technische universität
dortmund

Habilitationsschrift

Concentration inequalities in random Schrödinger operators

Christoph Schuhmacher

Preprint 2020-01 January 2020

Fakultät für Mathematik
Technische Universität Dortmund
Vogelpothsweg 87
44227 Dortmund

tu-dortmund.de/MathPreprints

Concentration inequalities in random Schrödinger operators

Habilitationsschrift

Der Fakultät für Mathematik
der Technischen Universität Dortmund
vorgelegt von

Dr. Christoph Schumacher

21. Januar 2019

Contents

Acknowledgments

First of all, I want to thank Ivan Veselić who supported me as teacher, mentor, coauthor, and employer for years. This work would not have been possible without him. Of course, I am indebted to my PhD advisor Andreas Knauf, who shaped me as mathematician and as person. I am also grateful to my further coauthors Fabian Schwarzenberger, Francisco Hoecker-Escuti, Sylvain Golénia, Thomas Kalmes, and Christoph Richard for sharing their knowledge and their time with me.

I am glad to thank the current and former members of the research groups I was privileged to be a part of, too. In particular my thanks go to Michela Egidi, Albrecht Seelmann, Matthias Täufer, Martin Tautenhahn, Christian Seifert, Max Kämper, Alexander Dicke, Rainer Brück, and Janine Textor for fruitful and joyous discussions and relaxing tee breaks.

Finally, I am endlessly grateful to Elisabeth Fischer for her seemingly unlimited love and support in every aspect of life. I am glad and excited about the prospect of sharing our future together and with our son Laurin.

Part I

Introduction

Chapter 1

Random Schrödinger operators

In statistical physics and in particular in solid state physics, one is interested in the behavior and properties of matter which consists of many atoms or molecules. The challenge is to derive macroscopic properties of a material like electrical conductivity or insulation from the interactions of its microscopic constituents. The proper description on the microscopic level is provided by quantum mechanics.

A moderately sized system involves 10^{23} and more particles, which have to be modeled on the configuration space $(\mathbb{R}^3)^{10^{23}}$. Naturally, one tries to simplify this model. From the point of view of one electron, the rest of the solid can be modeled as an external potential. This approach reduces the configuration space to $\mathbb{R}^3$, but neglects the influence of the electron on the crystal, which is a reasonable simplification. The external potential captures the influence of the nuclei and the remaining electrons on the electron and is often called background potential. We will also assume that the material does not change over time, i. e., the background potential will be constant in time. This simplification is justified by the fact that the nuclei are much heavier than the electron and accordingly move much slower.

Many solids are crystals, where the atoms are arranged in a periodic lattice structure, so periodic background potentials seem to be a good choice. To take full advantage of the lattice structure, the crystal is assumed to extend to infinity. The study of wave phenomena in such periodic potentials is called Bloch–Floquet theory. A fundamental result is that the energy the electron can have is restricted to certain intervals, called bands. This band structure allows to explain many physical properties of solids, including electrical conductivity of metals and even optical properties.

Of course not all physical effects are captured by this model. An example is superconductivity, which specifically needs the interactions within pairs of electrons. For many materials, the assumption of perfect periodicity is an oversimplification, too. In most crystals, the periodic arrangement of atoms is only a local property, the regular regions are separated by grain boundaries. And even in supposedly periodic domains, there are irregularities in form of impurities, foreign atoms or missing atoms.

To study the effects which are suppressed by ignoring the non-periodicity, Philip Warren Anderson introduced the so-called *Anderson model* in [And58]. The formerly periodic background potential is now perturbed by random impurities. The distribution of the impurities is chosen homogeneously to ensure that the material remains homogeneous. The task is now again to study the motion of one electron in this random potential. If we interpret the impurities as foreign atoms, we can model a mixture of different metals and thus alloys, too. Accordingly, the model is also known as *alloy-type model*. Anderson argued that enough randomness changes the electrical behavior of the matter under consideration from conductor to insulator. This phase transition caught the attention of many researchers in physics and in mathematics and lead to numerous publications in the field of random Schrödinger operators.

To continue the discussion on a slightly more technical level, we briefly review some standard notions and notation from quantum mechanics. A quantum particle is represented by its *wave function* φ, also called the *state* of the particle, which is a normalized vector in the Hilbert space $L^2(\mathbb{R}^3)$ with inner product $\langle\varphi_1, \varphi_2\rangle := \int \overline{\varphi_1(x)}\varphi_2(x)\,\mathrm{d}x$. For each measurable set $A \subseteq \mathbb{R}^3$, $\|\varphi \mathbf{1}_A\|^2 = \int_A |\varphi(x)|^2\,\mathrm{d}x$ is the probability of the particle to be found in the region A. Note that due to the normalization, the probability for the particle to be somewhere in space is $\int_{\mathbb{R}^3} |\varphi(x)|^2\,\mathrm{d}x = \|\varphi\|_2^2 = 1$. The state of the particle will change over time, so it is a function $\psi\colon \mathbb{R} \to L^2(\mathbb{R}^3)$. Given an initial state ψ_0 at time 0, the time evolution of the particle is governed by *Schrödinger's equation*

$$\mathrm{i}\dot{\psi} = H\psi, \quad \psi(0) = \psi_0,$$

where $\dot{\psi}$ is the time derivative of ψ and H is the *Hamiltonian*. We use physical units to suppress Planck's constant and the mass of the particle.

Mathematically, a Hamiltonian is a self-adjoint operator on $L^2(\mathbb{R}^3)$, and this immediately implies that the solution operator to Schrödinger's equation is a unitary and thus preserves the normalization of states. Physically, the Hamiltonian is the observable for the total energy of the particle. This means that the expectation value of the energy of a particle in the state φ is $\langle\varphi, H\varphi\rangle$. The total energy is the sum of the kinetic energy T and the potential energy V, i. e. $H = T + V$. We do not take spin and magnetic fields into account, so the kinetic energy is the Laplace operator $T = -\Delta$. The potential energy is a multiplication operator with the potential

$$V\colon L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3), \quad (V\varphi)(x) := V(x)\varphi(x).$$

As custom, we use the same symbol for the potential as a function of space and the potential as an operator on $L^2(\mathbb{R}^3)$. The domain of the multiplication operator has of course to be restricted if the potential is unbounded.

This is the point where the modeling discussed above enters the formalism. Let the contribution of one atom located at the origin to the potential be encoded in the *single site potential* $f \in L^2(\mathbb{R}^3)$. Then an infinite crystal with atoms of this type at every site of the lattice $\mathbb{Z}^3$ has the $\mathbb{Z}^3$-periodic potential

$$V_{\mathrm{per}}(x) := \sum_{k \in \mathbb{Z}^3} f(x-k).$$

Of course, we assume that this sum converges almost everywhere. In the *alloy-type model*, each lattice site is assigned an atom randomly:

$$V_\omega(x) := \sum_{k \in \mathbb{Z}^3} \lambda_k(\omega) f(x-k).$$

Here, f is the shape of an atom, but the charge of the nucleus is changed by the independent and identically distributed random coefficient $\lambda_k(\omega)$. There are many other random potentials considered in the literature. A Hamiltonian which is the sum of the Laplace operator and a random potential is called a *random Schrödinger operator*.

Formally, Schrödinger's equation is solved by

$$\psi(t) = \mathrm{e}^{-\mathrm{i}Ht}\psi_0.$$

But apart from the fact that $\mathrm{e}^{-\mathrm{i}Ht}$ is unitary, it is very hard to actually determine the behavior of the solution, especially for long times. The *RAGE theorem*, named after Ruelle [Rue69], Amrein and Georgescu [AG73], and Enss [Ens78], states that the long time behavior of solutions to Schrödinger's equation is intimately related to the spectrum of the Hamiltonian, see e. g. [Cyc+87]. More precisely, absolutely continuous spectrum corresponds to scattering states, while point spectrum indicates localized eigenstates. Thus, roughly speaking, absolutely continuous spectrum corresponds to a material that allows the electron to travel, a behavior which makes the material an electrical conductor, while localized eigenstates trap the electron and make the material an insulator. There are several precise notions of *Anderson localization* but the intricate details are beyond the scope of this introduction.

According to the RAGE theorem, it is vital to study the spectrum of random Schrödinger operators. For periodic potentials, the spectrum consists of intervals of absolutely continuous spectrum. Therefore, if a periodic crystal provides electrons in such a band, the material is a conductor, which is the case for metals. For random Schrödinger operators like the alloy-type model, the situation is more complicated.

One might expect that the spectrum of a random operator is random. But for the Anderson model and in fact many random Schrödinger operators, spatial homogeneity makes the spectrum almost surely constant, meaning that there is a closed set $\Sigma \subseteq \mathbb{R}$ such that, for almost all realizations of the random potential, this set Σ is the spectrum of the corresponding Schrödinger operator. From an ergodic theory point of view, this is understandable as follows. The homogeneity of the material makes the operator family ergodic. In the alloy-type model, the choice of i. i. d. random variables λ_k is the reason for this. Now, similar to the fact that invariant $\mathbb{R}$-valued random variables on an ergodic dynamical system are almost surely constant, the invariance of the spectrum, i. e. a set-valued random variable, under translations guarantees that it is deterministic, too. By the same reasoning, the components of the spectral measure according to the Lebesgue decomposition are deterministic, too.

To study the spectrum and its types, one needs efficient tools. One such tool is the *integrated density of states* (IDS). The IDS is a function that assigns each energy threshold the number of quantum mechanical states per unit volume with energy below its argument. To fill this description with life, we describe one common way to construct the IDS rigorously. First, we fix a large rectangular box in the configuration space and define the operator $H_\omega^L := -\Delta + V_\omega$ on (a domain in) $L^2(\Lambda_L)$ using suitable boundary conditions, for example Dirichlet or Neumann boundary conditions, such that H_ω^L is a self-adjoint operator. We call the operator H_ω^L the *restriction* of the operator H_ω to the box Λ_L with Dirichlet or Neumand boundary conditions and use the notation $H_\omega^L := H_\omega|_{\Lambda_L}$. We also refer to H_ω^L as *finite volume operator* or *finite volume approximation* of H_ω. Next, it is well-known that the finite volume operator has discrete spectrum consisting purely of eigenvalues. We count the eigenvalues below a given energy threshold with multiplicity, and the resulting number is the value of the *eigenvalue counting function* corresponding to this finite volume operator. Then, we normalize the eigenvalue counting function with the volume of the box we restricted the original Hamiltonian to. Finally, we take the limit $L \to \infty$ of box size to infinity. This procedure gives a limiting function, which is the IDS, and illustrates the interpretation of the IDS given above.

By construction, the normalized eigenvalue counting functions are monotonically increasing, and the IDS is too. Its Stieltjes derivative is a positive measure, called the *density of states measure* (DoS). The topological support of the DoS is the spectrum of the Hamiltonian, so the IDS encodes spectral information, even though it cannot distinguish between the spectral types.

A natural question to ask is whether or not the IDS can distinguish between random and non-random operators. For a deterministic Schrödinger operator like the Laplace operator without a potential or with a periodic potential, the IDS usually behaves like a polynomial close to the infimum of the spectrum. For example, straightforward Fourier analysis shows that the IDS of the Laplace operator $-\Delta$

on $\mathbb{R}^d$ is

$$\mathbb{R} \ni E \mapsto \frac{\tau_d}{(2\pi)^d} E^{d/2} \mathbf{1}_{(0,\infty)}(E),$$

where τ_d is the volume of the unit ball in $\mathbb{R}^d$, see [RS78, p. XIII.15].

In contrast, for random Schrödinger operators like the Anderson model, the IDS behaves radically different at the infimum of the spectrum. Typically, the IDS of a random Schrödinger operator with sufficiently strong randomness is, close to the bottom of the spectrum, exponentially small. This phenomenon is called *Lifshitz tails* in honor of [Lif65].

To understand this change of behavior of the IDS at the bottom of the spectrum, we return to the definition of the IDS. The first step in order to count eigenvalues is to restrict the Hamiltonian to a large but finite box with, say, Neumann boundary conditions. For Lifshitz tails, we are interested in small energies. We study the ground state of the random Schrödinger operator $H_\omega^L = (-\Delta + V_\omega)|_{\Lambda_L}$ restricted to the box $\Lambda_L := [0, L)^d$ of side length L with the minimax principle:

$$E_1(H_\omega^L) = \inf_{\|\varphi\|_2=1} \langle \varphi, H_\omega^L \varphi \rangle = \inf_{\|\varphi\|_2=1} \big(\langle \varphi, -\Delta\varphi \rangle + \langle \varphi, V_\omega \varphi \rangle \big),$$

where φ ranges over all normalized functions in the domain of the Laplace operator which are supported in Λ_L. For the sum to be small, both the expected kinetic and potential energy have to be small. In order for the kinetic energy $\langle \varphi, -\Delta\varphi \rangle$ to be small, φ has to be close to the ground state of $-\Delta$. Let us try the ground state $\varphi = |\Lambda_L|^{-1/2} \mathbf{1}_{\Lambda_L}$ as a test function. If the support of the single site potential f is contained in Λ_1, the potential energy is

$$\langle \varphi, V_\omega \varphi \rangle = \frac{1}{|\Lambda_L|} \int_{\Lambda_L} V_\omega(x)\, \mathrm{d}x = \frac{1}{|\Lambda_L|} \sum_{k \in \Lambda_L \cap \mathbb{Z}^d} \lambda_k(\omega) \int f(x)\, \mathrm{d}x, \qquad (1.1)$$

which is the average of the potential over the box Λ_L. Now we isolated the cause of the phenomenon of Lifshitz tails. Since, for averages of i. i. d. random variables, the probability concentrates around the expectation, the probability of the average to be small, in particular smaller than its expectation, is exponentially small. Morally, this is the reason why states with small energy are exponentially rare, and this translates to Lifshitz tails of the IDS.

When we add a non-negative potential, we shift the spectrum upwards, and the IDS decreases. In order to prove Lifshitz tails at the appropriate location, we have to identify the bottom of the spectrum. In the i. i. d. alloy-type model, a sufficient condition on the random variables λ_k in order to not move the infimum of the spectrum is that the topological support of the law of λ_k contains 0. This condition can be rephrased as follows: each neighborhood of 0 contains λ_k with positive probability. With the help of independence and the second Borel–Cantelli lemma, one finds a sequence of increasing boxes in configuration space with very low potential. A Weyl sequence argument on these boxes shows that the infimum of the spectrum remains unperturbed, i. e. $\inf \sigma(H_\omega) = 0$ almost surely.

This qualitative argument can be strengthened quantitatively. In fact, one can give a lower bound on the IDS matching the upper bound on a logarithmic scale. Of course, quantitative conditions on the random variables are needed. A typical result states that if the probability of λ_k being less than a threshold $\varepsilon > 0$ is at least polynomial in ε, then the limit

$$\nu := \lim_{E \searrow E_0} \frac{\log|\log N(E)|}{\log(E - E_0)}$$

exists almost surely and is strictly negative. This means in a sense that $N(E)$ behaves like $\exp(-(E - E_0)^\nu)$ as $E \searrow E_0$.

In most of this thesis, the quantum mechanic setup is simplified once more using the tight-binding approximation. There, instead of the configuration space $\mathbb{R}^3$, one considers the discrete lattices $\mathbb{Z}^d$, $d \geq 2$, or more general discrete groups, see Chapter 2. Let us explain the quantum mechanical construction for $\mathbb{Z}^d$ here. The wave functions will be normalized vectors in the Hilbert space $\ell^2(\mathbb{Z}^d)$. The observable for the kinetic energy is the discrete Laplace operator $-\Delta\colon \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$ given by

$$(-\Delta\varphi)(z) := \sum_{y \in \mathbb{Z}^d, \|y-z\|_1 = 1} \big(f(z) - f(y)\big).$$

This bounded operator mimics the negative sum of second derivatives on $\mathbb{R}^d$ because

$$(\Delta\varphi)(z) = \sum_{j=1}^{d} \Big(\frac{f(z + e_j) - f(z)}{1} - \frac{f(z) - f(z - e_j)}{1}\Big)/1,$$

where the standard basis vectors $e_1, \ldots, e_d$.

Potentials are functions $V\colon \mathbb{Z}^d \to \mathbb{R}$, and the corresponding observables their multiplication operators on $\ell^2(\mathbb{Z}^d)$. The discrete alloy-type potential is, as before, composed of a single site potential $f\colon \mathbb{Z}^d \to \mathbb{R}$ and independent and identically distributed random variables λ_k, $k \in \mathbb{Z}^d$, via

$$V_\omega(x) := \sum_{k \in \mathbb{Z}^d} \lambda_k(\omega) f(x - k)$$

The Schrödinger operator $H = -\Delta + V$ is, as before, the sum of the kinetic and the potential energy. The IDS is obtained as the limit of the normalized eigenvalue counting functions of H restricted to finite boxes $\Lambda_L := [0, L)^d \cap \mathbb{Z}^d$, $L \in \mathbb{N}$. The restricted operator $H|_{\Lambda_L}$ acts on the finite dimensional Hilbert space

$\ell^2(\Lambda_L)$. We usually use the basis $(\delta_v)_{v\in\Lambda_L}$ of Kronecker delta functions to describe the finite volume operators by their representing matrices. The total number of eigenvalues is the volume of Λ_L with respect to the counting measure. Thus, in the discrete setting, all normalized eigenvalue counting functions map into the interval $[0,1]$, and so does the IDS. These effects and the fact that H is bounded are the main technical advantages of the discrete model with respect to the continuous one. Luckily, the physical phenomena we are interested in also appear in discrete models. The interested reader can consult e. g. [Kir07] for more details on random Schrödinger operators including a discussion of boundary conditions for restrictions of discrete operators.

There is another way to express the IDS which is particularly simple in the discrete setting. Consider the *spectral distribution function* (SDF), which is defined as follows. Given a random Schrödinger operator H_ω, denote by $\mathbf{1}_{(-\infty,E]}(H_\omega)$ its spectral projection on the interval $(-\infty, E]$. The SDF is then the function $\mathcal{N}\colon \mathbb{R} \to \mathbb{R}$,

$$\mathcal{N}(E) := \mathbb{E}[\langle \delta_0, \mathbf{1}_{(-\infty,E]}(H_\omega)\delta_0\rangle],$$

where $\delta_0 \in \ell^2(\mathbb{Z}^d)$ is the Kronecker delta function on 0. There is an analogue formula for the continuous setting, too, see e. g. Chapter 8. The famous *Pastur–Shubin formula* states that the IDS equals the SDF, see [Pas71; Shu79; PF92], and connects the infinite operator and its finite volume approximations. Note that the SDF does not rely on the choices made in the construction of the IDS like the boundary conditions for the finite volume approximations or the sequence of bounded cubes ($[-L, L]$ or $[0, L]$), but only on the operator H_ω. When we generalize the setting in Chapter 6 to more general groups than $\mathbb{Z}^d$, the Pastur–Shubin formula serves as a test that we found the correct function.

Properties and the behavior of the IDS at the bottom of the spectrum have also been studied for discrete models. For Laplace operators on discrete graphs see e. g. [PT18], for a comparison of periodic and random operators with Lifshitz tails see [AV09].

Chapter 2

Groups

One key feature of the configuration space $\mathbb{R}^3$ of quantum mechanics is that $\mathbb{R}^3$ is also a group and acts on itself by translations. The laws of physics are invariant under translations, see e. g. [APP18], a fact that leads to conservation of momentum and the theory of relativity. We also use the group structure of $\mathbb{R}^3$, or rather the group action of its subgroup $\mathbb{Z}^3$ on $\mathbb{R}^3$ in the construction of periodic and random potentials. Under this point of view, the simplification to $\mathbb{Z}^3$, or more generally $\mathbb{Z}^d$, introduced at the end of the previous section, is quite natural.

The underlying geometry has of course some impact on the physical systems built on it. From a physical point of view, $\mathbb{Z}^d$ with $d \in \{1, 2, 3\}$ are the most relevant configuration spaces. On the other hand, many results of ergodic theory have been generalized to more general group actions. It is valuable to study if and how results for $\mathbb{Z}^d$ carry over to different geometries and to see which features of $\mathbb{Z}^d$ are responsible for which physical properties. Here is how one has to modify the basic definitions, in particular the Laplace operator, on finitely generated groups. A group G is *finitely generated* if there is a finite subset $S \subseteq G$ such that all $g \in G$ can be expressed as a product of elements of S. See Figure 2.1 for an illustration with $G = \mathbb{Z}^2$. Note that it is common practice to denote the group operation of an abelian group as sum and the group operation of a non-abelian group as product.

Usually, we assume without loss of generality that S is symmetric with respect to the group inversion: $S = S^{-1} := \{s^{-1} \mid s \in S\}$, and that S does not contain the identity element $\mathrm{id} \in G$. The *Cayley graph* $\Gamma(G, S)$ of a finitely generated group G with respect to a symmetric set of generators S has as a vertex set the group G itself, and two vertices $g, h \in G$ are connected with an edge if and only if $gh^{-1} \in S$, or equivalently, if there is an $s \in S$ such that $g = sh$. The edge (g, h) is then *labeled* with the generator $s = gh^{-1} \in S$. Consequently, the set of neighbors of g is $S^{-1}g = Sg$. Note that Cayley graphs are *regular*, which means that every node has the same *degree*, i. e. the same number of neighbors. Cayley graphs are directed graphs, which is necessary for the labeling. But since S is symmetric, all directed

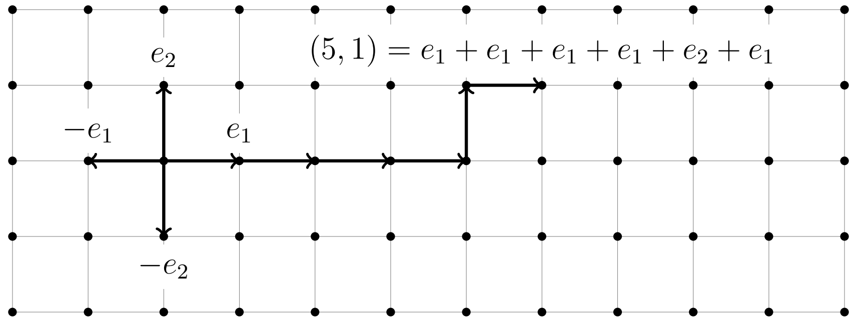


Figure 2.1: $\mathbb{Z}^2$ is generated by $\{(1,0),(0,1),(-1,0),(0,-1)\}$

edges occur in pairs with opposing orientation, which can be easily visualized as undirected edges, if we decide to ignore the labels in an appropriate context.

Each undirected graph comes with a metric on its set of vertices, which measures the length of the shortest path in the graph between points, counting each edge with length one. In the case of Cayley graphs with respect to symmetric generating sets, this metric is called *word metric*, because the distance between $g,h \in G$ is the length of the shortest word representing gh^{-1}:

$$d^S(g,h) = \inf\Big\{r \in \mathbb{N}_0 \mid \exists s_1,\dots,s_r \in S\colon gh^{-1} = \prod\nolimits_{j=1}^{r} s_j\Big\}.$$

We also denote $|g|_S := d^S(\mathrm{id},g)$.

As an example for a Cayley graph, consider $G = \mathbb{Z}^d$ and $S = \{\pm e_1,\dots,\pm e_d\}$, where e_j is the j-th standard basis unit vector, which is indicated in Figure 2.1 for $d = 2$. Figure 2.2 shows a finite part of the Cayley graph of $\mathbb{Z}^2$ with respect to the generator $\{\pm e_1, \pm e_2, \pm(e_1 - e_2)\}$ with directed edges.

Regular tree graphs, see Figure 2.3, are examples for Cayley graphs, too. If the degree $d = 2k$ of the tree is even, then it is the Cayley graph of the free group F_k generated by the symmetric set $S = \{a_1^{\pm 1},\dots,a_k^{\pm 1}\}$, see Figure 2.4. For odd degree $d = 2k-1$, one can use the group generated by k generators $a_1,\dots,a_k$ (and their inverses), but subject to the relation $a_k = a_k^{-1}$. In the resulting Cayley graph, the two branches labeled with a_k and a_k^{-1} are identified, so the degree is $2k-1 = d$. Another possible construction is as follows. Consider the group generated by d generators $a_1,\dots,a_d$ subject to the relations $a_j^2 = \mathrm{id}$ for all $j \in \{1,\dots,d\}$. This last approach actually works for even and odd degree d.

There are many definitions of Laplace operators on graphs, but for regular graphs like Cayley graphs, most of them are equivalent to each other. We will focus on the following straightforward generalization of the previous definition on $\ell^2(\mathbb{Z}^d)$.

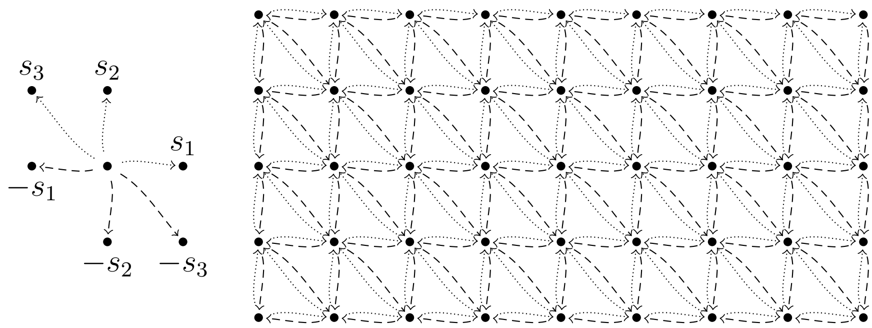


Figure 2.2: The Cayley graph of $\mathbb{Z}^2$ with respect to the generating set $\{s_1, s_2, s_3, -s_1, -s_2, -s_3\}$ with $s_1 = (1, 0)$, $s_2 = (0, 1)$, $s_3 = (-1, 1)$. The labels of the edges are indicated by the arrows: horizontally and dotted is the label s_1, while diagonally and dashed is labeled $-s_3$, for example.

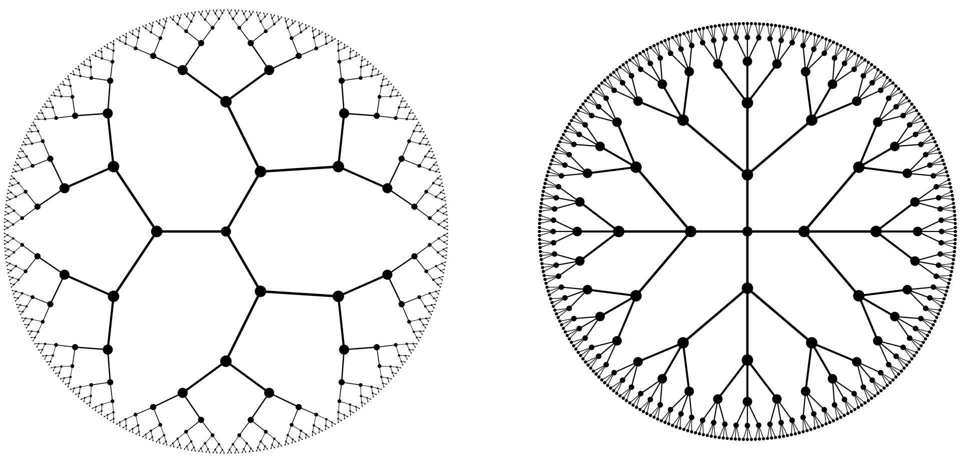


Figure 2.3: Regular trees with degree $d \in \{3, 4\}$.

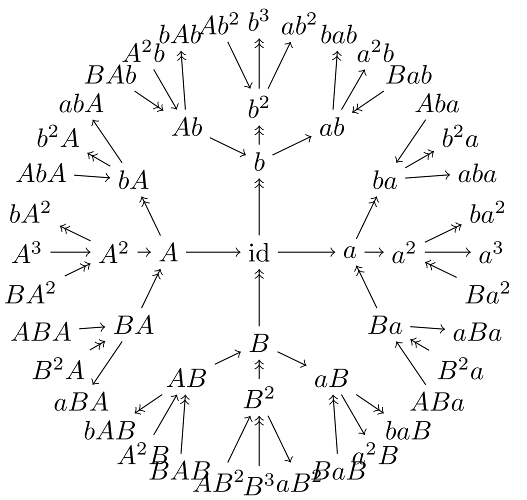


Figure 2.4: The Cayley graph of the ball B_3 of radius 3 in F_2, with $a := a_1$, $b := a_2$, $A := a^{-1}$ and $B := b^{-1}$. The arrows indicate the corresponding generator: $x \to y$ means $ax = y$, $x \twoheadrightarrow y$ is synonymous for $bx = y$.

For a finitely generated group G with symmetric generator S, we will call

$$\Delta\colon \ell^2(G) \to \ell^2(G), \quad (\Delta f)g := (\Delta_S f)g := \sum_{h \in Sg} \big(f(h) - f(g)\big)$$

the *Laplace operator* on $\ell^2(G)$ or on G, for short. Of course, it depends on S, too, but we often suppress S in the notation. It is now straightforward to generalize the alloy-type model and the SDF to finitely generated groups. For the IDS, we also need ergodicity and an analogue for cubes.

The group action of G on itself from the right, i. e.,

$$G \times G \to G, \quad (g, h) \mapsto hg^{-1},$$

extends to the Cayley graph: $G \times \Gamma(G, S) \to \Gamma(G, S)$. On vertices, the action is $(g, h) \mapsto hg^{-1}$, while on edges, it is $(g, (h, h')) \mapsto (hg^{-1}, h'g^{-1})$. Note that for all $g, h, h' \in G$, we have $(h'g^{-1})(hg^{-1})^{-1} = h'h^{-1}$, so edges map to edges. Even more, the group G acts transitively on the vertices and on the labeled edges of its Cayley graph, and all graph isomorphisms which preserve the labels stem from the group action. Note also, the Laplace operator commutes with the group action, which can be of great help to understand the spectrum of the Laplace operator.

Ergodic theory provides a large toolbox to deal with group actions. Especially amenable groups are well understood. A group G is *amenable* if it contains finite subsets with arbitrarily small boundary to volume ratio. In a discrete group, there

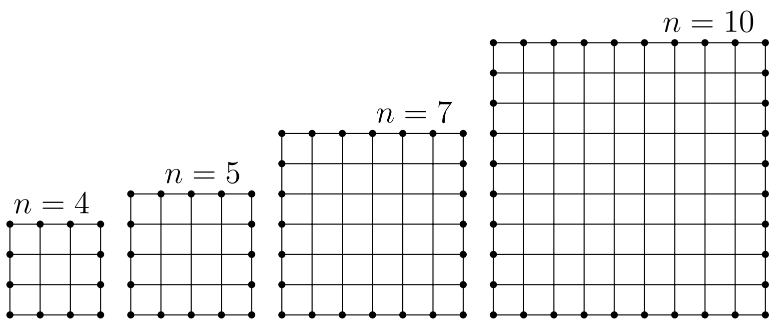


Figure 2.5: In $\mathbb{Z}^d$, the interior S-boundary of a box F_n of side length n satisfies $|\partial^S_{\mathrm{int}} F_n| = 4(n-1)$. The boxes form a Følner sequence: $\frac{|\partial^S_{\mathrm{int}} F_n|}{|F_n|} = \frac{4(n-1)}{n^2} \xrightarrow{n\to\infty} 0$.

are different notions of boundary. For a finite set $K \subseteq G$, the *K-boundary* of a set $F \subseteq G$ is the union of its *interior* and its *exterior K-boundary*:

$$\partial^K_{\mathrm{int}} F := \bigcup\nolimits_{g\in K} \big(F \setminus (gF)\big), \quad \partial^K_{\mathrm{ext}} F := \bigcup_{g\in K} \big((gF) \setminus F\big) = (KF) \setminus F,$$
$$\partial^K F := (\partial^K_{\mathrm{int}} F) \cup (\partial^K_{\mathrm{ext}} F).$$

A *Følner sequence* is a sequence of non-empty finite subsets $F_n \subseteq G$, $n \in \mathbb{N}$, such that

$$\lim_{n\to\infty} \frac{|\partial^S_{\mathrm{int}} F_n|}{|F_n|} = 0$$

where S is a finite generating set. A group is amenable if and only if it contains a Følner sequence, see [Føl55]. We will see soon that whether or not a sequence is Følner is independent of the choice of the generating set S. The Euclidean lattices $\mathbb{Z}^d$ are examples of amenable groups. The cubes $F_n := ([0,n) \cap \mathbb{Z})^d$ can serve as a Følner sequence, see Figure 2.5.

For all finite $K \subseteq G$, the K-boundaries along Følner sequence $(F_n)_n$ get small:

$$\lim_{n\to\infty} \frac{|\partial^K F_n|}{|F_n|} = 0. \tag{2.1}$$

In particular, this shows that amenability is independent of the choice of the symmetric finite generating set S. To see (2.1), note that $A \setminus C \subseteq (A\setminus B) \cup (B \setminus C)$ for arbitrary sets A, B, C, so that, by induction, $(gF)\setminus F \subseteq \bigcup_{k=1}^{|g|_S} \big(\big(\prod_{j=1}^{k-1} s_j\big)((s_k F)\setminus F)\big)$ for all $g = \prod_{j=1}^{|g|_S} s_j \in K$ with $s_j \in S$. We estimate for $F \subseteq G$

$$|(gF) \setminus F| \leq \sum_{k=1}^{|g|_S} |F \setminus (s_k^{-1} F)| \leq |g|_S |\partial^S_{\mathrm{int}} F|,$$

and analogously

$$|F \setminus (gF)| \leq \sum_{k=1}^{|g|_S} \left| \left(\prod\nolimits_{j=1}^{k-1} s_j\right)(F \setminus (s_k F)) \right| \leq \sum_{k=1}^{|g|_S} |F \setminus (s_k F)| \leq |g|_S |\partial^S_{\mathrm{int}} F|.$$

Combined, we get

$$|\partial^K F| \leq \sum_{g \in K} \big(|F \setminus (gF)| + |(gF) \setminus F|\big) \leq 2|\partial^S_{\mathrm{int}} F| \sum_{g \in K} |g|_S,$$

which proves (2.1).

Another interpretation for a Følner sequence is the following. For $g \in G$, the boundary $\partial^{\{g\}} F_n = (gF_n) \triangle F_n$ is the symmetric difference between F and F shifted by g. We can thus rephrase (2.1) as follows: A Følner sequence is asymptotically invariant with respect to the shift by $g \in G$.

Studying amenable groups is useful in order to see how far one can get only with this one feature of $\mathbb{Z}^d$. Here are some properties of amenable groups and their actions. A finitely generated group G is amenable if and only if for each compact space X and every measurable G-action on X, there exists a G-invariant probability measure on X, see [Wei00]. Amenable groups are the natural setting for Birkhoffs pointwise ergodic theorem, see [Lin99]: Assume that an amenable group G acts measure preservingly on a probability space X. Lindenstrauss' theorem states that for a (tempered) Følner sequence $(F_n)_n$ and an integrable observable $f \in L^1(X)$, the averages

$$x \mapsto \frac{1}{|F_n|} \sum_{g \in F_n} f(xg^{-1})$$

converge for almost all $x \in X$. This theorem is extraordinarily useful when dealing with averages, for example in equation (1.1).

In amenable groups, Følner sequences are the natural generalization of cubes in $\mathbb{Z}^d$, and they can be used to define the IDS. As in $\mathbb{Z}^d$, one restricts the random Schrödinger operator to the set F_n with small boundary and counts the eigenvalues of the resulting matrix. The limit of the eigenvalue counting functions along a Følner sequence exists, see Chapter 5.

As mentioned above, Anderson predicted transport for low randomness and insulation for large randomness. While the latter has been proven in a variety of settings, the former is still out of reach, at least on Euclidean geometries like $\mathbb{R}^3$ or $\mathbb{Z}^d$. But on regular trees, [Kle98] showed that the alloy-type model exhibits absolutely continuous spectrum for low randomness. This is another reason to consider different geometries, it might help to develop tools in random Schrödinger operators applicable in Euclidean lattices as well.

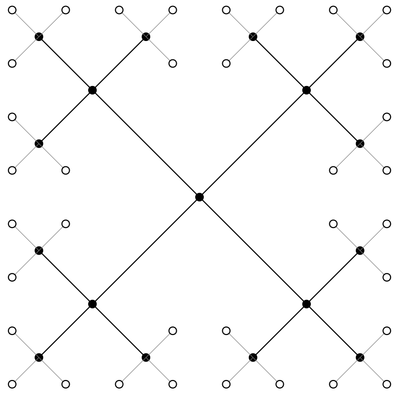


Figure 2.6: The S-boundary of balls in trees is large.

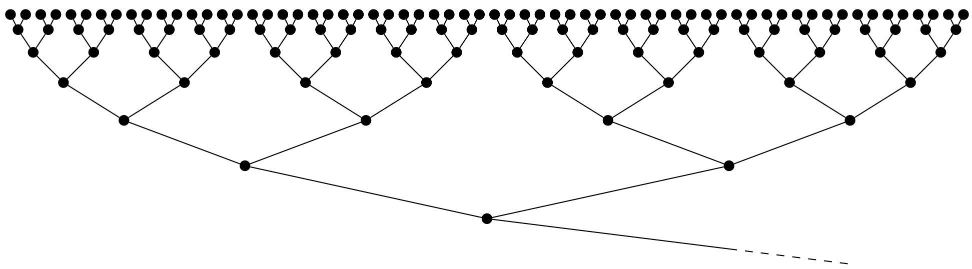


Figure 2.7: The canopy tree with branching number 2 has many leaves. The next zoom level extends along the dashed line.

Unfortunately, free groups are not amenable, and regular trees allow no Følner sequences. For example, the interior S-boundary of balls in the 4-regular tree graph contains more than half of the balls vertices, see Figure 2.6. Of course, there might be more cleverly chosen subsets with smaller boundaries. To eliminate this possibility, let us consider the spectrum of the Laplace operator on amenable groups. The ℓ^2-normalized indicator functions of sets of a Følner sequence form a Weyl sequence, and its Rayleigh ratio converges to 0. Since $-\Delta$ is a non-negative operator, the infimum of its spectrum on amenable groups is 0: $\inf \sigma(-\Delta) = 0$. This property actually characterizes amenable groups, see [Kes59a]. But on a tree of degree $d \geq 3$, the infimum of the spectrum of the operator $-\Delta$ is $(\sqrt{d-1}-1)^2 > 0$, see [Bro91; War13]. Therefore, free groups can not be amenable.

In [AW06], the authors study the eigenvalue counting functions on the balls of a regular tree and their limit. The limiting function equals the SDF on an infinite graph that they call *canopy tree*, see Figure 2.7. The canopy tree is isomorphic to a horoball of the regular tree, a concept which is analogous to horoballs in hyperbolic

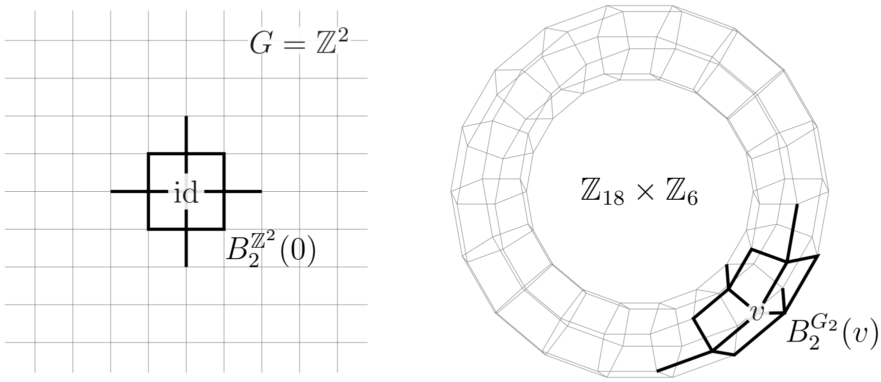


Figure 2.8: The discrete torus $\mathbb{Z}_{18} \times \mathbb{Z}_6$ is the image of a homomorphism that separates the points in all balls of radius 2.

geometry, see [Woe00, (12.13)]. We saw in Figure 2.6 that each ball in the tree contains many leaves, and that means that the leaves are still visible in the limiting graph. In order to define the IDS on free groups and regular trees, we need a different concept.

A group G is *residually finite* if the group homomorphisms from G onto finite groups separate points in G. This means, for each $g, g' \in G$, $g \neq g'$, there is a finite group $H_{g,g'}$ and a homomorphism $h_{g,g'} : G \to H_{g,g'}$ such that $h_{g,g'}(g) \neq h_{g,g'}(g')$ in $H_{g,g'}$. For each finite subset $K \subseteq G$, the product homomorphism $\widetilde{h}_K := \prod_{g,g' \in K, g \neq g'} h_{g,g'} : G \to \prod_{g,g' \in K, g \neq g'} H_{g,g'}$ is one-to-one on K. Let $H_K := \widetilde{h}_K(G)$ be the image of G under $\widetilde{h}_K$, which is a finite group, and $h_K : G \to H_K$, $h_K(g) := \widetilde{h}_K(g)$. For a generator S of G, $h_K(S)$ is a generator of H_K, because each h_K is onto. The Cayley graph of H_K with respect to $h_K(S)$ induces a graph of $h_K(K)$, and the Cayley graph of G with respect to S induces a graph on K. By construction, these two induced graphs are isomorphic. This allows us to use the Laplace operators on H_{K_n} with $K_n \nearrow G$ as a substitute for the restrictions of Laplace operators on a Følner sequence. The construction is illustrated by the example of $\mathbb{Z}^2$ and finite tori $(\mathbb{Z}/n\mathbb{Z}) \times (\mathbb{Z}/n'\mathbb{Z})$ as quotients, see Figure 2.8.

Free groups happen to be residually finite. For example, each element s of a generator S and $n \in \mathbb{N}$, $n \geq 2$, give a homomorphism $G \to \mathbb{Z}/(n\mathbb{Z})$ as follows. Write g as a word in the generator, count how often s occurs and substract the number of occurrences of s^{-1}. The difference modulo n is well defined and will be the image of g. These homomorphisms separate already all group elements which can not be written as words with the same letters in different order. Explicit constructions to separate all words in F_k can be found in Chapter 6. The notion of residually finite groups allows us to define the IDS of the Laplace operator on

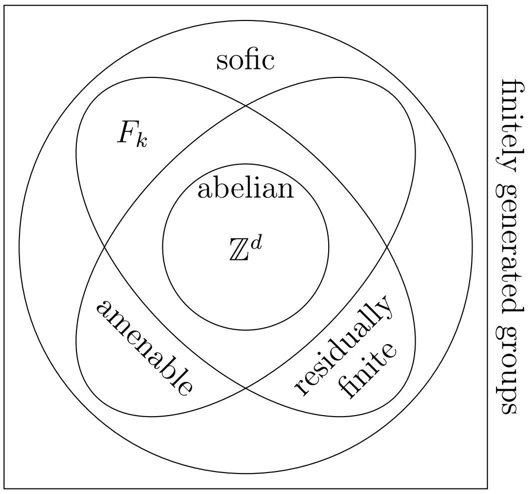


Figure 2.9: Amenable Gruppen und residuell endliche Gruppen sind sofisch.

trees, and indeed, with this construction, the Pastur–Shubin formula holds true, see also Chapter 6.

As we saw, free groups are not amenable but residually finite. There are also examples of amenable groups which are not residually finite. In [Gro99], Gromov introduced a class of groups which contains all amenable and all residually finite groups and many more. Weiss coined the term *sofic groups* for this class in [Wei00], see Figure 2.9. A finitely generated group G with generating set S is sofic if its S-edge-labeled Cayley graph is well approximated by finite S-edge-labeled graphs $\Gamma = (V, E)$, in the following sense. Consider a vertex $v \in V$ and a non-negative integer r. The balls of radius r around v in Γ and around $\mathrm{id} \in G$ in the Cayley graph $\Gamma(G, S)$ induce S-edge-labeled subgraphs of Γ and $\Gamma(G, S)$, respectively. Let us call v an *r-inner vertex* if these subgraphs are isomorphic with an isomorphism that preserves the labels. The r-boundary $\partial^r\Gamma$ consists of all vertices of Γ which are not r-inner vertices, see Figure 2.10. The group is *sofic* if there are finite S-edge-labeled graphs $\Gamma_n = (V_n, E_n)$, $n \in \mathbb{N}$, such that, for all $r \in \mathbb{N}$,

$$\lim_{n\to\infty} \frac{|\partial^r\Gamma_n|}{|V_n|} = 0.$$

For amenable groups, the approximating finite S-edge-labeled graphs are induced by the Cayley graph of G on the sets of a Følner sequence. For residually finite groups, the approximating graphs are the Cayley graphs of the finite groups H_K from above, where we choose K as the ball of radius r around $\mathrm{id} \in G$. In this case, the r-boundary is empty since H_K acts transitively on the vertices of its Cayley graph and preserves the labels, so all r-balls are isomorphic as labeled graphs.

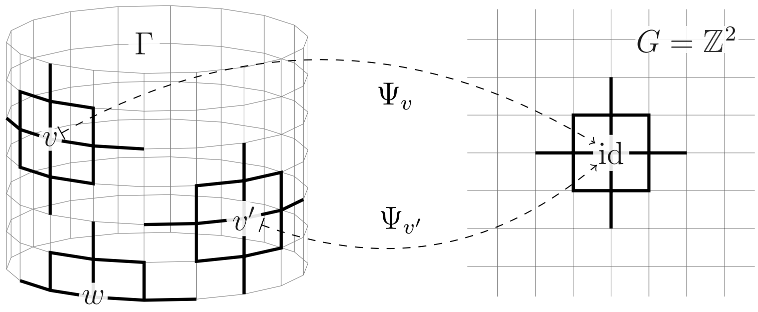


Figure 2.10: Two 2-inner vertices v and v'', while w is in the interior 2-boundary.

The class of sofic groups is quite large. In fact, so far no discrete group has been identified to be not sofic, to the authors best knowledge. With the local graph isomorphisms, the approximation of the Cayley graphs can be lifted to approximate operators on sofic groups by matrices on the finite S-edge-labeled graphs. We explain this strategy to define the IDS on sofic groups in the next Section 3.1. See also Chapter 6 for more details.

Chapter 3

Concentration inequalities

3.1 McDiarmid's inequality and sofic groups

We summarize the results of Chapter 6 and highlight the use of McDiarmid's concentration inequality. For this section let G denote a sofic group. As detailed above, this encompasses amenable and residually finite groups. We consider a deterministic and translation invariant operator $A\colon \ell^2(G) \to \ell^2(G)$ and ask how to construct approximating matrices. Recall that on a sofic group G, there are approximating graphs with many interior points. For each interior point v of the approximating graph $\Gamma = (V, E)$, there is a local graph isomorphism Ψ_v that maps v to $\mathrm{id} \in G$ and the, say, r-ball around v to the r-ball of id in G while preserving the S-labels on the edges, see Figure 2.10. With the local isomorphisms, we copy the matrix elements of A in order to define an operator on $\ell^2(V)$. By construction, this resulting resembles A for nearby vertices.

Of course, each matrix element can be copied with many local isomorphisms, so there is a question of well-definedness to address. Luckily, translation invariance comes to the rescue and resolves this issue. The local isomorphisms locally preserve the group structure, as the paths from h to g are preserved by all local isomorphisms which preserve large enough balls. The labels effectively prohibit the local isomorphisms from rotating the graph and from swapping vertical and horizontal edges, see Figure 3.1. By translation invariance, the matrix element $\langle \delta_g, A\delta_h \rangle$ between two vertices $g, h \in G$ depends only on gh^{-1}, so it is safe to use any local isomorphism. The eigenvalue counting functions of the matrices obtained by this construction indeed converge and define the IDS. The limit equals the SDF, that is: The Pastur–Shubin formula holds true. For more details, including precise formulas, see Chapter 6.

For random operators, the situation is more complicated because the matrix elements are random variables themselves and their values are no longer translation

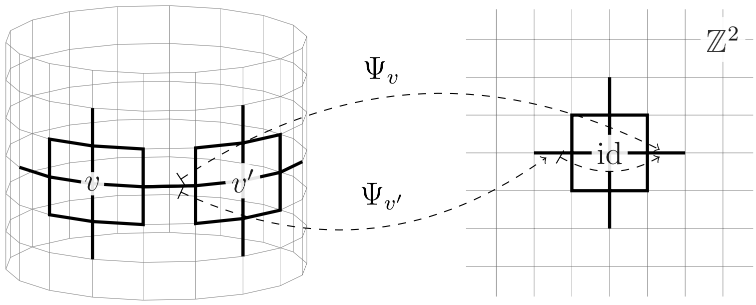


Figure 3.1: Labels and translation invariance guarantee well-defined matrices.

invariant. Only the probability distribution of the random variables is invariant. So, instead of copying the values of the matrix elements, we have to copy the distribution of the matrix elements. We use the distributions to independently sample new matrix elements for the finite dimensional approximation, see Chapter 6 for the exact procedure. By this construction, the eigenvalue counting functions of the operators on the finite graphs are independent of the matrix elements of the original operator. But the expectation of the eigenvalue counting functions converges, and the limiting function equals the SDF.

The next step is to improve the convergence of the expectations to almost sure convergence of the random variables themselves. This is where the phenomenon of concentration of probability enters. Consider the eigenvalue counting functions with a fixed argument as functions of the random entries of the matrix to $\mathbb{R}$. We need that the eigenvalue counting function viewed in this way concentrates its image measure around its expectation. The suitable tool is the following concentration inequality due to McDiarmid.

Theorem 3.1.1. *Let $X = (X_1, \ldots, X_n)$ be a family of independent random variables with values in $\mathbb{R}$, and let $f\colon \mathbb{R}^n \to \mathbb{R}$ satisfy the bounded differences assumption, i. e., there is a constant $c \in \mathbb{R}$ such that, whenever $x, x' \in \mathbb{R}^n$ differ only in one coordinate, we have*

$$|f(x) - f(x')| \leq c.$$

Then, for $\mu := \mathbb{E}[f(X)]$ and any $\epsilon \geq 0$,

$$\mathbb{P}(|f(X) - \mu| \geq \epsilon) \leq 2 \exp\left(-\frac{2\epsilon^2}{nc^2}\right).$$

Different proofs can be found in numerous places, e. g. [McD89, Lemma 1.2], [McD98, Theorem 3.1], and [BLM13, Theorem 6.2]. For our application at hand, we need that the eigenvalue counting functions meet the bounded differences assumption. The following proposition allows to control how much eigenvalue counting functions change by the rank of a perturbation.

