2(n+1)R\big\}=\big\{j\in\mathcal{X}\,\big|\,d(j,\mathcal{S}_0)\le 2(n+1)R\big\}. \tag{C9}$$

which implies

$$\begin{aligned}\mathcal{S}_{\sum_{k=1}^{n}[\hat{\mathcal{B}},[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k]_{n-k}} &\subset \bigcup_{k=1}^{n} \mathcal{S}_{[\hat{\mathcal{B}},[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k]_{n-k}} \\ &\subset \bigcup_{k=1}^{n} \Big\{ i \in \mathcal{X} \,\Big|\, d(i, \mathcal{S}_{[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k}) \leq 2(n-k)R \Big\} \\ &\subset \big\{ i \in \mathcal{X} \,\big|\, d(i, \mathcal{S}) \leq 2(n-1)R \big\} .\end{aligned} \tag{C12}$$

Hence, the support of $\hat{Z}^{(m+1)}(0)$ is contained in

$$\begin{aligned}\mathcal{S}_{\hat{Z}^{(m+1)}(0)} \subset & \bigcup_{n=0}^{m-1} \mathcal{S}_{\hat{Z}^{(n)}(0)} \cup \big\{ i \in \mathcal{X} \,\big|\, d(i, \mathcal{S}) \leq 2(m-n-1)R \big\} \\ \subset & \big\{ i \in \mathcal{X} \,\big|\, d(i, \mathcal{S}) \leq 2(m-1)R \big\} \cup \bigcup_{n=0}^{m-1} \mathcal{S}_{\hat{Z}^{(n)}(0)} \\ \subset & \cdots \subset \big\{ i \in \mathcal{X} \,\big|\, d(i, \mathcal{S}) \leq 2(m-1)R \big\} .\end{aligned} \tag{C13}$$

3. Operator norm

Consider

$$
\begin{aligned}
&\sum_{i\in\mathcal{X}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots]\\
&\quad= \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)\le 2R}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots] + \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)> 2R}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots]\\
&\quad= \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)\le 2R}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots] + \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)> 2R\\ d(i,i_2)\le 2R}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots]\\
&\qquad\quad+ \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)> 2R\\ d(i,i_2)> 2R}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots]\\
&\quad=\cdots= \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)\le 2R}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots] + \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)> 2R\\ d(i,i_2)\le 2R}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots]\\
&\qquad\quad+ \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)> 2R\\ d(i,i_2)> 2R\\ d(i,i_3)\le 2R}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots] +\cdots+ \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)> 2R\\ d(i,i_2)> 2R\\ \vdots\\ d(i,i_{k-1})> 2R\\ d(i,i_k)\le 2R}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots]\\
&\qquad\quad+ \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)> 2R\\ d(i,i_2)> 2R\\ \vdots\\ d(i,i_k)> 2R}}[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots],
\end{aligned}
\tag{C14}
$$

where the last term is zero, i.e., denoting $\alpha = 2yc_D(2R)^D$, we have

$$
\begin{aligned}
\sum_{i\in\mathcal{X}}\big\|[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots]\big\| &\le \sum_{\substack{i\in\mathcal{X}\\ d(i,i_1)\le 2R}}\big\|[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots]\big\|\\
&\qquad+\cdots+ \sum_{\substack{i\in\mathcal{X}\\ d(i,i_k)\le 2R}}\big\|[\hat{Y}_i,[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots]\big\|\\
&\le \alpha k\big\|[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{k-1}},\hat{Y}_{i_k}]]\cdots]\big\|.
\end{aligned}
\tag{C15}
$$

Hence,

$$\begin{aligned}\big\|[\hat{\mathcal{B}},[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k]_{n-k}\big\| &\leq \sum_{i_1\in\mathcal{X}}\cdots\sum_{i_{n-k}\in\mathcal{X}}\big\|[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{n-k}},[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k]\cdots]]\big\| \\ &\leq \sum_{i_1\in\mathcal{X}}\cdots\sum_{i_{n-k}\in\mathcal{X}}\sum_{j_1\in\mathcal{X}}\cdots\sum_{j_{k-1}\in\mathcal{X}}\sum_{(i,j)\in\mathcal{C}}\big\|[\hat{Y}_{i_1},[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{n-k}},[\hat{Y}_{j_1},[\cdots[\hat{Y}_{j_{k-1}},[\hat{Y}_i,\hat{Y}_j]]\cdots]\big\| \\ &\leq \alpha n\sum_{i_2\in\mathcal{X}}\cdots\sum_{i_{n-k}\in\mathcal{X}}\sum_{j_1\in\mathcal{X}}\cdots\sum_{j_{k-1}\in\mathcal{X}}\sum_{(i,j)\in\mathcal{C}}\big\|[\hat{Y}_{i_2},[\cdots[\hat{Y}_{i_{n-k}},[\hat{Y}_{j_1},[\cdots[\hat{Y}_{j_{k-1}},[\hat{Y}_i,\hat{Y}_j]]\cdots]\big\| \\ &\leq\cdots\leq\alpha^{n-1}n!\sum_{(i,j)\in\mathcal{C}}\|[\hat{Y}_i,\hat{Y}_j]\|.\end{aligned} \tag{C16}$$