Proposition 3.1.2. *Self-adjoint $n \times n$-matrices A and C satisfy*

$$\|N_{A+C} - N_A\|_\infty \leq \operatorname{rg} C/n.$$

The proof in [LSV11, Proposition 7.1] relies on the minimax principle for eigenvalues of self-adjoint operators, see e. g. [RS78, Theorem XIII.1]. To apply the proposition in our context, we consider the eigenvalue counting functions as functions of the random matrix elements. Each such matrix element accounts for a perturbation of at most rank 2, because of the symmetry of self-adjoint matrices, so that eigenvalue counting functions satisfy the bounded differences assumption and McDiarmid's inequality applies. From here, standard calculations reveal the almost sure convergence of the eigenvalue counting functions, see Chapter 6.

3.2 Glivenko–Cantelli theory and uniform convergence

In the following, we want to motivate the results of Chapters 4 and 5. For the Euclidean lattices, the IDS of the Laplacian can be calculated, see [PT18], and it turns out to be continuous. For the regular tree with degree $k+1$, the DoS of the Laplacian is explicitly known as

$$\mathbb{R} \ni E \mapsto \mathbf{1}_I(E) \frac{k+1}{2\pi} \frac{\sqrt{4k - (E-k-1)^2}}{(k+1)^2 - (E-k-1)^2}, \tag{3.1}$$

where $I = [(\sqrt{k}-1)^2, (\sqrt{k}+1)^2]$ is the spectrum of the Laplacian, see [Kes59b; McK81]. Its antiderivative, the IDS, is thus continuous, too, see Figure 3.2. A simple argument in [DS84] shows that the IDS is continuous for many ergodic random Schrödinger operators on amenable groups. Also, Wegner estimates can be used to see the continuity of the IDS. But there are also random Schrödinger operators with discontinuous IDS. Examples are the Laplacians on percolation graphs and certain Anderson models on percolation graphs, see [SSV19].

Continuity of the IDS is important for many reasons, in particular, for continuous IDS the limit in the construction of the IDS is actually better than pointwise: The pointwise limit of probability distribution functions on $\mathbb{R}$ to a continuous limiting probability distribution function f is actually uniform, see [Bau92, 30.13 Satz]. The argument uses finitely many sampling points and interpolates between them with the monotony of probability distribution functions. Unfortunately, this strategy

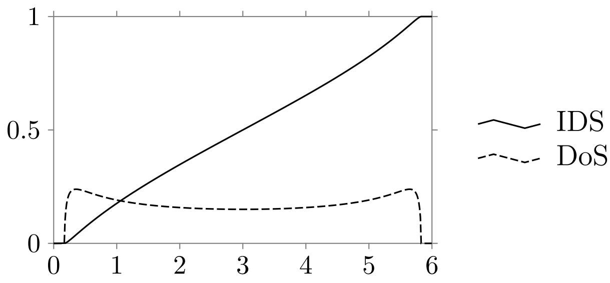


Figure 3.2: The denstity of states and the integrated density of states of the Laplacian on a regular tree of degree 3

seems to be not suitable to derive error estimates and a speed of convergence of the IDS.

In Chapters 4 and 5, we prove almost sure uniform convergence of the eigenvalue counting functions for a large class of random Schrödinger operators on amenable groups without taking advantage of continuity of the IDS. In fact, our results apply to random Schrödinger operators with discontinuous IDS, too. To explain the strategy, let us first consider the Anderson model with independent matrix elements and increasingly relax our assumptions.

Our key ingredient is Glivenko–Cantelli theory. Recall the classical Glivenko–Cantelli theorem:

Theorem 3.2.1 (Glivenko–Cantelli)**.** *Let $X, X_j \colon \Omega \to \mathbb{R}$, $j \in \mathbb{N}$, be independent and identically distributed random variables with probability distribution function $F_\infty := \mathbb{P}(X \leq \cdot\,) \colon \mathbb{R} \to [0,1]$. The empirical distribution functions are $F_n := \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{[X_j,\infty)} \colon \mathbb{R} \to [0,1]$. Then, we have $\mathbb{E}[F_n] = F_\infty$ and*

$$\|F_n - F_\infty\|_\infty \xrightarrow{n\to\infty} 0 \quad \textit{almost surely.}$$

The quantitative version of this classical theorem is a concentration inequality:

Theorem 3.2.2 (Dvoretzky–Kiefer–Wolfowitz [DKW56; Mas90])**.** *Under the assumptions of Theorem 3.2.1 and for all $\varepsilon > 0$, there is a sequence of events $\Omega_{\varepsilon,n}$, $n \in \mathbb{N}$, such that*

$$\mathbb{P}(\Omega_{\varepsilon,n}) \geq 1 - 2\mathrm{e}^{-2n\varepsilon^2} \quad \textit{and} \quad \|F_n - F_\infty\|_\infty \mathbf{1}_{\Omega_{\varepsilon,n}} \leq \varepsilon \quad (n \in \mathbb{N}).$$

If one can make sure that the set $\{\omega \in \Omega \mid \|F_n - F_\infty\|_\infty \leq \varepsilon\}$ is measurable, one can write this shorter as

$$\mathbb{P}(\|F_n - F_\infty\|_\infty \leq \varepsilon) \geq 1 - 2\mathrm{e}^{-2n\varepsilon^2}.$$

For a direct application of these two classical results, we consider a diagonal random operator $V\colon \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$, $(V\psi)(z) := V(z)\psi(z)$, where the matrix elements $V(z)$, $z \in \mathbb{Z}^d$, are independent and identically distributed. This is actually the potential of the classical Anderson model. Its eigenvalues are exactly the values $V(z)$, and the eigenfunctions are δ_z, $z \in \mathbb{Z}^d$. The eigenvalue counting functions along a sequence $(\Lambda_L)_L$ are thus the empirical distribution functions of the i. i. d. family of random variables $\{V(z) \mid z \in \Lambda_L\}$, and Theorem 3.2.2 provides almost sure uniform convergence and an estimate for the speed of convergence, even for discontinuous limiting functions.

In order to treat more general random Schrödinger operators, for example the Anderson model with the random Schrödinger operator $-\Delta + V_\omega$, we have off-diagonal matrix elements to take into account, and the eigenfunctions are not as localized as in the previous example.

Consider the one-dimensional case $d = 1$ and a finite discrete interval $\Lambda_L = \{0, 1, \dots, L-1\}$, $L \in \mathbb{Z}$, $L \geq 2$. Instead of considering each element of Λ_L individually, we form larger chunks Λ_k of length $k \in \mathbb{Z}$, $k < L$ and partition $\Lambda_L = \biguplus_{t \in k\mathbb{Z}} \Lambda_k(t)$ with $\Lambda_k(t) := \Lambda_L \cap (\Lambda_k + t)$. On each $\Lambda_k(t)$, we count the eigenvalues, and then we compare the sum of these with the one of Λ_L. Proposition 3.1.2 allows to bound the difference to the order k^{-1}.

Now we need a more powerful Glivenko–Cantelli-type theorem to deal with the sample of independent eigenvalue counting functions.

Theorem 3.2.3 ([DeH71; Wri81]). *Let $X, X_j\colon \Omega \to \mathbb{R}^k$, $j \in \mathbb{N}$, be i. i. d. random variables with independent components and*

$$\mathcal{M} := \{f\colon \mathbb{R}^k \to [-1,1] \mid f \text{ is monotone in each coordinate of its argument}\}.$$

Then, almost surely,

$$\sup_{f \in \mathcal{M}} \Big| \frac{1}{n} \sum_{j=1}^{n} f(X_j) - \mathbb{E}[f(X)] \Big| \xrightarrow{n \to \infty} 0.$$

Furthermore, for all $\varepsilon > 0$, there are constants $a_\varepsilon, b_\varepsilon > 0$ and a sequence of events $(\Omega_{\varepsilon,n})_n$ such that

$$\mathbb{P}(\Omega_{\varepsilon,n}) \geq 1 - b_\varepsilon \mathrm{e}^{-a_\varepsilon n} \quad \text{and} \quad \sup_{f \in \mathcal{M}} \Big| \frac{1}{n} \sum_{j=1}^{n} f(X_j) - \mathbb{E}[f(X)] \Big| \mathbf{1}_{\Omega_{\varepsilon,n}} \leq \varepsilon.$$

To recognize Theorem 3.2.3 as a generalization of Theorems 3.2.1 and 3.2.2, consider the case $k = 1$ and note that $\mathcal{M}_1 := \{\mathbf{1}_{(-\infty,t]} \mid t \in \mathbb{R}\}$ is a subset of $\mathcal{M}$ and that

$$\|F_n - F\|_\infty = \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| = \sup_{f \in \mathcal{M}_1} \Big| \frac{1}{n} \sum_{j=1}^{n} f(X_j) - \mathbb{E}[f(X)] \Big|.$$

In the case $k \geq 2$, the assumption that the components of the random vectors are independent can be weakened but not dropped. There are counterexamples that show that some condition of this kind is necessary, see Chapter 4.

For our application we have to check that eigenvalue counting functions are monotone in the random potential. This can be done with the minimax principle. In fact, increasing the potential means increasing the Rayleigh quotient. This monotony survives the minimum and the maximum and thus carries over to the eigenvalues.

Up to now, we know how to compare the eigenvalue counting function on Λ_L to the expected eigenvalue counting function on Λ_k, and that, in the limit $L \to \infty$, the error we make is of the order $1/k$. The last and final step towards almost sure uniform convergence with error estimate is to perform the limit $k \to \infty$ and to see that the expectation of the eigenvalue counting function on Λ_k converges. In fact, we prove that the expected eigenvalue counting functions on Λ_k form a Cauchy sequence in the Banach space of bounded and right continuous functions, see Chapter 4. To identify the limit as the spectral distribution function, we refer back to the results about sofic groups.

In higher dimensions $d \geq 2$, the ordering of the base $(\delta_v)_{v\in\mathbb{Z}^d}$ is not given. As a consequence, the block diagonal structure of the representing matrix of the sum of the operators on $\Lambda_k(t)$ is not that obvious. Nonetheless, the generalization to $d \geq 2$ is straightforward. We cover large cubes $\Lambda_L = \{0, \dots, L-1\}^d$ with smaller cubes Λ_k with an error estimate of the order of $1/k$ which corresponds to the boundary to volume ratio of cubes. By Theorem 3.2.3, the average of the eigenvalue counting functions on the smaller cubes converges uniformly to a their expectation with high probability. The expectations form a Cauchy sequence, and the limit is identified as above.

The results in Chapters 4 and 5 are actually more general than indicated above. We use an abstract framework of what we call almost additive fields which includes the setting of the IDS of alloy type model but also, for example, Laplacians of site percolation graphs and their IDS.

In Chapter 5, we generalize the almost sure uniform convergence to amenable groups and face various geometric and probabilistic challenges. As outlined in Chapter 2, the substitute for cubes in Z^d are Følnersequences. Unfortunately, we used more properties of cubes than their small boundary to volume ratio: We also used cubes as tiles to cover larger cubes without overlap. It is unknown if there is a Følner sequence with the additional property that every set in the sequence covers the group in a similar fashion as the cubes in $\mathbb{Z}^d$. Instead, we utilize quasi tilings as introduced in [OW87] and studied quantitatively in [PS16]. A quasi tiling relaxes the properties of tilings in several ways. Instead of one tile which covers each group element exactly once, one uses several tiles of different sizes to cover

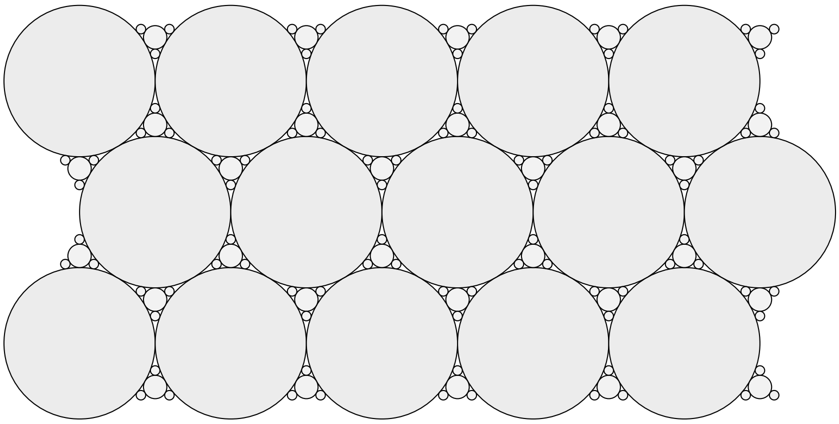


Figure 3.3: One circle covers 90.69%, two circles cover 95.03%, and three circles cover 97.17% of the plane. The radii are 1, 0.15, and $6.28 \cdot 10^{-2}$.

most of the group, and one even allows for some controlled overlap between certain shifted tiles. An impression of how much of the plane can be covered with only a few circular tiles is given in Figure 3.3.

To deal with quasi tilings is more laborious than working with tilings. The possible overlap of certain tiles destroys the independence of the sample. If we remove the overlap from the sample, we loose the identical distribution. The solution is to independently resample the overlap conditioned on the rest of the tile. The geometric control of the size of the overlap allows us to estimate the error. For details, see Chapter 5.

Our results do not only apply to the IDS. Instead, we employ a framework which is suitable to treat the IDS but also different intensive quantities like relative clustersize in percolation theory.

3.3 Hoeffding's inequality and Lifshitz tails of the Anderson model on the Bethe lattice

In Chapter 7 we prove Lifshitz tails of the Anderson model on regular trees. In the physics literature, a regular tree graph is called Bethe lattice in honor of Hans Bethe. The conjecture about Lifshitz tails on the Bethe lattice was made in [KH85].

As described above, for Lifshitz tails the concentration inequality is a central part of the argument, even more, concentration of probability is the moral source for the phenomenon of Lifshitz tails. But due to the geometry of the Bethe lattice, the application of the actual inequality is buried deep in the proof. We will unpack some of the stumbling blocks here. This will also explain why it took so long to implement a rigorous proof even though Lifshitz tails were studied before in a wide variety of settings including Cayley graphs of exponential volume growth, see e. g. [AV09].

The exponential growth of the balls and the non-amenability of the Bethe lattice create major problems for the traditional reasoning behind Lifshitz tails. The heuristic outlined in Chapter 1 works only in an amenable setting. One compares the restriction of the operators with Dirichlet and with Neumann boundary conditions along a Følner sequence. Because the boundary is negligible in the limit when compared to the volume, the choice of boundary conditions does not change the limiting object. The contrary happens on the Bethe lattice. The fraction of nodes in the boundary is bounded from below, so the choice of boundary conditions manifests itself in the limit. One says that the Dirichlet-Neumann bracket does not close.

Another efficient tool in the amenable setting is perturbation theory. A key requirement for this to work is a spectral gap between the ground state energy and the second eigenvalue, which does not close too rapidly with the growing size of the box. In Euclidean settings, the gap is of the order of L^{-2}, while the distance between the infimum of the spectrum of the unrestricted Schrödinger operator and the ground state energy of the restricted operator itself is of order of L^{-2}, too. On the Bethe lattice, the gap is of the order of L^{-3}, and the shift of the ground state energy with respect to the infimum of the unrestricted operator is L^{-2}. This discrepancy fails perturbation techniques.

Instead, we reformulate the problem of Lifshitz tails with the Laplace transform of the IDS. From there, we can relate the problem to the location of the ground state energy on the finite box using techniques from mathematical physics. Instead of a box, though, we use balls of the Bethe lattice, see Figure 2.4, or, to be precise, symmetric rooted trees, which are balls that miss one of the branches from their center, see Figure 3.4. The symmetric rooted trees have the advantage that one can write down an orthonormal basis of eigenfunctions of $-\Delta$. The ground state, see Figure 3.5, has an interpretation for the symmetric random walk with killing terms in the root and on the leaves. The symmetry of the trees helps to split the degeneration of the eigenvalues.

In the light of Chapter 6, our usage of balls in the Bethe lattice might seem surprising. The result there suggests to use the sofic approximations in order to deal with the IDS of the Bethe lattice instead of the canopy tree, see Figure 2.7. In Chapters 4 to 6, we use inequalities involving the rank of matrices. This approach is to blunt in the current setting. Instead, we rely on spectral information. Close to the leaves, the eigenfunctions with low eigenvalue are exponentially small in the height of the tree. In fact, the eigenfunctions suppress the boundary of the tree so that, at least at the bottom of the spectrum, the spectral behavior of the Bethe

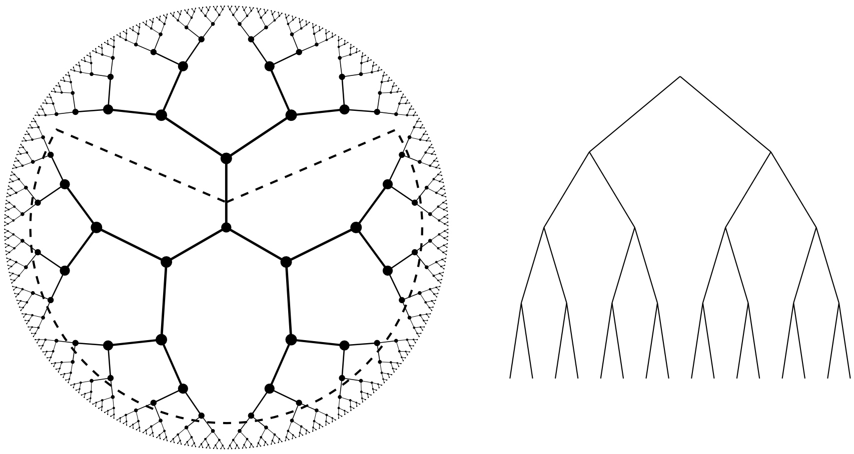


Figure 3.4: A symmetric rooted tree, left: embedded in the Bethe lattice and right: with the root at the top and the generations below

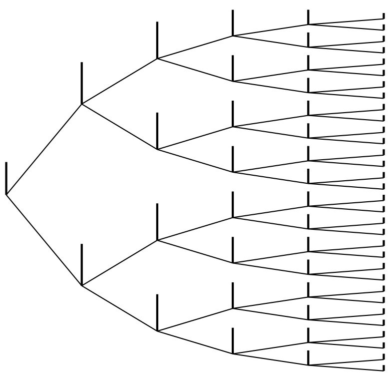


Figure 3.5: The ground state on the symmetric rooted tree.

lattice and the canopy tree coincide.

After having dealt with all these difficulties, we have to, similarly to the amenable case outlined in Chapter 1, bound the probability that the potential energy $\langle\varphi, V_\omega\varphi\rangle$ is smaller than its expectation for states φ with low kinetic energy. This is where the Hoeffding's concentration inequality enters. In general, Hoeffding's inequality states that the sum S_n of independent random variables X_j with values in bounded intervals $[a_j, b_j]$, $j \in \{1, \dots, n\}$, satisfies

$$\mathbb{P}(S_n \geq \kappa) \leq \exp\left(-\frac{2(\kappa - \mathbb{E}[S_n])^2}{\sum_j (b_j - a_j)^2}\right)$$

for all $\kappa \geq \mathbb{E}[S_n]$. As a random variable, $\langle\varphi, V_\omega\varphi\rangle$ has expectation $\mu := \mathbb{E}\langle\varphi, V_\omega\varphi\rangle = \|\varphi\|_2^2 \cdot \mathbb{E}[V_\omega]$ and variance $\operatorname{Var}\langle\varphi, V_\omega\varphi\rangle = \|\varphi\|_4^4 \cdot \operatorname{Var} V_\omega$. Furthermore, Hoeffding's inequality specializes as follows. For all $\kappa \geq \mu$, we have

$$\mathbb{P}\big(\langle\varphi, V_\omega\varphi\rangle \geq \kappa\big) \leq \exp\left(-\frac{(\kappa-\mu)^2}{2\|\varphi\|_4^4 \cdot \|V_\omega - \mathbb{E}[V_\omega]\|_{L^\infty(\mathbb{P})}}\right).$$

Here, we see the influences of three different ingredients on the exponential decay: the distance to the expectation $\kappa - \mu$, the size of the potential $\|V_\omega - \mathbb{E}[V_\omega]\|_{L^\infty(\mathbb{P})}$, and the term $\|\varphi\|_4^4$ which is proportional to the variance of the potential energy of φ. States with low kinetic energy are more or less flat, and indeed, the smallest value of $\|\varphi\|_4^4$ of all normalized $\varphi \in \ell^2(1, \dots, n)$ is attained at the constant wave function $\varphi_{\mathrm{const}} = (n^{-1/2})_j$, with the value $\|\varphi_{\mathrm{const}}\|_4^4 = n^{-1}$. In symmetric rooted trees, the number n of random variables grows exponentially in the height of the tree. Therefore, small potential energy is double exponentially unlikely in the height of the tree. This strong result is barely enough to prove Lifshitz tails for the Anderson model on the Bethe lattice. For more details please refer to Chapter 7.

3.4 Bernstein's inequality and Lifshitz tails in Euclidean space

Chapter 8 establishes Lifshitz tails in full generality for continuous random Schrödinger operators on $\mathbb{R}^d$ with non-negative random potential. First, we consider the *breather model*, where the single site potential is given by

$$f_\lambda := \mathbf{1}_{\lambda \cdot A}$$

for a measurable set $A \subseteq (-\frac{1}{2}, \frac{1}{2})^d \subseteq \mathbb{R}^d$ and a real parameter $\lambda \in [0, 1]$. The *random breather potential*

$$V_\omega(x) := \sum_{k \in \mathbb{Z}^d} f_{\lambda_k(\omega)}(x - k)$$

is not linear in the randomness. In fact, its derivative in the sense of distributions is not even a measure. Classical results about Lifshitz tails employ Temple's inequality, which treats the randomness as perturbation, see [KM83; Sto01; KV10]. It is unclear whether this approach can be implemented for the breather model. The perturbation result which is robust enough to be applied to the breather model is an inequality by Thirring. It allows us to extract the average over the random variables λ_k. We then apply Bernstein's concentration inequality, which states that the sum S_n of n non-negative i. i. d. random variables with positive and finite expectation satisfies

$$\mathbb{P}\big(S_n \leq \tfrac{1}{2}\mathbb{E}[S_n]\big) \leq \exp(-Cn)$$

for a suitable constant $C > 0$ independent of n.

Despite its apparent simplicity, the breather model turns out to include all major obstacles one has to tackle to address Lifshitz tails in a positive random potential. The Thirring argument does in fact not rely on the breathing structure. The important property to shift the eigenvalues and produce Lifshitz behavior is that the potential is strictly positive on a set with positive measure with a positive probability. The lower bound for the potential translates into an upper bound for the IDS and is used in Chapter 8 to prove Lifshitz tails. Said strategy applies to general non-negative random potentials. The formulation of conditions for a lower bound is more involved because the random potential is more general. A suitable version is contained in Chapter 8.

About this thesis

This thesis is a collection of five original research articles, which form the remaining chapters. For convenience, we give a short overview.

Chapter 4 is devoted to uniform convergence of the IDS on lattices, cf. Section 3.2. It is taken without changes from [SSV17].

Chapter 5 generalizes Chapter 4 to the setting of amenable groups and coincides with [SSV18].

Chapter 6 establishes the IDS on sofic groups, see Section 3.1. It has appeared in [SS15].

Chapter 7 proves Lifshitz tails behavior for the IDS of the Anderson Hamiltonian on the Bethe lattice, cf. Section 3.3. It has not been published yet, but a preprint can be found in [HS14].

Chapter 8 deals with Lifshitz tails of continuous random Schrödinger operators with monotone random potentials, as described in Section 3.4. Part of this material has been published in [SV17], the whole work [SV] is in preparation.

Bibliography

[AG73] W. O. Amrein and V. Georgescu. “On the characterization of bound states and scattering states in quantum mechanics”. *Helvetica Physica Acta. Physica Theoretica. Societatis Physicae Helveticae Commentaria Publica* 46 (1973), pp. 635–658. ISSN: 0018-0238 (cit. on p. 11).

[And58] P. W. Anderson. “Absence of Diffusion in Certain Random Lattices”. *Phys. Rev.* 109 (5 1958), pp. 1492–1505. DOI: `10.1103/PhysRev.109.1492` (cit. on p. 10).

[APP18] N. Ashby, T. E. Parker, and B. R. Patla. “A null test of general relativity based on a long-term comparison of atomic transition frequencies”. *Nature Physics* (2018), pp. 1–6. DOI: `10.1038/s41567-018-0156-2` (cit. on p. 17).

[AV09] T. Antunović and I. Veselić. “Equality of Lifshitz and van Hove exponents on amenable Cayley graphs”. *Journal de Mathématiques Pures et Appliquées* 92.4 (2009), pp. 342–362. ISSN: 0021-7824. DOI: `10.1016/j.matpur.2009.05.014`. arXiv: `0706.2844 [math.SP]` (cit. on pp. 15, 34).

[AW06] M. Aizenman and S. Warzel. “The Canopy Graph and Level Statistics for Random Operators on Trees”. *Mathematical Physics, Analysis and Geometry* 9.4 (2006), pp. 291–333. ISSN: 1572-9656. DOI: `10.1007/s11040-007-9018-3` (cit. on p. 23).

[Bau92] H. Bauer. *Mass- und Integrationstheorie*. German. 2nd ed. De Gruyter Lehrbuch. Walter de Gruyter, 1992. ISBN: `9783110136265`. Google Books: `7iAchGmcGXkC` (cit. on p. 29).

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities. A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. ISBN: `978-0-19-876765-7`. DOI: `10.1093/acprof:oso/9780199535255.001.0001` (cit. on p. 28).

[Bro91] R. Brooks. "The Spectral Geometry of k-regular Graphs". *Journal d'Analyse Mathématique* 57.1 (1991), pp. 120–151. ISSN: 1565-8538. DOI: 10.1007/BF03041067 (cit. on p. 23).

[Cyc+87] H. L. Cycon et al. *Schrödinger Operators. with Application to Quantum Mechanics and Global Geometry.* 1st ed. Theoretical and Mathematical Physics. Springer-Verlag Berlin Heidelberg, 1987, pp. XI, 329. ISBN: 978-3-540-16758-7. DOI: 10.1007/978-3-540-77522-5 (cit. on p. 11).

[DeH71] J. DeHardt. "Generalizations of the Glivenko-Cantelli Theorem". *The Annals of Mathematical Statistics* 42.6 (1971), pp. 2050–2055. Project Euclid: aoms/1177693073 (cit. on p. 31).

[DKW56] A. Dvoretzky, J. C. Kiefer, and J. Wolfowitz. "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator". *The Annals of Mathematical Statistics* 27.3 (1956), pp. 642–669. ISSN: 00034851. DOI: 10.1214/aoms/1177728174. Project Euclid: aoms/1177728174 (cit. on p. 30).

[DS84] F. Delyon and B. Souillard. "Remark on the Continuity of the Density of States of Ergodic Finite Difference Operators". *Communications in Mathematical Physics* 94.2 (1984), pp. 289–291. Project Euclid: cmp/1103941286 (cit. on p. 29).

[Ens78] V. Enss. "Asymptotic completeness for quantum mechanical potential scattering. I. Short range potentials". *Communications in Mathematical Physics* 61.3 (1978), pp. 285–291. ISSN: 0010-3616. DOI: 10.1007/BF01940771. Project Euclid: cmp/1103904257 (cit. on p. 11).

[Føl55] E. Følner. "On groups with full Banach mean value". *MATHEMATICA SCANDINAVICA* 3 (1955), pp. 243–254. DOI: 10.7146/math.scand.a-10442 (cit. on p. 21).

[Gro99] M. Gromov. "Endomorphisms of symbolic algebraic varieties". *Journal of the European Mathematical Society (JEMS)* 1.2 (1999), pp. 109–197. ISSN: 1435-9855. DOI: 10.1007/PL00011162 (cit. on p. 25).

[HS14] F. Hoecker-Escuti and C. Schumacher. *The Anderson model on the Bethe lattice: Lifshitz tails*. preprint. 2014. hal-01074905 (cit. on p. 37).

[Kes59a] H. Kesten. "Full Banach Mean Values on Countable groups". *Mathematica Scandinavica* 7 (1959), pp. 146–156. ISSN: 1903-1807. DOI: 10.7146/math.scand.a-10568 (cit. on p. 23).

[Kes59b] H. Kesten. "Symmetric random walks on groups". *Transactions of the American Mathematical Society* 92 (1959), pp. 336–354. DOI: 10.1090/S0002-9947-1959-0109367-6 (cit. on p. 29).

[KH85] Y. Kim and A. B. Harris. "Density of states of the random-hopping model on a Cayley tree". *Physical Review B* 31 (11 1985), pp. 7393–7407. DOI: 10.1103/PhysRevB.31.7393 (cit. on p. 33).

[Kir07] W. Kirsch. *An Invitation to Random Schrödinger Operators*. 2007. 129 pp. arXiv: 0709.3707 [math-ph] (cit. on p. 15).

[Kle98] A. Klein. "Extended States in the Anderson Model on the Bethe Lattice". *Advances in Mathematics* 133.1 (1998), pp. 163–184. ISSN: 0001-8708. DOI: 10.1006/aima.1997.1688 (cit. on p. 22).

[KM83] W. Kirsch and F. Martinelli. "Large deviations and Lifshitz singularity of the integrated density of states of random Hamitonians". *Communications in Mathematical Physics* 89 (1983), pp. 27–40. DOI: 10.1007/BF01219524. Project Euclid: cmp/1103922589 (cit. on p. 37).

[KV10] W. Kirsch and I. Veselić. "Lifshitz Tails for a Class of Schrödinger Operators with Random Breather-Type Potential". *Letters in Mathematical Physics* 94 (1 2010), pp. 27–39. DOI: 10.1007/s11005-010-0417-1. arXiv: math-ph/0609014 (cit. on p. 37).

[Lif65] I. M. Lifshitz. "Energy spectrum structure and quantum states of disordered condensed systems". *Soviet Physics Uspekhi* 7.4 (1965), pp. 549–573. DOI: 10.1070/PU1965v007n04ABEH003634 (cit. on p. 13).

[Lin99] E. Lindenstrauss. "Pointwise theorems for amenable groups". *Electronic Research Announcements of the American Mathematical Society* 5 (1999), pp. 82–90. DOI: 10.1007/s002220100162 (cit. on p. 22).

[LSV11] D. Lenz, F. Schwarzenberger, and I. Veselić. "A Banach space-valued ergodic theorem and the uniform approximation of the integrated density of states". *Geometriae Dedicata* 150 (1 2011), pp. 1–34. DOI: 10.1007/s10711-010-9491-x. arXiv: 1003.3620 [math-ph] (cit. on p. 29).

[Mas90] P. Massart. "The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality". *The Annals of Probability* 18.3 (1990), pp. 1269–1283. ISSN: 00911798. DOI: 10.1214/aop/1176990746. Project Euclid: aop/1176990746 (cit. on p. 30).

[McD89] C. McDiarmid. "On the method of bounded differences". In: *Surveys in Combinatorics, 1989: Invited Papers at the Twelfth British Combinatorial Conference*. Ed. by J. Siemons. London Mathematical Society Lecture Note Series. Cambridge University Press, 1989, pp. 148–188. DOI: 10.1017/CBO9781107359949.008 (cit. on p. 28).

[McD98] C. McDiarmid. "Concentration". In: *Probabilistic Methods for Algorithmic Discrete Mathematics*. Ed. by M. Habib, C. McDiarmid, and R. B. Ramirez-Alfonsin Jorge. Vol. 16. Algorithms and Combinatorics. Springer, Berlin, Heidelberg, 1998, pp. 195–248. ISBN: 978-3-642-08426-3. DOI: 10.1007/978-3-662-12788-9_6 (cit. on p. 28).

[McK81] B. D. McKay. "The expected eigenvalue distribution of a large regular graph". *Linear Algebra and its Applications* 40 (1981), pp. 203–216. ISSN: 0024-3795. DOI: 10.1016/0024-3795(81)90150-6 (cit. on p. 29).

[OW87] D. S. Ornstein and B. Weiss. "Entropy and isomorphism theorems for actions of amenable groups". *Journal d'Analyse Mathématique* 48 (1987), pp. 1–141. ISSN: 0021-7670. DOI: 10.1007/BF02790325 (cit. on p. 32).

[Pas71] L. A. Pastur. "Selfaverageability of the number of states of the Schrödinger equation with a random potential". *Mat. Fiz. i Funkcional. Anal.* 238.2 (1971), pp. 111–116 (cit. on p. 15).

[Pas80] L. A. Pastur. "Spectral properties of disordered systems in the one-body approximation". *Communications in Mathematical Physics* 75.2 (1980), pp. 179–196. ISSN: 1432-0916. DOI: 10.1007/BF01222516. Project Euclid: cmp/1103908097.

[PF92] L. A. Pastur and A. Figotin. *Spectra of Random and Almost-Periodic Operators*. 1st ed. Grundlehren der mathematischen Wissenschaften 297. Springer, 1992, pp. VIII, 587. ISBN: 0-387-50622-5 (cit. on p. 15).

[PS16] F. Pogorzelski and F. Schwarzenberger. "A Banach space-valued ergodic theorem for amenable groups and applications". *Journal d'Analyse Mathématique* 130 (2016), pp. 19–69. ISSN: 0021-7670. DOI: 10.1007/s11854-016-0028-6. arXiv: 1205.3649 [math.SP] (cit. on p. 32).

[PT18] N. Peyerimhoff and M. Täufer. "The IDS on Archimedian lattices?" (2018). arXiv: 1807.06406 [math-ph] (cit. on pp. 15, 29).

[RS78] M. Reed and B. Simon. *Methods of modern mathematical physics. Analysis of operators*. Methods of Modern Mathematical Physics 4. Academic Press, 1978. ISBN: 9780125850049. Google Books: WwApAQAAMAAJ (cit. on pp. 13, 29).

[Rue69] D. Ruelle. "A remark on bound states in potential-scattering theory". *Nuovo Cimento A (10)* 61 (1969), pp. 655–662 (cit. on p. 11).

[Shu79] M. A. Shubin. "Spectral theory and the index of elliptic operators with almost-periodic coefficients". *Russian Mathematical Surveys* 34.2 (1979), pp. 109–157. DOI: 10.1070/RM1979v034n02ABEH002908 (cit. on p. 15).

[SS15] C. Schumacher and F. Schwarzenberger. "Approximation of the Integrated Density of States on Sofic Groups". *Annales Henri Poincaré* 16 (4 2015), pp. 1067–1101. DOI: `10.1007/s00023-014-0342-4`. arXiv: `1212.5207 [math.SP]` (cit. on p. 37).

[SSV17] C. Schumacher, F. Schwarzenberger, and I. Veselić. "A Glivenko–Cantelli theorem for almost additive functions on lattices". *Stochastic Processes and their Applications* 127.1 (2017), pp. 179–208. ISSN: 0304-4149. DOI: `10.1016/j.spa.2016.06.005`. arXiv: `1606.07664 [math.PR]` (cit. on p. 37).

[SSV18] C. Schumacher, F. Schwarzenberger, and I. Veselić. "Glivenko-Cantelli Theory, Ornstein-Weiss quasi-tilings, and uniform Ergodic Theorems for distribution-valued fields over amenable groups". *The Annals of Applied Probability* 28.4 (2018), pp. 2417–2450. DOI: `10.1214/17-AAP1361`. arXiv: `1711.07525 [math-ph]` (cit. on p. 37).

[SSV19] C. Schumacher, F. Schwarzenberger, and I. Veselić. "Uniform existence of the IDS on lattices and groups". In: *Proceedings to the Conference "Analysis and Geometry on Graphs and Manifolds" (31.07.2017–04.08.2017, in Potsdam)*. accepted. 2019 (cit. on p. 29).

[Sto01] P. Stollmann. *Caught by Disorder. Bound States in Random Media.* 1st ed. Progress in Mathematical Physics 20. Birkhäuser Basel, 2001, pp. XVII, 166. ISBN: `978-1-4612-6644-0`. DOI: `10.1007/978-1-4612-0169-4` (cit. on p. 37).

[SV] C. Schumacher and I. Veselić. *Lifshitz asymptotics in case of monotone random breather potentials.* in preparation (cit. on p. 37).

[SV17] C. Schumacher and I. Veselić. "Lifshitz tails for Schrödinger operators with random breather potential". *Comptes Rendus Mathematique* 355.12 (2017), pp. 1307–1310. ISSN: 1631-073X. DOI: `10.1016/j.crma.2017.11.007`. arXiv: `1711.07318 [math-ph]` (cit. on p. 37).

[War13] S. Warzel. "Surprises in the phase diagram of the Anderson model on the Bethe lattice". *XVIIth International Congress on Mathematical Physics* (2013), pp. 239–253. DOI: `10.1142/9789814449243_0014`. arXiv: `1212.4367` (cit. on p. 23).

[Wei00] B. Weiss. "Sofic groups and dynamical systems". *Sankhya. The Indian Journal of Statistics. Series A* 62.3 (2000). Ergodic theory and harmonic analysis (Mumbai, 1999), pp. 350–359. ISSN: 0581-572X. URL: `http://sankhya.isical.ac.in/search/62a3/eh06fnl.html` (visited on 06/06/2018) (cit. on pp. 22, 25).

[Woe00] W. Woess. *Random Walks on Infinite Graphs and Groups*. Cambridge Tracts in Mathematics 138. Cambridge University Press, 2000. ISBN: 9780511470967. DOI: 10.1017/CBO9780511470967 (cit. on p. 24).

[Wri81] F. T. Wright. "The Empirical Discrepancy Over Lower Layers and a Related Law of Large Numbers". *The Annals of Probability* 9.2 (1981), pp. 323–329. Project Euclid: aop/1176994475 (cit. on p. 31).

Part II

Published papers and preprints

Chapter 4

A Glivenko–Cantelli theorem for almost additive functions on lattices

Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 127 (2017) 179–208

stochastic processes and their applications

www.elsevier.com/locate/spa

A Glivenko–Cantelli theorem for almost additive functions on lattices

Christoph Schumacher*, Fabian Schwarzenberger[1], Ivan Veselić

Fakultät für Mathematik, Technische Universität Chemnitz, 09107 Chemnitz, Germany

Received 11 September 2015; received in revised form 3 May 2016; accepted 5 June 2016
Available online 11 June 2016

Abstract

We develop a Glivenko–Cantelli theory for monotone, almost additive functions of i.i.d. sequences of random variables indexed by $\mathbb{Z}^d$. Under certain conditions on the random sequence, short range correlations are allowed as well. We have an explicit error estimate, consisting of a probabilistic and a geometric part. We apply the results to yield uniform convergence for several quantities arising naturally in statistical physics.
© 2016 Elsevier B.V. All rights reserved.

MSC: 60F99; 60B12; 62E20; 60K35

Keywords: Glivenko–Cantelli theory; Uniform convergence; Empirical measures; Large deviations; Statistical mechanics

1. Introduction

The classical Glivenko–Cantelli theorem states that the empirical cumulative distribution functions of an increasing set of independent and identically distributed random variables converge *uniformly* to the cumulative population distribution function almost surely. Due to its importance to applications, e.g. statistical learning theory, the Glivenko–Cantelli theorem

* Corresponding author.
E-mail addresses: Christoph.Schumacher@mathematik.tu-chemnitz.de (C. Schumacher), fabian.schwarzenberger@imise.uni-leipzig.de (F. Schwarzenberger), ivan.veselic@mathematik.tu-chemnitz.de (I. Veselić).
[1] Current address: Fakultät Informatik/Mathematik, Hochschule für Technik und Wirtschaft Dresden, 01069 Dresden, Germany.

http://dx.doi.org/10.1016/j.spa.2016.06.005
0304-4149/

is also called the "fundamental theorem of statistics". The theorem has initiated the study of so-called Glivenko–Cantelli classes as they feature, for instance, in the Vapnik–Chervonenkis theory [24]. Generalizations of the fundamental theorem rewrite the uniform convergence with respect to the real variable as convergence of a *supremum over a family (of sets or functions)* and widen the family over which the supremum is taken, making the statement "more uniform". However, there are limits to this uniformization: For instance, if the original distribution is continuous, there is no convergence if the supremum is taken w.r.t. the family of finite subsets of the reals. Thus, a balance has to be found between the class over which the supremum is taken and the distribution of the random variables, the details of which are often dictated by the application in mind. Another important extension are multivariate Glivenko–Cantelli theorems, where the i.i.d. random variables are generalized to i.i.d. random vectors with possibly dependent coordinates. Such results have been obtained e.g. in [18,22,2,30]. In contrast to the classical one-dimensional Glivenko–Cantelli theorem, where no assumptions on the underlying distribution is necessary, in the higher dimensional case, one has to exclude certain singular continuous measures, cf. Theorem 5.3. The multidimensional version of the Portmanteau theorem provides a hint why such conditions are necessary. We apply these results in Section 5.

To avoid confusion, let us stress that uniform convergence in the classical Glivenko–Cantelli Theorem and in our result involves discontinuous functions, so it is quite different to uniform convergence of differentiable functions, as it is encountered e.g. with power series.

In many models of statistical physics one shows that certain random quantities are self-averaging, i.e. possess a well defined non-random thermodynamic limit. This is not only true for random operators of Schrödinger type, cf. e.g. [21,16,27], but also for spin systems, cf. e.g. [5,6,28,29,1]. Note however that the latter papers, studying the free energy (and derived quantities), heavily use specific properties of the exponential function (entering the free energy) like convexity and smoothness. We lack these properties in the Glivenko–Cantelli setting and are thus dealing with a completely different situation. The geometric ingredients of the proof of the thermodynamic limit can be traced back to papers by Van Hove [23] and Følner [4]. This is why the exhaustion sets used in the thermodynamic limit are associated with their names.

While standard statistical problems concern i.i.d. samples, an independence assumption quickly appears unnatural in statistical physics. Neighboring entities in solid state models (such as atoms or spins) are unlikely to not influence each other. In order to treat physically relevant scenarios one introduces a geometry to encode location and adjacency relations between the random variables, which in turn are used to allow dependencies between close random variables. In the present paper we choose $\mathbb{Z}^d$ as our model of physical space, although our methods should apply to amenable groups as well, at least with an additional monotile condition. The focus on $\mathbb{Z}^d$ allows us to avoid technicalities of amenable groups with monotiles and can thus present our results in a simpler, more transparent manner. Furthermore, we can achieve more explicit error bounds due to the simple geometry of $\mathbb{Z}^d$.

Our main result is Theorem 2.6, which is a Glivenko–Cantelli type theorem for a class of monotone, almost additive functions and suitable distributions of the random variables, allowing spatial dependencies. Our precise hypotheses are spelled out in Assumption 2.1 and Definition 2.3. The theorem can be interpreted as a multi-dimensional ergodic theorem with values in the Banach space of right continuous and bounded functions with sup-norm, i.e. a uniform convergence result. Under slightly strengthened assumptions we obtain an explicit error term for the convergence, which is a sum of a geometric and a probabilistic part, cf. Theorem 2.8. While earlier Banach space valued ergodic theorems, e.g. [10,11], have been restricted to a finite set of colors, we are able to treat the real-valued case. To do this, we have to assume a monotonicity property, which is satisfied in most cases of interest. We obtain a more explicit convergence estimate than [10], as well. This is due to the fact that we assume a short range correlation condition, while [10] assumes the existence of limiting frequencies. The Glivenko–Cantelli result is applied to several examples from statistical physics in Sections 7 and 8. The flexibility and generality of our probabilistic model is displayed in the Appendix.

For the proof we use two sets of ideas. The first one concerns geometric approximation and tiling arguments for almost additive functions based on the amenability of the group $\mathbb{Z}^d$ going back to the mentioned seminal papers of Van Hove [23] and Følner [4]. In the context of Banach space valued ergodic theorems they have been used for instance in [9,12,13,10,11,17]. The second ingredient of the proof is multivariate Glivenko–Cantelli theory, as developed in [18,22,2,30]. Our Theorem 5.5 shows that in our setting a large deviations type estimate derived by Wright can be applied. The latter is a modification of the Dvoretzky–Kiefer–Wolfowitz inequality [3,14].

The structure of the paper is as follows: In Section 2 we present our notation and the two main theorems. Section 3 contains an intuitive sketch of the proof in the case $\mathbb{Z}^d = \mathbb{Z}$, Section 4 geometric tiling and approximation arguments, Section 5 multivariate Glivenko–Cantelli theory, Section 6 the proof of the main theorem, and Sections 7 and 8 examples.

2. Notation and main results

The geometric setting of this paper is given via $\mathbb{Z}^d$, which gives in a natural way rise to a graph $(\mathbb{Z}^d, \mathcal{E})$. Here, the set of edges $\mathcal{E}$ is the subset of the power set of $\mathbb{Z}^d$, consisting exactly of those $\{x, y\} \subseteq \mathbb{Z}^d$ which satisfy $\|y - x\|_1 = 1$. As usual $\|x\|_1 = \sum_{i=1}^d |x_i|$ denotes the ℓ^1-norm in $\mathbb{Z}^d$. By $\mathcal{F}$ we denote the (countable) set which consists of all finite subsets of $\mathbb{Z}^d$. For $\Lambda \in \mathcal{F}$, we write $|\Lambda|$ for the number of elements in Λ. The metric on the set of vertices $\mathfrak{d}\colon \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{N}_0$ is defined via the ℓ^1-norm, i.e. for $x, y \in \mathbb{Z}^d$ we set $\mathfrak{d}(x, y) := \|y - x\|_1$. For two sets $\Lambda_1, \Lambda_2 \subseteq \mathbb{Z}^d$ we write $\mathfrak{d}(\Lambda_1, \Lambda_2) := \min\{\mathfrak{d}(x, y) \mid x \in \Lambda_1, y \in \Lambda_2\}$. In the case that $\Lambda_1 = \{x\}$ contains only one element we write $\mathfrak{d}(x, \Lambda_2)$ for $\mathfrak{d}(\{x\}, \Lambda_2)$.

For $\Lambda \subseteq \mathbb{Z}^d$ we write $\Lambda + z := \{x + z \mid x \in \Lambda\}$. A *cube of side length* $n \in \mathbb{N}$ is a set which is given by $([0, n)^d \cap \mathbb{Z}^d) + z$ for some $z \in \mathbb{Z}^d$.

Using the metric $\mathfrak{d}$, we define for $r \in \mathbb{N}_0$ the r-boundary of a set $\Lambda \subseteq \mathbb{Z}^d$ by

$$\partial^r(\Lambda) := \{x \in \Lambda \mid \mathfrak{d}(x, \mathbb{Z}^d \setminus \Lambda) \leqslant r\} \cup \{x \in \mathbb{Z}^d \setminus \Lambda \mid \mathfrak{d}(x, \Lambda) \leqslant r\}.$$

Moreover, we set

$$\Lambda^r := \Lambda \setminus \partial^r(\Lambda) = \{x \in \Lambda \mid \mathfrak{d}(x, \mathbb{Z}^d \setminus \Lambda) > r\}. \tag{2.1}$$

If $(\Lambda_n)_{n\in\mathbb{N}}$ (or short (Λ_n)) is a sequence of subsets of $\mathbb{Z}^d$, we write $(\Lambda_n^r)_{n\in\mathbb{N}}$ or (Λ_n^r) instead of $((\Lambda_n)^r)_{n\in\mathbb{N}}$.

Note that for a cube Λ_n of side length n and $r \leqslant n/2$ we have

$$|\Lambda_n| = n^d, \qquad |\Lambda_n^r| = (n - 2r)^d \quad \text{and} \quad |\partial^r(\Lambda_n)| = (n + 2r)^d - (n - 2r)^d.$$

In the following we introduce colorings of the elements of $\mathbb{Z}^d$. To this end, let $\mathcal{A} \subseteq \mathbb{R}$ be the set of possible colors. The sample set, which describes the set of all possible colorings of $\mathbb{Z}^d$ is given by

$$\Omega := \mathcal{A}^{\mathbb{Z}^d} := \{\omega = (\omega_z)_{z\in\mathbb{Z}^d} \mid \omega_z \in \mathcal{A}\} \subseteq \mathbb{R}^{\mathbb{Z}^d}.$$

For each $z \in \mathbb{Z}^d$ we define the translation

$$\tau_z\colon \Omega \to \Omega, \qquad (\tau_z\omega)_x = \omega_{x+z}, \quad (z \in \mathbb{Z}^d), \tag{2.2}$$

i.e. $\mathbb{Z}^d$ acts on Ω via translations. For $\Lambda \in \mathcal{F}$ we set $\Omega_\Lambda := \mathcal{A}^\Lambda := \{\omega = (\omega_z)_{z\in\Lambda} \mid \omega_z \in \mathcal{A}\}$ and define $\Pi_\Lambda\colon \Omega \to \Omega_\Lambda$ by

$$\Pi_\Lambda(\omega) := \omega_\Lambda := (\omega_z)_{z\in\Lambda} \quad \text{for } \omega = (\omega_z)_{z\in\mathbb{Z}^d} \in \Omega.$$

We simplify $\Pi_z := \Pi_{\{z\}}$ for $z \in \mathbb{Z}^d$. As usual, $\mathcal{A}$ is equipped with the Borel σ-algebra $\mathcal{B}(\mathcal{A})$ inherited from $\mathbb{R}$. Let $\mathcal{B}(\Omega)$ be the product σ-algebra on Ω. Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{B}(\Omega))$ satisfying:

Assumption 2.1. (M1) *Translation invariance*: For each $z \in \mathbb{Z}^d$ we have $\mathbb{P} \circ \tau_z^{-1} = \mathbb{P}$.

(M2) *Existence of densities*: There are σ-finite measures μ_z, $z \in \mathbb{Z}^d$, on $(\mathcal{A}, \mathcal{B}(\mathcal{A}))$ such that for each $\Lambda \in \mathcal{F}$ the measure $\mathbb{P}_\Lambda := \mathbb{P} \circ \Pi_\Lambda^{-1}$ is absolutely continuous with respect to $\mu_\Lambda := \bigotimes_{z\in\Lambda} \mu_z$ on $(\Omega_\Lambda, \mathcal{B}(\Omega_\Lambda))$. We denote the density function by $\rho_\Lambda := \frac{\mathrm{d}\mathbb{P}_\Lambda}{\mathrm{d}\mu_\Lambda}$. The measure $\mathbb{P}_\Lambda$ is called a *marginal measure* of $\mathbb{P}$. It is defined on $(\Omega_\Lambda, \mathcal{B}(\Omega_\Lambda))$, where $\mathcal{B}(\Omega_\Lambda)$ is again the product σ-algebra.

(M3) *Independence at a distance*: There exists $r \geqslant 0$ such that for all $n \in \mathbb{N}$ and non-empty $\Lambda_1, \ldots, \Lambda_n \subseteq \mathbb{Z}^d$ with $\min\{\mathfrak{d}(\Lambda_i, \Lambda_j) \mid i \neq j\} > r$ we have $\rho_\Lambda = \prod_{j=1}^n \rho_{\Lambda_j}$, where $\Lambda = \bigcup_{j=1}^n \Lambda_j$.

Remark 2.2.
- The constant $r \geqslant 0$ in (M3) can be interpreted as *correlation length*. In particular, if $r = 0$ this property implies that the colors of the vertices are independent.
- Conditions (M2) and (M3) are trivially satisfied, if $\mathbb{P}$ is a product measure.
- For examples of measures $\mathbb{P}$ satisfying (M1), (M2) and (M3) we refer to Appendix.

In the following we deal with partial orderings on Ω and Ω_Λ, $\Lambda \in \mathcal{F}$. We write $\omega \leqslant \omega'$ for $\omega, \omega' \in \Omega$ if we have $\omega_z \leqslant \omega'_z$ for all $z \in \mathbb{Z}^d$, and analogously for Ω_Λ.

We consider the Banach space

$$\mathbb{B} := \{F\colon \mathbb{R} \to \mathbb{R} \mid F \text{ right-continuous and bounded}\},$$

which is equipped with supremum norm $\|\cdot\| := \|\cdot\|_\infty$.

We now introduce a certain class of $\mathbb{B}$-valued functions.

Definition 2.3. A function $f\colon \mathcal{F} \times \Omega \to \mathbb{B}$ is called *admissible* if the following conditions are satisfied

(i) translation invariance: For $\Lambda \in \mathcal{F}$, $z \in \mathbb{Z}^d$ and $\omega \in \Omega$ we have

$$f(\Lambda + z, \omega) = f(\Lambda, \tau_z\omega).$$

(ii) locality: For all $\Lambda \in \mathcal{F}$ and $\omega, \omega' \in \Omega$ satisfying $\Pi_\Lambda(\omega) = \Pi_\Lambda(\omega')$ we have

$$f(\Lambda, \omega) = f(\Lambda, \omega').$$

(iii) almost additivity: There exists a function $b = b_f\colon \mathcal{F} \to [0, \infty)$ such that for arbitrary $\omega \in \Omega$, pairwise disjoint $\Lambda_1, \ldots, \Lambda_n \in \mathcal{F}$ and $\Lambda := \bigcup_{i=1}^n \Lambda_i$ we have

$$\Big\| f(\Lambda, \omega) - \sum_{i=1}^n f(\Lambda_i, \omega) \Big\| \leqslant \sum_{i=1}^n b(\Lambda_i),$$

and b satisfies

- $b(\Lambda) = b(\Lambda + z)$ for arbitrary $\Lambda \in \mathcal{F}$ and $z \in \mathbb{Z}^d$,
- $\exists D = D_f > 0$ with $b(\Lambda) \leqslant D|\Lambda|$ for arbitrary $\Lambda \in \mathcal{F}$,
- $\lim_{i\to\infty} b(\Lambda_i)/|\Lambda_i| = 0$, if $(\Lambda_i)_{i\in\mathbb{N}}$ is a sequence of cubes with strictly increasing side length.

(iv) coordinatewise monotonicity: There exists a sign vector $s \in \{-1, 1\}^{\mathbb{Z}^d}$ such that for all $\Lambda \in \mathcal{F}$ and all $\omega, \omega' \in \Omega$, $z \in \Lambda$ and $E \in \mathbb{R}$ we have

$$\omega_{\Lambda\setminus\{z\}} = \omega'_{\Lambda\setminus\{z\}}, \omega_z \leqslant \omega'_z$$
$$\Longrightarrow \quad \begin{cases} f(\Lambda, \omega)(E) \leqslant f(\Lambda, \omega)(E) & \text{if } s(z) = 1, \text{ whereas} \\ f(\Lambda, \omega)(E) \geqslant f(\Lambda, \omega)(E) & \text{if } s(z) = -1. \end{cases}$$

(v) boundedness: We have

$$\sup_{\omega\in\Omega} \| f(\{0\}, \omega)\| < \infty.$$

Remark 2.4. • Property (ii) can be formulated as follows: $f(\Lambda, \cdot)\colon \Omega \to \mathbb{B}$ is Π_Λ-measurable. Property (ii) also enables us to define $f_\Lambda\colon \Omega_\Lambda \to \mathbb{B}$ by $f_\Lambda(\omega_\Lambda) := f(\Lambda, \omega)$ with $\Lambda \in \mathcal{F}$ and $\omega \in \Omega$. If $|\Lambda| = 1$, we can identify $\Omega_\Lambda = \mathcal{A}^\Lambda$ and $\mathcal{A}$. With this notation, (v) above translates into

$$\sup_{a\in\mathcal{A}} \| f_{\{0\}}(a)\| < \infty.$$

- In our examples in Sections 7 and 8, $b(\Lambda)$ from (iii) can be chosen proportional to $|\partial^1\Lambda|$, the size of the 1-boundary of $\Lambda \in \mathcal{F}$. Accordingly, we call the function $b = b_f$ *boundary term* for f. For quantitative estimates it is handy to require additionally that there exists $r' = r'_f \in \mathbb{N}$ and $D' = D'_f > 0$ such that

  $$b(\Lambda) \leqslant D'|\partial^{r'}\Lambda|$$

  for all cubes $\Lambda \in \mathcal{F}$. We call such a function b a *proper boundary term*.
- It is natural to call f with the property

  $$f(\Lambda, \omega) = \sum_{i=1}^{n} f(\Lambda_i, \omega)$$

  additive with respect to the disjoint decomposition $(\Lambda_i)_{i=1,\dots,n}$ of $\Lambda \in \mathcal{F}$. Hence, it is again natural to call (iii) almost additive, since the error term $\sum_{i=1}^{n} b(\Lambda_i)$ is in some sense small. Alternatively, (iii) could be called *low complexity* or *semi-locality* of f. The information contained in $f(\Lambda_1), \dots, f(\Lambda_n)$ does not differ much from the information contained in $f(\Lambda)$.
- Our examples in Sections 7 and 8 deal with antitone admissible functions, i.e. (iv) is satisfied with $s(z) = -1$ for all $z \in \mathbb{Z}^d$.
- If f is admissible, then

  $$K_f := \sup\left\{ \frac{\|f(\Lambda,\omega)\|}{|\Lambda|} \;\middle|\; \omega \in \Omega, \Lambda \in \mathcal{F}\setminus\{\emptyset\} \right\} < \infty. \tag{2.3}$$

To see this, we choose $\Lambda \in \mathcal{F}$ and $\omega \in \Omega$ arbitrarily and calculate as follows:

$$\begin{aligned}\|f(\Lambda,\omega)\| &\leqslant \Big\| f(\Lambda,\omega) - \sum_{z\in\Lambda} f(\{z\},\omega)\Big\| + \Big\|\sum_{z\in\Lambda} f(\{z\},\omega)\Big\| \\ &\leqslant \sum_{z\in\Lambda} b(\{z\}) + \sum_{z\in\Lambda} \|f(\{z\},\omega)\| \\ &\leqslant D|\Lambda| + \sum_{z\in\Lambda} \|f(\{0\},\tau_{-z}\omega)\| \leqslant \big(D + \sup_{\omega\in\Omega} \|f(\{0\},\omega)\|\big)|\Lambda|.\end{aligned}$$

Thus, $K_f \leqslant D + \sup_{\omega\in\Omega} \|f(\{0\},\omega)\| < \infty$.

Definition 2.5. For $K, D, D' > 0$ and $r' \in \mathbb{N}$, we form the set

$$\mathcal{U}_{K,D,D',r'} = \{f : \mathcal{F}\times\Omega \to \mathbb{B} \mid f \text{ admissible with } K_f \leqslant K, D_f \leqslant D, D'_f \leqslant D' \text{ and } r'_f \leqslant r'\}$$

where K_f, D_f, D'_f and r'_f are the constants from Definition 2.3 associated to f.

Let us state the main theorem of this paper.

Theorem 2.6. *Let $\mathcal{A} \subseteq \mathbb{R}$, $\Omega := \mathcal{A}^{\mathbb{Z}^d}$ and $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ a probability space such that $\mathbb{P}$ satisfies* (M1), (M2) *and* (M3) *with correlation length $r \in \mathbb{N}_0$, and let $f : \mathcal{F}\times\Omega \to \mathbb{B}$ be an admissible function. Let further $\Lambda_n := [0, n) \cap \mathbb{Z}^d$ for $n \in \mathbb{N}$. Then there exists a set $\tilde{\Omega} \in \mathcal{B}(\Omega)$ of full measure and a function $f^* \in \mathbb{B}$ such that for every $\omega \in \tilde{\Omega}$ we have*

$$\lim_{n\to\infty} \left\| \frac{f(\Lambda_n,\omega)}{|\Lambda_n|} - f^* \right\| = 0. \tag{2.4}$$

Remark 2.7.
- The following special case illustrates the relation to the Glivenko–Cantelli theorem. Let $\mathbb{P} := \bigotimes_{z\in\mathbb{Z}} \mu$ be a product measure on $\times_{\mathbb{Z}} \mathbb{R}$, where μ is a probability measure on $\mathbb{R}$, and let $f(\Lambda,\omega)(E) := \sum_{z\in\Lambda} \chi_{(-\infty,E]}(\omega_z)$ for $\Lambda \in \mathcal{F}$, $\omega \in \Omega$ and $E \in \mathbb{R}$. Then f is an admissible function. The quantities $f(\Lambda_n,\omega)(E)/|\Lambda_n| = |\Lambda_n|^{-1} \sum_{z\in\Lambda_n} \delta_{\omega_z}((-\infty, E])$ turn out to be the distribution functions of empirical measures. Theorem 2.6 now states that the empirical distribution functions converge uniformly. The limit f^* is of course the distribution function of μ: $f^*(E) = \mu((-\infty, E])$ for all $E \in \mathbb{R}$.
- We emphasize that the statement of Theorem 2.6 does not contain the measurability of the set

  $$\left\{\omega \in \Omega \;\middle|\; \lim_{n\to\infty} \left\| \tfrac{f(\Lambda_n,\omega)}{|\Lambda_n|} - f^* \right\| = 0\right\}.$$

  Instead, the claim is that this set contains a measurable subset $\tilde{\Omega}$ of full measure. If the probability space was complete, the above set would be measurable, too. We write all almost sure statements in explicit fashion, in order to avoid a completeness assumption and measurability issues.
- The limit function f^* inherits the boundedness from f, since there exists $\omega \in \Omega$ such that

  $$\|f^*\| \leqslant \limsup_{n\to\infty} \left(\left\| \tfrac{f(\Lambda_n,\omega)}{|\Lambda_n|} - f^* \right\| + \left\| \tfrac{f(\Lambda_n,\omega)}{|\Lambda_n|} \right\| \right) \leqslant K_f.$$

- Note that Theorem 2.6 readily generalizes to absolutely convergent linear combinations of admissible functions in the following sense. Let $K, \alpha_j \in \mathbb{R}$, $j \in \mathbb{N}$ such that $\sum_{j\in\mathbb{N}} |\alpha_j| < \infty$

and f_j, $j \in \mathbb{N}$, admissible functions such that $K_{f_j} \leqslant K$ for all $j \in \mathbb{N}$. For each $j \in \mathbb{N}$, Theorem 2.6 provides a limit function f_j^*. We let $f := \sum \alpha_j f_j$ and $f^* := \sum \alpha_j f_j^*$. Now fix $\varepsilon > 0$, find $J \in \mathbb{N}$ such that $\sum_{j=J}^{\infty} |\alpha_j| < \frac{\varepsilon}{2K}$ and note that by the triangle inequality for all $\omega \in \tilde{\Omega}$ we have

$$\begin{aligned}\limsup_{n\to\infty}\left\|\frac{f(\Lambda_n,\omega)}{|\Lambda_n|} - f^*\right\| &\leqslant \limsup_{n\to\infty}\Bigg(\sum_{j=1}^{J-1}|\alpha_j|\left\|\frac{f_j(\Lambda_n,\omega)}{|\Lambda_n|} - f_j^*\right\| \\ &\quad + \sum_{j=J}^{\infty}|\alpha_j|\left(\left\|\frac{f_j(\Lambda_n,\omega)}{|\Lambda_n|}\right\| + \|f_j^*\|\right)\Bigg) \\ &< \varepsilon.\end{aligned}$$

This shows that the coordinate-wise monotonicity (iv) can be somewhat weakened.

- By a Borel–Cantelli argument employing Theorem 5.3, the sequence of cubes Λ_n can in fact be replaced by an arbitrary van Hove sequence $(\Lambda_n)_{n\in\mathbb{N}}$, as long as for each Λ_n there exists a collection of translates which tiles $\mathbb{Z}^d$. The set $\tilde{\Omega}$ will depend on the sequence, of course.

Next, we state that for functions with proper boundary terms the convergence in Theorem 2.6 can be quantified by error terms. For the definition of the empirical measure $L_{m,n}^{r,\omega}$ see (4.6) and the notation $\langle f, \nu\rangle$ is introduced in (4.7).

Theorem 2.8. *Let $\mathcal{A} \subseteq \mathbb{R}$, $\Omega := \mathcal{A}^{\mathbb{Z}^d}$ and $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ a probability space such that $\mathbb{P}$ satisfies* (M1), (M2) *and* (M3) *with correlation length $r \in \mathbb{N}_0$, and let $\Lambda_n := [0, n) \cap \mathbb{Z}^d$ for $n \in \mathbb{N}$. Let $K, D, D' > 0$ and $r' \in \mathbb{N}$. Then, there exists a set $\tilde{\Omega} \in \mathcal{B}(\Omega)$ of full probability such that, for each $m, n \in \mathbb{N}$ with $n > 2m > 4r$ and $\omega \in \tilde{\Omega}$, we have*

$$\begin{aligned}&\sup_{f\in\mathcal{U}_{K,D,D',r'}}\left\|\frac{f(\Lambda_n,\omega)}{|\Lambda_n|} - f^*\right\| \\ &\quad\leqslant 2^{2d+1}\left(\frac{(2K+D)m^d + D'r'^d}{n-2m} + \frac{2(K+D)r^d + 3D'r'^d}{m-2r}\right) \\ &\qquad + \sup_{f\in\mathcal{U}_{K,D,D',r'}}\frac{\left\|\langle f_{\Lambda_m^r}, L_{m,n}^{r,\omega} - \mathbb{P}_{\Lambda_m^r}\rangle\right\|}{|\Lambda_m|},\end{aligned}$$

where f^* is the limit given by Theorem 2.6 applied to f. Furthermore, for all $m \in \mathbb{N}$ and $\omega \in \tilde{\Omega}$,

$$\lim_{n\to\infty}\sup_{f\in\mathcal{U}_{K,D,D',r'}}\left\|\langle f_{\Lambda_m^r}, L_{m,n}^{r,\omega} - \mathbb{P}_{\Lambda_m^r}\rangle\right\| = 0.$$

Even more, for each $\varepsilon > 0$ there exist $a = a(\varepsilon, m, K) > 0$ and $b = b(\varepsilon, m, K)$ such that for all $n \in \mathbb{N}$ there is a measurable set $\Omega(\varepsilon, n)$ with $\mathbb{P}(\Omega(\varepsilon, n)) \geqslant 1 - b\exp(-a\lfloor n/m\rfloor^d)$ and

$$\sup_{f\in\mathcal{U}_{K,D,D',r'}}\left\|\langle f_{\Lambda_m^r}, L_{m,n}^{r,\omega} - \mathbb{P}_{\Lambda_m^r}\rangle\right\| \leqslant \varepsilon \quad \text{for all } \omega \in \Omega(\varepsilon, n).$$

Remark 2.9. • It would be interesting to find an optimal explicit expression for a and b in terms of ε, m, and K.

- As before, the monotonicity can be weakened, see Remark 2.7. Note in particular, that a convex combination of functions in $\mathcal{U}_{K,D,D',r'}$ still obeys the quantitative estimates given by K, D, D' and r'. The statement of Theorem 2.8 remains valid for the convex combination since the geometric part is derived without the use of monotonicity and the argument from Remark 2.7 applies to the probabilistic part.

Corollary 2.10. *Under the conditions of Theorem 2.8, there exists a set* $\tilde{\Omega} \in \mathcal{B}(\Omega)$ *with* $\mathbb{P}(\tilde{\Omega}) = 1$ *such that for all* $\omega \in \tilde{\Omega}$*, we have*

$$\sup_{f \in \mathcal{U}_{K,D,D',r'}} \sup_{E \in \mathbb{R}} \left| \frac{f(\Lambda_n, \omega)(E)}{|\Lambda_n|} - f^*(E) \right| \xrightarrow{n\to\infty} 0. \tag{2.5}$$

If furthermore for an admissible f, $f(\Lambda_n, \omega) : \mathbb{R} \to \mathbb{R}$ *is an isotone function for all* Λ_n *and* $\omega \in \tilde{\Omega}$*, then the limit function* $f^* \in \mathbb{B}$ *is isotone, too. In particular, cumulative distribution functions are preserved.*

Proof. By Theorem 2.8, we have

$$0 \leqslant \lim_{n\to\infty} \sup_{f \in \mathcal{U}_{K,D,D',r'}} \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| \leqslant 2^{2d+1} \frac{2(K + D)r^d + 3Dr'^d}{m - 2r} \xrightarrow{m\to\infty} 0.$$

Recall that the norm in $\mathbb{B}$ is the sup norm $\|\cdot\| = \sup_{E \in \mathbb{R}} |\cdot(E)|$ to see (2.5).

If the functions $f_{n,\omega} := f(\Lambda_n, \omega)/|\Lambda_n| : \mathbb{R} \to \mathbb{R}$ are increasing, then for all $E, E' \in \mathbb{R}$ with $E < E'$ and $\varepsilon > 0$ we find $n \in \mathbb{N}$ such that $\|f_{n,\omega} - f^*\| < \varepsilon/2$ and

$$f^*(E) \leqslant f_n(E) + \varepsilon/2 \leqslant f_n(E') + \varepsilon/2 \leqslant f^*(E') + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, f^* is increasing, too. □

3. Illustration of the idea of proof

Let us consider the exemplary situation of dimension $d = 1$ and independently chosen colors, i.e., the constant r from (M3) equals 0. In this case, the idea of the proof of Theorem 2.6 is illustrated in the following line:

$$\frac{1}{mk} f\big([0, mk), \omega\big) \overset{(1)}{\approx} \frac{1}{m} \langle f_m, L^\omega_{m,mk} \rangle \overset{(2)}{\approx} \frac{1}{m} \langle f_m, \mathbb{P}_m \rangle \overset{(3)}{\approx} f^*, \tag{3.1}$$

where $0 \ll m \ll k$. Assume that $n = mk$ and $\Lambda_n = [0, n)$. Then the left hand side in (3.1) equals the approximant in Theorem 2.6. The function $f_m : \Omega_{[0,m)} \to \mathbb{B}$ is defined by $f_m(\omega) := f([0, m), \omega')$ for $\omega' \in \Pi^{-1}_{[0,m)}(\{\omega\})$, cf. Remark 2.4. $L^\omega_{m,n}(B)$ is the empirical probability measure counting the number of occurrences of elements of $B \in \mathcal{B}(\Omega_{[0,m)})$ at the positions $[jm, (j + 1)m)$, $j = 0, 1, \ldots, k - 1$ in ω, i.e.

$$L^\omega_{m,n} : \mathcal{B}(\Omega_{[0,m)}) \to [0, 1], \qquad L^\omega_{m,n} := \frac{1}{k} \sum_{j=0}^{k-1} \delta_{(\tau_{jm}\omega)_{[0,m)}} = \frac{1}{k} \sum_{j=0}^{k-1} \delta_{\Pi_{[0,m)}(\tau_{jm}\omega)}. \tag{3.2}$$

We use the shortcut notation

$$\langle f_m, L^\omega_{m,n} \rangle := \int_{\Omega_{[0,m)}} f_m(\omega')\, \mathrm{d}L^\omega_{m,n}(\omega') = \frac{1}{k} \sum_{j=0}^{k-1} f\big([jm, (j + 1)m), \omega\big).$$

Let us discuss the three approximation steps separately.

(1) The "$\overset{(1)}{\approx}$" means that the two expressions are getting close to each other if m increases. To see this we use almost additivity of an admissible function, cf. (iii) of Definition 2.3. The detailed calculations will be presented in Section 4.
(2) In the second step we compare the empirical measure $L^{\omega}_{m,mk}$ with the marginal measure $\mathbb{P}_m := \mathbb{P}_{[0,m)}$. The method of choice is a multivariate Glivenko–Cantelli theorem, which we apply in a version of DeHardt and Wright. In this particular situation it shows that for increasing k the expression $\langle f_m, L^{\omega}_{m,mk}\rangle$ converges to $\langle f_m, \mathbb{P}_m\rangle$ almost surely. This approximation step is explicitly discussed in Section 5.
(3) In the last step we make intensive use of almost additivity of f in order to obtain that $(\langle f_m, \mathbb{P}_m\rangle/m)_m$ is Cauchy sequence in $\mathbb{B}$. As $\mathbb{B}$ is a Banach space, we can identify the limit with an element $f^* \in \mathbb{B}$. The details are found in Section 6.

Remark 3.1 (*Frequencies*)**.** From the discussion of step (1) above it is clear that the empirical measure counts occurrences of patterns at the positions $[jm, (j+1)m)$ for $j = 0, 1, \ldots, k-1$. Thus, the corresponding sets are disjoint and their union covers the whole interval $[0, n)$, $n = mk$. In this sense, the present technique of counting occurrences differs from the counting in certain papers. For instance in [10,11], the authors count occurrences of patterns at each possible consecutive position, ignoring the fact that they may overlap. In our setting, this would correspond to the situation where the empirical measure is defined to count occurrences at all positions $[j, j+m)$, $j = 0, 1, \ldots, m(k-1)$, i.e.,

$$\bar{L}^{\omega}_{m,n} : \mathcal{B}(\Omega_{[0,m)}) \to [0,1], \qquad \bar{L}^{\omega}_{m,n} := \frac{1}{m(k-1)+1} \sum_{j=0}^{m(k-1)} \delta_{(\tau_j\omega)_{[0,m)}}. \tag{3.3}$$

However, both ways of counting can be related to each other. The link can be seen best by comparing with the average

$$\frac{1}{m(k-1)} \sum_{j=1}^{m(k-1)} \delta_{(\tau_j\omega)_{[0,m)}} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{k-1} \sum_{j=0}^{k-2} \delta_{(\tau_{jm+i}\omega)_{[0,m)}}, \tag{3.4}$$

where the first observation $\delta_{\omega_{[0,m)}}$ is discarded. Indeed, for large $n = mk$, the difference between $\bar{L}^{\omega}_{m,n}$ and (3.4) vanishes. The right hand side of (3.4) shows that $\bar{L}^{\omega}_{m,n}$ is essentially a convex combination of empirical measures of the type (3.2). As $k \to \infty$, all the empirical measures of type (3.2) in (3.4) converge to the same limit $\mathbb{P}_m$, rendering the convex combination harmless. Recall that in the approximation first the limit $k \to \infty$ and afterwards the limit $m \to \infty$ is performed. This shows that the empirical measure defined in (3.3) converges to the same limit as the empirical measures in (3.2).

4. Approximation via the empirical measure

In the following we show how to estimate an admissible function f in terms of the empirical measure. As in Theorem 2.6, let $\Lambda_n = ([0,n) \cap \mathbb{Z})^d$ for each $n \in \mathbb{N}$.

Our aim is to approximate for $m \ll n$ the set Λ_n using translates of the set Λ_m. To this end, we define the grid

$$T_{m,n} := \{t \in m\mathbb{Z}^d \mid \Lambda_m + t \subseteq \Lambda_n\}. \tag{4.1}$$

Thus, we have $|T_{m,n}| = \lfloor n/m \rfloor^d$, $\Lambda_{\lfloor n/m \rfloor m} = \dot{\bigcup}_{t \in T_{m,n}} (\Lambda_m + t) = \Lambda_m + T_{m,n}$, and

$$\Lambda_n \backslash \Lambda_{\lfloor n/m \rfloor m} \subseteq \partial^m(\Lambda_n) \quad \text{or equivalently} \quad \Lambda_n^m \subseteq \Lambda_{\lfloor n/m \rfloor m}. \tag{4.2}$$

As in Remark 2.4, define for an admissible f and $\Lambda \in \mathcal{F}$ the function

$$f_\Lambda \colon \Omega_\Lambda \to \mathbb{B}, \qquad f_\Lambda(\omega) := f(\Lambda, \omega') \quad \text{where } \omega' \in \Pi_\Lambda^{-1}(\{\omega\}). \tag{4.3}$$

By locality (ii) of Definition 2.3, f_Λ is well-defined. In the case $\Lambda = \Lambda_n$, we write

$$f_n := f_{\Lambda_n} \quad \text{and} \quad f_n^m := f_{\Lambda_n^m} \tag{4.4}$$

for $m \in \mathbb{N}_0$. Next, we introduce the empirical measure $L_{m,n}^\omega$ by setting for $\omega \in \Omega$ and $m, n \in \mathbb{N}$:

$$L_{m,n}^\omega \colon \mathcal{B}(\Omega_{\Lambda_m}) \to [0,1], \qquad L_{m,n}^\omega = \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \delta_{(\tau_t \omega)_{\Lambda_m}}. \tag{4.5}$$

Here, $\delta_\omega \colon \mathcal{B}(\Omega_{\Lambda_m}) \to [0,1]$ is the point measure on $\omega \in \Omega_{\Lambda_m}$. In the same manner, we define $L_{m,n}^{r,\omega}$ as an adaption of $L_{m,n}^\omega$ which ignores the r-boundary of Λ_m. The precise definition is the following: for $r \in \mathbb{N}_0$ we set

$$L_{m,n}^{r,\omega} \colon \mathcal{B}(\Omega_{\Lambda_m^r}) \to [0,1], \qquad L_{m,n}^{r,\omega} = \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \delta_{(\tau_t \omega)_{\Lambda_m^r}}. \tag{4.6}$$

The variable r we used here will eventually be the constant from (M3), but here in Section 4 we do not need that specific value.

As illustrated before in Section 3 we use for $\Lambda \in \mathcal{F}$, a bounded and measurable $f \colon \Omega_\Lambda \to \mathbb{B}$, and a measure ν on $(\Omega_\Lambda, \mathcal{B}(\Omega_\Lambda))$ the notation

$$\langle f, \nu \rangle := \int_{\Omega_\Lambda} f(\omega)\, \mathrm{d}\nu(\omega). \tag{4.7}$$

Lemma 4.1. *Recall $\Lambda_n := ([0,n) \cap \mathbb{Z})^d$. For any admissible function $f \colon \mathcal{F} \times \Omega \to \mathbb{B}$ we have, for all $\omega \in \Omega$ and all $n, m, r \in \mathbb{N}$ with $n > 2m$,*

$$\begin{aligned} \left\| \frac{f(\Lambda_n, \omega)}{n^d} - \frac{\langle f_m^r, L_{m,n}^{r,\omega} \rangle}{m^d} \right\| \leqslant{}& \frac{b(\Lambda_{\lfloor n/m \rfloor m})}{|\Lambda_{\lfloor n/m \rfloor m}|} + \frac{(2K_f + D)|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|} \\ &+ \frac{b(\Lambda_m) + b(\Lambda_m^r) + (K_f + D)|\partial^r(\Lambda_m)|}{|\Lambda_m|}. \end{aligned} \tag{4.8}$$

Moreover,

$$\lim_{m \to \infty} \lim_{n \to \infty} \left\| \frac{f(\Lambda_n, \omega)}{n^d} - \frac{\langle f_m^r, L_{m,n}^{r,\omega} \rangle}{m^d} \right\| = 0. \tag{4.9}$$

Proof. Let $\omega \in \Omega$ and $n, m, r \in \mathbb{N}$ be given such that $n > 2m$. This condition ensures that $\Lambda_n^m \neq \varnothing$. First we verify (4.8), and afterwards we show that this implies (4.9). By the triangle inequality we obtain

$$\begin{aligned}\left\|\frac{f(\Lambda_n,\omega)}{n^d}-\frac{\langle f_m^r,L_{m,n}^{r,\omega}\rangle}{m^d}\right\| &\leqslant \left\|\frac{f(\Lambda_n,\omega)}{|\Lambda_n|}-\frac{f(\Lambda_n,\omega)}{|\Lambda_{\lfloor n/m\rfloor m}|}\right\| \\ &+\left\|\frac{f(\Lambda_n,\omega)-f(\Lambda_{\lfloor n/m\rfloor m},\omega)}{|\Lambda_{\lfloor n/m\rfloor m}|}\right\|+\left\|\frac{f(\Lambda_{\lfloor n/m\rfloor m},\omega)}{|\Lambda_{\lfloor n/m\rfloor m}|}-\frac{\langle f_m,L_{m,n}^{\omega}\rangle}{m^d}\right\| \\ &+\frac{\|\langle f_m,L_{m,n}^{\omega}\rangle-\langle f_m^r,L_{m,n}^{r,\omega}\rangle\|}{m^d}. \end{aligned} \tag{4.10}$$

We bound the four terms on the right hand side separately. To estimate the first term, we use $|\Lambda_{\lfloor n/m\rfloor m}|\geqslant|\Lambda_n^m|$, see (4.2), and obtain

$$0\leqslant\frac{1}{|\Lambda_{\lfloor n/m\rfloor m}|}-\frac{1}{|\Lambda_n|}\leqslant\frac{1}{|\Lambda_n^m|}-\frac{1}{|\Lambda_n|}=\frac{|\Lambda_n|-|\Lambda_n^m|}{|\Lambda_n||\Lambda_n^m|}\leqslant\frac{|\partial^m(\Lambda_n^m)|}{|\Lambda_n||\Lambda_n^m|}.$$

Applying the bound given by (2.3) in Remark 2.4, we get

$$\left\|\frac{f(\Lambda_n,\omega)}{|\Lambda_n|}-\frac{f(\Lambda_n,\omega)}{|\Lambda_{\lfloor n/m\rfloor m}|}\right\|\leqslant K_f\frac{|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|}. \tag{4.11}$$

In order to find an appropriate upper bound for the second term in (4.10) we use almost additivity (iii), the inclusion (4.2) and $\hat{\Lambda}_{m,n}:=\Lambda_n\setminus\Lambda_{\lfloor n/m\rfloor m}$ to obtain

$$\begin{aligned}\left\|\frac{f(\Lambda_n,\omega)-f(\Lambda_{\lfloor n/m\rfloor m},\omega)}{|\Lambda_{\lfloor n/m\rfloor m}|}\right\| &\leqslant\frac{b(\Lambda_{\lfloor n/m\rfloor m})}{|\Lambda_{\lfloor n/m\rfloor m}|}+\frac{b(\hat{\Lambda}_{n,m})}{|\Lambda_{\lfloor n/m\rfloor m}|}+\frac{\|f(\hat{\Lambda}_{n,m},\omega)\|}{|\Lambda_{\lfloor n/m\rfloor m}|} \\ &\leqslant\frac{b(\Lambda_{\lfloor n/m\rfloor m})}{|\Lambda_{\lfloor n/m\rfloor m}|}+\frac{D|\hat{\Lambda}_{n,m}|}{|\Lambda_{\lfloor n/m\rfloor m}|}+\frac{K_f|\hat{\Lambda}_{n,m}|}{|\Lambda_{\lfloor n/m\rfloor m}|} \\ &\leqslant\frac{b(\Lambda_{\lfloor n/m\rfloor m})}{|\Lambda_{\lfloor n/m\rfloor m}|}+\frac{(K_f+D)|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|}. \end{aligned} \tag{4.12}$$

To approximate the third term in (4.10), we calculate using translation invariance (i) of admissible functions

$$\begin{aligned}\langle f_m,L_{m,n}^{\omega}\rangle &=\int_{\Omega_{\Lambda_m}}f_m(\omega')\,\mathrm{d}L_{m,n}^{\omega}(\omega')=\frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}}\int_{\Omega_{\Lambda_m}}f_m(\omega')\,\mathrm{d}\delta_{(\tau_t\omega)_{\Lambda_m}}(\omega') \\ &=\frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}}f_m((\tau_t\omega)_{\Lambda_m})=\frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}}f(\Lambda_m+t,\omega). \end{aligned} \tag{4.13}$$

This and (iii) of Definition 2.3 give

$$\begin{aligned}\left\|\frac{f(\Lambda_{\lfloor n/m\rfloor m},\omega)}{|\Lambda_{\lfloor n/m\rfloor m}|}-\frac{\langle f_m,L_{m,n}^{\omega}\rangle}{|\Lambda_m|}\right\| &=\frac{1}{|T_{m,n}||\Lambda_m|}\left\|f(\Lambda_{\lfloor n/m\rfloor m},\omega)-\sum_{t\in T_{m,n}}f(\Lambda_m+t,\omega)\right\| \\ &\leqslant\frac{1}{|T_{m,n}||\Lambda_m|}\sum_{t\in T_{m,n}}b(\Lambda_m+t)=\frac{b(\Lambda_m)}{|\Lambda_m|}. \end{aligned} \tag{4.14}$$

Finally, we estimate the fourth term. In the same way as in (4.13) we obtain

$$\langle f_m^r,L_{m,n}^{r,\omega}\rangle=\frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}}f(\Lambda_m^r+t,\omega).$$

Application of the triangle inequality, $\Lambda_m \backslash \Lambda_m^r = \Lambda_m \cap \partial^r(\Lambda_m) \subseteq \partial^r(\Lambda_m)$ and (iii) of Definition 2.3 as well as (2.3) lead to

$$\begin{aligned}
\|\langle f_m, L_{m,n}^{\omega}\rangle - \langle f_m^r, L_{m,n}^{r,\omega}\rangle\| &\leqslant \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \|f(\Lambda_m + t, \omega) - f(\Lambda_m^r + t, \omega)\| \\
&\leqslant \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \big(b(\Lambda_m^r) + b(\Lambda_m \backslash \Lambda_m^r) + \|f((\Lambda_m \backslash \Lambda_m^r) + t, \omega)\|\big) \\
&\leqslant b(\Lambda_m^r) + (K_f + D)|\partial^r(\Lambda_m)|. \qquad (4.15)
\end{aligned}$$

It remains to combine (4.10) with the bounds (4.11), (4.12), (4.14) and (4.15) to obtain (4.8).

Let us turn to (4.9). As required, we first perform the limit $n \to \infty$. In (4.8), the bounding terms affected by this limit vanish, due to property (iii) and the fact that $\mathbb{Z}^d$ is amenable:

$$\lim_{n\to\infty} \left(\frac{b(\Lambda_{\lfloor n/m \rfloor m})}{|\Lambda_{\lfloor n/m \rfloor m}|} + \frac{(2K_f + D)|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|} \right) = 0.$$

Secondly, we let $m \to \infty$. Since $b(\Lambda_m^r)/|\Lambda_m| \leqslant b(\Lambda_m^r)/|\Lambda_m^r|$ for $m > 2r$, this takes care of the remaining terms of the upper bound in (4.8).

$$\lim_{m\to\infty} \frac{b(\Lambda_m) + b(\Lambda_m^r) + (K_f + D)|\partial^r(\Lambda_m)|}{|\Lambda_m|} = 0.$$

Thus, (4.9) follows. □

Remark 4.2. Let us emphasize that the statement of the lemma is not an “almost sure”-statement, but rather holds for all $\omega \in \Omega$.

5. Application of multivariate Glivenko–Cantelli theory

We briefly restate multivariate Glivenko–Cantelli results in Theorem 5.3 and apply this result to our setting in Theorem 5.6. To do so, we need some notions concerning monotonicity in $\mathbb{R}^k$.

Definition 5.1. Let $k \in \mathbb{N}$.

- Let $s \in \{-1, 1\}^k$. The closed *cone* $\mathcal{C}_s$ with *sign vector* s is the set
$$\mathcal{C}_s := \{x = (x_j)_{j=1,\dots,k} \in \mathbb{R}^k \mid \forall j \in \{1, \dots, k\} \colon x_j s_j \geqslant 0\}.$$
The closed cone with sign vector s and *apex* $x \in \mathbb{R}^k$ is $\mathcal{C}_s(x) := x + \mathcal{C}_s$.
- A function $f \colon \mathbb{R}^k \to \mathbb{R}$ is *monotone*, if it is monotone in each coordinate, i.e. there exists $s \in \{-1, 1\}^k$ such that, for all $x, y \in \mathbb{R}^k$,
$$y \in \mathcal{C}_s(x) \Longrightarrow f(y) \geqslant f(x).$$
- A set $\Upsilon \subseteq \mathbb{R}^k$ is a *monotone graph*, if there exists a sign vector $s \in \{-1, 1\}^k$ such that, for all $x \in \Upsilon$,
$$\Upsilon \cap \mathcal{C}_s(x) \subseteq \partial\mathcal{C}_s(x),$$
where ∂C denotes the boundary of C in $\mathbb{R}^k$.
- A set $\Upsilon \subseteq \mathbb{R}^k$ is a *strictly monotone graph*, if there exists a sign vector $s \in \{-1, 1\}^k$ such that, for all $x \in \Upsilon$,
$$\Upsilon \cap \mathcal{C}_s(x) = \{x\}.$$

Remark 5.2. • This notion of monotonicity is compatible with (iv) in Definition 2.3.

- We want to emphasize that in the above definition a second meaning of the notion of a *graph* was used. In Section 2 a graph was introduced as a pair consisting of a set of vertices and a set of edges. In contrast to that, Definition 5.1 states that a *monotone graph* is a certain subset of $\mathbb{R}^k$. In order to distinguish both meanings we will always insert the term *monotone* when speaking about subsets of $\mathbb{R}^k$.

The following theorem is proven in [30, Theorems 1 and 2]. Recall that the continuous part μ_c of a measure μ on $\mathbb{R}^k$ is given by $\mu_c(A) := \mu(A) - \sum_{x \in A} \mu\{x\}$ for all Borel sets $A \in \mathcal{B}(\mathbb{R}^k)$.

Theorem 5.3 (*DeHardt, Wright*). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_t : \Omega \to \mathbb{R}^k$, $t \in \mathbb{N}$, independent and identically distributed random variables with distribution μ, i.e., $\mu(A) := \mathbb{P}(X_1 \in A)$ for all $A \in \mathcal{B}(\mathbb{R}^k)$. For each $J \subseteq \{1, \dots, k\}$, $J \neq \varnothing$, let μ^J be the distribution of the vector $(X_1^j)_{j \in J}$ consisting of the coordinates $j \in J$ of the vector $X_1 = (X_1^j)_{j \in \{1,\dots,k\}}$, i.e. a marginal of μ. We denote by*

$$L_n : \Omega \times \mathcal{B}(\mathbb{R}^k) \to \mathbb{R}, \qquad L_n^{(\omega)}(A) := \frac{1}{n} \sum_{t=1}^{n} \delta_{X_t(\omega)}$$

the empirical distribution corresponding to the sample $(X_1, \dots, X_n)$, $n \in \mathbb{N}$. Fix further $M > 0$ and let

$$\mathcal{M} := \{f : \mathbb{R}^k \to \mathbb{R} \mid f \text{ monotone and } \sup |f(\mathbb{R}^k)| \leqslant M\}.$$

Then the following assertions are equivalent:

(i) *For all $J \subseteq \{1, \dots, k\}$, $J \neq \varnothing$, the continuous part μ_c^J of the marginal μ^J of μ vanishes on every strictly monotone graph $\Upsilon \subseteq \mathbb{R}^J$:*

$$\mu_c^J(\Upsilon) = 0.$$

(ii) *There exists a set $\Omega' \in \mathcal{A}$ of full probability $\mathbb{P}(\Omega') = 1$ such that, for all $\omega \in \Omega'$,*

$$\sup_{f \in \mathcal{M}} |\langle f, L_n^{(\omega)} - \mu \rangle| \xrightarrow{n \to \infty} 0.$$

(iii) *For all $\varepsilon > 0$, there are $a = a(\varepsilon) > 0$ and $b = b(\varepsilon) > 0$ such that for all $n \in \mathbb{N}$ there exists an $\Omega_{\varepsilon,n} \in \mathcal{A}$, such that for all $\omega \in \Omega_{\varepsilon,n}$, we have*

$$\sup_{f \in \mathcal{M}} |\langle f, L_n^{(\omega)} - \mu \rangle| \leqslant \varepsilon \quad \text{and} \quad \mathbb{P}(\Omega_{\varepsilon,n}) \geqslant 1 - b \exp(-an).$$

Remark 5.4. Note that if we knew that the set $\{\omega \in \Omega \mid \sup_{f \in \mathcal{M}} |\langle f, L_n^{(\omega)} - \mu \rangle| \geqslant \varepsilon\}$ was measurable, we could rephrase (iii) as follows. For all $\varepsilon > 0$, the probabilities $\mathbb{P}(\sup_{f \in \mathcal{M}} |\langle f, L_n^{(\omega)} - \mu \rangle| \geqslant \varepsilon)$ converge exponentially fast to zero as $n \to \infty$.

We provide a sufficient condition for (i) in Theorem 5.3 and apply the theorem to our setting. The idea to use product measures in the context of Glivenko–Cantelli type theorems appears already in [22].

Theorem 5.5. *Let μ be a measure on $\mathbb{R}^k$ which is absolutely continuous with respect to a product measure $\bigotimes_{j=1}^{k} \mu_j$ on $\mathbb{R}^k$, where μ_j, $j \in \{1, \dots, k\}$ are measures on $\mathbb{R}$. Then, for each strictly monotone graph $\Upsilon \subseteq \mathbb{R}^k$ we have $\mu_c(\Upsilon) = 0$, where μ_c is the continuous part of μ. Moreover,* (i) *from Theorem 5.3 is satisfied.*

Proof. Let ρ be the density of μ with respect to $\bigotimes_{j=1}^{k}\mu_j$. We define the set of atoms of μ,

$$S := \{x \in \mathbb{R}^k \mid \mu\{x\} > 0\}, \quad \text{and} \quad S_j := \{x_j \in \mathbb{R} \mid \mu_j\{x_j\} > 0\} \quad (j \in \{1, \ldots, k\}).$$

Then we have $S \subseteq S_1 \times \cdots \times S_k$, and for each $x = (x_1, \ldots, x_k) \in S_1 \times \cdots \times S_k$, we have

$$\mu\{x\} = \rho(x) \prod_{j=1}^{k} \mu_j\{x_j\}. \tag{5.1}$$

This implies in particular that for all $x \in S_1 \times \cdots \times S_k \setminus S$, we have $\rho(x) = 0$.

In order to prove $\mu_c(\Upsilon) = 0$ it is sufficient to show

$$\mu(\Upsilon) = \sum_{x \in S \cap \Upsilon} \mu\{x\}. \tag{5.2}$$

We will prove this by induction over k. If $k = 1$ then a strictly monotone graph is a singleton, i.e. $\Upsilon = \{x\}$ for some $x \in \mathbb{R}$. Thus, (5.2) holds true. In the case $k > 1$ we assume that (5.2) holds for $k - 1$ and proceed by disintegration. Note that, for $z \in \mathbb{R}$, the cross section $\Upsilon_z := \{y \in \mathbb{R}^{k-1} \mid (y, z) \in \Upsilon\}$ is itself a strictly monotone graph in $\mathbb{R}^{k-1}$. Using the cross section $\rho_z \colon \mathbb{R}^{k-1} \to \mathbb{R}$, $\rho_z(y) := \rho(y, z)$, $z \in \mathbb{R}$, of the density, we define the cross section $\mu_z := \rho_z \bigotimes_{j=1}^{k-1} \mu_j$ of the measure μ. By Fubini's Theorem, the disintegration of μ is

$$\mu(\mathrm{d}(y, z)) = \rho_z(y) \bigotimes_{j=1}^{k-1} \mu_j(\mathrm{d}y_j) \otimes \mu_k(\mathrm{d}z).$$

By the induction hypothesis we obtain

$$\begin{aligned} \mu(\Upsilon) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}^{k-1}} \chi_{\Upsilon_z}(y)\, \mu_z(\mathrm{d}y) \right) \mu_k(\mathrm{d}z) \\ &= \int_{\mathbb{R}} \mu_z(\Upsilon_z)\, \mu_k(\mathrm{d}z) = \int_{\mathbb{R}} \sum_{y \in \bar{S} \cap \Upsilon_z} \mu_z\{y\}\, \mu_k(\mathrm{d}z), \end{aligned} \tag{5.3}$$

where $\bar{S} := S_1 \times \cdots \times S_{k-1}$. The next aim is to show that the set $\mathcal{Z} := \{z \in \mathbb{R} \mid \bar{S} \cap \Upsilon_z \neq \varnothing\}$ is countable. To this end, we will use that $\bar{S}$ is countable, define two mappings

$$\varphi \colon \bar{S} \to (\bar{S} \times \mathbb{R}) \cap \Upsilon \quad \text{and} \quad \psi \colon (\bar{S} \times \mathbb{R}) \cap \Upsilon \to \mathcal{Z}$$

and show that they are surjective. We first define φ. Let $(y, z), (y, z') \in (\bar{S} \times \mathbb{R}) \cap \Upsilon$ be given and assume without loss of generality that $z \leqslant z'$. Let $s \in \{-1, 1\}^k$ be the sign vector of Υ from Definition 5.1, and, again without loss of generality, consider the case $s(k) = 1$. Then we have

$$\mathcal{C}_s(y, z) \cap \Upsilon = \{(y, z)\} \quad \text{and} \quad \mathcal{C}_s(y, z') \cap \Upsilon = \{(y, z')\}.$$

As $z \leqslant z'$ and $s(k) = 1$, we have $\mathcal{C}_s(y, z) \supseteq \mathcal{C}_s(y, z')$, such that we obtain

$$\{(y, z)\} = \mathcal{C}_s(y, z) \cap \Upsilon \supseteq \mathcal{C}_s(y, z') \cap \Upsilon = \{(y, z')\}.$$

This shows that if $y \in \bar{S}$ is such that there exists an element $z \in \mathbb{R}$ with $(y, z) \in \Upsilon$, then this z is unique. Let $h \in (\bar{S} \times \mathbb{R}) \cap \Upsilon$ be arbitrary but fixed and set

$$\varphi \colon \bar{S} \to (\bar{S} \times \mathbb{R}) \cap \Upsilon, \qquad \varphi(y) := \begin{cases} (y, z) & \text{if } (y, z) \in \Upsilon, \text{ and} \\ h & \text{if } (\{y\} \times \mathbb{R}) \cap \Upsilon = \varnothing. \end{cases}$$

This φ is well-defined and surjective. The mapping ψ is defined by

$$\psi : (\bar{S} \times \mathbb{R}) \cap \Upsilon \to \mathcal{Z}, \quad \psi(y, z) := z.$$

To check that ψ is surjective let $z \in \mathcal{Z}$ be given. Then there exists $y \in \bar{S} \cap \Upsilon_z$. Thus, by definition of Υ_z we have $(y, z) \in \Upsilon$ and $(y, z) \in \bar{S} \times \mathbb{R}$. This shows that (y, z) is in the domain of ψ and $\psi(y, z) = z$.

The surjectivity of φ and ψ and the fact that $\bar{S}$ is countable show that $\mathcal{Z}$ is countable. Therefore the last integral in (5.3) is actually a sum:

$$\mu(\Upsilon) = \int_{\mathbb{R}} \sum_{y \in \bar{S} \cap \Upsilon_z} \mu_z\{y\}\, \mu_k(\mathrm{d}z) = \sum_{z \in S_k} \sum_{y \in \bar{S} \cap \Upsilon_z} \mu_z\{y\}\, \mu_k\{z\} = \sum_{x \in S \cap \Upsilon} \mu\{x\}.$$

Here, the last equality follows from (5.1), $\bigcup_{z \in S_k} (\bar{S} \cap \Upsilon_z) \times \{z\} \supseteq S \cap \Upsilon$, and the fact that ρ vanishes on

$$\bigcup_{z \in S_k} (\bar{S} \cap \Upsilon_z) \times \{z\} \backslash (S \cap \Upsilon) \subseteq S_1 \times \cdots \times S_k \backslash S.$$

This finishes the induction and we obtained (5.2) and $\mu_c(\Upsilon) = 0$.

Let $J \subseteq \{1, \ldots, k\}$ such that $J \neq \varnothing$ and $J^c := \{1, \ldots, k\} \backslash J \neq \varnothing$. Define $\rho^J : \mathbb{R}^J \to \mathbb{R}$ via

$$\rho^J(x^J) := \int_{\mathbb{R}^{J^c}} \rho(x)\, \mathrm{d} \bigotimes_{j \in J^c} \mu_j(x^{J^c}),$$

where $x = (x^J, x^{J^c}) \in \mathbb{R}^J \times \mathbb{R}^{J^c}$. The function ρ^J is the density of the marginal μ^J of μ with respect to $\bigotimes_{j \in J} \mu_j$, since by Fubini for all $A \in \mathcal{B}(\mathbb{R}^J)$

$$\mu^J(A) = \int_{\mathbb{R}^k} \chi_A(x^J) \rho(x)\, \mathrm{d} \bigotimes_{j=1}^{k} \mu_j(x) = \int_A \rho^J(x^J)\, \mathrm{d} \bigotimes_{j \in J} \mu_j(x^J).$$

Thus, the above calculation applies for all marginals of μ, too. This shows (i) from Theorem 5.3. $\square$

Now we approximate the empirical measure $L^{r,\omega}_{m,n}$ using the measure $\mathbb{P}^r_m$, see step (2) in Section 3. The connection to Assumption 2.1 is established by Theorem 5.5. As announced before we apply the multivariate Glivenko–Cantelli Theorem 5.3 for the proof of Theorem 5.6.