Thus, for $m\in\mathbb{N}$, $m>0$,

$$\begin{aligned}\|\hat{Z}^{(m+1)}(0)\| &\leq \sum_{n=0}^{m-1}\binom{m}{n}\|\hat{Z}^{(n)}(0)\|\sum_{k=1}^{m-n}\binom{m-n}{k}\|[\hat{\mathcal{B}},[\hat{\mathcal{A}},\hat{\mathcal{B}}]_k]_{m-n-k}\| \\ &\leq \beta\sum_{n=0}^{m-1}\binom{m}{n}\|\hat{Z}^{(n)}(0)\|\sum_{k=1}^{m-n}\binom{m-n}{k}\alpha^{m-n-1}(m-n)! \\ &= 2^{m+1}(m+1)!\alpha^{m+1}\frac{\beta}{2\alpha^2}\frac{1}{m+1}\sum_{n=0}^{m-1}\frac{\|\hat{Z}^{(n)}(0)\|}{2^n n!\alpha^n}\big(1-2^{n-m}\big),\end{aligned} \tag{C17}$$

i.e.,

$$z_m := \frac{\|\hat{Z}^{(m)}(0)\|}{2^m m!\alpha^m} \leq \max\left\{1,\frac{\beta}{2\alpha^2}\right\}\frac{1}{m}\sum_{n=0}^{m-2}z_n =: \frac{\gamma}{m}\sum_{n=0}^{m-2}z_n, \tag{C18}$$

for which we have $z_0=1$, $z_1=0$, and $z_m\leq\gamma^{\lfloor\frac{m}{2}\rfloor}$ by induction.[6]

4. Final steps

We start by showing that $\hat{Z}(t)=\sum_{n=0}^{\infty}\frac{t^n}{n!}\hat{Z}^{(n)}(0)$. Let $\hat{a}_n$ and $\hat{b}_n$ be sequences of operators such that the limit $\hat{B}=\sum_{n=0}^{\infty}\hat{b}_n$ exists and such that $\sum_{n=0}^{\infty}\|\hat{a}_n\|<\infty$. Then[7]

$$\hat{A}\hat{B}=\sum_{n=0}^{\infty}\sum_{k=0}^{n}\hat{a}_k\hat{b}_{n-k}. \tag{C24}$$

---

[6] Let $m\geq 2$ and $z_n\leq\gamma^{\lfloor\frac{n}{2}\rfloor}$ for $n\leq m-2$. Then

$$z_m\leq\frac{\gamma}{m}\sum_{n=0}^{m-2}\gamma^{\lfloor\frac{n}{2}\rfloor}\leq\gamma\frac{m-1}{m}\gamma^{\lfloor\frac{m-2}{2}\rfloor}\leq\gamma^{\lfloor\frac{m}{2}\rfloor}. \tag{C19}$$

[7] This is basically Merten's theorem on the product of series. It might be obvious for matrices, the proof is the same: Write

$$\hat{A}_m=\sum_{n=0}^{m}\hat{a}_n,\quad \hat{B}_m=\sum_{n=0}^{m}\hat{b}_n,\quad \hat{C}_m=\sum_{n=0}^{m}\sum_{k=0}^{n}\hat{a}_k\hat{b}_{n-k}. \tag{C20}$$

Rearranging terms, one finds

$$\hat{C}_m=\sum_{n=0}^{m}\hat{a}_{m-n}\sum_{k=0}^{n}\hat{b}_k=\sum_{n=0}^{m}\hat{a}_{m-n}(\hat{B}_n-\hat{B})+\hat{A}_m\hat{B}, \tag{C21}$$

Hence,

$$\begin{aligned}\hat{Z}(t) &= \mathrm{e}^{-\mathrm{i}t(\hat{\mathcal{A}}+\hat{\mathcal{B}})}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{A}}}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}} \\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{n}\sum_{l=0}^{k}\frac{(-\mathrm{i}t)^l}{l!}(\hat{\mathcal{A}}+\hat{\mathcal{B}})^l\frac{(\mathrm{i}t)^{k-l}}{(k-l)!}\hat{A}^{k-l}\frac{(\mathrm{i}t)^{n-k}}{(n-k)!}\hat{B}^{n-k} \\ &= \sum_{n=0}^{\infty}\frac{t^n}{n!}\sum_{k=0}^{n}\sum_{l=0}^{k}\binom{n}{k}\binom{k}{l}[-\mathrm{i}(\hat{\mathcal{A}}+\hat{\mathcal{B}})]^l(\mathrm{i}\hat{A})^{k-l}(\mathrm{i}\hat{B})^{n-k} \\ &=: \sum_{n=0}^{\infty}\frac{t^n}{n!}\hat{Y}_n.\end{aligned} \tag{C25}$$