Theorem 5.6. *Let $\Lambda_n := [0, n) \cap \mathbb{Z}^d$, $n \in \mathbb{N}$, a set $\mathcal{A} \subseteq \mathbb{R}$, $\Omega := \mathcal{A}^{\mathbb{Z}^d}$, a probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ such that $\mathbb{P}$ satisfies* (M1), (M2) *and* (M3) *and an admissible function f be given. Besides this let for $m, n \in \mathbb{N}$ and $\omega \in \Omega$ the empirical measure $L^{r,\omega}_{m,n}$ be given as in (4.6) and let $\mathbb{P}^r_m := \mathbb{P}_{\Lambda^r_m}$ be the marginal measure, where r is the constant given by* (M3). *Then there exists a set $\tilde{\Omega} \in \mathcal{B}(\Omega)$ of full measure, such that for all $\omega \in \tilde{\Omega}$ and all $m \in \mathbb{N}$:*

$$\lim_{n \to \infty} \| \langle f^r_m, L^{r,\omega}_{m,n} - \mathbb{P}^r_m \rangle \| = 0. \tag{5.4}$$

Furthermore, for $K, D, D' > 0$ and $r' \in \mathbb{N}$, we have for all $\omega \in \tilde{\Omega}$ and $m \in \mathbb{N}$

$$\lim_{n \to \infty} \sup_{f \in \mathcal{U}_{K,D,D',r'}} \| \langle f^r_m, L^{r,\omega}_{m,n} - \mathbb{P}^r_m \rangle \| = 0. \tag{5.5}$$

Additionally, for each $\varepsilon > 0$ *there exist* $a = a(\varepsilon, m, K) > 0$ *and* $b = b(\varepsilon, m, K)$ *such that for all* $n \in \mathbb{N}$ *there is a measurable set* $\Omega(\varepsilon, n)$ *with* $\mathbb{P}(\Omega(\varepsilon, n)) \geqslant 1 - b\exp(-a\lfloor n/m\rfloor^d)$ *and*

$$\sup_{f \in \mathcal{U}_{K,D,D',r'}} \|\langle f_m^r, L_{m,n}^{r,\omega} - \mathbb{P}_m^r\rangle\| \leqslant \varepsilon \quad \text{for all } \omega \in \Omega(\varepsilon, n). \tag{5.6}$$

Proof. Let $m \in \mathbb{N}$ be given. We set $k := |\Lambda_m^r|$ and embed $\Omega_{\Lambda_m^r} \subseteq \mathbb{R}^k$. Fix an admissible function f. For each $E \in \mathbb{R}$, there exists a monotone and bounded function $g_{m,E}^r : \mathbb{R}^k \to \mathbb{R}$ which extends $f_m^r(\cdot)(E) : \Omega_{\Lambda_m^r} \to \mathbb{R}$, i.e. $f_m^r(\omega)(E) = g_{m,E}^r(\omega)$ for all $\omega \in \Omega_{\Lambda_m^r}$. In fact, $g_{m,E}^r$ can be bounded by kK_f, where K_f is the constant introduced in (2.3). Thus, the set $\mathcal{M}_f := \{g_{m,E}^r \mid E \in \mathbb{R}\}$ is monotone and bounded by kK_f, see Remark 2.4.

In order to apply the Glivenko–Cantelli Theorem 5.3, we enumerate $[0, \infty)^d \cap m\mathbb{Z}^d$ with a sequence $(t_\ell)_{\ell\in\mathbb{N}}$ such that, for all $q \in \mathbb{N}$,

$$\{t_1, \ldots, t_{q^d}\} = [0, mq)^d \cap m\mathbb{Z}^d.$$

Consider further for each $\ell \in \mathbb{N}$ the mapping

$$X_\ell^r : \Omega \to \Omega_{\Lambda_m^r} \subseteq \mathbb{R}^k, \qquad X_\ell^r(\omega) := \Pi_{\Lambda_m^r}(\tau_{t_\ell}^{-1}\omega).$$

By (M3) the random variables X_ℓ^r, $\ell \in \mathbb{N}$ are independent with respect to the measure $\mathbb{P}$ on $(\Omega, \mathcal{B}(\Omega))$. Moreover, applying (M1) shows that X_ℓ^r, $\ell \in \mathbb{N}$, are identically distributed. By definition, the empirical measure of X_ℓ^r, $\ell \in \{1, \ldots, |T_{m,n}|\}$, where $|T_{m,n}| = \lfloor n/m\rfloor^d$, is exactly the empirical measure $L_{m,n}^{r,\omega}$ given in (4.6). According to (M2), the measure $\mathbb{P}_m^r$ is absolutely continuous with respect to a product measure on $\Omega_{\Lambda_m^r}$. We trivially extend $\mathbb{P}_m^r$ and $L_{m,n}^{r,\omega}$ to measures on $\mathbb{R}^k$ (and use the same names for the extensions). This allows to apply Theorem 5.5, which gives (i) of Theorem 5.3. Thus, the Glivenko–Cantelli theorem implies that (for the $m \in \mathbb{N}$ chosen above) there is a set $\Omega_m \in \mathcal{B}(\Omega)$ of probability one such that for each $\omega \in \Omega_m$ we have

$$\left\|\langle f_m^r, L_{m,n}^{r,\omega} - \mathbb{P}_m^r\rangle\right\| = \sup_{g\in\mathcal{M}_f} \left|\langle g, L_{m,n}^{r,\omega} - \mathbb{P}_m^r\rangle\right| \xrightarrow{n\to\infty} 0,$$

since the supremum is bounded by the supremum in (ii) from Theorem 5.3. By the same token,

$$\sup_{f\in\mathcal{U}_{K,D,D',r'}} \left\|\langle f_m^r, L_{m,n}^{r,\omega} - \mathbb{P}_m^r\rangle\right\| = \sup_{f\in\mathcal{U}_{K,D,D',r'}} \sup_{g\in\mathcal{M}_f} \left|\langle g, L_{m,n}^{r,\omega} - \mathbb{P}_m^r\rangle\right| \xrightarrow{n\to\infty} 0.$$

In the light of that, the claimed convergences in (5.4) and (5.5) hold independently from m for all $\omega \in \tilde{\Omega} := \bigcap_{m\in\mathbb{N}} \Omega_m$. To obtain (5.6) we apply Theorem 5.3, (iii). □

6. Almost additivity and limits, Proof of Theorem 2.6

Next we investigate the expression $\langle f_m^r, \mathbb{P}_m^r\rangle$ for large m. This is the third and last step in our approximation scheme. Thus, this step brings us in the position to prove our main results, namely Theorems 2.6 and 2.8.

Lemma 6.1. *Let* $\mathcal{A} \subseteq \mathbb{R}$, $\Omega := \mathcal{A}^{\mathbb{Z}^d}$, *a probability space* $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ *such that* $\mathbb{P}$ *satisfies* (M1), (M2) *and* (M3), *an admissible function* f *and the sequence* (Λ_n) *with* $\Lambda_n = ([0, n) \cap \mathbb{Z})^d$, $n \in \mathbb{N}$ *be given. Besides this, let* r *be the constant from* (M3) *and let for* $m \in \mathbb{N}$ *the* *marginal measure* $\mathbb{P}^r_m := \mathbb{P}_{\Lambda^r_m}$ *and the function* f^r_m *be given as in* (4.4). *Then there exists a function* $f^* \in \mathbb{B}$ *with*

$$\lim_{m\to\infty}\left\|\frac{\langle f^r_m, \mathbb{P}^r_m\rangle}{m^d} - f^*\right\| = 0. \tag{6.1}$$

Furthermore, we have, with b *and* D *from* Definition 2.3 *and* K_f *from* Remark 2.4, *for all* $m \in \mathbb{N}$

$$\left\|\frac{\langle f^r_m, \mathbb{P}^r_m\rangle}{m^d} - f^*\right\| \leqslant \frac{b(\Lambda^r_m)}{m^d} + (K_f + D)\frac{|\partial^r(\Lambda_m)|}{m^d}.$$

Proof. Let us define $F : \mathcal{F} \to \mathbb{B}$ by setting for each $\Lambda \in \mathcal{F}$:

$$F(\Lambda) := \langle f_\Lambda, \mathbb{P}_\Lambda\rangle = \int_{\Omega_\Lambda} f_\Lambda(\omega)\,\mathrm{d}\mathbb{P}_\Lambda(\omega) = \int_\Omega f(\Lambda, \omega)\,\mathrm{d}\mathbb{P}(\omega).$$

With this notation, it is sufficient to show that $(F(\Lambda^r_m)/m^d)_{m\in\mathbb{N}}$ is a Cauchy sequence.

First, we note that F is translation invariant, i.e. $F(\Lambda + t) = F(\Lambda)$. To see this, use (M1) and (i) of Definition 2.3. Note also, that F is almost additive with the same b and D as f, see (iii) of the same definition. Furthermore, it follows from Remark 2.4 that F is bounded in the following sense: For all $\Lambda \in \mathcal{F}$, we have $F(\Lambda) \leqslant K_f|\Lambda|$ with the same constant K_f as in (2.3).

Next, assume that two integers m, M with $m \leqslant M$ are given. As in (4.1), set

$$T_{m,M} := \{t \in (m\mathbb{Z})^d \mid \Lambda_m + t \subseteq \Lambda_M\}.$$

We are interested in an estimate of the difference

$$\delta(m, M) := \left\|\frac{F(\Lambda^r_M)}{M^d} - \frac{F(\Lambda^r_m)}{m^d}\right\|. \tag{6.2}$$

To study this we use the triangle inequality and get

$$\delta(m, M) \leqslant \frac{\alpha(m, M)}{M^d} + \beta(m, M) \tag{6.3}$$

with

$$\alpha(m, M) := \left\|F(\Lambda^r_M) - \sum_{t\in T_{m,M}} F(\Lambda^r_m + t)\right\|,$$

$$\beta(m, M) := \left\|\frac{F(\Lambda^r_m)}{m^d} - \sum_{t\in T_{m,M}} \frac{F(\Lambda^r_m + t)}{M^d}\right\|.$$

In order to estimate $\alpha(m, M)$, note that

$$\Lambda^r_M = \dot{\bigcup}_{t\in T_{m,M}}(\Lambda^r_m + t)\,\dot{\cup}\Big(\Lambda^r_M \cap \dot{\bigcup}_{t\in T_{m,M}}((\Lambda_m \cap \partial^r(\Lambda_m)) + t)\Big)\dot{\cup}\Lambda^r_M\setminus(\Lambda_{\lfloor M/m\rfloor m}).$$

This and (iii) of Definition 2.3 yield

$$\begin{aligned}\alpha(m, M) \leqslant &\sum_{t\in T_{m,M}}\Big(b(\Lambda^r_m) + b\big(\Lambda^r_M \cap ((\Lambda_m \cap \partial^r(\Lambda_m)) + t)\big)\\ &+ \big\|F\big(\Lambda^r_M \cap ((\Lambda_m \cap \partial^r(\Lambda_m)) + t)\big)\big\|\Big)\\ &+ b(\Lambda^r_M\setminus\Lambda_{\lfloor M/m\rfloor m}) + \|F(\Lambda^r_M\setminus\Lambda_{\lfloor M/m\rfloor m})\|\\ \leqslant &\,|T_{m,M}|b(\Lambda^r_m) + (K_f + D)|T_{m,M}||\partial^r(\Lambda_m)| + (K_f + D)|\Lambda^r_M\setminus\Lambda_{\lfloor M/m\rfloor m}|.\end{aligned}$$

Here, we also used translation invariance of F and property (v) of Definition 2.3. Dividing this term by M^d and using $|T_{m,M}|m^d \leqslant M^d$ as well as (4.2) leads to

$$\frac{\alpha(m, M)}{M^d} \leqslant \frac{b(\Lambda_m^r)}{m^d} + (K_f + D)\frac{|\partial^r(\Lambda_m)|}{m^d} + (K_f + D)\frac{|\partial^m(\Lambda_M)|}{M^d}.$$

To estimate $\beta(m, M)$ we apply again translation invariance of F and obtain

$$\beta(m, M) = \left\| \frac{F(\Lambda_m^r)}{m^d} - \frac{|T_{m,M}|F(\Lambda_m^r)}{M^d} \right\| = \left(\frac{1}{m^d} - \frac{\lfloor M/m \rfloor^d}{M^d} \right) \|F(\Lambda_m^r)\|.$$

Using the properties of the boundary term b, the above bounds on $\alpha(m, M)$ and $\beta(m, M)$ yield

$$\lim_{m\to\infty} \lim_{M\to\infty} \delta(m, M) = 0. \tag{6.4}$$

This is equivalent to $(F(\Lambda_m^r)/m^d)_{m\in\mathbb{N}}$ being a Cauchy sequence. To see this in detail, choose $\varepsilon > 0$ arbitrarily. Then, by (6.4), there exists $m_0 \in \mathbb{N}$ such that $\lim_{M\to\infty} \delta(m_0, M) \leqslant \varepsilon/4$. Therefore, we find $M_0 \in \mathbb{N}$ satisfying $\delta(m_0, M) \leqslant \varepsilon/2$ for all $M \geqslant M_0$. Now, let $j, k \geqslant M_0$ be arbitrary. Then we obtain using the triangle inequality

$$\begin{aligned} \left\| \frac{F(\Lambda_j^r)}{j^d} - \frac{F(\Lambda_k^r)}{k^d} \right\| &\leqslant \left\| \frac{F(\Lambda_j^r)}{j^d} - \frac{F(\Lambda_{m_0}^r)}{m_0^d} \right\| + \left\| \frac{F(\Lambda_{m_0}^r)}{m_0^d} - \frac{F(\Lambda_k^r)}{k^d} \right\| \\ &= \delta(m_0, j) + \delta(m_0, k) \leqslant \varepsilon. \end{aligned}$$

This shows that $(F(\Lambda_m^r)/|\Lambda_m|)_{m\in\mathbb{N}}$ is a Cauchy sequence and hence convergent in the Banach space $\mathbb{B}$.

Now, that we know that the limit f^* exists, we can study the speed of convergence.

$$\begin{aligned} \left\| \frac{\langle f_m^r, \mathbb{P}_m^r \rangle}{m^d} - f^* \right\| &= \lim_{M\to\infty} \left\| \frac{\langle f_m^r, \mathbb{P}_m^r \rangle}{m^d} - \frac{\langle f_M^r, \mathbb{P}_M^r \rangle}{M^d} \right\| = \lim_{M\to\infty} \delta(m, M) \\ &\leqslant \lim_{M\to\infty} \left(\frac{\alpha(m, M)}{M^d} + \beta(m, M) \right) \leqslant \frac{b(\Lambda_m^r)}{m^d} + (K_f + D)\frac{|\partial^r(\Lambda_m)|}{m^d}. \quad \square \end{aligned}$$

Now we are in the position to prove the main theorem of this paper.

Proof of Theorems 2.6 and 2.8. The proof is basically a combination of Lemmas 4.1 and 6.1 and Theorem 5.6. We choose $\tilde{\Omega}$ as in Theorem 5.6, r as the constant from (M3) and $f^* \in \mathbb{B}$ according to Lemma 6.1. Then we have for arbitrary $m \in \mathbb{N}$ and $\omega \in \tilde{\Omega}$:

$$\begin{aligned} \left\| \frac{f(\Lambda_n, \omega)}{n^d} - f^* \right\| &\leqslant \left\| \frac{f(\Lambda_n, \omega)}{n^d} - \frac{\langle f_m^r, L_{m,n}^{r,\omega} \rangle}{m^d} \right\| + \left\| \frac{\langle f_m^r, L_{m,n}^{r,\omega} \rangle}{m^d} - \frac{\langle f_m^r, \mathbb{P}_m^r \rangle}{m^d} \right\| \\ &\quad + \left\| \frac{\langle f_m^r, \mathbb{P}_m^r \rangle}{m^d} - f^* \right\|. \end{aligned}$$

Each of the above mentioned results controls one of the error terms on the right hand side, which leads to

$$\left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| \leqslant G(m, n) + m^{-d} \left\| \langle f_m^r, L_{m,n}^{r,\omega} \rangle - \langle f_m^r, \mathbb{P}_m^r \rangle \right\|$$

with

$$G(m,n) := \frac{b(\Lambda_{\lfloor n/m\rfloor m})}{|\Lambda_{\lfloor n/m\rfloor m}|} + \frac{(2K+D)|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|} + \frac{2b(\Lambda_m^r) + b(\Lambda_m) + 2(K+D)|\partial^r(\Lambda_m)|}{|\Lambda_m|}.$$

Taking first the limit $n \to \infty$ and afterwards the limit $m \to \infty$ on both sides proves Theorem 2.6.

To establish Theorem 2.8, we use the additional hypotheses on the boundary term and estimate $G(m,n)$. First we note for $n \geqslant 2r$

$$|\partial^r(\Lambda_n)| = (n+2r)^d - (n-2r)^d = \sum_{j=0}^{d} \binom{d}{j} (1-(-1)^k)(2r)^j n^{d-j} \leqslant 2^{2d+1} r^d n^{d-1}.$$

Therefore,

$$\frac{b(\Lambda_n)}{|\Lambda_n|} \leqslant \frac{|\partial^{r'}(\Lambda_n)| D'}{|\Lambda_n|} \leqslant \frac{2^{2d+1} r'^d D'}{n}$$

holds for all $n \geqslant 2r'$. With $\Lambda_n^m = \Lambda_{n-2m} + (m, m, \dots, m)$, it is now straightforward to verify

$$G(m,n) \leqslant 2^{2d+1} \Big(\frac{(2K+D)m^d + D' r'^d}{n-2m} + \frac{2(K+D)r^d + 3D' r'^d}{m-2r} \Big).$$

The two claims about $\sup_{f \in \mathcal{U}_{K,D,D',r'}} \big\| \langle f_m^r, L_{m,n}^{r,\omega} \rangle - \langle f_m^r, \mathbb{P}_m^r \rangle \big\|$ follow from Theorem 5.6. $\square$

7. Eigenvalue counting functions for the Anderson model

In the following, we introduce the Anderson model on $\mathbb{Z}^d$ or, more precisely, on the graph with nodes $\mathbb{Z}^d$ and nearest neighbor bonds. For the corresponding Schrödinger operators we show that the associated eigenvalue counting functions almost surely converge uniformly.

The *Laplace operator* $\Delta \colon \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$ is given by

$$(\Delta\varphi)(z) = \sum_{x : \mathfrak{d}(x,z)=1} \big(\varphi(x) - \varphi(z)\big) \quad (z \in \mathbb{Z}^d).$$

In order to define a random potential, we introduce the corresponding probability space. We fix the canonical space $\Omega := \mathcal{A}^{\mathbb{Z}^d}$, where $\mathcal{A} \subseteq \mathbb{R}$ is an arbitrary subset of $\mathbb{R}$. As before we equip Ω with $\mathcal{B}(\Omega)$, the σ-algebra on Ω generated by the cylinder sets. Moreover, we chose a probability measure $\mathbb{P} \colon \mathcal{B}(\Omega) \to [0,1]$ satisfying (M1), (M2) and (M3). In particular, a product measure $\mathbb{P} = \prod_{z \in \mathbb{Z}} \mu$ is allowed, where $\mu \colon \mathcal{B}(\mathcal{A}) \to [0,1]$ is a measure on $(\mathcal{A}, \mathcal{B}(\mathcal{A}))$. An alternative way to specify such a product measure is to say that the projections $\Omega \ni (\omega_x)_{x \in \mathbb{Z}} \to \omega_z$, $z \in \mathbb{Z}$, are $\mathcal{A}$-valued i.i.d. random variables.

The *random potential* $V = (V_\omega)_{\omega \in \Omega}$ is now defined by setting for each $\omega = (\omega_z)_{z \in \mathbb{Z}^d} \in \Omega$:

$$V_\omega \colon \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d), \qquad (V_\omega\varphi)(z) = \omega_z \varphi(z) \quad (\varphi \in \ell^2(\mathbb{Z}^d), z \in \mathbb{Z}^d). \tag{7.1}$$

Together, the Laplace operator and the random potential form the *random Schrödinger operator* $H = (H_\omega)_{\omega \in \Omega}$:

$$H_\omega \colon \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d), \qquad H_\omega := -\Delta + V_\omega. \tag{7.2}$$

This operator is almost surely self-adjoint and ergodic by (M1) and (M3). Thus, the spectrum $\sigma(H_\omega)$ of H_ω is a non-random subset of $\mathbb{R}$, cf. [16]. In the following we are interested in the distribution of $\sigma(H_\omega)$ on $\mathbb{R}$. The function which describes this distribution is called *integrated density of states*.

Let us define finite dimensional restrictions of H. To this end, consider for a given set $\Lambda \in \mathcal{F}$ the projection

$$p_\Lambda \colon \ell^2(\mathbb{Z}^d) \to \ell^2(\Lambda), \qquad (p_\Lambda \varphi)(z) := \varphi(z), \quad (\varphi \in \ell^2(\mathbb{Z}^d), z \in \Lambda) \tag{7.3}$$

and the inclusion

$$i_\Lambda \colon \ell^2(\Lambda) \to \ell^2(\mathbb{Z}^d), \qquad (i_\Lambda \varphi)(z) := \begin{cases} \varphi(z) & \text{if } z \in \Lambda, \\ 0 & \text{otherwise,} \end{cases} \qquad (\varphi \in \ell^2(\Lambda), z \in \mathbb{Z}^d). \tag{7.4}$$

Now, for any $\omega \in \Omega$ and $\Lambda \in \mathcal{F}$ we set

$$H_\omega^\Lambda \colon \ell^2(\Lambda) \to \ell^2(\Lambda), \qquad H_\omega^\Lambda := p_\Lambda H_\omega i_\Lambda.$$

The corresponding eigenvalue counting function is given by

$$f(\Lambda, \omega) := \Big(\mathbb{R} \ni x \mapsto \operatorname{Tr}\big(\chi_{(-\infty,x]}(H_\omega^\Lambda)\big)\Big). \tag{7.5}$$

Here, $\chi_{(-\infty,x]}(H_\omega^\Lambda)$ denotes the spectral projection of H_ω^Λ on the interval $(-\infty, x]$. Thus, $f(\Lambda, \omega)(x)$ equals the number of eigenvalues (counted with multiplicities) of H_ω^Λ which do not exceed x.

Lemma 7.1. *The eigenvalue counting function $f \colon \mathcal{F} \times \Omega \to \mathbb{B}$ given by (7.5) is admissible in the sense of Definition 2.3. It admits a proper boundary term, and possible constants for f are $D = D' = 8$, $K = 9$ and $r' = 1$.*

Proof. We verify (i)–(v) of Definition 2.3.

(i) Let $\Lambda \in \mathcal{F}$ and $z \in \mathbb{Z}^d$ be given. Using the definitions of the potential V_ω in (7.1), the translation τ_z in (2.2), the projection p_Λ in (7.3) and the inclusion i_Λ in (7.4) we obtain

$$p_\Lambda V_{\tau_z \omega} i_\Lambda = p_{\Lambda+z} V_\omega i_{\Lambda+z}.$$

This generalizes to $H_{\tau_z\omega}^\Lambda = H_\omega^{\Lambda+z}$ and hence implies for each $x \in \mathbb{R}$

$$f(\Lambda, \tau_z\omega)(x) = \operatorname{Tr}\big(\chi_{(-\infty,x]}(H_{\tau_z\omega}^\Lambda)\big) = \operatorname{Tr}\big(\chi_{(-\infty,x]}(H_\omega^{\Lambda+z})\big) = f(\Lambda + z, \omega)(x).$$

(ii) Let $\Lambda \in \mathcal{F}$ be given. Obviously, for all $\omega, \omega' \in \Omega$ with $\Pi_\Lambda(\omega) = \Pi_\Lambda(\omega')$ we have $H_\omega^\Lambda = H_{\omega'}^\Lambda$. Thus, we obtain $f(\Lambda, \omega) = f(\Lambda, \omega')$.

(iii) In order to show almost additivity, we make use of the following estimate, which holds for $\Lambda' \subseteq \Lambda \in \mathcal{F}$ and arbitrary $\omega \in \Omega$:

$$\|f(\Lambda, \omega) - f(\Lambda', \omega)\| \leqslant 4|\Lambda \setminus \Lambda'|. \tag{7.6}$$

This bound can be verified using the min–max-principle, cf. appendix of [11]. Now let $n \in \mathbb{N}$, disjoint sets $\Lambda_i \in \mathcal{F}$, $i = 1, \ldots, n$ and $\Lambda := \bigcup_{i=1}^n \Lambda_i \in \mathcal{F}$ be given. With triangle inequality and (7.6) we obtain for each $\omega \in \Omega$:

$$\begin{aligned}
&\Big\| f(\Lambda, \omega) - \sum_{i=1}^{n} f(\Lambda_i, \omega) \Big\| \\
&\quad \leqslant \Big\| f(\Lambda, \omega) - f\Big(\bigcup_{i=1}^{n} \Lambda_i^1, \omega\Big) \Big\| + \Big\| f\Big(\bigcup_{i=1}^{n} \Lambda_i^1, \omega\Big) - \sum_{i=1}^{n} f\big(\Lambda_i, \omega\big) \Big\| \\
&\quad \leqslant 4 \sum_{i=1}^{n} |\partial^1(\Lambda_i)| + \Big\| f\Big(\bigcup_{i=1}^{n} \Lambda_i^1, \omega\Big) - \sum_{i=1}^{n} f\big(\Lambda_i^1, \omega\big) \Big\| \\
&\qquad + \sum_{i=1}^{n} \| f(\Lambda_i^1, \omega) - f(\Lambda_i, \omega) \| \\
&\quad \leqslant 8 \sum_{i=1}^{n} |\partial^1(\Lambda_i)| + \Big\| f\Big(\bigcup_{i=1}^{n} \Lambda_i^1, \omega\Big) - \sum_{i=1}^{n} f\big(\Lambda_i^1, \omega\big) \Big\|.
\end{aligned}$$

In order to deal with the last difference, we use that the operator in consideration has hopping range 1, which gives for $\tilde{\Lambda} := \bigcup_{i=1}^{n} \Lambda_i^1$:

$$H_\omega^{\tilde{\Lambda}} = \bigoplus_{i=1}^{n} H_\omega^{\Lambda_i^1}.$$

Thus, the eigenvalues of $H_\omega^{\tilde{\Lambda}}$ are exactly the union of the eigenvalues of the operators $H_\omega^{\Lambda_i^1}$, $i = 1, \dots, n$. This implies

$$f(\tilde{\Lambda}, \omega) = \sum_{i=1}^{n} f(\Lambda_i^1, \omega)$$

and hence

$$\Big\| f(\Lambda, \omega) - \sum_{i=1}^{n} f(\Lambda_i, \omega) \Big\| \leqslant 8 \sum_{i=1}^{n} |\partial^1(\Lambda_i)|.$$

We set $b \colon \mathcal{F} \to [0, \infty)$ and $b(\Lambda) := 8|\partial^1(\Lambda)|$. Let $\Lambda \in \mathcal{F}$ and $z \in \mathbb{Z}^d$. Then obviously $b(\Lambda + z) = b(\Lambda)$ and $b(\Lambda) \leqslant 8|\Lambda|$, and for any sequence of cubes (Λ_n) with increasing side length, we have $b(\Lambda_n)/|\Lambda_n| \to 0$ as $n \to \infty$.

(iv) For $\Lambda \in \mathcal{F}$ and $\omega \in \Omega$ we denote the $|\Lambda|$ eigenvalues of H_ω^Λ (counted with multiplicities) by $E_1(H_\omega^\Lambda) \leqslant \cdots \leqslant E_{|\Lambda|}(H_\omega^\Lambda)$. Choose $n \in \{1, \dots, |\Lambda|\}$ and $\omega \leqslant \omega'$, i.e. for each $z \in \mathbb{Z}^d$ we have $\omega_z \leqslant \omega'_z$. By the min–max-principle we get for the nth eigenvalue:

$$\begin{aligned}
E_n(H_\omega^\Lambda) &= \min_{\substack{U \subseteq \mathbb{R}^\Lambda, \\ \dim(U) = n}} \max_{\substack{\varphi \in U, \\ \|\varphi\| = 1}} \langle H_\omega^\Lambda \varphi, \varphi \rangle \\
&= \min_{\substack{U \subseteq \mathbb{R}^\Lambda, \\ \dim(U) = n}} \max_{\substack{\varphi \in U, \\ \|\varphi\| = 1}} \big(\langle H_{\omega'}^\Lambda \varphi, \varphi \rangle - \langle (V_{\omega'} - V_\omega) \varphi, \varphi \rangle \big) \leqslant E_n(H_{\omega'}^\Lambda).
\end{aligned}$$

Therefore, we have for each $x \in \mathbb{R}$ the inequality $f(\Lambda, \omega)(x) \geqslant f(\Lambda, \omega')(x)$.

(v) Let arbitrary $\omega \in \Omega$ be given. Since the operator $H_\omega^{\{0\}}$ has exactly one eigenvalue, we have $\|f(\{0\},\omega)\| = 1$. $\square$

Let us state the main result of this section.

Theorem 7.2. *Let $\Lambda_n := [0,n)^d \cap \mathbb{Z}^d$. Moreover, let $\mathcal{A} \subseteq \mathbb{R}$, $\Omega := \mathcal{A}^{\mathbb{Z}^d}$ and $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ be a probability space satisfying* (M1)–(M3). *Consider the random Schrödinger operator H defined in (7.2) and the associated f given in (7.5). Then there exists a set $\tilde{\Omega} \in \mathcal{B}(\Omega)$ of full measure, such that for all $\omega \in \tilde{\Omega}$:*

$$\lim_{n\to\infty} \left\| \frac{f(\Lambda_n,\omega)}{|\Lambda_n|} - f^* \right\| = 0, \tag{7.7}$$

where $f^* \in \mathbb{B}$ is given by

$$f^*(x) := \mathbb{E}(\langle \delta_0, \chi_{(-\infty,x]}(H_\omega)\delta_0\rangle). \tag{7.8}$$

Here, $\delta_0 \in \ell^2(\mathbb{Z}^d)$ is given by $\delta_0(0) = 1$ and $\delta_0(x) = 0$ for $x \neq 0$. Moreover, $\chi_{(-\infty,x]}(H_\omega)$ is the spectral projection of H_ω on the interval $(-\infty, x]$. The convergence is quantified by

$$\begin{aligned}\left\| \frac{f(\Lambda_n,\omega)}{|\Lambda_n|} - f^* \right\| \leqslant{}& 2^{d+1}\left(\frac{26m^d+8}{n-m} + \frac{34r^d+24}{m-r}\right)\\ &+ \sup_{f\in\mathcal{U}_{K,D,D',r'}} \frac{\|\langle f_{\Lambda_m^r}, L_{m,n}^{r,\omega} - \mathbb{P}_{\Lambda_m^r}\rangle\|}{|\Lambda_m|}\end{aligned} \tag{7.9}$$

for $n, m \in \mathbb{N}$, $m < n$.

Proof. By Lemma 7.1 we know that the eigenvalue counting function $f : \mathcal{F} \times \Omega \to \mathbb{B}$ is admissible. Thus we can apply Theorem 2.6 and obtain that there exists a function $\bar{f} \in \mathbb{B}$ and a set $\Omega_1 \in \mathcal{B}(\Omega)$ of full measure such that for each $\omega \in \Omega_1$ we have

$$\lim_{n\to\infty} \left\| \frac{f(\Lambda_n,\omega)}{|\Lambda_n|} - \bar{f} \right\| = 0. \tag{7.10}$$

Thus, it remains to show that $\bar{f}$ equals the function

$$f^* : \mathbb{R} \to [0,1], \qquad f^*(x) := \mathbb{E}(\langle \delta_0, \chi_{(-\infty,x]}^\omega \delta_0\rangle).$$

Here we use ergodicity of H and infer from [16, Theorem 4.8] that the there is a set $\Omega_2 \in \mathcal{B}(\Omega)$ of full measure such that for each $\omega \in \Omega_2$

$$\lim_{n\to\infty} \frac{f(\Lambda_n,\omega)(x)}{|\Lambda_n|} = f^*(x) \tag{7.11}$$

for all $x \in \mathbb{R}$ which are continuity points of f^*. By definition, this is weak convergence of distribution functions. Thus, as for all $\omega \in \Omega_1 \cap \Omega_2$ we have that $f(\Lambda_n,\omega)/|\Lambda_n|$ converges weakly to f^* and uniformly to $\bar{f}$, which implies $\bar{f} = f^*$. $\square$

Remark 7.3. • The limit f^* of the normalized eigenvalue counting functions is called the *integrated density of states* or *spectral distribution function* of the operator H. The fact that f^* can be expressed as the function given in (7.8) is often referred to as the *Pastur–Shubin trace* *formula*, named after the pioneering works [15,20]. For more recent results in the specific context we are treating here, c.f. [27,19,13] and the references therein.

- Let us also emphasize that the f^* is a deterministic function. On the one hand this is interesting as this implies that the normalized eigenvalue counting function converges for almost all realizations to same limit function. On the other hand this is not surprising as we mentioned that H is ergodic, and in this setting it is well-known that the spectrum (as a set) is deterministic, see for instance [16].
- The result is easily generalized to sequences of cubes $(\Lambda_n)_n$ of diverging side length with $\Lambda_n \subsetneq \Lambda_{n+1}$. The validity of the Pastur–Shubin formula shows that the limit f^* is independent of the specific choice sequence of cubes $(\Lambda_n)_n$.
- The statement of Theorem 7.2 has been obtained before in a different setting. In [10,13] ergodic random operators have been considered. The assumption of ergodicity concerns the measure $\mathbb{P}$ (in our notation) and is weaker than the assumptions (M1)–(M3) which we use here. With this regard the result of [13] is more general than the one obtained here. However, under the mere assumption of ergodicity it is not possible to obtain explicit error estimates as in (7.9). The paper [10] obtains an error estimate, similar to, but weaker then (7.9). There the setting is also different from ours here: $\mathcal{A}$ needs to be countable and instead of a probability measure properties of frequencies are used.
- Similar, but weaker results have been proven for Anderson-percolation Hamiltonians in [25, 26,13]. These models are particularly interesting since their integrated density of states exhibits typically an infinite set of discontinuities, which lie dense in the spectrum. The random variables entering the Hamiltonian may take uncountably many different values.

8. Cluster counting functions in percolation theory

We introduce briefly percolation on $\mathbb{Z}^d$. Percolation comes in two flavors, site and bond percolation. We focus on site percolation here. Part of the results have already been obtained in [17]. However, we go far beyond since we not only obtain convergence of densities, but are even able to identify the limit objects.

As before, we let $\Omega := \mathbb{R}^{\mathbb{Z}^d}$. We fix the alphabet $\mathcal{A} := \{0, 1\}$ and a probability measure $\mathbb{P}\colon \mathcal{B}(\Omega) \to [0, 1]$ which is supported in $\mathcal{A}^{\mathbb{Z}^d} \in \mathcal{B}(\Omega)$, i.e. $\mathbb{P}(\mathcal{A}^{\mathbb{Z}^d}) = 1$. A configuration $\omega \in \mathcal{A}^{\mathbb{Z}^d} \subseteq \Omega$ determines a *percolation graph* $\Gamma_\omega = (\mathbb{Z}^d, \mathcal{E}_\omega)$ as follows. The set of vertices of Γ_ω is $\mathbb{Z}^d$, and an edge connects two vertices if and only if they have distance 1 and are both "switched on" in the configuration $\omega = (\omega_z)_{z\in\mathbb{Z}^d}$:

$$\mathcal{E}_\omega := \Big\{\{x, y\} \subseteq \mathbb{Z}^d \;\Big|\; \mathfrak{d}(x, y) = 1, \omega_x = \omega_y = 1\Big\}.$$

By this, the percolation graph Γ_ω is well-defined for $\mathbb{P}$-almost all $\omega \in \Omega$, and Γ_ω is a random graph. For our purposes, we want $\mathbb{P}$ to satisfy (M1), (M2) and (M3). This setting includes but is not limited to the product measure $\mathbb{P} = \prod_{z\in\mathbb{Z}^d} \mu$, where $\mu\colon \mathcal{B}(\mathbb{R}) \to [0, 1]$ is any probability measure supported on $\mathcal{A}$.

We need some standard terminology of graph theory. Let $\Gamma = (V, \mathcal{E})$ be a graph. For each subset $\Lambda \subseteq V$ of the set of nodes, Γ induces a graph $\Gamma^\Lambda := (\Lambda, \mathcal{E}^\Lambda)$ by

$$\mathcal{E}^\Lambda := \{e \in \mathcal{E} \mid e \subseteq \Lambda\}.$$

A *walk* of length $n \in \mathbb{N} \cup \{0, \infty\}$ in the graph Γ is a sequence of nodes $(z_j)_{j=0}^n \in (\mathbb{Z}^d)^{n+1}$ such that $\{z_j, z_{j+1}\}$ is an edge of Γ, i.e. $\{z_j, z_{j+1}\} \in \mathcal{E}$, for all $j \in \mathbb{N} \cup \{0\}$, $j < n$. Note that a finite walk of length n contains n edges but $n + 1$ nodes.

If the walk $(z_j)_{j=0}^n$ has finite length $n < \infty$, we say that it *connects* the points z_0 and z_n. Being connected by a walk is an equivalence relation on the nodes. We denote the fact that two points $x, y \in \mathbb{Z}^d$ are connected in the graph Γ as $x \overset{\Gamma}{\leftrightsquigarrow} y$.

The equivalence classes of $\leftrightsquigarrow$ are called *clusters*. Let $\Lambda \subseteq \mathbb{Z}^d$ and $x \in \Lambda$. The cluster of x in the percolation graph Γ_ω^Λ restricted to Λ consists of all nodes which are connected to x by a walk in Γ_ω^Λ:

$$C_x^\Lambda(\omega) := \{y \in \mathbb{Z}^d \mid x \overset{\Gamma_\omega^\Lambda}{\leftrightsquigarrow} y\},$$

again for $\omega \in \mathcal{A}^{\mathbb{Z}^d}$, $\Lambda \subseteq \mathbb{Z}^d$ and $x, y \in \Lambda$.

8.1. *Convergence of cluster counting functions*

We now define a cumulative counting function for clusters. As before, let $\mathcal{F}$ be the set of finite subsets of $\mathbb{Z}^d$ and $\mathbb{B}$ the set of bounded functions from $\mathbb{R}$ to $\mathbb{R}$ which are continuous from the right. The function $f : \mathcal{F} \times \Omega \to \mathbb{B}$ counts the number of clusters in Γ_ω^Λ which are smaller then the given threshold:

$$f(\Lambda, \omega)(\lambda) := \left|\left\{C_z^\Lambda(\omega) \;\middle|\; z \in \Lambda, |C_z^\Lambda(\omega)| \leqslant \lambda\right\}\right|. \tag{8.1}$$

Note that f counts clusters and not vertices in clusters.

Lemma 8.1. *The cluster counting function $f : \mathcal{F} \times \Omega \to \mathbb{B}$ given by (8.1) is admissible in the sense of Definition 2.3 and permits a proper boundary term. Possible constants are $D = D' = 2$, $r' = 1$ and $K = 3$.*

Proof. We verify (i)–(v) of Definition 2.3.

(i) Let $\Lambda \in \mathcal{F}$ and $x, z \in \mathbb{Z}^d$ be given. The percolation graph Γ_ω is determined by the configuration $\omega \in \Omega$, for almost all $\omega \in \Omega$. The shifted configuration gives shifted clusters, i.e.

$$|C_x^\Lambda(\tau_z\omega)| = |C_{x+z}^{\Lambda+z}(\omega)|$$

for all $x, z \in \mathbb{Z}^d$. Accordingly, for all $\lambda \in \mathbb{R}$,

$$\begin{aligned} f(\Lambda, \tau_z\omega)(\lambda) &= \left| \left\{C_x^\Lambda(\tau_z\omega) \;\middle|\; x \in \Lambda, |C_x^\Lambda(\tau_z\omega)| \leqslant \lambda\right\} \right| \\ &= \left| \left\{C_{x+z}^{\Lambda+z}(\omega) \;\middle|\; x \in \Lambda, |C_{x+z}^{\Lambda+z}(\omega)| \leqslant \lambda\right\} \right| = f(\Lambda + z, \omega)(\lambda). \end{aligned}$$

(ii) Fix $\Lambda \in \mathcal{F}$ and $\omega, \omega' \in \Omega$ with $\omega_\Lambda = \omega'_\Lambda$, where $\omega_\Lambda := (\omega_x)_{x\in\Lambda}$ as before. The edges of Γ_ω^Λ are determined by ω_Λ. Hence, $\Gamma_\omega^\Lambda = \Gamma_{\omega'}^\Lambda$, thus $C_z^\Lambda(\omega) = C_z^\Lambda(\omega')$ for all $z \in \Lambda$, and we obtain $f(\Lambda, \omega) = f(\Lambda, \omega')$.

(iii) In order to show almost additivity, fix $\omega \in \Omega$, $n \in \mathbb{N}$ and disjoint sets $\Lambda_j \in \mathcal{F}$, $j \in \{1, \ldots, n\}$. We name the union $\Lambda := \bigcup_{j=1}^n \Lambda_j$. For $x \in \mathbb{R}$,

$$\sum_{j=1}^n f(\Lambda_j, \omega)(x)$$

is the total number of clusters of size not larger than x in the graphs $\Gamma_\omega^{\Lambda_j}$. Whenever $\mathfrak{d}(\Lambda_j, \Lambda_k) = 1$, the graph Γ_ω^Λ could contain edges connecting a point in Λ_j with a point in Λ_k, depending on ω. Each of these edges join two possibly different clusters, so for each edge, there are two less small clusters and one more large one. By this mechanism, the number of clusters below the threshold x changes at most by twice the number of added edges. We note

$$\left|\mathcal{E}_\omega^\Lambda \setminus \bigcup_{j=1}^n \mathcal{E}_\omega^{\Lambda_j}\right| \leqslant \sum_{j=1}^n |\partial^1 \Lambda_j|$$

and conclude

$$\left| f(\Lambda, \omega)(x) - \sum_{j=1}^n f(\Lambda_j, \omega)(x) \right| \leqslant 2 \sum_{j=1}^n |\partial^1 \Lambda_j|$$

for all $x \in \mathbb{R}$. The choice $b(\Lambda) := 2|\partial^1 \Lambda|$ for $\Lambda \in \mathcal{F}$ gives a proper boundary term for f, cf. Lemma 7.1.

(iv) Let $\Lambda \in \mathcal{F}$ and $\omega, \omega' \in \Omega$, $\omega \leqslant \omega'$. Then each edge of Γ_ω is also an edge in $\Gamma_{\omega'}$: $\mathcal{E}_\omega \subseteq \mathcal{E}_{\omega'}$. As reasoned in (iii), a new edge never increases the number of clusters below a threshold $x \in \mathbb{R}$, so

$$f(\Lambda, \omega)(x) \geqslant f(\Lambda, \omega')(x).$$

(v) For all $\omega \in \Omega$, $f(\{0\}, \omega)(x) = 0$ for $x < 1$ and $f(\{0\}, \omega)(x) = 1$ for $x \geqslant 1$. $\square$

Theorem 2.6 and Lemma 8.1 immediately give the following.

Corollary 8.2. *Let $\Lambda_n := [0, n)^d \cap \mathbb{Z}^d$ for $n \in \mathbb{N}$ and $f : \mathcal{F} \times \Omega \to \mathbb{B}$ be the cumulative cluster counting function given in (8.1). There exists a set $\tilde{\Omega} \subseteq \Omega$ of full measure and a function $f^* \in \mathbb{B}$ such that, for each $\omega \in \tilde{\Omega}$,*

$$\lim_{n\to\infty} \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| = 0.$$

For all $m, n \in \mathbb{N}$, $m < n$, we have

$$\left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| \leqslant 2^{d+1} \left(\frac{8m^d + 2}{n - m} + \frac{10r^d + 6}{m - r} \right) + \sup_{f \in \mathcal{U}_{K,D,D',r'}} \frac{\left\| \langle f_{\Lambda_m^r}, L_{m,n}^{r,\omega} - \mathbb{P}_{\Lambda_m^r} \rangle \right\|}{|\Lambda_m|}.$$

8.2. Identification of the limit

In the previous section we studied the convergence of the counting function in (8.1) normalized with $|\Lambda_n|$. Next, we give a brief overview on closely related convergence results. We sketch the proofs only briefly since these results are not in the main focus of this paper. The heart of the section is that we do not just give statements about convergence, but even present closed expressions of the limits.

We start with defining

$$K_\omega(\Lambda) := |\{C_x^\Lambda(\omega) \mid x \in \Lambda\}|, \tag{8.2}$$

which counts the number of all clusters in Γ_ω^Λ. Using this quantity we set:

$$a_n^{(m)}(\omega) := |\Lambda_n|^{-1}|\{C_x^{\Lambda_n}(\omega) \mid x \in \Lambda_n, |C_x^{\Lambda_n}(\omega)| = m\}|,$$
$$b_n^{(m)}(\omega) := K_\omega(\Lambda_n)^{-1}|\{C_x^{\Lambda_n}(\omega) \mid x \in \Lambda_n, |C_x^{\Lambda_n}(\omega)| = m\}|, \quad \text{and}$$
$$c_n^{(m)}(\omega) := |\Lambda_n|^{-1}|\{x \in \Lambda_n \mid |C_x^{\Lambda_n}(\omega)| = m\}|,$$

where again $\Lambda_n := [0, n)^d \cap \mathbb{Z}^d$ for $n \in \mathbb{N}$.

Lemma 8.3. *In the above setting, we have almost surely*

$$\lim_{n\to\infty} a_n^{(m)}(\omega) = \frac{1}{m}\mathbb{P}(|C_0| = m),$$
$$\lim_{n\to\infty} b_n^{(m)}(\omega) = \frac{1}{\kappa m}\mathbb{P}(|C_0| = m), \quad \textit{and}$$
$$\lim_{n\to\infty} c_n^{(m)}(\omega) = \mathbb{P}(|C_0| = m),$$

where $\kappa := \mathbb{E}(|C_0|^{-1})$.

Note that the existence of the limit in the case corresponding to $a_n^{(m)}$ was treated in Section 8.1. The existence of the limits in Lemma 8.3 has already been proved in [17] in the setting of bond percolation. However, the authors did not give explicit expressions for the limit objects. For the proof of Lemma 8.3 one may use Theorem 2.6 in combination with the d-dimensional version of Birkhoff's ergodic theorem, see [8], and the fact [7] that for almost all ω:

$$\lim_{n\to\infty} \frac{K_\omega(\Lambda_n)}{|\Lambda_n|} = \kappa.$$

The above convergence results can again be extend to the associated distribution functions. To formulate the corresponding result, we introduce for $n \in \mathbb{N}$ and $\omega \in \Omega$ the maps $\Theta_\omega^n, \Phi_\omega^n, \Psi_\omega^n : \mathbb{R} \to \mathbb{R}$ by setting for each $m \in \mathbb{N}$

$$\Theta_\omega^n(m) := \sum_{j=1}^{\lfloor m\rfloor} a_n^{(j)}(\omega) = \frac{|\{C_x^{\Lambda_n}(\omega) \mid x \in \Lambda_n, |C_x^{\Lambda_n}(\omega)| \leqslant m\}|}{|\Lambda_n|}$$
$$\Phi_\omega^n(m) := \sum_{j=1}^{\lfloor m\rfloor} b_n^{(j)}(\omega) = \frac{|\{C_x^{\Lambda_n}(\omega) \mid x \in \Lambda_n, |C_x^{\Lambda_n}(\omega)| \leqslant m\}|}{K_\omega(\Lambda_n)}, \quad \text{and}$$
$$\Psi_\omega^n(m) := \sum_{j=1}^{\lfloor m\rfloor} c_n^{(j)}(\omega) = \frac{|\{x \in \Lambda_n \mid |C_x(\omega)| \leqslant m\}|}{|\Lambda_n|}.$$

Moreover, we define the deterministic functions $\Theta, \Phi, \Psi : \mathbb{R} \to \mathbb{R}$ by

$$\Theta(m) := \sum_{j=1}^{\lfloor m\rfloor} \frac{1}{j}\mathbb{P}(|C_0| = j), \qquad \Phi(m) := \frac{1}{\kappa}\Theta(m), \quad \text{and} \quad \Psi(m) := \mathbb{P}(|C_0| \leqslant m) \tag{8.3}$$

for $m \in \mathbb{N}$.

Theorem 8.4. *In the above setting, we can find a set $\tilde{\Omega} \subseteq \Omega$ of full measure such that for all $\omega \in \tilde{\Omega}$ we have*

$$\lim_{n\to\infty} \|\Theta_\omega^n - \Theta\| = 0, \qquad \lim_{n\to\infty} \|\Phi_\omega^n - \Phi\| = 0 \quad \textit{and} \quad \lim_{n\to\infty} \|\Psi_\omega^n - \Psi\| = 0.$$

Here $\|\cdot\|$ denotes the supremum norm in $\mathcal{B}(\mathbb{R})$.

Let us give a brief sketch of the proof. The convergence of Θ_ω^n and Φ_ω^n follows rather direct from Theorem 2.6 and Lemma 8.3. However, in order to obtain the convergence of Ψ_ω^n one has to apply a different scheme, which was used in the context of the eigenvalue counting function in [13, Section 6]. The strategy consist of the following steps: One first verifies weak convergence of the distribution functions and second, shows that $\nu_\omega^n(\{\lambda\}) \to \nu(\{\lambda\})$ for each $\lambda \in \mathbb{R}$. Here ν and ν_ω^n are the measures corresponding to Ψ and Ψ_ω^n, respectively. Both steps together imply uniform convergence. To verify these convergences one applies again Lemma 8.3 as well as Birkhoff's ergodic theorem.

Remark 8.5. The first statement of Theorem 8.4 identifies the limit f^* from Corollary 8.2, namely it shows $f^* = \Theta$, where Θ is given in (8.3).

Appendix. Examples of measures

Let us discuss three classes of examples of measures $\mathbb{P}$ satisfying (M1), (M2) and (M3).

(a) *Countable colors*: Consider the case $d = 1$ and let $\mathcal{A} = \mathbb{N}_0$. Let $\Omega = \mathbb{R}^{\mathbb{Z}}$ and fix an arbitrary product measure $\tilde{\mathbb{P}}\colon \mathcal{B}(\Omega) \to [0,1]$ with support

$$\operatorname{supp} \tilde{\mathbb{P}} \subseteq \mathcal{A}^{\mathbb{Z}}.$$

We define a transformation of $\tilde{\mathbb{P}}$. To this end, let constants $c, \beta, \alpha_{-c}, \alpha_{-c+1}, \dots, \alpha_c \in \mathbb{N}_0$ be given and consider the function

$$\varphi\colon \Omega \to \Omega, \qquad (\varphi(\omega))_z := \beta + \sum_{k=-c}^{c} \alpha_k \omega_{z-k}.$$

We define $\mathbb{P} := \tilde{\mathbb{P}} \circ \varphi^{-1}$. Let us check the conditions (M1), (M2) and (M3) for $\mathbb{P}$. In order to check (M1) let $z \in \mathbb{Z}$ be given. Then, using stationarity of the product measure $\tilde{\mathbb{P}}$,

$$\begin{aligned}\mathbb{P} \circ \tau_z^{-1} &= \tilde{\mathbb{P}} \circ \varphi^{-1} \circ \tau_z^{-1} = \tilde{\mathbb{P}} \circ (\tau_z \circ \varphi)^{-1} \\ &= \tilde{\mathbb{P}} \circ (\varphi \circ \tau_z)^{-1} = \tilde{\mathbb{P}} \circ \tau_z^{-1} \circ \varphi^{-1} = \tilde{\mathbb{P}} \circ \varphi^{-1} = \mathbb{P}.\end{aligned}$$

Let us verify condition (M2) for $\mathbb{P}$. We define for each $\Lambda \in \mathcal{F}$ the function

$$\rho_\Lambda\colon \mathcal{A}^\Lambda \to \mathbb{R}, \qquad x = (x_z)_{z\in\Lambda} \mapsto \rho_\Lambda(x) = \mathbb{P}(\Pi_\Lambda^{-1}(\{x\})).$$

Then ρ_Λ is the density of the marginal measure $\mathbb{P}_\Lambda$ with respect to the counting measure on $\mathbb{N}_0$, since we have for each $\Lambda \in \mathcal{F}$ and $A \in \mathcal{B}(\mathcal{A}^\Lambda)$

$$\mathbb{P}_\Lambda(A) = \sum_{x\in A} \mathbb{P}(\Pi_\Lambda^{-1}(\{x\})) = \sum_{x\in A} \rho_\Lambda(x).$$

It remains to verify condition (M3). To this end, let $\Lambda_1, \dots, \Lambda_n \subseteq \mathbb{Z}$ with $\min\{\mathfrak{d}(\Lambda_i, \Lambda_j) \mid i \neq j\} > 2c$ be given. Then, using the definition of φ, we have for each $x = (x_z)_{z\in\Lambda} \in \Lambda := \bigcup_{i=1}^n \Lambda_i$

$$\mathbb{P}(\Pi_\Lambda^{-1}(\{x\})) = (\tilde{\mathbb{P}} \circ \varphi^{-1} \circ \Pi_\Lambda^{-1})(\{x\}) = \prod_{i=1}^n (\tilde{\mathbb{P}} \circ \varphi^{-1} \circ \Pi_{\Lambda_i}^{-1})(\{x\}),$$

which proves that $\rho_\Lambda = \prod_{i=1}^n \rho_{\Lambda_i}$.

(b) *Normal distribution*: Here, we treat the case $d = 1$, $\mathcal{A} = \mathbb{R}$, $\Omega = \mathbb{R}^{\mathbb{Z}}$ and set $\tilde{\mathbb{P}} := \bigotimes_{z\in\mathbb{Z}} \mathcal{N}(0,1) \colon \mathcal{B}(\Omega) \to [0,1]$, where $\mathcal{N}(0,1)$ is the standard normal distribution. For $c \in \mathbb{N}_0$ and $\beta, \alpha_{-c}, \alpha_{-c+1}, \dots, \alpha_c \in \mathbb{R}$ we use

$$\varphi \colon \Omega \to \Omega, \qquad (\varphi(\omega))_z = \beta + \sum_{k=-c}^{c} \alpha_k \omega_{z-k}$$

to define $\mathbb{P} := \tilde{\mathbb{P}} \circ \varphi^{-1}$. As before, the conditions (M1) and (M3) are implied by the choice of φ and the product structure of $\tilde{\mathbb{P}}$. For (M2), let $\Lambda \subseteq \mathbb{Z}$ be finite and first assume that $\Lambda = [a,b] \cap \mathbb{Z}$, $a, b \in \mathbb{Z}$. We define the matrix

$$A_\Lambda \in \mathbb{R}^{\Lambda \times \{a-c,\dots,b+c\}}, \qquad (A_\Lambda)_{i,j} = \alpha_{i-j},$$

where $\alpha_k := 0$ if $k \notin \{-c, \dots, c\}$. Recall that $\mathbb{P}_\Lambda = \tilde{\mathbb{P}} \circ \varphi^{-1} \circ \Pi_\Lambda^{-1} = \tilde{\mathbb{P}} \circ (\Pi_\Lambda \circ \varphi)^{-1}$. For $\omega \in \Omega$ we get

$$\Pi_\Lambda(\varphi(\omega)) = A_\Lambda \Pi_{[a-c,b+c]}(\omega) + \beta e_\Lambda,$$

where $e_\Lambda = (1, \dots, 1)^\top \in \mathbb{R}^\Lambda$. Now, it follows that $\mathbb{P}_\Lambda$ is normal distributed with mean βe_Λ and covariance matrix $A_\Lambda A_\Lambda^\top$. Note that $A_\Lambda A_\Lambda^\top$ is invertible since the rows of A_Λ are linearly independent. Thus, the measure $\mathbb{P}_\Lambda$ is absolutely continuous with respect to the multi-dimensional Lebesgue measure.

In the situation where Λ is not of the form $[a,b] \cap \mathbb{Z}$, consider the interval $I := [\min \Lambda, \max \Lambda] \cap \mathbb{Z}$. The measure $\mathbb{P}_\Lambda$ is a marginal measure of $\mathbb{P}_I$ and therefore has a density.

(c) *Abstract densities and finite range*: In the following we develop a more general example with densities. Again, we consider for simplicity reasons the case $d = 1$, however this is easily generalized to higher dimensions. Choose $\mathcal{A}, \mathcal{B} \in \mathcal{B}(\mathbb{R})$ and independent $\mathcal{B}$-valued random variables X'_x, $x \in \mathbb{Z}$ with density $g \colon \mathcal{A} \to \mathbb{R}_+$. We use the abbreviation $X_{[m,\ell]} := (X_m, \dots, X_\ell)$. We utilize a function $\varphi \colon \mathcal{B}^{k+1} \to \mathcal{A}$ to introduce the $\mathcal{A}$-valued random variables

$$X_x := \varphi(X'_{[x,x+k]}) \quad x \in \mathbb{Z}.$$

We require from φ, that there is a function $\psi \colon \mathcal{A} \times \mathcal{B}^k \to \mathcal{B}$ such that

$$\psi(\varphi(x_{[0,k]}), x_{[1,k]}) = x_0$$

for all $x_{[0,k]} \in \mathcal{B}^{k+1}$. Further, ψ shall be continuously differentiable w.r.t. its first argument: $\psi' := D_1\psi$. An example of such a pair of functions is

$$\varphi(x_{[0,k]}) := \frac{1}{k+1} \sum_{j=0}^{k} x_j, \qquad \psi(\xi_0, x_{[1,k]}) := (k+1)\xi_0 - \sum_{j=1}^{k} x_j,$$

where $\mathcal{A} := \mathcal{B} := [0,1]$ and $\psi'(\xi_0, x_{[1,k]}) = k+1$. In this example, $(X_x)_x$ is a moving average process. By suitable modifications, all moving average processes are seen to be included in our setting.

Proposition A.1. *Fix a finite set $\Lambda \subseteq \mathbb{Z}$. Under the specified circumstances, the joint distribution of $(X_x)_{x\in\Lambda}$, is absolutely continuous with respect to Lebesgue measure on $\mathcal{A}^{\Lambda}$.*

Proof. Without loss of generality, we treat only the case $\Lambda = \{1, \dots, \ell\}$. By construction, for $A_1, \dots, A_\ell \subseteq \mathcal{A}$ measurable,

$$\begin{aligned} p &:= \mathbb{P}(X_1 \in A_1, \dots, X_\ell \in A_\ell) \\ &= \int_{\mathcal{B}^{\ell+k}} \mathrm{d}x_{[1,\ell+k]} \prod_{m=1}^{\ell} \chi_{A_m}(\varphi(x_{[m,m+k]})) \cdot \prod_{m=1}^{\ell+k} g(x_m). \end{aligned}$$

By Fubini and induction on $j \in \{0, \dots, \ell\}$, we see

$$\begin{aligned} p = &\int_{A_1\times\cdots\times A_j} \mathrm{d}\xi_{[1,j]} \int_{\mathcal{B}^{\ell-j+k}} \mathrm{d}x_{[j+1,\ell+k]} \\ &\times \prod_{m=1}^{j} \left(g(\psi(\tilde{x}_m^{(j)}))|\psi'(\tilde{x}_m^{(j)})|\right) \cdot \prod_{m=j+1}^{\ell} \chi_{A_m}(\varphi(x_{[m,m+k]})) \cdot \prod_{m=j+1}^{\ell+k} g(x_m), \end{aligned}$$

where $\tilde{x}_j^{(j)} := (\xi_j, x_{[j+1,j+k]})$ and, for $1 \leqslant m < j$, the term $\tilde{x}_m^{(j)} := \tilde{x}_m^{(j-1)}|_{x_j \mapsto \psi(\xi_j, x_{[j+1,j+k]})}$ is generated from $\tilde{x}_m^{(j-1)}$ by substituting x_j by $\psi(\xi_j, x_{[j+1,j+k]})$. For the induction step use the substitution $\xi_j := \varphi(x_{[j,j+k]})$ or $x_j = \psi(\xi_j, x_{[j+1,j+k]})$ in

$$\begin{aligned} &\int_{\mathcal{B}} \mathrm{d}x_j \chi_{A_j}(\varphi(x_{[j,j+k]})) f_j(x_j) g(x_j) \\ &\quad = \int_{A_j} \mathrm{d}\xi_j f_j(\psi(\xi_j, x_{[j+1,j+k]})) g(\psi(\xi_j, x_{[j+1,j+k]})) |\psi'(\xi_j, x_{[j+1,j+k]})|, \end{aligned}$$

for any $x_{[j+1,j+k]} \in \mathcal{B}^k$ and suitable $f_j : \mathcal{B} \to \mathbb{R}_+$. For $j = \ell$, we conclude

$$p = \int_{A_1\times\cdots\times A_\ell} \mathrm{d}\xi_{[1,\ell]} \int_{\mathcal{B}^k} \mathrm{d}x_{[\ell+1,\ell+k]} \prod_{m=1}^{\ell} \left(g(\psi(\tilde{x}_m^{(\ell)}))|\psi'(\tilde{x}_m^{(\ell)})|\right) \cdot \prod_{m=\ell+1}^{\ell+k} g(x_m).$$

We hereby identified the density with respect to the product Lebesgue measure on $\mathcal{A}^\ell$. □

References

[1] A. Bovier, Statistical Mechanics of Disordered Systems. A Mathematical Perspective, in: Cambridge Series in Statistical and Probabilistic Mathematics, vol. 18, Cambridge University Press, Cambridge, 2006.

[2] J. DeHardt, Generalizations of the Glivenko–Cantelli theorem, Ann. Math. Statist. 42 (6) (1971) 2050–2055.

[3] A. Dvoretzky, J. Kiefer, J. Wolfowitz, Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, Ann. Math. Statist. 27 (3) (1956) 642–669.

[4] E. Følner, On groups with full Banach mean value, Math. Scand. 3 (1955) 243–254.

[5] R. Griffiths, A proof that the free energy of a spin system is extensive, J. Math. Phys. 5 (1964) 1215–1222.

[6] R. Griffiths, J. Lebowitz, Random spin systems: some rigorous results, J. Math. Phys. 9 (8) (1968) 1284–1292.

[7] G. Grimmett, Percolation, in: Grundlehren der Mathematischen Wissenschaften, vol. 321, Springer, Berlin, 1999.

[8] G. Keller, Equilibrium States in Ergodic Theory, in: London Mathematical Society Student Texts, vol. 42, Cambridge University Press, Cambridge, 1998.

[9] D. Lenz, Uniform ergodic theorems on subshifts over a finite alphabet, Ergodic Theory Dynam. Systems 22 (1) (2002) 245–255.

[10] D. Lenz, P. Müller, I. Veselić, Uniform existence of the integrated density of states for models on $\mathbb{Z}^d$, Positivity 12 (4) (2008) 571–589. arxiv.org/abs/math-ph/0607063.

[11] D. Lenz, F. Schwarzenberger, I. Veselić, A Banach space-valued ergodic theorem and the uniform approximation of the integrated density of states, Geom. Dedicata 150 (1) (2010) 1–34.
[12] D. Lenz, P. Stollmann, An ergodic theorem for Delone dynamical systems and existence of the integrated density of states, J. Anal. Math. 97 (2005) 1–24.
[13] D. Lenz, I. Veselić, Hamiltonians on discrete structures: Jumps of the integrated density of states and uniform convergence, Math. Z. 263 (2009) 813–835.
[14] P. Massart, The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality, Ann. Probab. 18 (3) (1990) 1269–1283.
[15] L.A. Pastur, Selfaverageability of the number of states of the Schrödinger equation with a random potential, Mat. Fiz. i Funkcional. Anal. Vyp. 2 (238) (1971) 111–116.
[16] L.A. Pastur, A.L. Figotin, Spectra of Random and Almost-Periodic Operators, Springer Verlag, Berlin, 1992.
[17] F. Pogorzelski, F. Schwarzenberger, A Banach space-valued ergodic theorem for amenable groups and applications, J. Anal. Math. (2012) in press. arXiv:1205.3649 [math.SP].
[18] R. Ranga Rao, Relations between weak and uniform convergence of measures with applications, Ann. Math. Statist. 33 (2) (1962) 659–680.
[19] F. Schwarzenberger, The integrated density of states for operators on groups (Ph.D. thesis), Technische Universität Chemnitz, 2013, http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-123241.
[20] M.A. Shubin, Spectral theory and the index of elliptic operators with almost-periodic coefficients, Uspekhi Mat. Nauk 34 (2 (206)) (1979) 95–135. [English translation: Russ. Math. Surveys, 34:109–157, 1979].
[21] T. Spencer, The Schrödinger equation with a random potential—a mathematical review, in: K. Osterwaler, R. Stora (Eds.), Phénomènes Critiques, Systèmes aléatoires, Théories de jauge, Part I, II, in: Les Houches, vol. 1984, North-Holland, Amsterdam, 1986, pp. 895–944.
[22] W. Stute, On a generalization of the Glivenko–Cantelli theorem, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 35 (1976) 167–175.
[23] L. Van Hove, Quelques proprietes generales de l integrale de configuration d un systeme de particules avec interaction, Physica 15 (1949) 951–961.
[24] V.N. Vapnik, A.Ya. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, Theory Probab. Appl. 16 (1971) 264–280.
[25] I. Veselić, Spectral analysis of percolation Hamiltonians, Math. Ann. 331 (4) (2005) 841–865.
[26] I. Veselić, Spectral properties of Anderson-percolation Hamiltonians, Oberwolfach Rep. 3 (1) (2006) 545–547.
[27] I. Veselić, Existence and Regularity Properties of the Integrated Density of States of Random Schrödinger Operators, in: Lecture Notes in Mathematics, vol. 1917, Springer, Berlin, 2008.
[28] P.A. Vuillermot, Thermodynamics of quenched random spin systems, and application to the problem of phase transitions in magnetic (spin) glasses, J. Phys. A: Math. Gen. 10 (8) (1977) 1319.
[29] J. Wehr, M. Aizenman, Fluctuations of extensive functions of quenched random couplings, J. Stat. Phys. 60 (3–4) (1990) 287–306.
[30] F.T. Wright, The empirical discrepancy over lower layers and a related law of large numbers, Ann. Probab. 9 (2) (1981) 323–329.

Chapter 5

Glivenko–Cantelli Theory, Ornstein–Weiss Quasi-Tilings, and Uniform Ergodic Theorems for Distribution Valued Fields over amenable groups

The Annals of Applied Probability
2018, Vol. 28, No. 4, 2417–2450
https://doi.org/10.1214/17-AAP1361

GLIVENKO–CANTELLI THEORY, ORNSTEIN–WEISS QUASI-TILINGS, AND UNIFORM ERGODIC THEOREMS FOR DISTRIBUTION-VALUED FIELDS OVER AMENABLE GROUPS

BY CHRISTOPH SCHUMACHER*, FABIAN SCHWARZENBERGER† AND IVAN VESELIĆ*

TU Dortmund and HTW Dresden†*

We consider random fields indexed by finite subsets of an amenable discrete group, taking values in the Banach-space of bounded right-continuous functions. The field is assumed to be equivariant, local, coordinate-wise monotone and almost additive, with finite range dependence. Using the theory of quasi-tilings we prove an uniform ergodic theorem, more precisely, that averages along a Følner sequence converge uniformly to a limiting function. Moreover, we give explicit error estimates for the approximation in the sup norm.

1. Introduction. Ergodic theorems for Banach space valued functions or fields have been studied among others in [6, 7, 11] in a combinatorial setting. The three quoted papers consider different group actions in increasing generality: the lattice $\mathbb{Z}^d$, monotilable amenable discrete groups and general amenable discrete groups, respectively. Note that amenability is a natural assumption for the validity of the ergodic theorem, as shown explicitly in [14]. Already before that combinatorial ergodic theorems for Banach space valued functions have been proven in the context of Delone dynamical systems; see [8] and the references therein.

The combinatorial framework offers the advantage of a minimum of probabilistic or measure theoretic assumptions, the necessary one being that frequencies or densities of finite patterns are well defined and can be approximated by an exhaustion (corresponding to a law of large numbers). A disadvantage of the combinatorial approach chosen is that the range of colors (or the alphabet corresponding to the values of the random variables) needs to be finite. Also, the derived ergodic theorems are in a sense conditional: The convergence bound depends on the speed of convergence of the pattern frequencies.

Our present research aims at dispensing with the finiteness condition on the set of colors. The price to pay is that we have to assume more probabilistic structure and in particular independence or at least finite range correlations. In return, this structure yields automatically quantitative approximation error bounds. No extra assumptions on the speed of convergence of the pattern frequencies are needed.

Received June 2017.
MSC2010 subject classifications. 60F99, 60B12, 62E20, 60K35.
Key words and phrases. Følner sequence, amenable group, quasi-tilings, Glivenko–Cantelli theory, Uniform convergence, Empirical measures.

For the case of fields defined over $\mathbb{Z}^d$ and $\mathbb{Z}^d$-actions, we have established such an ergodic theorem in [12], which takes on the form of a Glivenko–Cantelli theorem, and which we recall now in an informal way.

THEOREM A ([12]). *Let $\Lambda_n = [0,n)^d \cap \mathbb{Z}^d$, and $\omega = (\omega_g)_{g\in\mathbb{Z}^d} \in \mathbb{R}^{\mathbb{Z}^d}$ be an i.i.d. sequence of real random variables. Assume the field*

$$f : \mathcal{P}(\mathbb{Z}^d) \times \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{B} := \{D : \mathbb{R} \to \mathbb{R} \mid D \text{ right-continuous and bounded}\}$$

is $\mathbb{Z}^d$-equivariant, monotone in each coordinate ω_g, local and almost additive, that is, for disjoint $Q_1, \dots, Q_n \subseteq \mathbb{Z}^d$ and $Q := \bigcup_{i=1}^n Q_i$ we have

$$\left\| f(Q,\omega) - \sum_{i=1}^n f(Q_i,\omega) \right\|_\infty \le \sum_{i=1}^n |\partial Q_i|,$$

where ∂Q_i denotes the boundary set. Assume furthermore that $f_\infty := \sup_\omega \| f(\mathrm{id}, \omega)\|_\infty < \infty$.

Then there is a function $f^* : \mathbb{R} \to \mathbb{R}$ such that for each $m \in \mathbb{N}$, there exist $a(m), b(m) > 0$, such that for all $j \in \mathbb{N}$, $j > 2m$, there is an event $\Omega_{j,m} \subseteq \mathbb{R}^{\mathbb{Z}^d}$, with the properties

$$\mathbb{P}(\Omega_{j,m}) \ge 1 - b(m)\exp\bigl(-a(m)|\Lambda_j|\bigr)$$

and

$$\forall \omega \in \Omega_{j,m} : \qquad \left\| \frac{f(\Lambda_j,\omega)}{|\Lambda_j|} - f^* \right\|_\infty \le 2^{2d+1}\left(\frac{(6d+3+2f_\infty)m^d + 1}{j-2m} + \frac{4}{m} \right).$$

In particular, almost surely we have $\lim_{n\to\infty} \| \frac{f(\Lambda_n,\bullet)}{|\Lambda_n|} - f^*\|_\infty = 0$.

For a precise formulation of the properties of the field f, see Section 2. Let us note that in our theorem f takes values in the Banach space $\mathbb{B}$ of right continuous and bounded functions with sup-norm while in [6, 7, 11] an arbitrary Banach space was allowed. This restriction is due to our use of the Glivenko–Cantelli theory in the proof, and currently we do not know how to extend it to arbitrary Banach spaces.

Naturally, one asks whether the above result and its proof extend to general finitely generated amenable groups. In this case, obviously, the boundary has to be taken with respect to a generating set $S \subseteq G$, and the sequence of squares Λ_n has to be replaced by a Følner sequence. Indeed, if G satisfies additionally,

($\boxplus$) There exists a Følner sequence $(\Lambda_n)_{n\in\mathbb{N}}$ in G, and a sequence of symmetric *grids* $T_n = T_n^{-1} \subseteq G$ such that $G = \dot{\bigcup}_{t\in T_n} \Lambda_n t$ is a disjoint union.

the proofs of [12] apply with technical, but no strategic, modifications, as sketched in Appendix B.

However, it is *not clear* in which generality assumption (⊞) holds. In fact, the existence of tiling Følner sequences (for general amenable groups) has been investigated in several instances. It turned out that there exist useful additional conditions which imply the validity of (⊞); cf. [5, 16]. For instance, a group which is residually finite and amenable contains a tiling Følner sequence. Unfortunately, there is a lack of the complete picture: It is still an open question whether there exists a tiling Følner sequence in each amenable group.

Since this question seems hard to answer, Ornstein and Weiss invented in [10] the theory of ε-quasi tilings. The idea is to consider a tiling which is in several senses weaker as the one in (⊞). For a given $\varepsilon > 0$, one has the following properties:

- the group is not tiled with one element of a Følner sequence, but with finitely many elements of this sequence; the number of these elements depends on ε;
- the tiles are allowed to overlap, but the proportion of the part of any tile which is allowed to intersect other tiles is at most of size ε. This property is called ε-disjointness;
- each element of a Følner sequence with a sufficiently large index is, up to a proportion of size ε the union of ε-disjoint tiles.

The authors showed that each amenable group can be ε-quasi tiled. In [11], these ideas have been developed further in order to obtain quantitative estimates on the portion which is covered by translates of one specific element of the tiles. The proof of our main result, which we state now in an informal way, is based on these results on quasi tilings.

THEOREM B. *Let (Λ_n) be a Følner sequence in a finitely generated group G. Let $\omega = (\omega_g)_{g\in G} \in \mathbb{R}^G$ be an i.i.d. sequence of real random variables. Assume the field*

$$f : \mathcal{P}(G) \times \mathbb{R}^G \to \{D : \mathbb{R} \to \mathbb{R} \mid D \text{ right-continuous and bounded}\},$$

is G-equivariant, monotone in each coordinate ω_g, local and almost additive, that is, for disjoint $Q_1, \dots, Q_n \subseteq G$ and $Q := \bigcup_{i=1}^n Q_i$ we have

$$\left\| f(Q, \omega) - \sum_{i=1}^n f(Q_i, \omega) \right\|_\infty \le \sum_{i=1}^n |\partial Q_i|,$$

where ∂Q_i denotes the boundary relative to a set of generators $S \subseteq G$. Assume furthermore, that $f_\infty := \sup_\omega \| f(\mathrm{id}, \omega)\|_\infty < \infty$.

Then there is a function $f^* : \mathbb{R} \to \mathbb{R}$ such that for each $\delta \in (0, 1)$, there exists $a(\delta) > 0$, such that for all sufficiently large $j \in \mathbb{N}$, there is an event $\Omega_{j,\delta} \subseteq \mathbb{R}^G$, with the properties

$$\mathbb{P}(\Omega_{j,\delta}) \ge 1 - \exp\bigl(-a(\delta)|\Lambda_j|\bigr)$$

and

$$\forall \omega \in \Omega_{j,\delta} \qquad \left\| \frac{f(\Lambda_j, \omega)}{|\Lambda_j|} - f^* \right\|_\infty \le (37 f_\infty + 84|S| + 131)\delta.$$

In particular, almost surely we have $\lim_{n\to\infty} \| \frac{f(\Lambda_n,\bullet)}{|\Lambda_n|} - f^* \|_\infty = 0$.

For a precise formulation, see Definition 2.2 and Theorem 2.5. To achieve the error bound in the theorem, we work with an ε-quasi tiling with $\varepsilon = \delta^2$.

REMARK 1.1. Let us sketch the difference between the proof of Theorem B (see also Theorem 2.5 below) and the Theorem 2.8 of [12] sketched as Theorem A above. There we heavily relied on the fact that $\mathbb{Z}^d$ can be tiled exactly with any cube of integer length. Since a general discrete amenable group need not have such a tiling, we have to modify the *geometric parts* of the proof and use ε-quasi tilings as in [10, 11]. Since quasi tilings in general overlap, we lose independence of the corresponding random variables. This requires a change in the *probabilistic part* of the proof and in particular the use of resampling.

The structure of the paper is as follows. In Section 2, we precisely describe the model and our result. In Section 3, we summarize results about ε-quasi tilings, which are fundamental for our proof. The error estimate in the main theorem and the corresponding approximation procedure naturally split in three parts, which are treated consecutively in Sections 4 to 6. Section 4 is of geometric nature. Section 5 is based on multivariate Glivenko–Cantelli theory. Section 6 is geometric in spirit again. In the Appendix, we prove a resampling lemma and indicate how the proof of [12] could be adapted to cover monotileable amenable groups.

2. Model and main results. We start this section with the introduction of the geometric and probabilistic setting: We recall the notion of a Cayley graph of an amenable group G, introduce random colorings of vertices, and define so-called admissible fields, which are random functions mapping finite subsets of G to functions on $\mathbb{R}$ and satisfying a number of natural properties; cf. Definition 2.2. We are then in the position to formulate our main Theorem 2.5.

Let G be a finitely generated group and $S = S^{-1} \subseteq G \setminus \{\mathrm{id}\}$ a finite generating system. Obviously, G is countable. The set of all finite subsets of G is denoted by $\mathcal{F}$ and is countable as well. Throughout this paper, we will assume that G is amenable, that is, there exists a sequence $(\Lambda_n)_{n\in\mathbb{N}}$ of elements in $\mathcal{F}$ such that for each $K \in \mathcal{F}$ one has

$$\frac{|\Lambda_n \triangle K\Lambda_n|}{|\Lambda_n|} \xrightarrow{n\to\infty} 0. \tag{2.1}$$

Here, $K\Lambda_n := \{kg \mid k \in K, g \in \Lambda_n\}$ is the pointwise group multiplication of sets, $\Lambda_n \triangle K\Lambda_n$ denotes the symmetric difference between the sets Λ_n and $K\Lambda_n$ and $|A|$

denotes the cardinality of the finite set A. A sequence $(\Lambda_n)_{n\in\mathbb{N}}$ satisfying property (2.1) is called *Følner sequence*.

The pair (G, S) gives rise to an undirected graph $\Gamma(G, S) = (V, E)$ with vertex set $V := G$ and edge set $E := \{\{x, y\} \mid xy^{-1} \in S\}$. The graph $\Gamma(G, S)$ is known as the *Cayley graph* of G with respect to the generating system S. Note that by symmetry of S the edge set E is well defined. Let $d\colon G \times G \to \mathbb{N}_0$ denote the usual graph metric of $\Gamma(G, S)$. The distance between two nonempty sets $\Lambda_1, \Lambda_2 \subseteq G$ is given by

$$d(\Lambda_1, \Lambda_2) := \min\{d(x, y) \mid x \in \Lambda_1, y \in \Lambda_2\}.$$

In the case where $\Lambda_1 = \{x\}$ consists of only one element, we write $d(x, \Lambda_2)$ for $d(\{x\}, \Lambda_2)$. The diameter of a nonempty set $\Lambda \in \mathcal{F}$ is defined by $\operatorname{diam}(\Lambda) := \max\{d(x, y) \mid x, y \in \Lambda\}$.

Given $r \geq 0$, the *r-boundary* of a set $\Lambda \subseteq G$ is defined by

$$\partial^r(\Lambda) := \{x \in \Lambda \mid d(x, G \setminus \Lambda) \leq r\} \cup \{x \in G \setminus \Lambda \mid d(x, \Lambda) \leq r\}$$

and besides this we use the notation:

$$\Lambda^r := \Lambda \setminus \partial^r(\Lambda) = \{x \in \Lambda \mid d(x, G \setminus \Lambda) > r\}.$$

It is easy to verify that for a given Følner sequence $(\Lambda_n)_{n\in\mathbb{N}}$, or (Λ_n) for short, and $r \geq 0$ we have

$$\lim_{n\to\infty} \frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|} = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{|\Lambda_n^r|}{|\Lambda_n|} = 1. \tag{2.2}$$

Moreover, if (Λ_n) is a Følner sequence, then for arbitrary $r \geq 0$ the sequence (Λ_n^r) is a Følner sequence as well. Conversely, in order to show that a given sequence (Λ_n) is a Følner sequence, it is sufficient [1, 13] to show for $n \to \infty$ either

$$\frac{|\Lambda_n \triangle S\Lambda_n|}{|\Lambda_n|} \to 0 \quad \text{or} \quad \frac{|\partial^1(\Lambda_n)|}{|\Lambda_n|} \to 0. \tag{2.3}$$

Let us introduce colorings of the group G [or equivalently colorings of the vertices of $\Gamma(G, S)$]. We choose a (finite or infinite) set of possible colors $\mathcal{A} \in \mathcal{B}(\mathbb{R})$. The sample set,

$$\Omega = \mathcal{A}^G = \{\omega = (\omega_g)_{g\in G} \mid \omega_j \in \mathcal{A}\},$$

is the set of all possible colorings of G. Note that G acts in a natural way via translations on Ω. To be precise, we define for each $g \in G$

$$\tau_g\colon \Omega \to \Omega, \qquad (\tau_g\omega)_x = \omega_{xg} \qquad (x \in G). \tag{2.4}$$

Next, we introduce random colorings. As the σ-algebra, we choose $\mathcal{B}(\Omega)$, the product σ-algebra on Ω generated by cylinder sets. Oftentimes, we are interested in (finite) products of $\mathcal{A}$ embedded in the infinite product space Ω. To this end, we set for $\Lambda \subseteq G$

$$\Omega_\Lambda := \mathcal{A}^\Lambda := \{(\omega_g)_{g\in\Lambda} \mid \omega_g \in \mathcal{A}\}$$

and define

$$\Pi_\Lambda : \Omega \to \Omega_\Lambda \quad \text{by} \quad (\Pi_\Lambda(\omega))_g := \omega_g \qquad \text{for each } g \in \Lambda.$$

As shorthand notation, we write ω_Λ instead of $\Pi_\Lambda(\omega)$. Having introduced the measurable space $(\Omega, \mathcal{B}(\Omega))$, we choose a probability measure $\mathbb{P}$ with the following properties:

(M1) *equivariance*: For each $g \in G$, we have $\mathbb{P} \circ \tau_g^{-1} = \mathbb{P}$.

(M2) *existence of densities*: There is a σ-finite measure μ_0 on $(\mathcal{A}, \mathcal{B}(\mathcal{A}))$, such that for each $\Lambda \in \mathcal{F}$ the measure $\mathbb{P}_\Lambda := \mathbb{P} \circ \Pi_\Lambda^{-1}$ is absolutely continuous with respect to $\mu_\Lambda := \bigotimes_{g\in\Lambda} \mu_0$ on Ω_Λ. We denote the corresponding probability density function by ρ_Λ.

(M3) *independence condition*: There exists $r \geq 0$ such that for all $n \in \mathbb{N}$ and nonempty $\Lambda_1, \dots, \Lambda_n \in \mathcal{F}$ with $\min\{d(\Lambda_i, \Lambda_j) \mid i \neq j\} > r$ we have $\rho_\Lambda = \prod_{j=1}^n \rho_{\Lambda_j}$, where $\Lambda = \bigcup_{j=1}^n \Lambda_j$.

The measure $\mathbb{P}_\Lambda$ is called the *marginal measure* of $\mathbb{P}$. It is defined on $(\Omega_\Lambda, \mathcal{B}(\Omega_\Lambda))$, where again $\mathcal{B}(\Omega_\Lambda)$ is generated by the corresponding cylinder sets.

REMARK 2.1. (a) The constant $r \geq 0$ in (M3) can be interpreted as the correlation length. In particular, if $r = 0$ this property implies that the colors of the vertices are chosen independently.

(b) (M2) is trivially satisfied, if $\mathbb{P}$ is a product measure.

In the following, we consider partial orderings on Ω and on $\mathbb{R}^k$, respectively. Here, we write $\omega \leq \omega'$ for $\omega, \omega' \in \Omega$, if for all $g \in G$ we have $\omega_g \leq \omega'_g$. The notion $x \leq x'$ for $x, x' \in \mathbb{R}^k$ is defined in the same way. We consider the Banach space

$$\mathbb{B} := \{F : \mathbb{R} \to \mathbb{R} \mid F \text{ right-continuous and bounded}\},$$

which is equipped with supremum norm $\|\cdot\| := \|\cdot\|_\infty$.

DEFINITION 2.2. A field $f : \mathcal{F} \times \Omega \to \mathbb{B}$ is called *admissible* if the following conditions are satisfied:

(A1) equivariance: for $\Lambda \in \mathcal{F}$, $g \in G$ and $\omega \in \Omega$ we have

$$f(\Lambda g, \omega) = f(\Lambda, \tau_g \omega).$$

(A2) locality: for all $\Lambda \in \mathcal{F}$ and $\omega, \omega' \in \Omega$ satisfying $\Pi_\Lambda(\omega) = \Pi_\Lambda(\omega')$ we have

$$f(\Lambda, \omega) = f(\Lambda, \omega').$$

(A3) almost additivity: for arbitrary $\omega \in \Omega$, pairwise disjoint $\Lambda_1, \dots, \Lambda_n \in \mathcal{F}$ and $\Lambda := \bigcup_{i=1}^n \Lambda_i$ we have

$$\left\| f(\Lambda, \omega) - \sum_{i=1}^n f(\Lambda_i, \omega) \right\| \le \sum_{i=1}^n b(\Lambda_i),$$

where $b\colon \mathcal{F} \to [0, \infty)$ satisfies:

- $b(\Lambda) = b(\Lambda g)$ for arbitrary $\Lambda \in \mathcal{F}$ and $g \in G$,
- $\exists D_f > 0$ with $b(\Lambda) \le D_f |\Lambda|$ for arbitrary $\Lambda \in \mathcal{F}$,
- $\lim_{i\to\infty} b(\Lambda_i)/|\Lambda_i| = 0$, if $(\Lambda_i)_{i\in\mathbb{N}}$ is a Følner sequence.
- for $\Lambda, \Lambda' \in \mathcal{F}$, we have $b(\Lambda \cup \Lambda') \le b(\Lambda) + b(\Lambda')$, $b(\Lambda \cap \Lambda') \le b(\Lambda) + b(\Lambda')$, and $b(\Lambda \setminus \Lambda') \le b(\Lambda) + b(\Lambda')$.

(A4) monotonicity: f is antitone with respect to the partial orderings on $\Omega \subseteq \mathbb{R}^G$ and $\mathbb{B}$, that is, if $\omega, \omega' \in \Omega$ satisfy $\omega \le \omega'$, we have

$$f(\Lambda, \omega)(x) \ge f(\Lambda, \omega')(x) \qquad \text{for all } x \in \mathbb{R} \text{ and } \Lambda \in \mathcal{F}.$$

(A5) boundedness:

$$\sup_{\omega\in\Omega} \| f(\{\mathrm{id}\}, \omega)\| < \infty.$$

REMARK 2.3. • Locality (A2) can be formulated as follows: $f(\Lambda, \cdot)$ is $\sigma(\Pi_\Lambda)$-measurable. This enables us to define $f_\Lambda\colon \Omega_\Lambda \to \mathbb{B}$ by $f_\Lambda(\omega_\Lambda) := f(\Lambda, \omega)$ with $\Lambda \in \mathcal{F}$ and $\omega \in \Omega$.

- We call the function b in (A3) *boundary term*. Note that the fourth assumption on b in (A3) was not made in [12]. Indeed, this inequality is used to separate overlapping tiles and is unnecessary as soon as the group has the tiling property (⊞). This fourth point is used only in Lemmas 3.5 and 5.3.
- The antitonicity assumption in (A4) can be weakened. In particular, our proofs apply to fields which are monotone in each coordinate, where the direction of the monotonicity can be different for distinct coordinates. For simplicity reasons and as our main example (see [12]) satisfies (A4), we restrict ourselves to this kind of monotonicity.
- As shown in [12], a combination of (A1), (A3) and (A5) implies that the bound

$$\begin{aligned} K_f &:= \sup\{\| f(\Lambda, \omega)\|/|\Lambda| \mid \omega \in \Omega, \Lambda \in \mathcal{F}\} \\ &\le D_f + \sup_{\omega\in\Omega} \| f(\{\mathrm{id}\}, \omega)\| < \infty. \end{aligned} \tag{2.5}$$

DEFINITION 2.4. A set $\mathcal{U}$ of admissible fields is called *admissible set*, if their bound is uniform:

$$K_{\mathcal{U}} := \sup_{f\in\mathcal{U}} K_f < \infty$$

and each for each $f \in \mathcal{U}$ condition (A3) is satisfied with the same boundary term b. In this situation, we denote the constant in (A3) by $D_{\mathcal{U}}$.

Let us state the main theorem of this paper.

THEOREM 2.5. *Let G be a finitely generated amenable group with a Følner sequence (Λ_n). Further, let $\mathcal{A} \in \mathcal{B}(\mathbb{R})$ and $(\Omega = \mathcal{A}^G, \mathcal{B}(\Omega), \mathbb{P})$ a probability space such that $\mathbb{P}$ satisfies* (M1) *to* (M3). *Finally, let $\mathcal{U}$ be an admissible set.*

(a) *Then there exists an event $\tilde{\Omega} \in \mathcal{B}(\Omega)$ such that $\mathbb{P}(\tilde{\Omega}) = 1$ and for any $f \in \mathcal{U}$ there exists a function $f^* \in \mathbb{B}$, which does not depend on the specific Følner seqeunce (Λ_n), with*

$$\forall \omega \in \tilde{\Omega}: \qquad \lim_{n\to\infty} \left\| \frac{f(\Lambda_n, \omega)}{|\Lambda_n|} - f^* \right\| = 0.$$

(b) *Furthermore, for each $\varepsilon \in (0, 1/10)$, there exist $j_0(\varepsilon) \in \mathbb{N}$, independent of $K_{\mathcal{U}}$, and $a(\varepsilon, K_{\mathcal{U}}), b(\varepsilon, K_{\mathcal{U}}) > 0$, such that for all $j \in \mathbb{N}$, $j \geq j_0(\varepsilon)$, there is an event $\Omega_{j,\varepsilon,K_{\mathcal{U}}} \in \mathcal{B}(\Omega)$, with the properties*

$$\mathbb{P}(\Omega_{j,\varepsilon,K_{\mathcal{U}}}) \geq 1 - b(\varepsilon, K_{\mathcal{U}}) \exp\bigl(-a(\varepsilon, K_{\mathcal{U}})|\Lambda_j|\bigr)$$

and

$$\left\| \frac{f(\Lambda_j, \omega)}{|\Lambda_j|} - f^* \right\| \leq (37K_{\mathcal{U}} + 47D_{\mathcal{U}} + 47)\sqrt{\varepsilon} \qquad \text{for all } \omega \in \Omega_{j,\varepsilon,K_{\mathcal{U}}} \text{ and all } f \in \mathcal{U}.$$

For examples of measures $\mathbb{P}$ satisfying (M1) to (M3) and of admissible fields, we refer to [12]. The generalization of the geometry from the lattice $\mathbb{Z}^d$ to an amenable group G does not affect the examples. See also [9, 15] for a discussion of models giving rise to a discontinuous integrated density of states, which nevertheless can be uniformly approximated by almost additive fields.

3. Outline of $\boldsymbol{\varepsilon}$-quasi tilings. Let us give a brief introduction to the theory of ε-quasi tilings. The main ideas go back to Ornstein and Weiss in [10]. However, the specific results we use here are taken from [11]; see also [13].

Let (Q_n) be a Følner sequence. This sequence is called *nested*, if for all $n \in \mathbb{N}$ we have $\{\mathrm{id}\} \subseteq Q_n \subseteq Q_{n+1}$. Using translations and subsequences, it is easy to show that every amenable group contains a nested Følner sequence; cf. [11], Lemma 2.6.

We will use the elements of the nested Følner sequence (Q_n) to ε-quasi tile elements of a given Følner sequence (Λ_j) for (very) large index j. The next definition provides the notion of an α-covering, ε-disjointness and ε-quasi tiling.

DEFINITION 3.1. Let G be a finitely generated group, $\alpha, \varepsilon \in (0, 1)$ and I some index set.

- The sets $Q_i \in \mathcal{F}$, $i \in I$, are said to *α-cover* the set $\Lambda \in \mathcal{F}$, if:
  (i) $\bigcup_{i \in I} Q_i \subseteq \Lambda$, and
  (ii) $|\Lambda \cap \bigcup_{i \in I} Q_i| \geq \alpha |\Lambda|$.
- The sets $Q_i \in \mathcal{F}$, $i \in I$ are called *ε-disjoint*, if there are subsets $\mathring{Q}_i \subseteq Q_i$, $i \in I$, such that for all $i \in I$ we have:
  (i) $|Q_i \setminus \mathring{Q}_i| \leq \varepsilon |Q_i|$, and
  (ii) $\mathring{Q}_i$ and $\bigcup_{j \in I \setminus \{i\}} \mathring{Q}_j$ are disjoint.
- The $K_i \in \mathcal{F}$, $i \in I$, are said to *ε-quasi tile* $\Lambda \in \mathcal{F}$, if there exist $T_i \in \mathcal{F}$, $i \in I$, such that:
  (i) the elements of $\{K_i T_i \mid i \in I\}$ are pairwise disjoint,
  (ii) for each $i \in I$, the elements of $\{K_i t \mid t \in T_i\}$ are ε-disjoint, and
  (iii) the family $\{K_i T_i \mid i \in I\}$ $(1 - 2\varepsilon)$-covers Λ.

  The set T_i is called *center set* for the *tile* K_i, $i \in I$.

Actually, the details in this definition are adapted to our needs in this paper, as is the following theorem. The general and more technical versions as well as the proof of can be found [11]. See also [10] for earlier results.

Roughly speaking, the following theorem provides, in the setting of finitely generated amenable groups, ε-quasi covers for every set with small enough boundary compared to its volume. Additionally, the theorem also provides control over the fraction covered by different tiles with uniform almost densities. To quantify these densities, we use the standard notation $\lceil b \rceil := \inf\{z \in \mathbb{Z} \mid z \geq b\} = \inf \mathbb{Z} \cap [b, \infty)$ for the smallest integer above $b \in \mathbb{R}$ and define, for all $\varepsilon > 0$ and $i \in \mathbb{N}$,

$$N(\varepsilon) := \left\lceil \frac{\ln(\varepsilon)}{\ln(1 - \varepsilon)} \right\rceil \quad \text{and} \quad \eta_i(\varepsilon) := \varepsilon(1 - \varepsilon)^{N(\varepsilon) - i}. \tag{3.1}$$

THEOREM 3.2. *Let G be a finitely generated amenable group, (Q_n) a nested Følner sequence and $\varepsilon \in (0, 1/10)$. Then there is a finite and strictly increasing selection of sets $K_i \in \{Q_n \mid n \in \mathbb{N}\}$, $i \in \{1, \ldots, N(\varepsilon)\}$, with the following quasi tiling property. For each Følner sequence (Λ_j), there exists $j_0(\varepsilon) \in \mathbb{N}$ such that for all $j \geq j_0(\varepsilon)$, the sets K_i, $i \in \{1, \ldots, N(\varepsilon)\}$, are an ε-quasi tiling of Λ_j. Moreover, for all $j \geq j_0(\varepsilon)$ and all $i \in \{1, \ldots, N(\varepsilon)\}$, the proportion of Λ_j covered by the tile K_i satisfies*

$$\left| \frac{|K_i T_i^j|}{|\Lambda_j|} - \eta_i(\varepsilon) \right| \leq \frac{\varepsilon^2}{N(\varepsilon)}, \tag{3.2}$$

where T_i^j denotes the center set of the tile K_i for the ε-quasi cover of Λ_j.

To make full use of Theorem 3.2, we need some properties of the densities $\eta_i(\varepsilon)$.

LEMMA 3.3. *For $N(\varepsilon)$ and $\eta_i(\varepsilon)$ as in (3.1), the following hold true*:

(a) *For each $\varepsilon \in (0, 1)$, we have*

$$1-\varepsilon \le \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) = 1-(1-\varepsilon)^{N(\varepsilon)} \le 1.$$

(b) *For each $\varepsilon \in (0, 1/10)$ and $i \in \{1, \dots, N(\varepsilon)\}$, we have*

$$\frac{\varepsilon}{N(\varepsilon)} \le \eta_i(\varepsilon) \le \varepsilon.$$

(c) *For a bounded sequence $(\alpha_i)_{i\in\mathbb{N}}$ and $\varepsilon \in (0, 1/10)$, we have the inequality*

$$\left|\sum_{i=1}^{N(\varepsilon)} \alpha_i \eta_i(\varepsilon)\right| \le A\sqrt{\varepsilon} + A_\varepsilon,$$

where $A := \sup\{|\alpha_i| \mid i \in \mathbb{N}\}$ and $A_\varepsilon := \sup\{|\alpha_i| \mid i \in \mathbb{N}, i \ge \varepsilon^{-1/2}\}$. In particular,

$$\lim_{\varepsilon\searrow 0} \sum_{i=1}^{N(\varepsilon)} \alpha_i \eta_i(\varepsilon) \le \liminf_{i\to\infty} |\alpha_i|.$$

PROOF. Part (a) is an easy implication of the sum formula for the geometric series. We refer to [11], Remark 4.3, for the details.

Let us prove (b). By definition of $\eta_i(\varepsilon)$, we have $\eta_i(\varepsilon) \le \varepsilon$. In order to see the other inequality, we note that

$$\eta_i(\varepsilon) \ge \varepsilon(1-\varepsilon)^{N(\varepsilon)-1} \ge \varepsilon(1-\varepsilon)^{\frac{\ln(\varepsilon)}{\ln(1-\varepsilon)}} = \varepsilon^2.$$

Thus, it is sufficient to show that $\varepsilon \ge 1/N(\varepsilon)$. To this end, note that by definition of $N(\varepsilon)$ the following holds true:

$$\varepsilon N(\varepsilon) \ge \frac{\varepsilon \ln(\varepsilon)}{\ln(1-\varepsilon)}.$$

Using the assumption $\varepsilon \in (0, 1/10)$, a short and elementary calculation shows that the last expression is bounded from below by 1.

To verify part (c), set $N_\varepsilon^* := \lfloor \varepsilon^{-1/2} \rfloor := \sup \mathbb{Z} \cap (-\infty, \varepsilon^{-1/2}]$, and calculate as follows:

$$\left|\sum_{i=1}^{N(\varepsilon)} \alpha_i \eta_i(\varepsilon)\right| \le \left|\sum_{i=1}^{N_\varepsilon^*} \alpha_i \eta_i(\varepsilon)\right| + \left|\sum_{i=N_\varepsilon^*+1}^{N(\varepsilon)} \alpha_i \eta_i(\varepsilon)\right| \le A N_\varepsilon^* \varepsilon + A_\varepsilon \le A\sqrt{\varepsilon} + A_\varepsilon.$$

Note that it is easy to show that for $0 < \varepsilon < 1/10$ we have $N(\varepsilon) > N_\varepsilon^* > 0$, such that both sums are nonempty. $\square$

Next, we derive a useful corollary of Theorem 3.2.