Now, by the general Leibniz rule, we have

$$\begin{aligned}\hat{Z}^{(n)}(t) &= \sum_{k=0}^{n}\binom{n}{k}\left(\mathrm{e}^{-\mathrm{i}t(\hat{\mathcal{A}}+\hat{\mathcal{B}})}\mathrm{e}^{\mathrm{i}t\hat{\mathcal{A}}}\right)^{(k)}\left(\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}}\right)^{(n-k)} \\ &= \sum_{k=0}^{n}\sum_{l=0}^{k}\binom{n}{k}\binom{k}{l}\left(\mathrm{e}^{-\mathrm{i}t(\hat{\mathcal{A}}+\hat{\mathcal{B}})}\right)^{(l)}\left(\mathrm{e}^{\mathrm{i}t\hat{\mathcal{A}}}\right)^{(k-l)}\left(\mathrm{e}^{\mathrm{i}t\hat{\mathcal{B}}}\right)^{(n-k)},\end{aligned} \tag{C26}$$

i.e., $\hat{Y}_n = \hat{Z}^{(n)}(0)$ and therefore $\hat{Z}(t) = \sum_{n=0}^{\infty}\frac{t^n}{n!}\hat{Z}^{(n)}(0)$. Now let $\tau = 2t\alpha\sqrt{\gamma} \le 1/2$. Then, using the bound derived in the previous section,

$$\left\|\hat{Z}(t) - \sum_{n=0}^{M}\frac{t^n}{n!}\hat{Z}^{(n)}(0)\right\| \le \sum_{n=M+1}^{\infty}\frac{t^n}{n!}\|\hat{Z}^{(n)}(0)\| \le \sum_{n=M+1}^{\infty}\tau^n = \frac{\tau^{M+1}}{1-\tau} \le 2\tau^{M+1} \tag{C27}$$

and for $m \le M$

$$\begin{aligned}\left\|\Big(\sum_{n=0}^{M}\frac{t^n}{n!}\hat{Z}^{(n)}(0)\Big)^{(m)}\right\| &\le \sum_{n=m}^{M}\frac{t^{n-m}}{(n-m)!}\|\hat{Z}^{(n)}(0)\| \le \sum_{n=m}^{M}\frac{n!}{(n-m)!}2^n\alpha^n\gamma^{n/2}t^{n-m} \\ &= m!(2\alpha\gamma^{1/2})^m\sum_{n=m}^{M}\binom{n}{m}\tau^{n-m} \le m!(2\alpha\gamma^{1/2})^m\frac{1}{(1-\tau)^{m+1}} \le 2m!(4\alpha\gamma^{1/2})^m.\end{aligned} \tag{C28}$$

---

i.e.,

$$\|\hat{C}_m - \hat{A}\hat{B}\| \le \sum_{n=0}^{m}\|\hat{a}_{m-n}\|\|\hat{B}_n - \hat{B}\| + \|\hat{A}_m - \hat{A}\|\|\hat{B}\|. \tag{C22}$$

Let $\epsilon > 0$. As $\sum_{n=0}^{\infty}\|\hat{a}_n\| < \infty$ and $\lim_{n\to\infty}\|\hat{B}_n - \hat{B}\| = 0$, there is a $N \in \mathbb{N}$ such that $\|\hat{B}_n - \hat{B}\| \le \frac{\epsilon}{3}(1 + \sum_{n=0}^{\infty}\|\hat{a}_n\|)^{-1}$ for all $n \ge N$. Further, as $\lim_{n\to}\|a_n\| = 0$, there is a $M \in \mathbb{N}$ such that $\|a_n\| \le \frac{\epsilon}{3N}(1 + \max_{0\le k\le N-1}\|\hat{B}_k - \hat{B}\|)^{-1}$ for all $n \ge M$. Finally, as $\lim_{n\to\infty}\|\hat{A}_n - \hat{A}\| = 0$, there is a $L \in \mathbb{N}$ such that $\|\hat{A}_m - \hat{A}\| \le \frac{\epsilon}{3}(1 + \|\hat{B}\|)^{-1}$ for all $m \ge L$. Hence, for $m \ge \max\{N+M, L\}$

$$\begin{aligned}\|\hat{C}_m - \hat{A}\hat{B}\| &\le \sum_{n=0}^{N-1}\|\hat{a}_{m-n}\|\|\hat{B}_n - \hat{B}\| + \sum_{n=N}^{m}\|\hat{a}_{m-n}\|\|\hat{B}_n - \hat{B}\| + \|\hat{A}_m - \hat{A}\|\|\hat{B}\| \\ &\le \frac{\epsilon\sum_{n=0}^{N-1}\|\hat{B}_n - \hat{B}\|}{3N(1 + \max_{0\le k\le N-1}\|\hat{B}_k - \hat{B}\|)} + \frac{\epsilon\sum_{n=N}^{m}\|\hat{a}_{m-n}\|}{3 + 3\sum_{n=0}^{\infty}\|\hat{a}_n\|} + \frac{\epsilon\|\hat{B}\|}{3(1 + \|\hat{B}\|)} \le \epsilon.\end{aligned} \tag{C23}$$