COROLLARY 3.4. *Let a finitely generated group G, a subset $\Lambda \in \mathcal{F}$ and $\varepsilon \in (0, 1/2)$ be given. Assume furthermore that the sets $K_i \in \mathcal{F}$, $i \in \{1, \dots, N(\varepsilon)\}$ are an ε-quasi tiling of Λ with almost densities $\eta_i(\varepsilon)$ and center sets $T_i \in \mathcal{F}$, $i \in \{1, \dots, N(\varepsilon)\}$, satisfying (3.2). Then we have for each $i \in \{1, \dots, N(\varepsilon)\}$, the inequality estimating the "density" of the tile K_i:*

$$\left| \frac{|T_i|}{|\Lambda|} - \frac{\eta_i(\varepsilon)}{|K_i|} \right| \le 4\varepsilon \frac{\eta_i(\varepsilon)}{|K_i|}.$$

PROOF. We fix $i \in \{1, \dots, N(\varepsilon)\}$, employ ε-disjointness and the density estimate (3.2) and deduce

$$(1 - \varepsilon) \frac{|K_i||T_i|}{|\Lambda|} \le \frac{|K_i T_i|}{|\Lambda|} \le \eta_i(\varepsilon) + \frac{\varepsilon^2}{N(\varepsilon)}.$$

Therefore, with part (b) of Lemma 3.3, we get

$$\begin{aligned} \frac{|T_i|}{|\Lambda|} - \frac{\eta_i(\varepsilon)}{|K_i|} &\le \frac{\eta_i(\varepsilon) + \frac{\varepsilon^2}{N(\varepsilon)}}{(1-\varepsilon)|K_i|} - \frac{\eta_i(\varepsilon)}{|K_i|} \\ &= \frac{\varepsilon\eta_i(\varepsilon) + \frac{\varepsilon^2}{N(\varepsilon)}}{(1-\varepsilon)|K_i|} \\ &\le \frac{2\varepsilon\eta_i(\varepsilon)}{(1-\varepsilon)|K_i|} \le \frac{4\varepsilon\eta_i(\varepsilon)}{|K_i|}. \end{aligned}$$

Equation (3.2) gives also a bound for the other direction. To be precise, we use

$$\eta_i(\varepsilon) - \frac{\varepsilon^2}{N(\varepsilon)} \le \frac{|K_i T_i|}{|\Lambda|} \le \frac{|K_i||T_i|}{|\Lambda|} \tag{3.3}$$

and again part (b) of Lemma 3.3 to obtain

$$\frac{|T_i|}{|\Lambda|} - \frac{\eta_i(\varepsilon)}{|K_i|} \ge \frac{\eta_i(\varepsilon) - \frac{\varepsilon^2}{N(\varepsilon)}}{|K_i|} - \frac{\eta_i(\varepsilon)}{|K_i|} = -\frac{\varepsilon^2}{N(\varepsilon)|K_i|} \ge -\frac{\varepsilon\eta_i(\varepsilon)}{|K_i|}.$$

This implies the claimed bound. □

Finally, we provide a generalization of almost additivity for sets which are not disjoint, but only ε-disjoint. The proof can be found in [13], Lemma 5.23.

LEMMA 3.5. *Let G be a finitely generated group, f an admissible field with boundary term b and $\varepsilon \in (0, 1/2)$. Then for any ε-disjoint sets Q_i, $i \in \{1, \dots, k\}$, we have for each $\omega \in \Omega$*

$$\left\| f(Q, \omega) - \sum_{i=1}^{k} f(Q_i, \omega) \right\| \le \varepsilon(3K_f + 9D_f)|Q| + 3 \sum_{i=1}^{k} b(Q_i),$$

where $Q := \bigcup_{i=1}^{k} Q_i$ and D_f is the constant from (A3) of Definition 2.2.

4. Approximation via the empirical measure. Given some Følner sequence (Λ_j) and an admissible field f, the aim of this section is the approximation of the expression $\frac{f(\Lambda_j,\omega)}{|\Lambda_j|}$ using elements of a second Følner sequence (Q_n) and associated empirical measures; cf. Lemma 4.3. This second sequence needs to satisfy certain additional assumptions, namely we need that (Q_n) is nested and satisfies for the correlation length $r \in \mathbb{N}_0$ from (M3) that the sequences

$$\left(\frac{b(Q_n)}{|Q_n|}\right), \qquad \left(\frac{b(Q_n^r)}{|Q_n|}\right) \quad \text{and} \qquad \left(\frac{|\partial^r(Q_n)|}{|Q_n|}\right) \qquad \text{converge } \textit{monotonically} \text{ to } 0. \tag{4.1}$$

That these sequences converge to zero is clear by the fact that (Q_n) is a Følner sequence and b a boundary term in the sense of Definition 2.2. In order to obtain the monotonicity in (4.1), we choose a subsequence of (Q_n). These considerations show that each amenable group admits a nested Følner sequence (Q_n) which satisfies (4.1). These terms will be used in the error estimates in the approximations throughout this text. To abbreviate the notation, we define

$$\beta_n' := \max\left\{\frac{b(Q_n)}{|Q_n|}, \frac{b(Q_n^r)}{|Q_n|}, \frac{|\partial^r(Q_n)|}{|Q_n|}\right\} \quad \text{and} \quad \beta(\varepsilon) := \beta_1'\sqrt{\varepsilon} + \beta'_{\lceil 1/\sqrt{\varepsilon}\rceil} \tag{4.2}$$

for $n \in \mathbb{N}$ and $\varepsilon \in (0, 1/10)$. Note that $(\beta_n')_n$ is a monotone sequence and converges to 0, and that by Lemma 3.3(c)

$$\sum_{i=1}^{N(\varepsilon)} \beta_i' \eta_i(\varepsilon) \leq \beta(\varepsilon) \xrightarrow{\varepsilon \searrow 0} 0. \tag{4.3}$$

REMARK 4.1. For the proof of Theorem 2.5, we additionally have to ensure $\beta_n' \leq (2n)^{-1}$ for all $n \in \mathbb{N}$ while taking the subsequences above. We will track the boundary terms throughout the paper and use $\beta(\varepsilon)$ until the very end, where we simplify the result by applying

$$\beta(\varepsilon) = \beta_1'\sqrt{\varepsilon} + \beta'_{\lceil 1/\sqrt{\varepsilon}\rceil} \leq \frac{1}{2}\sqrt{\varepsilon} + \frac{1}{2\lceil 1/\sqrt{\varepsilon}\rceil} \leq \sqrt{\varepsilon}.$$

The cost of this additional condition on the boundary terms is that, via Theorem 3.2, $j_0(\varepsilon)$ in Theorem 2.5 will potentially increase. But up to here, we deal only with the geometry of G and still have that $j_0(\varepsilon)$ depends only on ε.

Moreover, let us emphasize that when considering an admissible set $\mathcal{U}$ the value $\sqrt{\varepsilon}$ gives a uniform bound on $\beta(\varepsilon)$ for all $f \in \mathcal{U}$, since in this situation all $f \in \mathcal{U}$ are almost additive with the same boundary term b.

Define for an admissible field f and $\Lambda \in \mathcal{F}$ the function

$$f_\Lambda : \Omega_\Lambda \to \mathbb{B}, f_\Lambda(\omega) := f(\Lambda, \omega') \qquad \text{where } \omega' \in \Pi_\Lambda^{-1}(\{\omega\}). \tag{4.4}$$

Note that by (A2) of Definition 2.2 we see that f_Λ is well defined (and measurable). In the situation where we insert elements of the Følner sequence (Λ_n) or (Λ_n^r), for some $r \in \mathbb{N}_0$, we write

$$f_n := f_{\Lambda_n} \qquad \text{or} \qquad f_n^r := f_{\Lambda_n^r}. \tag{4.5}$$

For given $K, T \in \mathcal{F}$ and $\omega \in \Omega$, we define the *empirical measure* by

$$L^\omega(K,T)\colon \mathcal{B}(\Omega_{KT}) \to [0,1], \qquad L^\omega(K,T) = \frac{1}{|T|}\sum_{t\in T} \delta_{(\tau_t\omega)_K}. \tag{4.6}$$

Given $\varepsilon \in (0, 1/10)$ and sequences (Λ_j) and (Q_j) as above, we obtain by Theorem 3.2 finite sets $K_i(\varepsilon)$, $i = 1, \dots, N(\varepsilon)$ and (for j large enough) center sets $T_i^j(\varepsilon)$ which form an ε-quasi tiling of Λ_j. In this setting, we use for given $\omega \in \Omega$, $\varepsilon \in (0, 1/10)$, $r > 0$, $i \in \{1, \dots, N(\varepsilon)\}$ and $j \in \mathbb{N}$ large enough the notation

$$L_{i,j}^\omega(\varepsilon) := L^\omega\bigl(K_i(\varepsilon), T_i^j(\varepsilon)\bigr) \quad \text{and} \quad f_i(\varepsilon) := f_{K_i(\varepsilon)} \tag{4.7}$$

as well as

$$L_{i,j}^{r,\omega}(\varepsilon) := L^\omega\bigl(K_i^r(\varepsilon), T_i^j(\varepsilon)\bigr) \quad \text{and} \quad f_i^r(\varepsilon) := f_{K_i^r(\varepsilon)}. \tag{4.8}$$

Here, the reader may recall that $K_i^r(\varepsilon) = K_i(\varepsilon) \setminus \partial^r(K_i(\varepsilon))$.

Moreover, we use for $\Lambda \in \mathcal{F}$, a measurable $f\colon \Omega_\Lambda \to \mathbb{B}$ and a measure ν on $(\Omega_\Lambda, \mathcal{B}(\Omega_\Lambda))$ the notation

$$\langle f, \nu\rangle := \int_{\Omega_\Lambda} f(\omega)\, \mathrm{d}\nu(\omega).$$

LEMMA 4.2. *Let f be an admissible field and let $K, T \in \mathcal{F}$ and $\omega \in \Omega$. Then*

$$\bigl\langle f_K, L^\omega(K,T)\bigr\rangle = \frac{1}{|T|}\sum_{t\in T} f(Kt, \omega).$$

PROOF. We calculate using linearity and (A1) of Definition 2.2:

$$\begin{aligned}
\bigl\langle f_K, L^\omega(K,T)\bigr\rangle &= \int_{\Omega_K} f_K(\omega')\, \mathrm{d}L^\omega(K,T)(\omega') \\
&= \frac{1}{|T|}\sum_{t\in T}\int_{\Omega_K} f_K(\omega')\, \mathrm{d}\delta_{(\tau_t\omega)_K}(\omega') \\
&= \frac{1}{|T|}\sum_{t\in T} f_K\bigl((\tau_t\omega)_K\bigr) \\
&= \frac{1}{|T|}\sum_{t\in T} f(Kt, \omega).
\end{aligned}$$

□

We proceed with the first approximation lemma.

LEMMA 4.3. *Let G be a finitely generated amenable group, let f be an admissible field and let (Λ_n) and (Q_n) be Følner sequences, where (Q_n) is additionally nested and satisfies* (4.1). *Then we have for all $\omega \in \Omega$ that*

$$\lim_{\varepsilon \searrow 0} \lim_{j\to\infty} \left\| \frac{f(\Lambda_j,\omega)}{|\Lambda_j|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), L_{i,j}^{r,\omega}(\varepsilon)\rangle}{|K_i(\varepsilon)|} \right\| = 0, \tag{4.9}$$

where $K_i(\varepsilon)$, $i \in \{1,\dots,N(\varepsilon)\}$ are given by Theorem 3.2. *Moreover, we have for arbitrary $\varepsilon \in (0,1/10)$ and $j \ge j_0(\varepsilon)$, with $j_0(\varepsilon)$ from Theorem* 3.2, *the inequality*

$$\begin{aligned} &\left\| \frac{f(\Lambda_j,\omega)}{|\Lambda_j|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), L_{i,j}^{r,\omega}(\varepsilon)\rangle}{|K_i(\varepsilon)|} \right\| \\ &\qquad \le (9K_f + 15D_f)\varepsilon + 12(2 + K_f + D_f)\beta(\varepsilon). \end{aligned}$$

PROOF. Let $\varepsilon \in (0,1/10)$ and $j \ge j_0(\varepsilon) \in \mathbb{N}$ be given, where $j_0(\varepsilon)$ is the constant given by Theorem 3.2. We estimate using the triangle inequality

$$\begin{aligned} &\left\| \frac{f(\Lambda_j,\omega)}{|\Lambda_j|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), L_{i,j}^{r,\omega}(\varepsilon)\rangle}{|K_i(\varepsilon)|} \right\| \\ &\qquad \le a(\varepsilon,j) + \sum_{i=1}^{N(\varepsilon)} b_i(\varepsilon,j) + \sum_{i=1}^{N(\varepsilon)} c_i(\varepsilon,j), \end{aligned} \tag{4.10}$$

where

$$a(\varepsilon,j) := \frac{1}{|\Lambda_j|} \left\| f(\Lambda_j,\omega) - \sum_{i=1}^{N(\varepsilon)} \sum_{t\in T_i^j(\varepsilon)} f\bigl(K_i(\varepsilon)t,\omega\bigr) \right\|,$$

$$b_i(\varepsilon,j) := \left\| \sum_{t\in T_i^j(\varepsilon)} \frac{f(K_i(\varepsilon)t,\omega)}{|\Lambda_j|} - \eta_i(\varepsilon) \frac{\langle f_i(\varepsilon), L_{i,j}^{\omega}(\varepsilon)\rangle}{|K_i(\varepsilon)|} \right\| \quad \text{and}$$

$$c_i(\varepsilon,j) := \frac{\eta_i(\varepsilon)}{|K_i(\varepsilon)|} \bigl\| \langle f_i(\varepsilon), L_{i,j}^{\omega}(\varepsilon)\rangle - \langle f_i^r(\varepsilon), L_{i,j}^{r,\omega}(\varepsilon)\rangle \bigr\|.$$

Here, the expressions $L_{i,j}^{\omega}(\varepsilon)$ and $f_i(\varepsilon)$ are given by (4.7). Let us estimate the term $a(\varepsilon,j)$. To this end, denote the part which is covered by translates of $K_i(\varepsilon)$, $i \in \{1,\dots,N(\varepsilon)\}$ by

$$R_i^j(\varepsilon) := \bigcup_{i=1}^{N(\varepsilon)} K_i(\varepsilon) T_i^j(\varepsilon) \subseteq \Lambda_j.$$

Then we have, using the properties of the ε-quasi tiling and part (a) of Lemma 3.3,

$$|R_i^j(\varepsilon)| = \sum_{i=1}^{N(\varepsilon)} |K_i(\varepsilon)T_i^j(\varepsilon)| \ge |\Lambda_j| \sum_{i=1}^{N(\varepsilon)} \left(\eta_i(\varepsilon) - \frac{\varepsilon^2}{N(\varepsilon)}\right) \ge (1-2\varepsilon)|\Lambda_j|,$$

which in turn gives $|\Lambda_j \setminus R_i^j(\varepsilon)| \le 2\varepsilon|\Lambda_j|$. We use this and Lemma 3.5 to calculate

$$\begin{aligned}
|\Lambda_j|a(\varepsilon, j) &\le (3K_f + 9D_f)\varepsilon|\Lambda_j| + 3b(\Lambda_j \setminus R_i^j(\varepsilon)) \\
&\quad + \|f(\Lambda_j \setminus R_i^j(\varepsilon))\| + 3\sum_{i=1}^{N(\varepsilon)} \sum_{t\in T_i^j(\varepsilon)} b(K_i(\varepsilon)t) \\
&\le (3K_f + 9D_f)\varepsilon|\Lambda_j| + (K_f + 3D_f)|\Lambda_j \setminus R_i^j(\varepsilon)| \\
&\quad + 3\sum_{i=1}^{N(\varepsilon)} |T_i^j(\varepsilon)| b(K_i(\varepsilon)) \\
&\le (5K_f + 15D_f)\varepsilon|\Lambda_j| + 3\sum_{i=1}^{N(\varepsilon)} |T_i^j(\varepsilon)| b(K_i(\varepsilon)).
\end{aligned}$$

By ε-disjointness and (3.2), we obtain

$$\begin{aligned}
\frac{1}{2}|K_i(\varepsilon)||T_i^j(\varepsilon)| &\le (1-\varepsilon)|K_i(\varepsilon)||T_i^j(\varepsilon)| \\
&\le |K_i(\varepsilon)T_i^j(\varepsilon)| \\
&\le \left(\eta_i(\varepsilon) + \frac{\varepsilon^2}{N(\varepsilon)}\right)|\Lambda_j|,
\end{aligned} \tag{4.11}$$

which together with (b) of Lemma 3.3 gives

$$\begin{aligned}
&\sum_{i=1}^{N(\varepsilon)} |T_i^j(\varepsilon)| b(K_i(\varepsilon)) \\
&\quad \le 2|\Lambda_j| \sum_{i=1}^{N(\varepsilon)} \left(\eta_i(\varepsilon) + \frac{\varepsilon^2}{N(\varepsilon)}\right)\frac{b(K_i(\varepsilon))}{|K_i(\varepsilon)|} \\
&\quad \le 4|\Lambda_j| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\frac{b(K_i(\varepsilon))}{|K_i(\varepsilon)|}.
\end{aligned}$$

This implies the following bound:

$$a(\varepsilon, j) \le (5K_f + 15D_f)\varepsilon + 12\sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\frac{b(K_i(\varepsilon))}{|K_i(\varepsilon)|}. \tag{4.12}$$

To estimate the second term in (4.10), we apply Lemma 4.2 to obtain

$$\sum_{t\in T_i^j(\varepsilon)} f(K_i(\varepsilon)t, \omega) = |T_i^j(\varepsilon)| \cdot \langle f_i(\varepsilon), L_{i,j}^\omega(\varepsilon)\rangle.$$

Thus, by Corollary 3.4 and the fact $\|\langle f_i(\varepsilon), L^{\omega}_{i,j}(\varepsilon)\rangle\| \le K_f|K_i(\varepsilon)|$, we have for each $i \in \{1,\dots,N(\varepsilon)\}$

$$\begin{aligned} b_i(\varepsilon, j) &= \left\| \frac{|T_i^j(\varepsilon)|\langle f_i(\varepsilon), L^{\omega}_{i,j}(\varepsilon)\rangle}{|\Lambda_j|} - \eta_i(\varepsilon)\frac{\langle f_i(\varepsilon), L^{\omega}_{i,j}(\varepsilon)\rangle}{|K_i(\varepsilon)|}\right\| \\ &= \left|\frac{|T_i^j(\varepsilon)|}{|\Lambda_j|} - \frac{\eta_i(\varepsilon)}{|K_i(\varepsilon)|}\right| \|\langle f_i(\varepsilon), L^{\omega}_{i,j}(\varepsilon)\rangle\| \\ &\le 4\frac{\varepsilon\eta_i(\varepsilon)}{|K_i(\varepsilon)|}K_f\big|K_i(\varepsilon)\big| = 4K_f\varepsilon\eta_i(\varepsilon). \end{aligned} \tag{4.13}$$

Let us finally estimate the term $c_i(\varepsilon, j)$. By Lemma 4.2, we have for each $i \in \{1,\dots,N(\varepsilon)\}$:

$$\begin{aligned} &\|\langle f_i(\varepsilon), L^{\omega}_{i,j}(\varepsilon)\rangle - \langle f_i^r(\varepsilon), L^{r,\omega}_{i,j}(\varepsilon)\rangle\| \\ &\quad\le \frac{1}{|T_i^j(\varepsilon)|}\sum_{t\in T_i^j(\varepsilon)} \|f\big(K_i(\varepsilon)t,\omega\big) - f\big(K_i^r(\varepsilon)t,\omega\big)\| \\ &\quad\le \frac{1}{|T_i^j(\varepsilon)|}\sum_{t\in T_i^j(\varepsilon)} b\big(K_i^r(\varepsilon)\big) + b\big(\partial^r\big(K_i(\varepsilon)\big)\cap K_i(\varepsilon)\big) \\ &\qquad + \|f\big(\partial^r\big(K_i(\varepsilon)\big)t \cap K_i(\varepsilon)t,\omega\big)\| \\ &\quad\le b\big(K_i^r(\varepsilon)\big) + (K_f + D_f)\big|\partial^r\big(K_i(\varepsilon)\big)\big|. \end{aligned} \tag{4.14}$$

Together with (4.10), the estimates for $a(\varepsilon, j)$ in (4.12), for $b_i(\varepsilon, j)$ in (4.13) and for $c_i(\varepsilon, j)$ in (4.14) yield

$$\begin{aligned} &\left\| \frac{f(\Lambda_j,\omega)}{|\Lambda_j|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\frac{\langle f_i^r(\varepsilon), L^{r,\omega}_{i,j}(\varepsilon)\rangle}{|K_i(\varepsilon)|}\right\| \\ &\quad\le (5K_f + 15D_f)\varepsilon + 12\sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\frac{b(K_i(\varepsilon))}{|K_i(\varepsilon)|} \\ &\qquad + \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\left(4K_f\varepsilon + \frac{b(K_i^r(\varepsilon)) + (K_f + D_f)|\partial^r(K_i(\varepsilon))|}{|K_i(\varepsilon)|}\right) \\ &\quad\le (9K_f + 15D_f)\varepsilon \\ &\qquad + 12\sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\frac{b(K_i(\varepsilon)) + b(K_i^r(\varepsilon)) + (K_f + D_f)|\partial^r(K_i(\varepsilon))|}{|K_i(\varepsilon)|}. \end{aligned}$$

To verify (4.9), recall that we assumed that (Q_n) satisfies (4.1). By the choice of $K_i(\varepsilon)$ in Theorem 3.2, this gives

$$\begin{aligned}
&\left\| \frac{f(\Lambda_j,\omega)}{|\Lambda_j|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), L_{i,j}^{r,\omega}(\varepsilon)\rangle}{|K_i(\varepsilon)|} \right\| \\
&\qquad \le (9K_f + 15D_f)\varepsilon \\
&\qquad\quad + 12 \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \underbrace{\frac{b(Q_i) + b(Q_i^r) + (K_f + D_f)|\partial^r(Q_i)|}{|Q_i|}}_{\le (2+K_f+D_f)\beta_i'} \\
&\qquad \le (9K_f + 15D_f)\varepsilon + 12(2 + K_f + D_f)\beta(\varepsilon).
\end{aligned}$$

The last inequality follows from (4.3). As this bound holds for arbitrary $\varepsilon \in (0, 1/10)$ and $j \ge j_0(\varepsilon)$, this particularly proves (4.9). $\square$

5. Approximation via Glivenko–Cantelli. In this section, we aim to apply a multivariate Glivenko–Cantelli theorem in order to approximate the empirical measure using the theoretical measure. Recall that a Glivenko–Cantelli theorem compares the empirical measure of a normalized sum of independent and identically distributed random variables with their distribution. At the end of this section, we will apply the following Glivenko–Cantelli theorem which was proved in [12] based on results by DeHardt and Wright; see [3, 17]. Monotone functions on $\mathbb{R}^k$ were defined in (A4).

THEOREM 5.1. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_t\colon \Omega \to \mathbb{R}^k$, $t \in \mathbb{N}$, independent and identically distributed random variables such that the distribution $\mu := \mathbb{P}(X \in \cdot\,)$ is absolutely continuous with respect to a product measure $\bigotimes_{\ell=1}^k \mu_\ell$ on $\mathbb{R}^k$, where μ_ℓ, $\ell \in \{1,\dots,k\}$, are σ-finite measures on $\mathbb{R}$. For each $n \in \mathbb{N}$, we denote by $L_n^{(\omega)} := \frac{1}{n}\sum_{t=1}^n \delta_{X_t}$ the empirical distribution of $(X_t)_{t\in\{1,\dots,n\}}$. Further, fix $M \in \mathbb{R}$ and let $\mathcal{M} := \{g\colon \mathbb{R}^k \to \mathbb{R} \mid g$ is monotone, and $\sup_{x\in\mathbb{R}^k} |g(x)| \le M\}$.*

Then, for all $\kappa > 0$, there are $a = a(\kappa, M) > 0$ and $b = b(\kappa, M) > 0$ such that for all $n \in \mathbb{N}$, there exists an event $\Omega_{\kappa,n,M} \in \mathcal{A}$ with large probability $\mathbb{P}(\Omega_{\kappa,n,M}) \ge 1 - b\exp(-an)$, such that for all $\omega \in \Omega_{\kappa,n,M}$, we have

$$\sup_{g\in\mathcal{M}} \left|\langle g, L_n^{(\omega)} - \mu\rangle\right| \le \kappa.$$

In particular, there exists a set $\Omega_0 \in \mathcal{A}$ with $\mathbb{P}(\Omega_0) = 1$ and $\sup_{g\in\mathcal{M}} |\langle g, L_n^{(\omega)} - \mu\rangle| \xrightarrow{n\to\infty} 0$ for all $\omega \in \Omega_0$.

In the present situation, we encounter several challenges when applying Theorem 5.1, caused by our tiling scheme:

- Each Λ_j is tiled using $N(\varepsilon)$ different shapes. Thus, the corresponding random variables (for different shapes) are *not identically distributed*.
- In an ε-quasi tiling, translates of the same shape K_i are allowed to overlap. Thus, the corresponding random variables are *not necessarily independent*.

The first point can be handled by applying Glivenko–Cantelli theory for each shape K_i separately. The second point is more challenging. The core of the following approach is the "generation of independence" by resampling of the overlapping areas using conditional probabilities and controlling errors introduced on the altered areas with their volume. Let us explain this in detail.

Fix $\varepsilon > 0$, $i \in \{1, \dots, N(\varepsilon)\}$ and $j \in \mathbb{N}$, $j \geq j_0(\varepsilon)$ (cf. Theorem 3.2), and consider Figure 1, which sketches a tile $K = K_i$, a finite set $\Lambda = \Lambda_j$, and the translations Kt, $t \in T := T_i^j(\varepsilon)$, of $K = K_i$ from an ε-quasi tiling. The sets

$$U^{i,j,t} := (K_i^r t) \setminus (K_i(T_i^j(\varepsilon) \setminus \{t\})) \subseteq G, \qquad t \in T, \tag{5.1}$$

are marked with stripes. Their distance is at least

$$d(U^{i,j,t}, U^{i,j,t'}) \geq d(K_i^r t, G \setminus K_i t) > r, \qquad t \neq t', \tag{5.2}$$

so the colors there are $\mathbb{P}$-independent from each other. Unfortunately, if we take only the values on $U^{i,j,t}$, $t \in T$, we will end up with an independent, but not identically distributed sample. We therefore resample independent colors in $K^r \setminus U^{i,j,t}$.

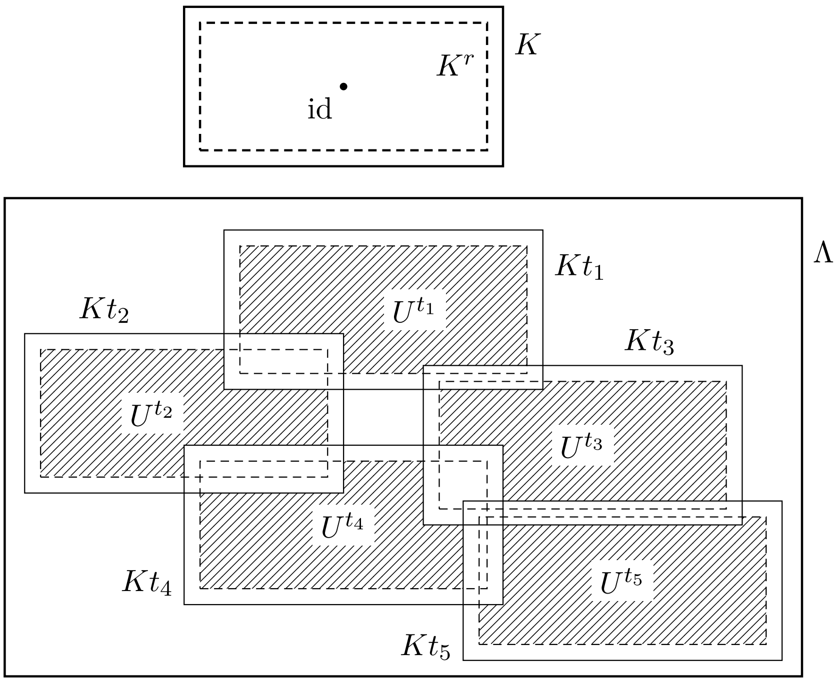


FIG. 1. *ε-covering and independence structure: The set $\Lambda = \Lambda_j \subseteq G$ is ε-quasi covered by copies of $K = K_i$ with centers in $T = T_i^j(\varepsilon) = \{t_1, \dots, t_5\}$. The sets $U^t = U^{i,j,t}$, $t \in T$, here marked by diagonal stripes, have at least distance r and satisfy $|U^t| \geq (1 - \varepsilon)|K|$.*

Fortunately, the sets $U^{i,j,t}$ are large enough to compensate this small random perturbation. The following lemma specifies the resampling procedure.

LEMMA 5.2. *Let $\varepsilon > 0$ and $I := \bigcup_{i=1}^{N(\varepsilon)} \bigcup_{j=j_0(\varepsilon)}^{\infty} \{(i,j)\} \times T_i^j(\varepsilon)$. There exists a probability space $(\underline{\Omega}, \mathcal{B}(\underline{\Omega}), \underline{\mathbb{P}})$ and random variables $X, X^{i,j,t} : \underline{\Omega} \to \Omega$, $(i,j,t) \in I$, such that for all $(i,j,t) \in I$:*

(i) *X and $X^{i,j,t}$ have distribution $\mathbb{P}$,*
(ii) *X and $X^{i,j,t}$ agree on $U^{i,j,t}$ $\underline{\mathbb{P}}$-almost surely, and*
(iii) *the random variables in the set $\{X^{i,j,t'}\}_{t' \in T_i^j(\varepsilon)}$ are $\underline{\mathbb{P}}$-independent.*

PROOF. Theorem A.1 solves the problem of resampling in an abstract setting. We apply the result here as follows. Since we use the canonical probability space in our construction, we apply Theorem A.1 with $(S, \mathcal{S}) := (\Omega, \mathcal{A})$, $X := \mathrm{id}_\Omega$, $I := \bigcup_{i=1}^{N(\varepsilon)} \bigcup_{j=j_0(\varepsilon)}^{\infty} \{(i,j)\} \times T_i^j(\varepsilon)$, and $\mathcal{Y}_{j'} := \sigma(\Pi_{U^{j'}})$, $j' \in I$. Theorem A.1 provides the following quantities, which we here want to use as $(\underline{\Omega}, \underline{\mathcal{A}}, \underline{\mathbb{P}}) := (\underline{\Omega}, \underline{\mathcal{A}}, \underline{\mathbb{P}})$, $X := X_0$, and $X^{i,j,t} := X_{j'}$ for all $j' = (i,j,t) \in I$. The properties (i) and (ii) follow directly from Theorem A.1(i), (ii). With (5.2), Theorem A.1(iv) implies (iii). $\square$

Next, we control the error we introduce by using our independent samples instead of the dependent ones.

LEMMA 5.3. *Fix $\varepsilon > 0$, an admissible f and $U \subseteq K \in \mathcal{F}$. For $\omega, \tilde{\omega} \in \Omega$ with $\omega_U = \tilde{\omega}_U$, we have*

$$\| f(\omega, K) - f(\tilde{\omega}, K) \| \le 2b(K) + 2(2D_f + K_f)|K \setminus U|.$$

In particular, in the notation from (4.4)–(4.8) and with the corresponding empirical measure

$$\underline{L}_{i,j}^{r,\underline{\omega}}(\varepsilon) := \frac{1}{|T_i^j(\varepsilon)|} \sum_{t \in T_i^j(\varepsilon)} \delta_{(\tau_t X^{i,j,t}(\underline{\omega}))_{K_i(\varepsilon)}} \qquad (\underline{\omega} \in \underline{\Omega}),$$

we have for $\underline{\mathbb{P}}$-almost all $\underline{\omega} \in \underline{\Omega}$ that

$$\| \langle f_i^r(\varepsilon), L_{i,j}^{r,X(\underline{\omega})}(\varepsilon) - \underline{L}_{i,j}^{r,\underline{\omega}}(\varepsilon) \rangle \| \le 2b\big(K_i^r(\varepsilon)\big) + 2(2D_f + K_f)\varepsilon \big|K_i^r(\varepsilon)\big|.$$

PROOF. The values of ω on U determine $f(\omega, K)$ up to

$$\begin{aligned} &\| f(\omega, K) - f(\omega, U) \| \\ &\quad \le \| f(\omega, K) - f(\omega, U) - f(\omega, K \setminus U) \| + \| f(\omega, K \setminus U) \| \\ &\quad \le b(U) + b(K \setminus U) + \| f(\omega, K \setminus U) \| \\ &\quad \le b(U) + (D_f + K_f)|K \setminus U|. \end{aligned}$$

With the fourth point in (A3), we can continue this estimate with

$$b(U) \leq b\bigl(K \setminus (K \setminus U)\bigr) \leq b(K) + b(K \setminus U) \leq b(K) + D_f|K \setminus U|.$$

We now employ the triangle inequality to show the first claim: For $\omega, \tilde{\omega} \in \Omega$ with $\omega_U = \tilde{\omega}_U$, we have

$$\begin{aligned}
&\|f(\omega, K) - f(\tilde{\omega}, K)\| \\
&\quad \leq \|f(\omega, K) - f(\omega, U)\| + \|f(\tilde{\omega}, U) - f(\tilde{\omega}, K)\| \\
&\quad \leq 2\bigl(b(K) + (2D_f + K_f)|K \setminus U|\bigr).
\end{aligned}$$

This calculation allows us to change ω on $K \setminus U$ to the independent values provided by Lemma 5.2. To implement this, observe that for $\underline{\mathbb{P}}$-almost all $\underline{\omega} \in \underline{\Omega}$ and all $i \in \{1, \dots, N(\varepsilon)\}$, $j \in \mathbb{N}$, $j \geq j_0(\varepsilon)$ and $t \in T_i^j(\varepsilon)$, the set $U^{i,j,t}$ from (5.1) exhausts $K_i^r(\varepsilon)t$ up to a fraction of ε: $|K_i^r(\varepsilon)t \setminus U^{i,j,t}| \leq \varepsilon|K_i^r(\varepsilon)|$. By construction, on $U^{i,j,t}$, the colors are preserved: $U^{i,j,t} \subseteq \{g \in K_i^r(\varepsilon)t \mid X_g(\underline{\omega}) = X_g^{i,j,t}(\underline{\omega})\}$. Together with Lemma 4.2 and the triangle inequality, this immediately implies for $\underline{\mathbb{P}}$-almost all $\underline{\omega} \in \underline{\Omega}$ that

$$\begin{aligned}
&\|\langle f_i^r(\varepsilon), L_{i,j}^{r,X(\underline{\omega})}(\varepsilon) - \underline{L}_{i,j}^{r,\tilde{\omega}}(\varepsilon)\rangle\| \\
&\quad \leq \frac{1}{|T_i^j(\varepsilon)|} \sum_{t \in T_i^j(\varepsilon)} \|f\bigl(K_i^r(\varepsilon)t, \omega\bigr) - f\bigl(K_i^r(\varepsilon)t, X^{i,j,t}(\underline{\omega})\bigr)\| \\
&\quad \leq 2b\bigl(K_i^r(\varepsilon)\bigr) + 2(2D_f + K_f)\varepsilon\bigl|K_i^r(\varepsilon)\bigr|. \qquad \square
\end{aligned}$$

The empirical measure $L_{i,j}^{r,X(\underline{\omega})}$ formed by independent samples should converge to

$$\mathbb{P}_i^r(\varepsilon) := \mathbb{P}_{K_i^r(\varepsilon)}.$$

The following result makes this notion precise. It is the main result of this section.

PROPOSITION 5.4. *Let G be a finitely generated amenable group, let $\mathcal{A} \in \mathcal{B}(\mathbb{R})$ and $(\Omega := \mathcal{A}^G, \mathcal{B}(\Omega), \mathbb{P})$ a probability space such that $\mathbb{P}$ satisfies (M1) to (M3). Moreover, let (Λ_n) and (Q_n) be Følner sequences, where (Q_n) is nested and satisfies (4.1). For given $\varepsilon \in (0, 1/10)$, let $K_i(\varepsilon)$, $i \in \{1, \dots, N(\varepsilon)\}$, and $j_0(\varepsilon)$ be given by Theorem 3.2. Furthermore, let $\mathcal{U}$ be an admissible set of admissible fields.*

Then, for all $\kappa > 0$, there exist $a(\varepsilon, \kappa, K_{\mathcal{U}}), b(\varepsilon, \kappa, K_{\mathcal{U}}) > 0$ such that for all $j \geq j_0(\varepsilon)$, there is an event $\Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}} \in \mathcal{B}(\Omega)$ with large probability

$$\mathbb{P}(\Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}}) \geq 1 - b(\varepsilon, \kappa, K_{\mathcal{U}}) \exp\bigl(-a(\varepsilon, \kappa, K_{\mathcal{U}})|\Lambda_j|\bigr)$$

and the property that for all $\omega \in \Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}}$ and $f \in \mathcal{U}$, it holds true that

$$\left\| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), L_{i,j}^{r,\omega}(\varepsilon)\rangle}{|K_i(\varepsilon)|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), \mathbb{P}_i^r(\varepsilon)\rangle}{|K_i(\varepsilon)|} \right\| \\ \leq 2\beta(\varepsilon) + 2(2D_f + K_f)\varepsilon + \kappa.$$

In particular, there is an event $\tilde{\Omega} \in \mathcal{B}(\Omega)$ with $\mathbb{P}(\tilde{\Omega}) = 1$ such that for all $\omega \in \tilde{\Omega}$, we have

$$\lim_{\varepsilon \searrow 0} \sup_{f \in \mathcal{U}} \left\| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), L_{i,j}^{r,\omega}(\varepsilon)\rangle}{|K_i(\varepsilon)|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), \mathbb{P}_i^r(\varepsilon)\rangle}{|K_i(\varepsilon)|} \right\| = 0.$$

PROOF. Fix $f \in \mathcal{U}$. For $\varepsilon \in (0, 1/10)$, $j \in \mathbb{N}$ and $\omega \in \Omega$, two applications of the triangle inequality give

$$\begin{aligned} \Delta_f(\varepsilon, \omega) &:= \left\| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), L_{i,j}^{r,\omega}(\varepsilon)\rangle}{|K_i(\varepsilon)|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), \mathbb{P}_i^r(\varepsilon)\rangle}{|K_i(\varepsilon)|} \right\| \\ &\leq \sum_{i=1}^{N(\varepsilon)} \frac{\eta_i(\varepsilon)}{|K_i(\varepsilon)|} \| \langle f_i^r(\varepsilon), L_{i,j}^{r,\omega}(\varepsilon) - \mathbb{P}_i^r(\varepsilon)\rangle \| \qquad (5.3) \\ &\leq \inf_{\underline{\omega} \in X^{-1}(\{\omega\})} \left(\sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\gamma_1(i, j, \varepsilon, \underline{\omega}) + \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\gamma_2(i, j, \varepsilon, \underline{\omega}) \right), \end{aligned}$$

where $\underline{\omega} \in \underline{\Omega}$ extends ω, that is, $X(\underline{\omega}) = \omega$ in the notation of Lemma 5.2, and

$$\gamma_1(i, j, \varepsilon, \underline{\omega}) := \frac{\| \langle f_i^r(\varepsilon), L_{i,j}^{r,\omega}(\varepsilon) - \underline{L}_{i,j}^{r,\underline{\omega}}(\varepsilon)\rangle \|}{|K_i(\varepsilon)|} \quad \text{and}$$

$$\gamma_2(i, j, \varepsilon, \underline{\omega}) := \frac{\| \langle f_i^r(\varepsilon), \underline{L}_{i,j}^{r,\underline{\omega}}(\varepsilon) - \mathbb{P}_i^r(\varepsilon)\rangle \|}{|K_i(\varepsilon)|}.$$

By Lemma 5.3 and assumption (4.1), we see that for all $\underline{\omega} \in \underline{\Omega}$ with $X(\underline{\omega}) = \omega$

$$\gamma_1(i, j, \varepsilon, \underline{\omega}) \leq \frac{2b(K_i^r(\varepsilon))}{|K_i^r(\varepsilon)|} + 2(2D_f + K_f)\varepsilon \leq \frac{2b(Q_i)}{|Q_i|} + 2(2D_f + K_f)\varepsilon.$$

With Lemma 3.3(a) and (4.3), we yield the deterministic upper bound

$$\sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\gamma_1(i, j, \varepsilon, \underline{\omega}) \leq 2\beta(\varepsilon) + 2(2D_f + K_f)\varepsilon$$

for all $\underline{\omega} \in X^{-1}(\omega) \subseteq \underline{\Omega}$. By now, our overall inequality (5.3) reads

$$\Delta_f(\varepsilon, \omega) \leq 2\beta(\varepsilon) + 2(2D_f + K_f)\varepsilon + \inf_{\underline{\omega} \in X^{-1}(\{\omega\})} \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\gamma_2(i, j, \varepsilon, \underline{\omega}). \qquad (5.4)$$

To deal with γ_2, recall that the norm on the Banach space $\mathbb{B}$ our admissible fields map into is the sup-norm. We translate the sup-norm into the Glivenko–Cantelli setting as follows. Let

$$\mathcal{M}_f := \{g^r_{i,E} : \mathbb{R}^{|K^r_i(\varepsilon)|} \to \mathbb{R}, g^r_{i,E}(\omega) := f^r_i(\omega)(E)/|K_i(\varepsilon)| \mid E \in \mathbb{R}\}.$$

Therefore, we can write

$$\gamma_2(i, j, \varepsilon, \underline{\omega}) = \sup_{g \in \mathcal{M}_f} |\langle g, \underline{L}^{r,\underline{\omega}}_{i,j}(\varepsilon) - \mathbb{P}^r_i(\varepsilon)\rangle| \le \sup_{f \in \mathcal{U}} \sup_{g \in \mathcal{M}_f} |\langle g, \underline{L}^{r,\underline{\omega}}_{i,j}(\varepsilon) - \mathbb{P}^r_i(\varepsilon)\rangle|.$$

From (2.5), we see that the fields in $\mathcal{M}_{\mathcal{U}} := \bigcup_{f \in \mathcal{U}} \mathcal{M}_f$ are bounded by $K_{\mathcal{U}}$. As assumed in (A4), the fields in $\mathcal{M}_{\mathcal{U}}$ are also monotone. By Lemma 5.2(iii), the samples are independent, also. This is crucial in order to invoke Theorem 5.1. We thus obtain that, for each $\kappa > 0$, $\varepsilon \in (0, 1/10)$, $i \in \{1, \dots, N(\varepsilon)\}$ and $j \in \mathbb{N}$, $j \ge j_0(\varepsilon)$, there are $a_i \equiv a(i, \varepsilon, \kappa, K_{\mathcal{U}}) > 0$, $b_i \equiv b(i, \varepsilon, \kappa, K_{\mathcal{U}}) > 0$ and $\underline{\Omega}_{i,j} \equiv \underline{\Omega}_{i,j,\varepsilon,\kappa,K_{\mathcal{U}}} \in \mathcal{B}(\underline{\Omega})$ such that

$$\underline{\mathbb{P}}(\underline{\Omega}_{i,j}) \ge 1 - b_i \exp(-a_i |T^j_i(\varepsilon)|) \quad \text{and} \quad \sup_{\underline{\omega} \in \underline{\Omega}_{i,j}} \gamma_2(i, j, \varepsilon, \underline{\omega}) \le \kappa.$$

We need this estimate for all $i \in \{1, \dots, N(\varepsilon)\}$ simultaneously and consider

$$\underline{\Omega}_j \equiv \underline{\Omega}_{j,\varepsilon,\kappa,K_{\mathcal{U}}} := \bigcap_{i=1}^{N(\varepsilon)} \underline{\Omega}_{i,j}.$$

To estimate the probability of $\underline{\Omega}_j$ is the next step. From (3.3) and Lemma 3.3(b), we note that

$$|T^j_i(\varepsilon)| \ge \left(\eta_i(\varepsilon) - \frac{\varepsilon^2}{N(\varepsilon)}\right) \frac{|\Lambda_j|}{|K_i(\varepsilon)|} \ge \frac{(1-\varepsilon)\varepsilon}{N(\varepsilon)|K_i(\varepsilon)|} |\Lambda_j|.$$

With the definition

$$a \equiv a_{\varepsilon,\kappa,K_{\mathcal{U}}} := \frac{(1-\varepsilon)\varepsilon}{N(\varepsilon)} \min_{i \in \{1,\dots,N(\varepsilon)\}} \frac{a_i}{|K_i(\varepsilon)|} \quad \text{and} \quad b \equiv b_{\varepsilon,\kappa,K_{\mathcal{U}}} := 2 \sum_{i=1}^{N(\varepsilon)} b_i,$$

we get $\underline{\mathbb{P}}(\underline{\Omega}_{i,j}) \ge 1 - b_i \exp(-a|\Lambda_j|)$ and

$$\underline{\mathbb{P}}(\underline{\Omega}_j) = 1 - \underline{\mathbb{P}}\left(\bigcup_{i=1}^{N(\varepsilon)} \underline{\Omega} \setminus \underline{\Omega}_{i,j}\right) \ge 1 - \sum_{i=1}^{N(\varepsilon)} \underline{\mathbb{P}}(\underline{\Omega} \setminus \underline{\Omega}_{i,j}) \ge 1 - \frac{b \exp(-a|\Lambda_j|)}{2}.$$

Next, we should transition from $(\underline{\Omega}, \mathcal{B}(\underline{\Omega}), \underline{\mathbb{P}})$ to $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$. The set $X(\underline{\Omega}_j) \subseteq \Omega$ seems to be a good candidate, because for all $\omega \in X(\underline{\Omega}_j)$, there exists $\underline{\omega} \in X^{-1}(\{\omega\}) \cap \bigcap_{i=1}^{N(\varepsilon)} \underline{\Omega}_{i,j}$, and thus we can estimate

$$\inf_{\underline{\omega} \in X^{-1}(\{\omega\})} \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \gamma_2(i, j, \varepsilon, \underline{\omega}) \le \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \kappa \le \kappa.$$

Together with (5.4), this inequality shows the claimed bound on $\Delta_f(\varepsilon, \omega)$ for all $\omega \in X(\underline{\Omega}_j)$.

Unfortunately, the image of a measurable set under a measurable map is not necessarily measurable, but only analytic; see [2], Theorem 10.23. At least the outer measure of our candidate is bounded from below by

$$\begin{aligned}\mathbb{P}^*\big(X(\underline{\Omega}_j)\big) &:= \inf_{B\in\mathcal{B}(\Omega), X(\underline{\Omega}_j)\subseteq B} \mathbb{P}(B) \\ &= \inf_{B\in\mathcal{B}(\Omega), X(\underline{\Omega}_j)\subseteq B} \underline{\mathbb{P}}(X \in B) \\ &\geq \inf_{B\in\mathcal{B}(\Omega), X(\underline{\Omega}_j)\subseteq B} \underline{\mathbb{P}}(\underline{\Omega}_j) \\ &= \underline{\mathbb{P}}(\underline{\Omega}_j) \geq 1 - b\exp\big(-a|\Lambda_j|\big)/2.\end{aligned}$$

From [2], Lemma 10.36, we learn that $\mathbb{P}^*$ is a nice capacity, and the Choquet capacity theorem [2], Theorem 10.39, states for the analytic set $X(\underline{\Omega}_j)$ that

$$\mathbb{P}^*\big(X(\underline{\Omega}_j)\big) = \sup_{K\subseteq X(\underline{\Omega}_j)\text{ compact}} \mathbb{P}(K).$$

Thus, there exists a compact subset $\Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}} \subseteq X(\underline{\Omega}_j)$ with probability at least $1 - b\exp(-a|\Lambda_j|)$.

We complete the proof with a standard Borel–Cantelli argument to show that $\tilde{\Omega}$ exists as claimed. For all $\kappa > 0$, the events

$$A_\kappa := \bigcup_{n=j_0(\varepsilon)}^{\infty} \bigcap_{j=n}^{\infty} \Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}}$$

have probability 1, since

$$\sum_{j=j_0(\varepsilon)}^{\infty} \mathbb{P}(\Omega \setminus \Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}}) \leq \sum_{j=j_0(\varepsilon)}^{\infty} b\exp\big(-a|\Lambda_j|\big) \leq b \sum_{j=j_0(\varepsilon)}^{\infty} \exp(-a)^j < \infty.$$

Note that by (5.4), $\beta(\varepsilon) \to 0$ and by construction of A_k, for all $\omega \in A_\kappa$, we have

$$\lim_{\varepsilon\searrow 0} \sup_{f\in\mathcal{U}} \Delta_f(\varepsilon, \omega) \leq \kappa.$$

Thus, the event $\tilde{\Omega} := \bigcap_{k\in\mathbb{N}} A_{1/k}$ has full probability $\mathbb{P}(\tilde{\Omega}) = 1$, and for all $\omega \in \tilde{\Omega}$, we have $\lim_{\varepsilon\searrow 0} \sup_{f\in\mathcal{U}} \Delta_f(\varepsilon, \omega) = 0$. □

6. Almost additivity and Cauchy sequences. The following calculations are devoted to a Cauchy sequence argument to obtain the desired limit function f^*.

LEMMA 6.1. *Let G be a finitely generated amenable group, let $\mathcal{A} \in \mathcal{B}(\mathbb{R})$ and $(\Omega = \mathcal{A}^G, \mathcal{B}(\Omega), \mathbb{P})$ a probability space such that $\mathbb{P}$ satisfies (M1) to (M3). Moreover, let f be an admissible field and (Q_n) a nested Følner sequence satisfying (4.1). Then there exists $f^* \in \mathbb{B}$ with*

$$\lim_{\varepsilon \searrow 0} \left\| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\langle f_i^r(\varepsilon), \mathbb{P}_i^r(\varepsilon) \rangle}{|K_i(\varepsilon)|} - f^* \right\| = 0,$$

where for $k \in \mathbb{N}$ and $\varepsilon \in (1/(k+1), 1/k)$ the sets $K_i(\varepsilon)$, $i \in \{1, \dots, N(\varepsilon)\}$ are extracted from the sequence $(Q_{n+k})_n$ via Theorem 3.2. The approximation error is bounded by

$$\left\| \sum_{j=1}^{N(\varepsilon)} \eta_j(\varepsilon) \frac{\langle f_j^r(\varepsilon), \mathbb{P}_j^r(\varepsilon) \rangle}{|K_j(\varepsilon)|} - f^* \right\| \leq (9K_f + 11D_f)\varepsilon + 5(4 + K_f + D_f)\beta(\varepsilon).$$

PROOF. In order to prove the existence of f^*, we study for $\varepsilon, \delta \in (0, 1/10)$ the difference

$$\mathcal{D}(\varepsilon, \delta) := \left\| \sum_{j=1}^{N(\varepsilon)} \eta_j(\varepsilon) \frac{\langle f_j^r(\varepsilon), \mathbb{P}_j^r(\varepsilon) \rangle}{|K_j(\varepsilon)|} - \sum_{i=1}^{N(\delta)} \eta_i(\delta) \frac{\langle f_i^r(\delta), \mathbb{P}_i^r(\delta) \rangle}{|K_i(\delta)|} \right\|.$$

Our aim is to show $\lim_{\delta \searrow 0} \lim_{\varepsilon \searrow 0} \mathcal{D}(\varepsilon, \delta) = 0$. To prove this, we insert terms which interpolate between the minuend and the subtrahend. These terms will be given using Theorem 3.2. For each $\varepsilon \in (1/(k+1), 1/k]$, we apply Theorem 3.2 to choose the sets $K_j(\varepsilon)$, $j = 1, \dots, N(\varepsilon)$, from the Følner sequence $(Q_{n+k})_{n \in \mathbb{N}}$. The particular choice of the sets $K_j(\varepsilon)$, $j = 1, \dots, N(\varepsilon)$, as elements of the sequence $(Q_{n+k})_n$ ensures that for given $\delta > 0$ we find $\varepsilon_0 > 0$ such that for arbitrary $\varepsilon \in (0, \varepsilon_0)$ each $K_j(\varepsilon)$, $j = 1, \dots, N(\varepsilon)$, can be δ-quasi tiled with the elements $K_i(\delta)$, $i = 1, \dots, N(\delta)$. As in Theorem 3.2, we denote the associated center sets by $T_i^j(\delta)$, where we emphasize the dependence on the parameter δ.

For $K \in \mathcal{F}$, we use the notation

$$F(K) := \langle f_K, \mathbb{P}_K \rangle \tag{6.1}$$

and hence for the tiles $K_j(\varepsilon)$, $i = 1, \dots, N(\varepsilon)$, we write $F(K_i^r(\varepsilon)) := \langle f_i^r(\varepsilon), \mathbb{P}_i^r(\varepsilon) \rangle$. The function F is translation invariant, that is, for all $K \in \mathcal{F}$ and $t \in G$ we have $F(Kt) = F(K)$.

With the convention (6.1) and using the triangle inequality, we obtain $\mathcal{D}(\varepsilon, \delta) \leq \mathcal{D}_1(\varepsilon, \delta) + \mathcal{D}_2(\varepsilon, \delta)$, where

$$\mathcal{D}_1(\varepsilon, \delta) := \left\| \sum_{j=1}^{N(\varepsilon)} \eta_j(\varepsilon) \frac{F(K_j^r(\varepsilon)) - \sum_{i=1}^{N(\delta)} |T_i^j(\delta)| F(K_i^r(\delta))}{|K_j(\varepsilon)|} \right\|, \quad \text{and}$$

$$\mathcal{D}_2(\varepsilon, \delta) := \left\| \sum_{j=1}^{N(\varepsilon)} \eta_j(\varepsilon) \frac{\sum_{i=1}^{N(\delta)} |T_i^j(\delta)| F(K_i^r(\delta))}{|K_j(\varepsilon)|} - \sum_{i=1}^{N(\delta)} \eta_i(\delta) \frac{F(K_i^r(\delta))}{|K_i(\delta)|} \right\|.$$

The translation invariance of F and the triangle inequality yield

$$\mathcal{D}_1(\varepsilon,\delta) \le \sum_{j=1}^{N(\varepsilon)} \frac{\eta_j(\varepsilon)}{|K_j(\varepsilon)|} \left\| F\big(K_j^r(\varepsilon)\big) - \sum_{i=1}^{N(\delta)} \sum_{t\in T_i^j(\delta)} F\big(K_i^r(\delta)t\big) \right\|. \tag{6.2}$$

We decompose $K_j^r(\varepsilon)$ in the following way:

$$\begin{aligned}
K_j^r(\varepsilon) = {} & \bigcup_{i=1}^{N(\delta)} \bigcup_{t\in T_i^j(\delta)} K_i^r(\delta)t \,\dot\cup\, K_j^r(\varepsilon) \setminus \bigcup_{i=1}^{N(\delta)} K_i(\delta)T_i^j(\delta) \\
& \dot\cup \left(\left(K_j^r(\varepsilon) \Big\backslash \bigcup_{i=1}^{N(\delta)} K_i^r(\delta)T_i^j(\delta) \right) \cap \bigcup_{i=1}^{N(\delta)} \big(K_i(\delta) \cap \partial^r\big(K_i(\delta)\big)\big)T_i^j(\delta) \right) \\
=: {} & \alpha_1 \,\dot\cup\, \alpha_2 \,\dot\cup\, \alpha_3.
\end{aligned}$$

By definition of the function F, the almost additivity of the admissible field f inherits to F. Note that δ-disjointness of the sets $K_i t$, $t \in T_i^j(\delta)$ implies δ-disjointness of the sets $K_i^r t$, $t \in T_i^j(\delta)$. Therefore, applying almost additivity, Lemma 3.5 and the properties of admissible fields and the boundary term we obtain

$$\begin{aligned}
& \left\| F\big(K_j^r(\varepsilon)\big) - \sum_{i=1}^{N(\delta)} \sum_{t\in T_i^j(\delta)} F\big(K_i^r(\delta)t\big) \right\| \\
& \quad \le \left\| F\big(K_j^r(\varepsilon)\big) - \sum_{i=1}^{3} F(\alpha_i) \right\| + \left\| F(\alpha_1) - \sum_{i=1}^{N(\delta)} \sum_{t\in T_i^j(\delta)} F\big(K_i^r(\delta)\big) \right\| \\
& \qquad + \|F(\alpha_2)\| + \|F(\alpha_3)\| \\
& \quad \le \sum_{i=1}^{3} b(\alpha_i) + \delta(3K_f + 9D_f)\big|K_j(\varepsilon)\big| \\
& \qquad + 3\sum_{i=1}^{N(\delta)} \sum_{t\in T_i^j} b\big(K_i^r(\delta)\big) + K_f|\alpha_2| + K_f|\alpha_3| \\
& \quad \le \delta(3K_f + 9D_f)\big|K_j(\varepsilon)\big| \\
& \qquad + 4\sum_{i=1}^{N(\delta)} \sum_{t\in T_i^j(\delta)} b\big(K_i^r(\delta)\big) + (K_f + D_f)|\alpha_2| + (K_f + D_f)|\alpha_3|.
\end{aligned}$$

Next, we estimate the sizes of α_2 and α_3. For α_3, we drop some of the intersections in its definition. In order to give a bound on the size of α_2, we use that $K_j^r(\varepsilon)$ is

$(1-2\varepsilon)$-covered by $\{K_i^r(\delta) \mid i\}$, more specifically, part (iii) in Definition 3.1. We obtain

$$|\alpha_2| \le 2\delta\left|K_j(\varepsilon)\right| \quad \text{and} \quad |\alpha_3| \le \sum_{i=1}^{N(\delta)} \left|T_i^j(\delta)\right|\left|\partial^r\left(K_i(\delta)\right)\right|,$$

and therewith achieve

$$\begin{aligned}
&\left\| F\left(K_j^r(\varepsilon)\right) - \sum_{i=1}^{N(\delta)} \sum_{t \in T_i^j(\delta)} F\left(K_i^r(\delta)t\right) \right\| \\
&\qquad \le \delta(5K_f + 11D_f)\left|K_j(\varepsilon)\right| \\
&\qquad\quad + \sum_{i=1}^{N(\delta)} \left|T_i^j(\delta)\right|\left(4b\left(K_i^r(\delta)\right) + (K_f + D_f)\left|\partial^r\left(K_i(\delta)\right)\right|\right).
\end{aligned}$$

This together with (6.2) and part (a) of Lemma 3.3 yields

$$\begin{aligned}
&\mathcal{D}_1(\varepsilon,\delta) \\
&\quad \le \sum_{j=1}^{N(\varepsilon)} \Bigg(\delta(5K_f + 11D_f)\eta_j(\varepsilon) \\
&\qquad + \sum_{i=1}^{N(\delta)} \frac{\eta_j(\varepsilon)|T_i^j(\delta)|}{|K_j(\varepsilon)|}\left(4b\left(K_i^r(\delta)\right) + (K_f + D_f)\left|\partial^r\left(K_i(\delta)\right)\right|\right)\Bigg) \\
&\quad \le \delta(5K_f + 11D_f) \\
&\qquad + \sum_{j=1}^{N(\varepsilon)} \sum_{i=1}^{N(\delta)} \frac{\eta_j(\varepsilon)|T_i^j(\delta)|}{|K_j(\varepsilon)|}\left(4b\left(K_i^r(\delta)\right) + (K_f + D_f)\left|\partial^r\left(K_i(\delta)\right)\right|\right).
\end{aligned}$$

As δ is assumed to be smaller than $1/10$, we can apply Corollary 3.4, which gives for arbitrary $i \in \{1,\dots,N(\delta)\}$ and $j \in \{1,\dots,N(\varepsilon)\}$

$$\frac{|T_i^j(\delta)|}{|K_j(\varepsilon)|} \le \frac{\eta_i(\delta)}{|K_i(\delta)|} + 4\frac{\delta\eta_i(\delta)}{|K_i(\delta)|} \le 5\frac{\eta_i(\delta)}{|K_i(\delta)|}.$$

Inserting this in the last estimate for $\mathcal{D}_1(\varepsilon,\delta)$ implies together with part (a) of Lemma 3.3 that

$$\begin{aligned}
\mathcal{D}_1(\varepsilon,\delta) &\le \delta(5K_f + 11D_f) \\
&\quad + \sum_{i=1}^{N(\delta)} \frac{5\eta_i(\delta)}{|K_i(\delta)|}\left(4b\left(K_i^r(\delta)\right) + (K_f + D_f)\left|\partial^r\left(K_i(\delta)\right)\right|\right).
\end{aligned}$$

Now, we use the monotonicity assumption in (4.1), which allows to replace the elements $K_i^r(\delta)$ and $K_i(\delta)$ by Q_i^r and Q_i, respectively,

$$\begin{aligned}\mathcal{D}_1(\varepsilon,\delta) &\le \delta(5K_f+11D_f)\\ &\quad + \sum_{i=1}^{N(\delta)} \frac{5\eta_i(\delta)}{|Q_i|}\bigl(4b(Q_i^r)+(K_f+D_f)\bigl|\partial^r(Q_i)\bigr|\bigr).\end{aligned} \tag{6.3}$$

Let us proceed with the estimation of $\mathcal{D}_2(\varepsilon,\delta)$:

$$\mathcal{D}_2(\varepsilon,\delta) = \left\| \sum_{i=1}^{N(\delta)} F\bigl(K_i^r(\delta)\bigr)\left(\sum_{j=1}^{N(\varepsilon)} \eta_j(\varepsilon)\frac{|T_i^j(\delta)|}{|K_j(\varepsilon)|} - \frac{\eta_i(\delta)}{|K_i(\delta)|}\right)\right\|. \tag{6.4}$$

With the triangle inequality, Corollary 3.4 and part (a) of Lemma 3.3 we obtain

$$\begin{aligned}&\left|\sum_{j=1}^{N(\varepsilon)} \eta_j(\varepsilon)\frac{|T_i^j(\delta)|}{|K_j(\varepsilon)|} - \frac{\eta_i(\delta)}{|K_i(\delta)|}\right|\\ &\quad\le \sum_{j=1}^{N(\varepsilon)} \eta_j(\varepsilon)\left|\frac{|T_i^j(\delta)|}{|K_j(\varepsilon)|} - \frac{\eta_i(\delta)}{|K_i(\delta)|}\right| + \left|\sum_{j=1}^{N(\varepsilon)} \eta_j(\varepsilon) - 1\right|\frac{\eta_i(\delta)}{|K_i(\delta)|}\\ &\quad\le \sum_{j=1}^{N(\varepsilon)} \eta_j(\varepsilon)\frac{4\delta\eta_i(\delta)}{|K_i(\delta)|} + \frac{\varepsilon\eta_i(\delta)}{|K_i(\delta)|} \le \frac{4\delta\eta_i(\delta)}{|K_i(\delta)|} + \frac{\varepsilon\eta_i(\delta)}{|K_i(\delta)|}.\end{aligned}$$

This together with (6.4) gives the bound

$$\mathcal{D}_2(\varepsilon,\delta) \le \sum_{i=1}^{N(\delta)} K_f\bigl|K_i^r(\delta)\bigr|\left(\frac{4\delta\eta_i(\delta)}{|K_i(\delta)|} + \frac{\varepsilon\eta_i(\delta)}{|K_i(\delta)|}\right) \le 4K_f\delta + K_f\varepsilon. \tag{6.5}$$

Thus, the estimates of $\mathcal{D}_1(\varepsilon,\delta)$ and $\mathcal{D}_2(\varepsilon,\delta)$ in (6.3) and (6.5) together yield

$$\begin{aligned}\mathcal{D}(\varepsilon,\delta) &\le K_f\varepsilon + \delta(9K_f+11D_f)\\ &\quad + \sum_{i=1}^{N(\delta)} \frac{5\eta_i(\delta)}{|Q_i|}\bigl(4b(Q_i^r)+(K_f+D_f)\bigl|\partial^r(Q_i)\bigr|\bigr)\end{aligned} \tag{6.6}$$

for all $\delta > 0$ and $\varepsilon \in (0, \varepsilon_0(\delta))$. Applying part (c) of Lemma 3.3, we see

$$\lim_{\delta\searrow 0}\lim_{\varepsilon\searrow 0} \mathcal{D}(\varepsilon,\delta) = 0.$$

Using a Cauchy argument and the fact that $\mathbb{B}$ is a Banach space, we obtain that there exists an element $f^* \in \mathbb{B}$ with

$$\lim_{\varepsilon\searrow 0}\left\| \sum_{j=1}^{N(\varepsilon)} \eta_j(\varepsilon)\frac{\langle f_j^r(\varepsilon), \mathbb{P}_j^r(\varepsilon)\rangle}{|K_j(\varepsilon)|} - f^*\right\| = 0.$$

In order to get the error estimate for finite $\delta > 0$, we use (6.6), Lemma 3.3(c) and (4.1) as follows:

$$\begin{aligned}
&\left\| \sum_{j=1}^{N(\delta)} \eta_j(\delta) \frac{\langle f_j^r(\delta), \mathbb{P}_j^r(\delta)\rangle}{|K_j(\delta)|} - f^* \right\| \\
&\qquad = \lim_{\varepsilon \searrow 0} \mathcal{D}(\varepsilon, \delta) \\
&\qquad \le (9K_f + 11D_f)\delta + \sum_{i=1}^{N(\delta)} \frac{5\eta_i(\delta)}{|Q_i|} \bigl(4b(Q_i^r) + (K_f + D_f)|\partial^r(Q_i)|\bigr) \\
&\qquad \le (9K_f + 11D_f)\delta + 5(4 + K_f + D_f)\beta(\delta). \qquad \square
\end{aligned}$$

7. Proof of the main theorem. We will prove a slightly more explicit statement which tracks the geometric error in terms of ε and the probabilistic error in terms of κ separately. Theorem 2.5 is implied by the choice $\kappa := \sqrt{\varepsilon}$. Recall that $\mathbb{B}$ is the Banach space of bounded and right-continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

THEOREM 7.1. *Let G be a finitely generated amenable group. Further, let $\mathcal{A} \in \mathcal{B}(\mathbb{R})$ and $(\Omega = \mathcal{A}^G, \mathcal{B}(\Omega), \mathbb{P})$ a probability space such that $\mathbb{P}$ satisfies (M1) to (M3). Finally, let $\mathcal{U}$ be an admissible set of admissible fields with common bound $K_{\mathcal{U}}$; cf. Definition 2.2.*

Then there exists a limit element $f^* \in \mathbb{B}$ with the following properties. For each Følner sequence (Λ_n), $\varepsilon \in (0, 1/10)$ and $\kappa > 0$, there exist $j_0(\varepsilon) \in \mathbb{N}$, which is independent of κ and $K_{\mathcal{U}}$, and $a(\varepsilon, \kappa, K_{\mathcal{U}}), b(\varepsilon, \kappa, K_{\mathcal{U}}) > 0$, such that for all $j \in \mathbb{N}$, $j \ge j_0(\varepsilon)$, there is an event $\Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}} \in \mathcal{B}(\Omega)$ with the properties

$$\mathbb{P}(\Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}}) \ge 1 - b(\varepsilon, \kappa, K_{\mathcal{U}}) \exp\bigl(-a(\varepsilon, \kappa, K_{\mathcal{U}})|\Lambda_j|\bigr)$$

and

$$\begin{aligned}
&\left\| \frac{f(\Lambda_j, \omega)}{|\Lambda_j|} - f^* \right\| \\
&\qquad \le (37K_f + 47D_f + 46)\sqrt{\varepsilon} + \kappa \qquad \text{for all } \omega \in \Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}} \text{ and all } f \in \mathcal{U}.
\end{aligned}$$

PROOF. We follow the path prescribed in the previous chapters and:

- quasi tile Λ_j, $j \ge j_0(\varepsilon)$, with $K_i(\varepsilon)$, $i = 1, \dots, N(\varepsilon)$ (see Theorem 3.2),
- approximate $|\Lambda_j|^{-1} f(\Lambda_j, \omega)$ with the empirical measures $L_{i,j}^{r,\omega}(\varepsilon)$; cf. (4.8) and Lemma 4.3,
- express the empirical measures by their limiting counterparts $\mathbb{P}_i^r(\varepsilon)$ with Lemma 5.4, and
- use the Cauchy property of the remaining terms to obtain a limiting function f^*; see Lemma 6.1.

To confirm the error estimate, we employ the triangle inequality

$$\begin{aligned}
&\left\|\frac{f(\Lambda_j,\omega)}{|\Lambda_j|}-f^*\right\| \\
&\quad\le \left\|\frac{f(\Lambda_j,\omega)}{|\Lambda_j|}-\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\langle f_i^r(\varepsilon),L_{i,j}^{r,\omega}(\varepsilon)\rangle}{|K_i(\varepsilon)|}\right\| \\
&\qquad+\left\|\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\langle f_i^r(\varepsilon),L_{i,j}^{r,\omega}(\varepsilon)\rangle}{|K_i(\varepsilon)|}-\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\langle f_i^r(\varepsilon),\mathbb{P}_i^r(\varepsilon)\rangle}{|K_i(\varepsilon)|}\right\| \\
&\qquad+\left\|\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\langle f_i^r(\varepsilon),\mathbb{P}_i^r(\varepsilon)\rangle}{|K_i(\varepsilon)|}-f^*\right\| =: \Delta(\varepsilon,j,\omega).
\end{aligned}$$

By Lemmas 6.1 and 4.3 and Proposition 5.4, we immediately get that there is an event $\tilde{\Omega}\in\mathcal{B}(\Omega)$ with full probability $\mathbb{P}(\tilde{\Omega})=1$ such that $\lim_{\varepsilon\searrow 0}\lim_{j\to\infty}\Delta(\varepsilon, j,\omega)=0$ for all $\omega\in\tilde{\Omega}$. Furthermore, Lemma 5.4 provides the event $\Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}}$ with probability as large as claimed, and by collecting all the error terms and by Remark 4.1, we see that for all $\varepsilon\in(0,1/10)$, $j\ge j_0(\varepsilon)$, $\kappa>0$, $f\in\mathcal{U}$ and $\omega\in\Omega_{j,\varepsilon,\kappa,K_{\mathcal{U}}}$ (see Lemma 5.4),

$$\begin{aligned}
\left\|\frac{f(\Lambda_j,\omega)}{|\Lambda_j|}-f^*\right\| &\le (20K_f+30D_f)\varepsilon+(17K_f+17D_f+46)\beta(\varepsilon)+\kappa \\
&\le (37K_f+47D_f+46)\sqrt{\varepsilon}+\kappa.
\end{aligned}$$

Note the uniformity of the last inequality for all $f\in\mathcal{U}$ is also discussed in Remark 4.1.

To see that the limit f^* does not depend on the specific choice of (Λ_j) use the following argument: Every two Følner sequences can be combined two one Følner sequence, which yields by our theory a limit $f^*\in\mathbb{B}$. As the two original sequences are subsequences, they lead to the same limit function f^*. □

APPENDIX A: CONDITIONAL RESAMPLING

In Lemma 5.2, we need to remove the dependent parts of samples. We achieve this by resampling the critical parts of the samples, keeping the large enough already independent parts. This is done by augmenting the probability space to provide room for more random variables. The problem of resampling turned out to be treatable in a much broader setting, so a general tool is provided here.

THEOREM A.1 (Resampling). *Let $(\Omega,\mathcal{A},\mathbb{P})$ be a Borel probability space, $(S,\mathcal{S})$ a Borel space and $X\colon\Omega\to S$ an S-valued random variable with distribution $\mathbb{P}_X:=\mathbb{P}\circ X^{-1}\colon\mathcal{S}\to[0,1]$. Further, let I be an index set, and for each $j\in I$, let $\mathcal{Y}_j\subseteq\mathcal{S}$ be a σ-algebra.*

Then there is a probability space $(\underline{\Omega}, \underline{\mathcal{A}}, \underline{\mathbb{P}})$, *such that for all* $j \in I$, *maps as indicated in the following diagram exist and are measure preserving*, *and all the diagrams commute almost surely*:

$$\begin{tikzcd}
(\Omega,\mathcal{A},\mathbb{P}) \arrow[d, "X"'] & (\underline{\Omega},\underline{\mathcal{A}},\underline{\mathbb{P}}) \arrow[l, "\Pi_0"'] \arrow[dl, "X_0"] \arrow[dr, "X_j"] & \\
(S,\mathcal{S},\mathbb{P}_X) \arrow[r, "\mathrm{id}_S"'] & (S,\mathcal{Y}_j,\mathbb{P}_X|_{\mathcal{Y}_j}) & (S,\mathcal{S},\mathbb{P}_X) \arrow[l, "\mathrm{id}_S"]
\end{tikzcd}$$

This means in particular that Π_0 *is measure preserving*, *and that*, *for all* $j \in I$:

(i) *the random variable* X_j *has distribution* $\mathbb{P}_X$,

(ii) *for each measure space* $(T, \mathcal{T})$ *and each* $\mathcal{Y}_j$-$\mathcal{T}$*-measurable map* $g: (S, \mathcal{Y}_j) \to (T, \mathcal{T})$, *we have* $g(X_0) = g(X_j)$ $\underline{\mathbb{P}}$*-almost surely.*

Furthermore, *the joint distribution of* $(X_j)_{j\in I}$ *has the following properties*:

(iii) *For each finite subset* $F \subseteq I$ *and* $A_F = \times_{j\in F} A_j$, *where* $A_j \in \mathcal{S}$, *we have* $\mathbb{P}_X$*-almost surely that*

$$\underline{\mathbb{P}}(X_F \in A_F \mid X_0 = \cdot\,) = \prod_{j\in F} \underline{\mathbb{P}}(X_j \in A_j \mid X_0 = \cdot\,) = \prod_{j\in F} \mathbb{P}_X(A_j \mid \mathcal{Y}_j).$$

In particular, *the random variables* X_j, $j \in I$, *are independent when conditioned on* X_0.

(iv) *If*, *for a* (*not necessarily finite*) *subset* $J \subseteq I$, *the* σ*-algebras* $\mathcal{Y}_j$, $j \in J$, *are* $\mathbb{P}_X$*-independent*, *then the random variables* X_j, $j \in J$, *are* $\underline{\mathbb{P}}$*-independent.*

Since Π_0 is measure preserving, $(\underline{\Omega}, \underline{\mathcal{A}}, \underline{\mathbb{P}})$ extends $(\Omega, \mathcal{A}, \mathbb{P})$. Property (i) justifies the name resampling. Statement (ii) says that in X_j the information contained in $\mathcal{Y}_j$ is preserved throughout the resampling, $j \in I$. Point (iii) states that the new random variables copied only the information from $\mathcal{Y}_j$, $j \in I$, and not more. In (iv), we learn how to provide independence of the resampling random variables.

PROOF. We define the spaces and maps as follows:

$$\begin{aligned}
\underline{\Omega} &:= \Omega \times S^I, & \underline{\mathcal{A}} &:= \mathcal{A} \otimes \mathcal{S}^{\otimes I},\\
\Pi_0 &: \underline{\Omega} \to \Omega, & \Pi_0\bigl(\omega, (s_j)_{j\in I}\bigr) &:= \omega,\\
X_0 &: \underline{\Omega} \to S, & X_0\bigl(\omega, (s_j)_{j\in I}\bigr) &:= X(\omega),\\
X_j &: \underline{\Omega} \to S, & X_j\bigl(\omega, (s_k)_{k\in I}\bigr) &:= s_j.
\end{aligned}$$

We now define the measure $\underline{\mathbb{P}}$ via Kolmogorov's extension theorem; see [4], Theorem 14.36. We need a consistent family of probability measures. For a more unifying notation, we augment $I_0 := \{0\} \dot\cup I$. Fix a finite subset $F \subseteq I_0$. If $0 \in F$, we define a probability measure $\mathbb{P}^F : \mathcal{A} \otimes \mathcal{S}^{\otimes F \setminus \{0\}} \to [0,1]$. In the case $0 \notin F$, we define a probability measure $\mathbb{P}^F : \mathcal{S}^{\otimes F} \to [0,1]$. If $0 \in F$, then choose $A_0 \in \mathcal{A}$; otherwise, let $A_0 := \Omega$. For all $j \in F \setminus \{0\}$, we let $A_j \in \mathcal{S}$. Now let $A_F := \times_{j \in F} A_j$ and

$$\mathbb{P}^F(A_F) := \mathbb{E}\Big[\mathbf{1}_{A_0} \prod_{j \in F \setminus \{0\}} \mathbb{P}_X(A_j \mid \mathcal{Y}_j) \circ X\Big]. \tag{A.1}$$

Here, $\mathbb{E}$ denotes integration with respect to $\mathbb{P}$. By the extension theorem for measures (see [4], Theorem 1.53), (A.1) defines a probability measure. The family $(\mathbb{P}^F)_{F \subseteq I \text{ finite}}$ is consistent. For example, for finite subsets $0 \notin F \subseteq J \subseteq I$ with the projection $\Pi^J_F : S^J \to S^F$ and $A_F = \times_{j \in F} A_j$ with $A_j \in \mathcal{S}$, we have $(\Pi^J_F)^{-1}(A_F) = A_F \times \times_{j \in J \setminus F} S$. Thus,

$$\mathbb{P}^J\big((\Pi^J_F)^{-1}(A_F)\big) = \mathbb{E}_X\Big[\prod_{j \in F} \mathbb{P}_X(A_j \mid \mathcal{Y}_j) \prod_{j \in J \setminus F} \mathbb{P}_X(S \mid \mathcal{Y}_j)\Big] = \mathbb{P}^F(A_F),$$

where $\mathbb{E}_X$ is integration with respect to $\mathbb{P}_X$. The remaining cases $0 \in F \subseteq J$, and $0 \notin F$ but $0 \in J$ work analogously. By Kolmogorov's extension theorem, we have exactly one measure $\underline{\mathbb{P}} := \varprojlim_{F \subseteq I} \mathbb{P}^F : \underline{\mathcal{A}} \to [0,1]$.

We now verify the properties of $\underline{\mathbb{P}}$. Let us first check that Π_0 is measure preserving. Indeed, for $A \in \mathcal{A}$, we have

$$\underline{\mathbb{P}}(\Pi_0 \in A) = \mathbb{P}^{\{0\}}(A) = \mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A).$$

Now we already know that $X_0 = X \circ \Pi_0$ is measure preserving, also.

Ad (i): For all $j \in I$ and $B \in \mathcal{S}$, we have

$$\underline{\mathbb{P}}(X_j \in B) = \mathbb{P}^{\{j\}}(B) = \mathbb{E}_X\big[\mathbb{P}_X(B \mid \mathcal{Y}_j)\big] = \mathbb{E}_X[\mathbf{1}_B] = \mathbb{P}_X(B).$$

Ad (ii): Let $j \in I$, $(T, \mathcal{T})$ be a measure space and $g : S \to T$ be $\mathcal{Y}_j$-$\mathcal{T}$-measurable. We determine the joint distribution of X and X_j. By (A.1), we have, for $B, B' \in \mathcal{T}$, that $A := g^{-1}(B) \in \mathcal{Y}_j$ as well as $A' := g^{-1}(B') \in \mathcal{Y}_j$, and

$$\begin{aligned}
\underline{\mathbb{P}}\big(g(X_0) \in B, g(X_j) \in B'\big) &= \underline{\mathbb{P}}\big(X_0 \in A, X_j \in A'\big) \\
&= \mathbb{P}^{\{0,j\}}\big(X^{-1}(A) \times A'\big) \\
&= \mathbb{E}\big[\mathbf{1}_{X^{-1}(A)} \mathbb{P}_X\big(A' \mid \mathcal{Y}_j\big) \circ X\big] \\
&= \mathbb{E}_X[\mathbf{1}_A \mathbf{1}_{A'}] \\
&= \mathbb{P}_X\big(A \cap A'\big) \\
&= \underline{\mathbb{P}}\big(X_0 \in A \cap A'\big) \\
&= \underline{\mathbb{P}}\big(g(X_0) \in B \cap B'\big),
\end{aligned} \tag{A.2}$$

where in the last line, we used that $A \cap A' = g^{-1}(B) \cap g^{-1}(B') = g^{-1}(B \cap B')$. Now, since the rectangles $\{B \times B' \mid B, B' \in \mathcal{T}\}$ are stable under intersections and generate $\mathcal{T} \otimes \mathcal{T}$, equation (A.2) determines the distribution of $(g(X_0), g(X_j))\colon \underline{\Omega} \to T^2$. Note that the measure which is concentrated on the diagonal $\{(t,t) \mid t \in T\}$ with both marginals equal to $\mathbb{P}_X \circ g^{-1}$ satisfies (A.2), also. Therefore, $\underline{\mathbb{P}}(g(X_0) = g(X_j)) = 1$.

Ad (iii): Fix a finite subset $F \subseteq I$ and $A_j \in \mathcal{S}$ for $j \in F$, and let $A_F := \bigtimes_{j \in F} A_j$. For all $B \in \mathcal{S}$, we have

$$\begin{aligned}
\underline{\mathbb{E}}[\mathbf{1}_{\{X_0 \in B\}} \underline{\mathbb{P}}(X_F \in A_F \mid X_0)] &= \underline{\mathbb{E}}[\mathbf{1}_{\{X_0 \in B\}} \underline{\mathbb{E}}[\mathbf{1}_{\{X_F \in A_F\}} \mid X_0]] \\
&= \underline{\mathbb{E}}[\mathbf{1}_{\{X_0 \in B\}} \mathbf{1}_{\{X_F \in A_F\}}] \\
&= \underline{\mathbb{P}}[X_0 \in B, X_F \in A_F] \\
&= \mathbb{P}^{\{0\} \cup F}\big(X^{-1}(B) \times A_F\big) \\
&= \mathbb{E}\bigg[\mathbf{1}_{X^{-1}(B)} \prod_{j \in F} \mathbb{P}_X(A_F \mid \mathcal{Y}_j) \circ X\bigg] \\
&= \underline{\mathbb{E}}\bigg[\mathbf{1}_{\{X_0 \in B\}} \prod_{j \in F} \mathbb{P}_X(A_F \mid \mathcal{Y}_j) \circ X_0\bigg].
\end{aligned}$$

Since $\sigma(X_0) = \{\{X_0 \in B\} \mid B \in \mathcal{S}\}$, this proves

$$\underline{\mathbb{P}}(X_F \in A_F \mid X_0) = \prod_{j \in F} \mathbb{P}_X(X_j \in A_j \mid \mathcal{Y}_j) \circ X_0$$

$\underline{\mathbb{P}}$-almost surely. For $F = \{j\}$, we get $\underline{\mathbb{P}}(X_j \in A_j \mid X_0) = \mathbb{P}_X(X_j \in A_j \mid \mathcal{Y}_j)$, also. The claim is the factorized version of these statements, which exist because $(S, \mathcal{S})$ is a Borel space.

Ad (iv): For $F \subseteq J$ finite and $A_F = \bigtimes_{j \in F} A_j$ with $A_j \in \mathcal{S}$, we use (iii) to get

$$\begin{aligned}
\underline{\mathbb{P}}(X_F \in A_F) &= \underline{\mathbb{E}}[\underline{\mathbb{P}}(X_F \in A_F \mid X_0)] \\
&= \underline{\mathbb{E}}\bigg[\prod_{j \in F} \mathbb{P}_X(A_j \mid \mathcal{Y}_j) \circ X_0\bigg] \\
&= \mathbb{E}_X\bigg[\prod_{j \in F} \mathbb{P}_X(A_j \mid \mathcal{Y}_j)\bigg].
\end{aligned}$$

The σ-algebras $\mathcal{Y}_j$, $j \in F \subseteq J$, are $\mathbb{P}_X$-independent. This independence is inherited by $\mathcal{Y}_j$-measurable functions like $\mathbb{P}_X(A_j \mid \mathcal{Y}_j)$. We can therefore continue the calculation with

$$\underline{\mathbb{P}}(X_F \in A_F) = \prod_{j \in F} \mathbb{E}_X[\mathbb{P}_X(A_j \mid \mathcal{Y}_j)] = \prod_{j \in F} \mathbb{P}_X(A_j) = \prod_{j \in F} \underline{\mathbb{P}}(X_j \in A_j).$$

Since the cylinder sets generate $\mathcal{S}^{\otimes J}$, this is the claimed $\underline{\mathbb{P}}$-independence. $\square$

APPENDIX B: PROOF SUMMARY FOR MONTILABLE AMENABLE GROUPS

The proofs of [12] concerning the case $G = \mathbb{Z}^d$ can be generalized to apply to a finitely generated amenable group G if it satisfies the tiling property (⊞).

We list the major changes which are necessary for this purpose:

(a) Instead of defining the set $T_{m,n}$ using multiples of m (cf. equation (4.1) in [12]), we employ the grid T_m, namely, we set

$$T_{m,n} := \{t \in T_m \mid \Lambda_m t \subseteq \Lambda_n\} \tag{B.1}$$

Thus, $T_{m,n}$ contains the elements of T_m which correspond to translates of Λ_m which are completely contained in Λ_n. Using this definition, the empirical measures are $L^{\omega}_{m,n}$ and $L^{\omega,r}_{m,n}$ are given accordingly.

(b) One needs to verify the following basic result. Given a tiling Følner sequence (Λ_n), we have:

(i) for each $m \in \mathbb{N}$, the sequence $(\Lambda_m T_{m,n})_{n\in\mathbb{N}}$ is a Følner sequence;

(ii) for each $m, n \in \mathbb{N}$, we have $\Lambda_n \subseteq \partial^{\rho(m)}(\Lambda_n) \cup \Lambda_m T_{m,n}$, where $\rho(m) = \operatorname{diam}(\Lambda_m)$; and

(iii) for each $m \in \mathbb{N}$ we have $\lim_{n\to\infty} |\Lambda_n|/|T_{m,n}| = |\Lambda_m|$.

(c) Points (a) and (b) allow to prove an equivalent version of Lemma 4.1 of [12] in the situation of amenable groups with property (⊞), by following exactly the steps of the proof presented therein.

(d) Besides Lemma 4.1 in [12], also Lemma 6.1 in [12] needs to be slightly changed. In fact, again by using (a) and (b) the proof can directly be adapted to the situation where G is amenable and (Λ_n) is a tiling Følner sequence.

(e) In the end, the proof of the main theorem reduces basically to an application of the triangle inequality, the new versions of Lemma 4.1 and Lemma 6.1 as well as Theorem 5.6 in [12]. Note that Theorem 5.6 need not to be adapted as it is independent of the geometry.

REFERENCES

[1] ADACHI, T. (1993). A note on the Følner condition for amenability. *Nagoya Math. J.* **131** 67–74. MR1238633

[2] ALIPRANTIS, C. D. and BORDER, K. C. (2006). *Infinite Dimensional Analysis*: *A Hitchhiker's Guide*, 3rd ed. Springer, Berlin. MR2378491

[3] DEHARDT, J. (1971). Generalizations of the Glivenko–Cantelli theorem. *Ann. Math. Stat.* **42** 2050–2055.

[4] KLENKE, A. (2008). *Probability Theory. Number* 223 *in Universitext*, 1st ed. Springer, London.

[5] KRIEGER, F. (2009). Le lemme d'Ornstein–Weiss d'aprés Gromov. *Dyn.*, *Ergod. Theory Geom.* **54** 99–112.

[6] LENZ, D., MÜLLER, P. and VESELIĆ, I. (2008). Uniform existence of the integrated density of states for models on $\mathbb{Z}^d$. *Positivity* **12** 571–589. MR2448749

[7] LENZ, D., SCHWARZENBERGER, F. and VESELIĆ, I. (2010). A Banach space-valued ergodic theorem and the uniform approximation of the integrated density of states. *Geom. Dedicata* **150** 1–34.

[8] LENZ, D. and STOLLMANN, P. (2005). An ergodic theorem for Delone dynamical systems and existence of the integrated density of states. *J. Anal. Math.* **97** 1–24. MR2274971

[9] LENZ, D. and VESELIĆ, I. (2009). Hamiltonians on discrete structures: Jumps of the integrated density of states and uniform convergence. *Math. Z.* **263** 813–835.

[10] ORNSTEIN, D. and WEISS, B. (1987). Entropy and isomorphism theorems for actions of amenable groups. *J. Anal. Math.* **48** 1–141.

[11] POGORZELSKI, F. and SCHWARZENBERGER, F. (2016). A Banach space-valued ergodic theorem for amenable groups and applications. *J. Anal. Math.* **130** 19–69.

[12] SCHUMACHER, C., SCHWARZENBERGER, F. and VESELIĆ, I. (2017). A Glivenko–Cantelli theorem for almost additive functions on lattices. *Stochastic Process. Appl.* **127** 179–208.

[13] SCHWARZENBERGER, F. (2013). The integrated density of states for operators on groups. Ph.D. thesis, Technische Universität Chemnitz. Available at http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-123241.

[14] TAO, T. (2015). Failure of the L^1 pointwise and maximal ergodic theorems for the free group. *Forum Math. Sigma* **3** 19.

[15] VESELIĆ, I. (2005). Spectral analysis of percolation Hamiltonians. *Math. Ann.* **331** 841–865.

[16] WEISS, B. (2001). Monotileable amenable groups. In *Topology, Ergodic Theory, Real Algebraic Geometry. Amer. Math. Soc. Transl. Ser.* 2 **202** 257–262. Amer. Math. Soc., Providence, RI. MR1819193

[17] WRIGHT, F. T. (1981). The empirical discrepancy over lower layers and a related law of large numbers. *Ann. Probab.* **9** 323–329. MR0606996

C. SCHUMACHER
I. VESELIĆ
FAKULTÄT FÜR MATHEMATIK
TU DORTMUND
44221 DORTMUND
GERMANY
E-MAIL: christoph.schumacher@math.tu-dortmund.de
ivan.veselic@math.tu-dortmund.de

F. SCHWARZENBERGER
FAKULTÄT FÜR INFORMATIK/MATHEMATIK
HTW DRESDEN
01069 DRESDEN
GERMANY
E-MAIL: fabian.schwarzenberger@htw-dresden.de

Chapter 6

Approximation of the Integrated Density of States on Sofic Groups

Ann. Henri Poincaré 16 (2015), 1067–1101

1424-0637/15/041067-35
published online June 26, 2014
DOI 10.1007/s00023-014-0342-4

Annales Henri Poincaré

Approximation of the Integrated Density of States on Sofic Groups

Christoph Schumacher and Fabian Schwarzenberger

Abstract. In this paper, we study spectral properties of self-adjoint operators on a large class of geometries given via sofic groups. We prove that the associated integrated densities of states can be approximated via finite volume analogues. This is investigated in the deterministic as well as in the random setting. In both cases, we cover a wide range of operators including in particular unbounded ones. The large generality of our setting allows one to treat applications from long-range percolation and the Anderson model. Our results apply to operators on $\mathbb{Z}^d$, amenable groups, residually finite groups and therefore in particular to operators on trees. All convergence results are established without an ergodic theorem at hand.

1. Introduction

The study of self-adjoint operators on discrete structures has a long history in mathematical physics, both in the deterministic as well as in the random case. The investigation of spectral properties of such operators is motivated as essential features of solutions of differential equations are encoded in the spectrum of the corresponding operator. However, it is in many cases hard to obtain results on the spectrum by directly studying the operator. One way to overcome this difficulty is to study the *integrated density of states* (IDS) as a rather simple object which still carries much information about the spectrum of the operator.

In order to define the IDS, one chooses a sequence of finite-dimensional self-adjoint operators approximating the original operator in a suitable sense and considers their eigenvalue counting functions. For each $\lambda \in \mathbb{R}$, this function returns the number of eigenvalues of the approximating operator (counting multiplicity) not larger than λ. The IDS is then defined as the pointwise limit of the normalized eigenvalue counting functions, if the limit exists. In this situation, it is in many cases possible to show that the IDS equals the so-called *spectral distribution function* (SDF), given via a trace of certain projections,

Birkhäuser

see (2.3). This equality is called the *Pastur–Shubin trace formula*. Depending on the context, the SDF is sometimes called von Neumann trace, see for instance [23], and in other situations, the associated measure is known as the Plancherel measure or Kesten spectral measure, see e.g. [9].

Two questions arise:

(a) Does the limit of the eigenvalue counting functions exist?
(b) Does the Pastur–Shubin trace formula hold?

The investigation of these questions has a long history. In the seminal papers, [33, 42], the existence of the limit was first rigorously studied. The authors studied random ergodic and almost periodic operators in Euclidean space. To the present day, many results were obtained in random as well as in the deterministic settings and for various geometries. Convergence results on manifolds for random and periodic Schrödinger operators are studied in [2, 22, 37, 44, 45] and in the discrete case for finite difference operators on periodic graphs in [9, 13, 30, 31, 47]. Note that the approximability of the zeroth ℓ^2-Betti number can be interpreted as the evaluation of the IDS at one single point. Therefore, it is important to mention the works [14–16, 24, 27], where this problem was studied.

In the present paper, we study the questions (a) and (b) in a very broad background. First, the class of operators we treat is very large, in the deterministic as well as in the random setting. The operators are defined on $\ell^2(G)$, where G is a finitely generated group, which is detailed in the following paragraph. We assume the deterministic operators to be self-adjoint and translation invariant, and we require that the finitely supported functions are a core. In the random situation, we require translation invariance in distribution, some independence and a classical moment condition (3.2) on the matrix elements. These assumptions imply essential self-adjointness of the random operators on the compactly supported functions. This class of *random Hamiltonians* includes famous random models from mathematical physics such as the Anderson model on Cayley graphs and the Laplacian of percolation graphs. Note that our methods allow for unboundedness of the operators and of their hopping range in the random and in the non-random case.

Second, our geometry is very general: We can treat any finitely generated sofic group, cf. Definition 1.1. The notion of sofic groups goes back to Gromov [19] and Weiss [48] and was later on studied for instance in [7, 8, 11, 17, 34, 43, 46]. This class of groups contains all amenable groups, all residually finite groups and therefore especially all groups of sub-exponential growth as well as some exponentially growing groups, as for instance the free group. It was shown in [11] that there exist sofic groups which are not limits of amenable groups. Moreover, there is no group which is known to be non-sofic.

The Cayley graph of a sofic group with a finite generating set can be approximated with a sequence of finite graphs. We extend this approximation to the level of operators, leading to an appropriate definition of the IDS. We establish the Pastur–Shubin trace formula in the deterministic and in the random settings. Hence, we give positive answers to the questions (a) and (b) in very general situations. The deterministic result Theorem 2.4 should be compared to [27, Theorem 2.3.1]. While Lück covers residually finite groups and bounded operators, we treat the more general class of sofic groups and allow for unboundedness of the operators. The random results have, to the best of our knowledge, never been achieved in this generality.

With stronger assumptions on the geometry, uniform convergence of the eigenvalue counting functions can be obtained. For example, sufficient conditions are $G = \mathbb{Z}^d$ [21] or amenability of G [26,36]. For a survey on uniform convergence for operators on groups and additional references, we recommend [41]. Similar results are possible for operators on Delone sets [25].

The physical relevance of our results is underlined by Theorem 3.10, which states that the topological support of the measure associated with the SDF equals the spectrum of the operator. For a detailed study of related results, we refer to [23] and the references therein.

As mentioned, our results apply to the free group and therefore to regular trees. The approximation of trees via finite volume graphs is an intensively studied problem, see e.g. [4] and references therein.

The main obstacle is the non-amenability of trees, i.e., the average over a ball depends drastically on the contribution of the boundary sphere of the ball. The sphere of the ball has nodes of altered degree and that prevents good approximation properties. As a result, instead of the Cayley graph of the free group, [4] approximated the canopy tree, which highlights the leaves of degree 1. The analogous phenomenon was encountered in [44,45] in the continuous setting. In order to construct good approximating graphs for regular trees, one resorts to regular graphs. As shown by [29], the correct strategy in order to approximate the regular tree is to avoid large quantities of small cycles. Other possibilities to improve the approximation properties of balls are studied in [18]. There the authors insert weighted edges connecting the boundary elements.

Our results show that the definition of sofic groups gives a natural criterion for the choice of the approximating finite objects. We hereby open the way to explore phenomena like eigenvalue statistics, which depend by definition on suitable approximations, for a wide variety of models. In [4], the authors study Poisson statistics vs. level repulsion for the canopy tree. We expect sofic approximations to be a good starting point to treat this topic in more models with absolutely continuous spectrum.

Let us describe the content of the paper in detail. First, we detail the setting, give the definition of sofic groups and fix notation. Section 2 is devoted to the proof of Theorem 2.4, the convergence result for deterministic operators. In Sect. 3, we study questions (a) and (b) for random operators on sofic groups. We define the class of random operators we study henceforth on sofic groups. Section 3.1 is the version of Sect. 2 for random Hamiltonians. Here, we define the finite-dimensional approximating operators and show the convergence of their eigenvalue counting functions in expectation, see Theorem 3.5. In Sect. 3.2, we improve the convergence of the expectation values to almost sure convergence. The means of choice here is a well-known concentration inequality by McDiarmid. Note that all of our convergence results are established without an ergodic theorem at hand.

Section 4 contains an example, namely long-range percolation on sofic groups, in Sect. 4.1. For free groups, the IDS is known explicitly. We give two explicit constructions of deterministic sofic approximations for the free group in Sect. 4.2. We illustrate our deterministic convergence result for the free group with two generators with numerical implementations.

Four appendices finish the paper. Thanks to M. Hansmann, Appendix A presents a class of deterministic, unbounded operators on $\mathbb{Z}$ which fit in the setting of Sect. 2. In Appendix C, we explicitly construct random self-adjoint operators on countable groups, which are translation invariant in distribution, employing techniques from [35]. The proofs in Appendices B and D are rather folklore. We include them since we did not find appropriate references.

1.1. Setting and Notation

Let G be a group and $S \subseteq G$ a finite and symmetric set of generators. The *Cayley graph* $\Gamma = \Gamma(G, S)$ is the graph with vertices G and a directed edge from $g \in G$ to $h \in G$, if $gh^{-1} \in S$. We label the edge between g and h with gh^{-1}. For any graph (V, E) the *graph distance* $d^{(V,E)} \colon V \times V \to \mathbb{N}_0$ is given as the length of the shortest path between the arguments, ignoring the direction of the edges. We denote the ball around $v \in V$ of radius $r \geq 0$ with respect to $d^{(V,E)}$ by $B_r^V(v)$. If $(V, E) = \Gamma(G, S)$, we write $B_r^G := B_r^V(\mathrm{id})$, where id is the identity of G.

Definition 1.1 (cf. [48]). In the above setting, G is *sofic*, if for all $\epsilon > 0$ and $r \in \mathbb{N}$ there is a finite directed graph $(V_{r,\epsilon}, E_{r,\epsilon})$, edge labeled by S, which has a finite subset $V_{r,\epsilon}^{(0)} \subseteq V_{r,\epsilon}$ such that:

(S1) sofic]S1 for all $v \in V_{r,\epsilon}^{(0)}$ the r-ball around v in the graph distance of $(V_{r,\epsilon}, E_{r,\epsilon})$ is isomorphic as a labeled graph to $\Gamma|_{B_r^G}$,

(S2) sofic]S2 $|V_{r,\epsilon}^{(0)}| \geq (1 - \epsilon)|V_{r,\epsilon}|$.

The approximating graphs $(V_{r,\epsilon}, E_{r,\epsilon})$ are called *sofic approximations.* Note that the property of being sofic is independent of the specific choice of the symmetric generating system S, cf. [48]. The class of sofic groups is quite large. In fact, there is no known example for a (finitely generated) group which fails to be sofic. As already proven by Weiss, amenable as well as residually finite groups are sofic. In particular, finitely generated free groups are residually finite and hence sofic.

Except otherwise mentioned, we assume that the group G is infinite and sofic. In order to simplify notation, we choose some function $\epsilon \colon \mathbb{N} \to (0, \infty)$ with $\lim_{r\to\infty} \epsilon(r) = 0$ and fix for each $r \in \mathbb{N}$ a graph $(V_{r,\epsilon(r)}, E_{r,\epsilon(r)})$ and a subset $V_{r,\epsilon(r)}^{(0)}$ satisfying (S1) and (S2). We abbreviate

$$\Gamma_r = (V_r, E_r) = (V_{r,\epsilon(r)}, E_{r,\epsilon(r)}), \quad V_r^{(0)} = V_{r,\epsilon(r)}^{(0)}, \quad d_r = d^{(V_r,E_r)}. \tag{1.1}$$

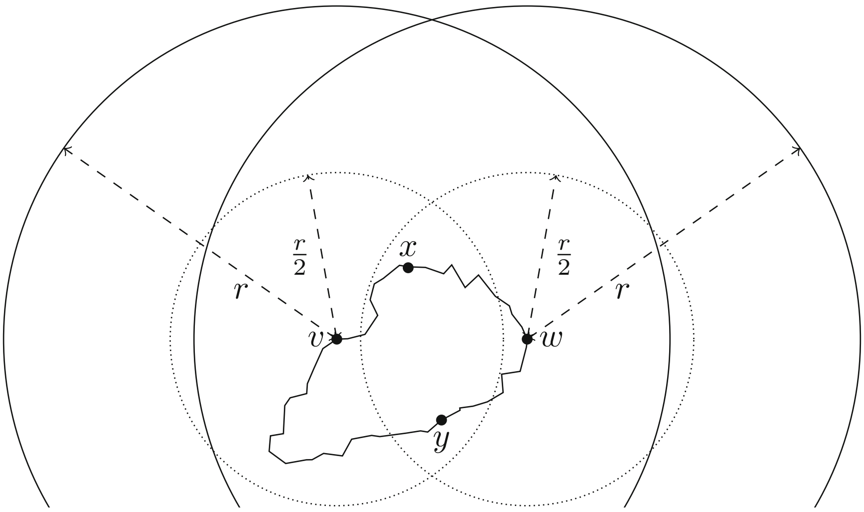


Figure 1. Illustration of Lemma 2.2. Note that all paths stay inside the solid balls

Throughout the paper, we deal with the Hilbert space $\ell^2(G)$ of square summable functions on G. We denote the Kronecker delta $\delta_g \in \ell^2(G)$ at $g \in G$ by $\delta_g \colon G \to \{0,1\}$, i.e., $\delta_g(h) = 1$ iff $g = h$. The set of compactly supported functions on G is $D_0 := \operatorname{lin}\{\delta_g \mid g \in G\}$. In this sense, a function $\phi \in \ell^2(G)$ is called compactly supported (or finitely supported) if the set $\operatorname{spt}(\phi) := \{x \in G \mid \phi(x) \neq 0\}$ is finite.

2. Deterministic Approximation Results

Let $A \colon D \subseteq \ell^2(G) \to \ell^2(G)$ be a self-adjoint operator satisfying

(A1) assumption]core The set of compactly supported functions D_0 is a core for A.
(A2) assumption]transl A is translation invariant, i.e. $a(g,h) = a(gh', hh')$ for all $g, h, h' \in G$,

where $a(g,h) := \langle \delta_g, A\delta_h \rangle$ is the matrix element of A at $g, h \in G$. Note that Assumption (A1) implies $\|A\delta_g\|_2^2 = \sum_{h \in G} |a(g,h)|^2 < \infty$ for all $g \in G$.

Remark 2.1. The operators fulfilling Assumptions (A1) and (A2) can be unbounded. An example of such an operator on $\mathbb{Z}$ is constructed in Appendix A.

Since G is sofic, we have, for each $v \in V_r^{(0)}$, an isomorphism of labeled graphs

$$\Psi_{r,v} \colon B_r^{V_r}(v) \to B_r^G. \tag{2.1}$$

Note that for $v, w \in V_r^{(0)}$ with $d_r(v,w) < r$, we have

$$\Psi_{r,v}(w) = (\Psi_{r,w}(v))^{-1}, \tag{2.2}$$

since the labels along a path from v to w are preserved and equal the inverse labels of the reversed path. In particular, we have $\Psi_{r,v}(v) = \mathrm{id} \in G$ for all v.

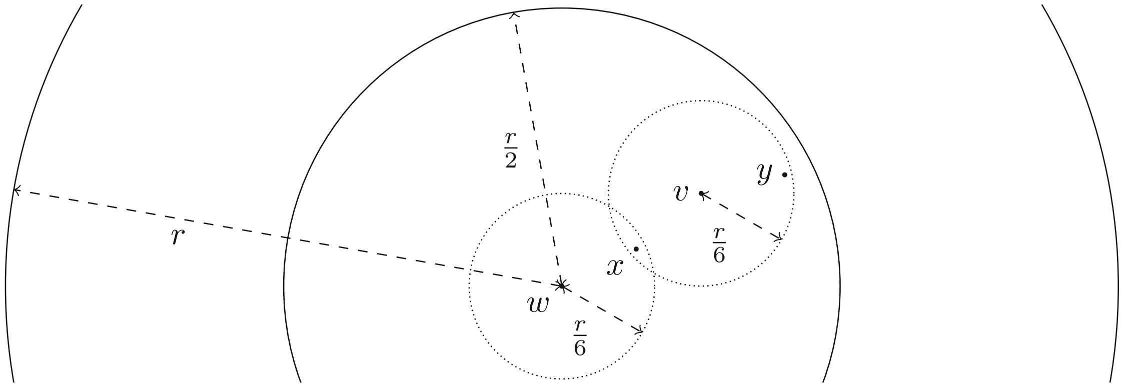


FIGURE 2. Illustration of Remark 2.3

Lemma 2.2. *Let $r \in \mathbb{N}$. If $x, y \in V_r$ and $v, w \in V_r^{(0)}$ fulfill $x, y \in B_{r/2}^{V_r}(v) \cap B_{r/2}^{V_r}(w)$, then we have*

$$\Psi_{r,v}(x)(\Psi_{r,v}(y))^{-1} = \Psi_{r,w}(x)(\Psi_{r,w}(y))^{-1}.$$

Proof. Let $x, y \in V_r$ and $v, w \in V_r^{(0)}$ be such that $x, y \in B_{r/2}^{V_r}(v) \cap B_{r/2}^{V_r}(w)$. Then $k := d_r(x, y) \leq r$ and hence all shortest paths in Γ_r connecting x and y are completely contained in $B_r^{V_r}(v)$ as well as in $B_r^{V_r}(w)$ (Fig. 1). We consider one of these shortest (directed) paths from x to y. Let $(s_1, \ldots, s_k)$ be the vector of the labels of this path. Then, we have, by the properties of $\Psi_{r,v}$,

$$\Psi_{r,v}(x)(\Psi_{r,v}(y))^{-1} = s_1 \cdots s_k(\Psi_{r,v}(y))(\Psi_{r,v}(y))^{-1} = s_1 \cdots s_k.$$

As we also have $\Psi_{r,w}(x) = s_1 \cdots s_k(\Psi_{r,w}(y))$, the claim follows. $\square$

We define the approximating operator $A_r \colon \ell^2(V_r) \to \ell^2(V_r)$ on the graph Γ_r by

$$(A_r f)(x) := \sum_{y \in V_r} a_r(x, y) f(y), \text{ where}$$

$$a_r(x, y) := \begin{cases} a(\Psi_{r,v}(x), \Psi_{r,v}(y)) & \text{if } \exists v \in V_r^{(0)} : x, y \in B_{r/6}^{V_r}(v) \\ 0 & \text{otherwise.} \end{cases}$$

This operator is well-defined by Lemma 2.2. Note that A_r is a symmetric and hence self-adjoint operator on $\ell^2(V_r)$.

Remark 2.3. The reason why we use $r/6$ instead of $r/2$ is the following. In the proofs of Theorems 2.4 and 3.5, we need to ensure that $y \in B_{r/2}^{V_r}(w)$ whenever $w \in V_r^{(0)}$, $x \in B_{r/6}^{V_r}(w)$ and $a_r(x, y) \neq 0$, see Fig. 2.

We can then use Lemma 2.2 and Assumption (A2) to change the reference point v in the definition of $a_r(x, y)$ to w. As $B_r^{V_r}(w)$ is the domain of $\Psi_{r,w}$, we have a corresponding statement in B_r^G. Note that $r/6$ is an upper bound. Any function $\rho(r) \leq r/6$ with $\lim_{r\to\infty} \rho(r) = \infty$ is permissible, see Eq. (3.4).

Define for each $r \in \mathbb{N}$ the normalized eigenvalue counting function N_r of A_r by

$$N_r : \mathbb{R} \to [0,1], \quad N_r(\lambda) := \frac{|\{\text{eigenvalues of } A_r \text{ not larger than } \lambda\}|}{|V_r|},$$

where the eigenvalues are counted with multiplicity. If the limit of these functions for $r \to \infty$ exists (in an appropriate sense), this limit is called the *integrated density of states* (IDS) of A. We denote by E_λ the spectral projection of the operator A on the interval $(-\infty, \lambda]$. With its help, we define

$$N : \mathbb{R} \to [0,1], \quad N(\lambda) := \langle \delta_{\mathrm{id}}, E_\lambda \delta_{\mathrm{id}} \rangle. \tag{2.3}$$

This is a distribution function for a probability measure and is called the *spectral distribution function* (SDF) of A. The next theorem shows that the integrated density of states exists and equals the spectral distribution function. In other words, we prove the *Pastur–Shubin trace formula*.

Theorem 2.4. *Let N and N_r be given as above. Then*

$$\lim_{r \to \infty} N_r(\lambda) = N(\lambda)$$

at all continuity points λ of N.

In the proof of Theorem 2.4, we will use the following well-known lemmas.

Lemma 2.5. *Let $H : D(H) \to \ell^2(G)$ be a self-adjoint operator with domain $D(H) \subseteq \ell^2(G)$, and assume that $D_0 \subseteq D(H)$ is a core of H. Then, for all $z \in \mathbb{C} \backslash \mathbb{R}$,*

$$(z - H)(D_0) = \{\psi \in \ell^2(G) \mid (z - H)^{-1}\psi \in D_0\}$$

is dense in $\ell^2(G)$.

Proof. Since D_0 is a core of H, it is dense in $D(H)$ with respect to the graph norm $\xi \mapsto \|\xi\|_H := \|\xi\| + \|H\xi\|$. The map

$$z - H : (D(H), \|\cdot\|_H) \to (\ell^2(G), \|\cdot\|_2)$$

is continuous and subjective. The statement follows. □

Lemma 2.6. *Let $N, N_1, N_2, \ldots : \mathbb{R} \to [0,1]$ be probability distribution functions. Then the following are equivalent.*

(i) $\lim_{r\to\infty} N_r(\lambda) = N(\lambda)$ *for all continuity points λ of N.*
(ii) $\lim_{r\to\infty} \int (z-\lambda)^{-1} dN_r(\lambda) = \int (z-\lambda)^{-1} dN(\lambda)$ *for all $z \in \mathbb{C}\backslash\mathbb{R}$.*

For the proof of Lemma 2.6 we refer to Appendix B.

Remark 2.7. If one of the assertions holds true, the sequence N_r is said to *converge weakly* to N. The function $\mathbb{C}\backslash\mathbb{R} \to \mathbb{C}$, $z \mapsto \int (z-\lambda)^{-1} dN(\lambda)$ is called *Stieltjes transform* of the distribution function N. One can rephrase *(ii)* as follows. The Stieltjes transforms of N_r converge pointwise toward the Stieltjes transform of N.

Proof of Theorem 2.4. Fix $z \in \mathbb{C}\setminus\mathbb{R}$ and define

$$D_r := \Big| \int_{\mathbb{R}} (z-\lambda)^{-1} dN_r(\lambda) - \int_{\mathbb{R}} (z-\lambda)^{-1} dN(\lambda) \Big|.$$

By Lemma 2.6 it suffices to show that $\lim_{r\to\infty} D_r = 0$. As a first step, we denote the multiplicity of $\lambda \in \sigma(A_r)$ by m_λ and calculate

$$\begin{aligned}\int_{\mathbb{R}} (z-\lambda)^{-1} dN_r(\lambda) &= \frac{1}{|V_r|} \sum_{\lambda\in\sigma(A_r)} m_\lambda (z-\lambda)^{-1} = \frac{1}{|V_r|} \operatorname{Tr}((z-A_r)^{-1}) \\ &= \frac{1}{|V_r|} \sum_{v\in V_r} \langle \delta_v, (z-A_r)^{-1}\delta_v\rangle. \end{aligned} \tag{2.4}$$

On the other hand, the spectral theorem gives

$$\int_{\mathbb{R}} (z-\lambda)^{-1} dN(\lambda) = \langle \delta_{\mathrm{id}}, (z-A)^{-1}\delta_{\mathrm{id}}\rangle.$$

Therefore, by the Cauchy–Schwarz inequality, the bound $\|(z-A)^{-1}\| \leq |\Im z|^{-1}$ and Condition (S2), we obtain

$$\begin{aligned} D_r &= \Big| \frac{1}{|V_r|} \sum_{v\in V_r} \langle \delta_v, (z-A_r)^{-1}\delta_v\rangle - \langle \delta_{\mathrm{id}}, (z-A)^{-1}\delta_{\mathrm{id}}\rangle \Big| \\ &\leq \frac{1}{|V_r|} \sum_{v\in V_r^{(0)}} |\langle \delta_v, (z-A_r)^{-1}\delta_v\rangle - \langle \delta_{\mathrm{id}}, (z-A)^{-1}\delta_{\mathrm{id}}\rangle| \\ &\quad + \frac{1}{|V_r|} \sum_{v\in V_r\setminus V_r^{(0)}} |\langle \delta_v, (z-A_r)^{-1}\delta_v\rangle - \langle \delta_{\mathrm{id}}, (z-A)^{-1}\delta_{\mathrm{id}}\rangle| \\ &\leq \sup_{v\in V_r^{(0)}} |\langle \delta_v, (z-A_r)^{-1}\delta_v\rangle - \langle \delta_{\mathrm{id}}, (z-A)^{-1}\delta_{\mathrm{id}}\rangle| + \frac{2\epsilon(r)}{|\Im z|}. \end{aligned} \tag{2.5}$$

Note that the resolvents live on different spaces, which makes their matrix elements difficult to compare. We overcome this difficulty by introducing local analogues of A_r on $\ell^2(G)$. To this end, we extend for each $v \in V_r^{(0)}$ the graph isomorphism form (2.1) to an injective map

$$\Psi'_{r,v} \colon V_r \to G.$$

Note that we do not require that $\Psi'_{r,v}$ is a graph isomorphism but an injection of the set V_r into the set G. This map induces a projection

$$\Phi_{r,v} \colon \ell^2(G) \to \ell^2(V_r), \quad \Phi_{r,v}(f) := f \circ \Psi'_{r,v}. \tag{2.6}$$

We use this projection to transport A_r to $\ell^2(G)$ from the point of view of v:

$$\hat{A}_{r,v} := \Phi^*_{r,v} A_r \Phi_{r,v} \colon \ell^2(G) \to \ell^2(G),$$

and set $\hat{a}_{r,v}(g,h) := \langle \delta_g, \hat{A}_{r,v}\delta_h\rangle$ for $g,h \in G$. The operator $z - \hat{A}_{r,v}$ has block structure, and with $\Psi'_{r,v}(v) = \mathrm{id}$, one easily verifies

$$\langle \delta_v, (z-A_r)^{-1}\delta_v\rangle = \langle \delta_{\mathrm{id}}, (z-\hat{A}_{r,v})^{-1}\delta_{\mathrm{id}}\rangle.$$

We plug this into (2.5) and insert $\psi \in \ell^2(G)$ for later optimization. Use the second resolvent identity to deduce

$$\begin{aligned}
D_r &\leq \sup_{v \in V_r^{(0)}} |\langle \delta_{\mathrm{id}}, \big((z - \hat{A}_{r,v})^{-1} - (z - A)^{-1}\big)\psi\rangle| + \frac{2(\epsilon(r) + \|\delta_{\mathrm{id}} - \psi\|_2)}{|\Im z|} \\
&\leq \sup_{v \in V_r^{(0)}} |\langle \delta_{\mathrm{id}}, (z - \hat{A}_{r,v})^{-1}(A - \hat{A}_{r,v})(z - A)^{-1}\psi\rangle| + \frac{2(\epsilon(r) + \|\delta_{\mathrm{id}} - \psi\|_2)}{|\Im z|} \\
&\leq \frac{1}{|\Im z|} \sup_{v \in V_r^{(0)}} \|(A - \hat{A}_{r,v})(z - A)^{-1}\psi\|_2 + \frac{2(\epsilon(r) + \|\delta_{\mathrm{id}} - \psi\|_2)}{|\Im z|}. \qquad (2.7)
\end{aligned}$$

Let us choose ψ appropriately. For an arbitrary $\kappa > 0$, Lemma 2.5 provides $\psi \in \ell^2(G)$ with

$$\|\delta_{\mathrm{id}} - \psi\|_2 < \kappa \quad \text{and} \quad \phi := (z - A)^{-1}\psi \in D_0.$$

For $r \geq 6\,\mathrm{diam}(\mathrm{spt}\,\phi)$, we continue to estimate, using the properties of the approximation $\hat{A}_{r,v}$, $v \in V_r^{(0)}$, and the Cauchy–Schwarz as well as the triangle inequality:

$$\begin{aligned}
\|(A - \hat{A}_{r,v})\phi\|_2 &= \Bigg(\sum_{g \in G \setminus B^G_{r/6}} \Big| \sum_{h \in \mathrm{spt}\,\phi} \langle \delta_g, (A - \hat{A}_{r,v})\delta_h\rangle \phi(h)\Big|^2 \Bigg)^{1/2} \\
&\leq \|\phi\|_2 \Big(\sum_{g \in G \setminus B^G_{r/6}} \sum_{h \in \mathrm{spt}\,\phi} |a(g,h) - \hat{a}_{r,v}(g,h)|^2 \Big)^{1/2} \\
&\leq \|\phi\|_2 \Bigg(\Big(\sum_{\substack{g \in G \setminus B^G_{r/6} \\ h \in \mathrm{spt}\,\phi}} |a(g,h)|^2 \Big)^{1/2} + \Big(\sum_{\substack{g \in G \setminus B^G_{r/6} \\ h \in \mathrm{spt}\,\phi}} |\hat{a}_{r,v}(g,h)|^2 \Big)^{1/2} \Bigg). \qquad (2.8)
\end{aligned}$$

By Remark 2.3, $\hat{a}_{r,v}(g,h) \neq 0$ with $h \in \mathrm{spt}\,\phi \subseteq B_{r/6}$ implies $g \in B^G_{r/2}$. Therefore, employing Lemma 2.2, we can estimate

$$\begin{aligned}
&\sum_{h \in \mathrm{spt}\,\phi} \sum_{g \in G \setminus B^G_{r/6}} |\hat{a}_{r,v}(g,h)|^2 \\
&\quad = \sum_{y \in \Psi^{-1}_{r,v}(\mathrm{spt}\,\phi)} \sum_{x \in B^{V_r}_{r/2}(v) \setminus B^{V_r}_{r/6}(v)} |a_r(x,y)|^2 \\
&\quad \leq \sum_{y \in \Psi^{-1}_{r,v}(\mathrm{spt}\,\phi)} \sum_{x \in B^{V_r}_{r/2}(v) \setminus B^{V_r}_{r/6}(v)} |a(\Psi_{r,v}(x), \Psi_{r,v}(y))|^2 \\
&\quad \leq \sum_{h \in \mathrm{spt}\,\phi} \sum_{g \in G \setminus B^G_{r/6}} |a(g,h)|^2. \qquad (2.9)
\end{aligned}$$

These considerations lead to

$$\|(A - \hat{A}_{r,v})\phi\|_2 \leq 2\|\phi\|_2 \Big(\sum_{h \in \mathrm{spt}\,\phi} \sum_{g \in G \setminus B^G_{r/6}} |a(g,h)|^2 \Big)^{1/2}.$$

By Assumption (A1) and $|\operatorname{spt}\phi| < \infty$, the last expression converges to 0 as $r \to \infty$. According to (2.7),

$$\limsup_{r\to\infty} D_r \leq 2\kappa/|\Im z|.$$

Since κ was chosen arbitrary, we obtain $\lim_{r\to\infty} D_r = 0$. $\square$

3. Approximation Results in the Random Setting

We study the existence of the IDS for random operators on sofic groups. First, we introduce the random operators and state relevant properties. Detailed definitions are given in Appendix C. Then, we investigate the eigenvalue counting functions of suitable approximating matrix operators. Proceeding in two steps, we first show convergence in mean, see Sect. 3.1, and improve this to almost sure convergence in Sect. 3.2. Through this, we again obtain a Pastur–Shubin trace formula.

For the construction of the random operators, we proceed as follows. Let G be a finitely generated sofic group. Further, let $\mathcal{P}_{1,2} := \{e \subseteq G \mid |e| \in \{1,2\}\}$ be the set of all edges of the complete undirected graph with vertex set G. Moreover, fix independent random variables $X_e\colon \Omega \to \mathbb{R}$, $e \in \mathcal{P}_{1,2}$, such that for each $x \in G$ the random variables in

$$\{X_{\{g,h\}} \mid g,h \in G, x \in \{gh^{-1}, hg^{-1}\}\} \tag{3.1}$$

are identically distributed. Note that in (3.1), $x = \mathrm{id} \in G$ gives the set of all random variables $X_{\{g\}} = X_{\{g,g\}}$, $g \in G$. We require further

$$\mathbb{E}\Big[\Big(\sum_{g\in G}|X_{\{\mathrm{id},g\}}|\Big)^2\Big] < \infty. \tag{3.2}$$

Using these random variables, we define

$$a(g,h) := X_{\{g,h\}} - \alpha\delta_g(h) \sum_{g'\in G\setminus\{g\}} X_{\{h,g'\}} \tag{3.3}$$

for $g,h \in G$ and a parameter $\alpha \in \mathbb{R}$. In the following lemma, we use these random variables $a(g,h)$ to define an adapted random operator. Afterward, we discuss the particular choice of the matrix elements $a(g,h)$ in (3.3).

Lemma 3.1. *There exists a random operator $A = (A^{(\omega)})_{\omega\in\Omega}$ with the following properties.*

(i) For all $\omega \in \Omega$, the operator $A^{(\omega)}\colon D(A^{(\omega)}) \to \ell^2(G)$ is self-adjoint, and the compactly supported functions D_0 are a core for $A^{(\omega)}$.

(ii) For almost all $\omega \in \Omega$, the matrix elements of $A^{(\omega)}$ are given by

$$\langle\delta_g, A^{(\omega)}\delta_h\rangle = a^{(\omega)}(g,h).$$

(iii) We have

$$\mathbb{E}\big[\|A\delta_{\mathrm{id}}\|_1^2\big] = \mathbb{E}\Big[\Big(\sum_{g\in G}|a(g,\mathrm{id})|\Big)^2\Big] < \infty.$$

(iv) *The operator A is translation invariant in distribution.*

(v) *The resolvents $\omega \to (z - A^{(\omega)})^{-1}$, $z \in \mathbb{C}\backslash\mathbb{R}$, are strongly measurable.*

Remark 3.2. By *(ii)*, the operator A is, for almost all $\omega \in \Omega$, uniquely determined by the matrix elements (3.3).

Proof. The moment condition (3.2) implies

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{x\in G}|a(x,\mathrm{id})|\Big)^2\Big] &\le \mathbb{E}\Big[\Big(\sum_{x\in G}|X_{\{\mathrm{id},x\}}| + \alpha\sum_{z\in G}|X_{\{\mathrm{id},z\}}|\Big)^2\Big]\\ &= (1+\alpha)^2\mathbb{E}\Big[\Big(\sum_{x\in G}|X_{\{\mathrm{id},x\}}|\Big)^2\Big] < \infty.\end{aligned}$$

Thus, the lemma follows from Appendix C. □

In the case $\alpha = 0$, the operator is a version of the adjacency matrix on graphs with vertices in G and random edge weights. For $\alpha = 1$ and $X_{\{g\}} = 0$ a.s. we obtain randomly weighted Laplace operators on such graphs. More generally, one interprets the diagonal terms $X_{\{g\}}$ as random potential. This well-studied setting is known as *Anderson model*. For all these reasons, we call the random operators $(A^{(\omega)})_{\omega\in\Omega}$ *random Hamiltonians*.

Remark 3.3. The spectrum of our random Hamiltonians is almost surely deterministic. In e. g. [10,23,35], statements of this kind are a consequence of ergodicity. Since our operators share their probability space with the finite approximations, they are not ergodic. However, with an adapted choice for the probability space, the operators of Lemma 3.1 are seen to be ergodic. For the case of $G = \mathbb{Z}^d$, this is implemented in [35]. Unfortunately, such a probability space is to small to support the independent approximating operators. Of course, the spectral properties of the operators persist when represented on a different probability space.

Note that the argument to establish deterministic spectrum does not require an ergodic theorem but only the fact that invariant functions are almost surely constant. One can adapt the proof of [23, Theorem 5.1] to our setting of non-ergodic operators, too. The group G acts equivariant on $\ell^2(G)$ and the probability space. As a substitute for ergodicity of the action of G on the probability space, one uses that all G-invariant random variables which are measurable with respect to the σ-Algebra generated by the matrix elements of A are almost surely constant. For more details see [23] or the references therein.

The next well-known lemma gives sufficient conditions for boundedness and for unboundedness of the operators in consideration. The proof is deferred to Appendix D.

Lemma 3.4. *Let A be the random Hamiltonian defined above with the random variables $X_{\{g,h\}}$, $g, h \in G$, and $D := \sup_{g\in G}\|X_{\{\mathrm{id},g\}}\|_\infty \in [0,\infty]$.*

(i) *If $D = \infty$, then A is unbounded.*

(ii) *If $D < \infty$ and A is of finite hopping range, i.e. $X_{\{g,h\}} = 0$ whenever $d(g,h) \ge R$ for some fixed $R \in \mathbb{N}$, then A is bounded.*

3.1. Approximation in Mean

Let $A = (A^{(\omega)})_{\omega\in\Omega}$ be a random Hamiltonian, as outlined in Lemma 3.1. We consider the approximating graphs Γ_r, $r \in \mathbb{N}$, and use the simplified notation (1.1). Recall the graph isomorphisms $\Psi_{r,v} \colon B_r^{V_r}(v) \to B_r^G$, $v \in V_r$, from (2.1), too.

Let us define finite-dimensional approximations $A_r \colon \ell^2(V_r) \to \ell^2(V_r)$ of A. For Sect. 3.2, we generalize $r \mapsto r/6$ from Remark 2.3 to $\rho\colon \mathbb{N} \to \mathbb{R}$ satisfying

$$\rho(r) \le \frac{r}{6} \quad \text{and} \quad \rho(r) \xrightarrow{r\to\infty} \infty. \tag{3.4}$$

Let

$$C_r(e) := \left\{ v \in V_r^{(0)} \;\middle|\; x, y \in B_{\rho(r)}^{V_r}(v) \right\}$$

for each $e = \{x, y\} \in \mathcal{P}_{1,2}^{(r)} := \{e \subseteq V_r \mid |e| \in \{1, 2\}\}$. For each such edge $e = \{x, y\}$, we fix a random variable X_e^r. If there exists a vertex $v \in C_r(e)$, let X_e^r have the same distribution as $X_{\{\Psi_{r,v}(x),\Psi_{r,v}(y)\}}$. Otherwise, set $X_e^r := 0$. We also require all random variables in

$$\left\{ X_e \;\middle|\; e \in \mathcal{P}_{1,2} \right\} \cup \left\{ X_e^r \;\middle|\; r \in \mathbb{N}, e \in \mathcal{P}_{1,2}^{(r)} \right\}$$

to be independent. Use Lemma 2.2 and (3.1) to see that the distribution of X_e^r does not depend on the choice of $v \in C_r(e)$.

We are now in position to define the approximating operator $A_r^{(\omega)}\colon \ell^2(V_r) \to \ell^2(V_r)$, $\omega \in \Omega$, by its matrix elements. With the same $\alpha \in \mathbb{R}$ as in (3.3), we set

$$a_r^{(\omega)}(x, y) := X_{\{x,y\}}^r(\omega) - \alpha\delta_x(y) \sum_{z \in V_r \setminus \{x\}} X_{\{x,z\}}^r(\omega)$$

for all $x, y \in V_r$. This operator is symmetric and hence self-adjoint. Note, too, that $A_r^{(\omega)}$ has hopping range $2\rho(r)$, i.e. $a_r(x, y) = 0$, as soon as $d_r(x, y) > 2\rho(r)$.

As in Sect. 2, we define eigenvalue counting functions. For each $\omega \in \Omega$ and $r \in \mathbb{N}$, we set

$$N_r^{(\omega)}\colon \mathbb{R} \to [0, 1], \quad N_r^{(\omega)}(\lambda) := \frac{|\{\text{eigenvalues of } A_r^{(\omega)} \text{ not larger than } \lambda\}|}{|V_r|},$$

where the eigenvalues are counted with multiplicity. Similarly as before, we define

$$N^{(\omega)}\colon \mathbb{R} \to [0, 1], \quad N^{(\omega)}(\lambda) := \langle \delta_{\mathrm{id}}, E_\lambda^{(\omega)} \delta_{\mathrm{id}} \rangle, \tag{3.5}$$

where $E_\lambda^{(\omega)}$ is the spectral projection of $A^{(\omega)}$ on the interval $(-\infty, \lambda]$.

We also need their expectation values, i.e., the functions $\bar{N}_r, \bar{N} \colon \mathbb{R} \to [0, 1]$,

$$\bar{N}(\lambda) = \mathbb{E}[N(\lambda)] \quad \text{and} \quad \bar{N}_r(\lambda) = \mathbb{E}[N_r(\lambda)]. \tag{3.6}$$

The function $\bar{N}$ is called *spectral distribution function* of the random operator A.

Theorem 3.5. *Let $\bar{N}_r, \bar{N}\colon \mathbb{R} \to [0,1]$ be as above. Then*

$$\lim_{r\to\infty} \bar{N}_r(\lambda) = \bar{N}(\lambda)$$

at all continuity points λ of $\bar{N}$.

The proof of Theorem 3.5 largely follows the lines of the proof of Theorem 2.4. Lemma 2.6 is ready for use, but Lemma 2.5 needs a rather technical upgrade. Lemma 3.6 states in essence that the compactly supported random vectors constitute a core of A.

Lemma 3.6. *Let $A = (A^{(\omega)})_{\omega\in\Omega}$ be a random operator such that all realizations $A^{(\omega)}$ are self-adjoint and share D_0 as a core. Let further $\kappa > 0$ and $z \in \mathbb{C}\backslash\mathbb{R}$. Then there exist a radius $n \in \mathbb{N}$ and a random vector $\psi\colon \Omega \to \ell^2(G)$, such that*

$$\mathbb{E}\big[\|\psi - \delta_{\mathrm{id}}\|_2\big] \leq \kappa \quad \textit{and} \quad \mathrm{spt}\big((z - A^{(\omega)})^{-1}\psi(\omega)\big) \subseteq B_n^G$$

for all $\omega \in \Omega$.

Proof. Fix $\kappa > 0$. For each $n \in \mathbb{N}$, we define the set

$$M_{n,\kappa} := \big\{\omega \in \Omega \bigm| \exists f \in \ell^2(G)\colon \mathrm{spt}\big((z - A^{(\omega)})^{-1} f\big) \subseteq B_n^G, \|f - \delta_{\mathrm{id}}\|_2 \leq \tfrac{\kappa}{2}\big\}.$$

The measurability of $M_{n,\kappa} \subseteq \Omega$ follows from the strong measurability of the resolvents of $A = (A^{(\omega)})_{\omega\in\Omega}$ and

$$M_{n,\kappa} = \bigcap_{m\in\mathbb{N}} \bigcup_{f\in B} \bigcap_{g\in G\backslash B_n^G} \{\omega \in \Omega \mid |\langle \delta_g, (z - A^{(\omega)})^{-1} f\rangle| < \tfrac{1}{m}\}, \tag{3.7}$$

where $B \subseteq \{f \in \ell^2(G) \mid \|f - \delta_{\mathrm{id}}\|_2 < \frac{\kappa}{2} + \frac{1}{m}\}$ is countable and dense. To prove (3.7), we note that the inclusion "$\subseteq$" holds because of the continuity of $f \mapsto \langle \delta_g, (z - A^{(\omega)})^{-1} f\rangle$ uniformly in g. The reverse inclusion "$\supseteq$" requires more care. For each ω in the set on the right-hand side of (3.7), we find for all $m \in \mathbb{N}$ a function $f_m \in B$, such that $|((z - A^{(\omega)})^{-1} f_m)(g)| < \frac{1}{m}$ for all $g \in G\backslash B_n^G(\mathrm{id})$. The condition $f_m \in B$ implies in particular

$$\sup_{m\in\mathbb{N}} \|f_m\|_\infty \leq \sup_{m\in\mathbb{N}} \|f_m\|_2 \leq \sup_{m\in\mathbb{N}} \|f_m - \delta_{\mathrm{id}}\|_2 + \|\delta_{\mathrm{id}}\|_2 \leq 2 + \tfrac{\kappa}{2}.$$

Thus, $(f_m(g))_{m\in\mathbb{N}}$ is a bounded sequence for all $g \in G$. A diagonal argument extracts a subsequence $(f_{m_k})_{k\in\mathbb{N}}$ for which $(f_{m_k}(g))_{k\in\mathbb{N}}$ converges for all $g \in G$. Let $f := \lim_{k\to\infty} f_{m_k}$ be the pointwise limit of these functions. Then, we obtain by Fatou's Lemma that $f \in \ell^2(G)$ and

$$\|f - \delta_{\mathrm{id}}\|_2 \leq \tfrac{\kappa}{2} \quad \text{and} \quad \mathrm{spt}\big((z - A^{(\omega)})^{-1} f\big) \subseteq B_n^G.$$

This shows $\omega \in M_{n,\kappa}$ and (3.7).

For each $\omega \in \Omega$, the compactly supported functions D_0 are a core for $A^{(\omega)}$. Hence, for each $\omega \in \Omega$, Lemma 2.5 provides a function $f \in \ell^2(G)$ with

$$\|f - \delta_{\mathrm{id}}\|_2 < \tfrac{\kappa}{2} \quad \text{and} \quad (z - A^{(\omega)})^{-1} f \in D_0.$$

We rephrase this as

$$\bigcup_{n\in\mathbb{N}} M_{n,\kappa} = \Omega.$$

In fact, the sequence $(M_{n,\kappa})_{n\in\mathbb{N}}$ increases, so we conclude $\mathbb{P}(M_{n,\kappa}) \xrightarrow{n\to\infty} 1$. This enables us to choose $n \in \mathbb{N}$ large enough such that $\mathbb{P}(M_{n,\kappa}^c) < \frac{\kappa}{2}$.

For the choice of ψ, note that the set

$$\begin{aligned}
&\widetilde{M_{n,\kappa}} \\
&\quad := \big\{(\omega, f) \in M_{n,\kappa} \times \ell^2(G) \bigm| \operatorname{spt}\big((z - A^{(\omega)})^{-1} f\big) \subseteq B_n^G, \|f - \delta_{\mathrm{id}}\|_2 < \tfrac{\kappa}{2}\big\} \\
&\quad = \bigcap_{g\in G\setminus B_n^G} \big\{(\omega, f) \in \Omega \times \ell^2(G) \bigm| \langle(\bar{z} - A^{(\omega)})^{-1}\delta_g, f\rangle = 0, \|f - \delta_{\mathrm{id}}\|_2 < \tfrac{\kappa}{2}\big\}
\end{aligned}$$

is measurable, since the factors of the scalar product are measurable, see Proposition C.3. The measurable choice theorem by R. J. Aumann, see [1, 18.27 Corollary], hands us a measurable section, i.e., a random vector $\psi' \colon M_{n,\kappa} \to \ell^2(G)$ whose graph is contained in $\widetilde{M_{n,\kappa}}$. We extend ψ' to $\psi \colon \Omega \to \ell^2(G)$ by 0 on $M_{n,\kappa}$, which has the required properties:

$$\mathbb{E}\big[\|\psi - \delta_{\mathrm{id}}\|_2\big] \le \mathbb{E}\big[\tfrac{\kappa}{2} \cdot \chi_{M_{n,\kappa}}\big] + \mathbb{E}\big[\chi_{M_{n,\kappa}^c}\big] = \kappa$$

and $\operatorname{spt}\big((z - A^{(\omega)})^{-1}\psi(\omega)\big) \subseteq B_n^G$ for all $\omega \in \Omega$. □

Equipped with this keen edge tool, we proceed to the proof of Theorem 3.5.

Proof of Theorem 3.5. Fix $z \in \mathbb{C}\setminus\mathbb{R}$ and let

$$D_r := \Big|\int_{\mathbb{R}} (z - \lambda)^{-1} d\bar{N}_r(\lambda) - \int_{\mathbb{R}} (z - \lambda)^{-1} d\bar{N}(\lambda)\Big|.$$

By Lemma 2.6, we need to show $\lim_{r\to\infty} D_r = 0$.

The following integral is in fact a finite sum. Thus, by linearity of $\mathbb{E}$ and the calculation (2.4),

$$\begin{aligned}
\int_{\mathbb{R}} (z - \lambda)^{-1} d\bar{N}_r(\lambda) &= \mathbb{E}\left[\int_{\mathbb{R}} (z - \lambda)^{-1} dN_r(\lambda)\right] \\
&= \mathbb{E}\Big[\frac{1}{|V_r|} \sum_{v\in V_r} \langle\delta_v, (z - A_r)^{-1}\delta_v\rangle\Big].
\end{aligned}$$

By definition, the Riemann-Stieltjes integral is the limit of a sum. The boundedness of $\lambda \mapsto (z - \lambda)^{-1}$ and dominated convergence allow us to interchange limit and expectation. We thus obtain

$$\int_{\mathbb{R}} (z - \lambda)^{-1} d\bar{N}(\lambda) = \mathbb{E}\left[\int_{\mathbb{R}} (z - \lambda)^{-1} dN(\lambda)\right] = \mathbb{E}\big[\langle\delta_{\mathrm{id}}, (z - A)^{-1}\delta_{\mathrm{id}}\rangle\big]. \quad (3.8)$$

The last equality follows from the spectral theorem. Analogous to (2.5), we achieve

$$D_r \le \sup_{v\in V_r^{(0)}} \Big|\mathbb{E}\big[\langle\delta_v, (z - A_r)^{-1}\delta_v\rangle\big] - \mathbb{E}\big[\langle\delta_{\mathrm{id}}, (z - A)^{-1}\delta_{\mathrm{id}}\rangle\big]\Big| + \frac{2\epsilon(r)}{|\Im z|}. \quad (3.9)$$

As before, we use $\Phi_{r,v}$ from (2.6) to transport $A_r^{(\omega)}$ to $\ell^2(G)$ from the point of view of v:

$$A_{r,v}^{(\omega)} := \Phi_{r,v}^* A_r^{(\omega)} \Phi_{r,v} \colon \ell^2(G) \to \ell^2(G).$$

Again, we have, for all $v \in V_r^{(0)}$,

$$\langle \delta_v, (z - A_r^{(\omega)})^{-1} \delta_v \rangle = \langle \delta_{\mathrm{id}}, (z - A_{r,v}^{(\omega)})^{-1} \delta_{\mathrm{id}} \rangle.$$

The matrix elements of $A_{r,v}$ and A are independent and will thus not cancel each other at the end of the proof. But we arranged that corresponding matrix elements have the same distribution. Under the expectation and from the point of view of v, we can adapt the approximation to A. Define a substitute approximating operator $\hat{A}_{r,v}^{(\omega)} \colon \ell^2(G) \to \ell^2(G)$ of $A^{(\omega)}$ by copying matrix elements from $A^{(\omega)}$:

$$\hat{a}_{r,v}^{(\omega)}(g,h) := \begin{cases} a^{(\omega)}(g,h) & \text{if } g,h \in B_{\rho(r)}^G, g \neq h \\ X_{\{g\}}(\omega) - \alpha \sum_{h' \in G \setminus \{g\}} \hat{a}_{r,v}^{(\omega)}(g,h') & \text{if } g = h \in B_{\rho(r)}^G \\ a_{r,v}^{(\omega)}(g,h) & \text{otherwise.} \end{cases}$$

Here, $X_{\{g\}}(\omega)$ is the first summand in the definition of the matrix element $a^{(\omega)}(g,g)$, see (3.3), and $a_{r,v}^{(\omega)}(g,h) := \langle \delta_g, A_{r,v}^{(\omega)} \delta_h \rangle$ is the matrix element of $A_{r,v}^{(\omega)}$. By construction, the distribution of $a_{r,v}(g,h)$ equals the distribution of $\hat{a}_{r,v}(g,h)$ for $g,h \in G$. Thereby, as intended,

$$\mathbb{E}[\langle \delta_{\mathrm{id}}, (z - A_{r,v})^{-1} \delta_{\mathrm{id}} \rangle] = \mathbb{E}[\langle \delta_{\mathrm{id}}, (z - \hat{A}_{r,v})^{-1} \delta_{\mathrm{id}} \rangle].$$

This is the point, where Lemma 3.6 enters. We fix $\kappa > 0$ arbitrarily and get a radius $n \in \mathbb{N}$ and a random vector $\psi \colon \Omega \to \ell^2(G)$ such that

$$\mathbb{E}[\|\psi - \delta_{\mathrm{id}}\|_2] \le \kappa.$$

Furthermore, $\phi(\omega) := (z - A^{(\omega)})^{-1} \psi(\omega)$, $\omega \in \Omega$, is supported in B_n^G and bounded by

$$\|\phi(\omega)\|_2 \le \|(z - A^{(\omega)})^{-1} \psi(\omega)\|_2 \le (1+\kappa)/|\Im z|.$$

Following the strategy from (2.7), we introduce ψ into (3.9) and apply the second resolvent identity:

$$\begin{aligned} D_r &\le \sup_{v \in V_r^{(0)}} \bigl|\mathbb{E}\bigl[\langle \delta_{\mathrm{id}}, (z - \hat{A}_{r,v})^{-1} \delta_{\mathrm{id}} \rangle\bigr] - \mathbb{E}\bigl[\langle \delta_{\mathrm{id}}, (z - A)^{-1} \delta_{\mathrm{id}} \rangle\bigr]\bigr| + \frac{2\epsilon(r)}{|\Im z|} \\ &\le \sup_{v \in V_r^{(0)}} \bigl|\mathbb{E}\bigl[\langle \delta_{\mathrm{id}}, \bigl((z - \hat{A}_{r,v})^{-1} - (z - A)^{-1}\bigr)\psi \rangle\bigr]\bigr| + 2\frac{\epsilon(r) + \mathbb{E}[\|\psi - \delta_{\mathrm{id}}\|_2]}{|\Im z|} \\ &\le \frac{1}{|\Im z|} \sup_{v \in V_r^{(0)}} \mathbb{E}\bigl[\|(A - \hat{A}_{r,v})(z - A)^{-1}\psi\|_2\bigr] + 2\frac{\epsilon(r) + \kappa}{|\Im z|}. \end{aligned} \tag{3.10}$$

We study the expectation in the last line. Use the subadditivity of the square root in

$$\mathbb{E}\bigl[\|(A - \hat{A}_{r,v})\phi\|_2\bigr] = \mathbb{E}\Bigl[\Bigl(\sum_{g \in G} \bigl|(A - \hat{A}_{r,v})\phi(g)\bigr|^2\Bigr)^{1/2}\Bigr] \le T_1^v(r) + |\alpha| T_2^v(r)$$

to separate the off-diagonal terms in $T_1^v(r)$ and the diagonal terms in $T_2^v(r)$. The expressions $T_1^v(r)$ and $T_2^v(r)$ are detailed in the subsequent paragraphs.

For $r > n$, bound $T_1^v(r)$ as in (2.8):

$$\begin{aligned}
T_1^v(r) &:= \mathbb{E}\Big[\Big(\sum_{g\in G\setminus B^G_{\rho(r)}} \Big|\sum_{h\in B^G_n} \big(a(g,h)-\hat{a}_{r,v}(g,h)\big)\phi(h)\Big|^2\Big)^{1/2}\Big] \\
&\le \mathbb{E}\Big[\|\phi\|_2\Big(\sum_{g\in G\setminus B^G_{\rho(r)}} \sum_{h\in B^G_n} \big|a(g,h)-\hat{a}_{r,v}(g,h)\big|^2\Big)^{1/2}\Big] \\
&\le \frac{1+\kappa}{|\Im z|}\Bigg(\Big(\sum_{\substack{g\in G\setminus B^G_{\rho(r)}\\ h\in B^G_n}} \mathbb{E}\big[|a(g,h)|^2\big]\Big)^{1/2} + \Big(\sum_{\substack{g\in G\setminus B^G_{\rho(r)}\\ h\in B^G_n}} \mathbb{E}\big[|\hat{a}_{r,v}(g,h)|^2\big]\Big)^{1/2}\Bigg).
\end{aligned}$$

In the last step, we used Jensen's inequality. We proceed as in (2.9), using $\mathbb{E}[\hat{a}_{r,v}(g,h)] = \mathbb{E}[a_{r,v}(g,h)]$, and obtain

$$\sup_{v\in V_r^{(0)}} T_1^v(r) \le 2\frac{1+\kappa}{|\Im z|}\Big(\sum_{h\in B^G_n} \mathbb{E}\Big[\sum_{g\in G\setminus B^G_{\rho(r)}} \big|a(g,h)\big|^2\Big]\Big)^{1/2} \xrightarrow{r\to\infty} 0$$

by dominated convergence, since $\mathbb{E}\big[\|A\delta_g\|_2^2\big] < \infty$.

For $\alpha \neq 0$, the diagonal term

$$T_2^v(r) := |\alpha|^{-1}\mathbb{E}\Big[\Big(\sum_{h\in B^G_n} \big|\big(a(h,h)-\hat{a}_{r,v}(h,h)\big)\phi(h)\big|^2\Big)^{1/2}\Big]$$

is dealt with as follows, using Jensen's inequality and $\|\phi\|_\infty \le \|\phi\|_2$,

$$\begin{aligned}
T_2^v(r) &\le \mathbb{E}\Big[\|\phi\|_\infty\Big(\sum_{h\in B^G_n}\Big|\sum_{g\in G\setminus B^G_{\rho(r)}} \big(\hat{a}_{r,v}(h,g)-a(h,g)\big)\Big|^2\Big)^{1/2}\Big] \\
&\le \frac{1+\kappa}{|\Im z|}\mathbb{E}\Big[\Big(\sum_{h\in B^G_n}\Big(\sum_{g\in G\setminus B^G_{\rho(r)}} |\hat{a}_{r,v}(h,g)| + \sum_{g\in G\setminus B^G_{\rho(r)}} |a(h,g)|\Big)^2\Big)^{1/2}\Big] \\
&\le \sqrt{2}\frac{1+\kappa}{|\Im z|}\Bigg(\sum_{h\in B^G_n}\Big(\mathbb{E}\Big[\Big(\sum_{g\in G\setminus B^G_{\rho(r)}} |\hat{a}_{r,v}(h,g)|\Big)^2 + \Big(\sum_{g\in G\setminus B^G_{\rho(r)}} |a(h,g)|\Big)^2\Big]\Big)\Bigg)^{1/2}.
\end{aligned}$$

Broadening the argument from (2.9) and taking advantage of the independence of the involved matrix elements, we see for $h \in B^G_n$:

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sum_{g\in G\setminus B^G_{\rho(r)}} |\hat{a}_{r,v}(h,g)|\Big)^2\Big] &= \mathbb{E}\Big[\sum_{g,g'\in B^G_r\setminus B^G_{\rho(r)}} |\hat{a}_{r,v}(h,g)||\hat{a}_{r,v}(h,g')|\Big] \\
&= \sum_{g\neq g'\in B^G_r\setminus B^G_{\rho(r)}} \mathbb{E}\big[|\hat{a}_{r,v}(h,g)|\big]\mathbb{E}\big[|\hat{a}_{r,v}(h,g')|\big] + \sum_{g\in B^G_r\setminus B^G_{\rho(r)}} \mathbb{E}\big[|\hat{a}_{r,v}(h,g)|^2\big]
\end{aligned}$$

$$\leq \sum_{g\neq g'\in G\setminus B^G_{\rho(r)}} \mathbb{E}\big[|a(h,g)|\big]\mathbb{E}\big[|a(h,g')|\big] + \sum_{g\in G\setminus B^G_{\rho(r)}} \mathbb{E}\big[|a(h,g)|^2\big]$$
$$= \mathbb{E}\Big[\Big(\sum_{g\in G\setminus B^G_{\rho(r)}} |a(h,g)|\Big)^2\Big].$$

We deduce

$$\sup_{v\in V_r^{(0)}} T_2^v(r) \leq 2\sqrt{2}\frac{1+\kappa}{|\Im z|}\Big(\sum_{h\in B^G_n} \mathbb{E}\Big[\Big(\sum_{g\in G\setminus B^G_{\rho(r)}} \big|a(h,g)\big|\Big)^2\Big]\Big)^{1/2} \xrightarrow{r\to\infty} 0,$$

again by Lebesgue, this time with $\mathbb{E}\big[\|A\delta_h\|_1^2\big] < \infty$. Now conclude from (3.10)

$$\limsup_{r\to\infty} D_r \leq \frac{2\kappa}{|\Im z|}.$$

Since $\kappa > 0$ was arbitrary, we reached $D_r \xrightarrow{r\to\infty} 0$. $\square$

3.2. Almost Sure Weak Convergence

In Theorem 3.5, we achieved weak convergence of the eigenvalue counting functions in expectation. Here, we improve this result to weak convergence for almost all realizations. Again, this results in a Pastur–Shubin trace formula.

The following concentration inequality from [28, Theorem 3.1] is our main tool.

Theorem 3.7 ([28, Theorem 3.1]). *Let $X = (X_1,\ldots,X_n)$ be a family of independent random variables with values in $\mathbb{R}$, and let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a function, such that whenever $x, x' \in \mathbb{R}^n$ differ only in one coordinate, we have*

$$|f(x) - f(x')| \leq c.$$

Then, for $\mu := \mathbb{E}[f(X)]$ and any $\epsilon \geq 0$,

$$\mathbb{P}(|f(X) - \mu| \geq \epsilon) \leq 2\exp\Big(-\frac{2\epsilon^2}{nc^2}\Big).$$

We need to strengthen the assumptions (3.4) on ρ, namely, ρ must not grow too fast:

$$\rho(r) \leq \frac{\ln r}{4\ln|S|} - 1 \quad \text{and} \quad \rho(r) \xrightarrow{r\to\infty} \infty. \tag{3.11}$$

This is a conservative approach. See Remark 3.9 for possible relaxations in more restricted settings.

Theorem 3.8. *Let N_r and $\bar{N}$ be as in (3.5) and (3.6) and ρ according to (3.11). Then there is a set $\tilde{\Omega} \in \mathcal{A}$ with full probability, $\mathbb{P}(\tilde{\Omega}) = 1$, and*

$$\lim_{r\to\infty} N_r^{(\omega)}(\lambda) = \bar{N}(\lambda)$$

at continuity points λ of $\bar{N}$, for all $\omega \in \tilde{\Omega}$.

Proof. Denote the set of continuity points of $\bar{N}$ by $\mathcal{C}$ and let $\lambda \in \mathcal{C}$ and $\epsilon > 0$. For r large enough, we have by Theorem 3.5

$$\begin{aligned} \mathbb{P}\big(|N_r(\lambda) - \bar{N}(\lambda)| \geq \epsilon\big) &\leq \mathbb{P}\big(|N_r(\lambda) - \bar{N}_r(\lambda)| \geq \epsilon - |\bar{N}_r(\lambda) - \bar{N}(\lambda)|\big) \\ &\leq \mathbb{P}\big(|N_r(\lambda) - \bar{N}_r(\lambda)| \geq \epsilon/2\big). \end{aligned} \tag{3.12}$$

We want to apply Theorem 3.7. Each random variable on a bond of the graph Γ_r in a sofic approximation has bounded effect on the eigenvalue counting function, namely

$$|N_r^{(\omega)}(\lambda) - N_r^{(\omega')}(\lambda)| \leq c_r := 2/|V_r|$$

for all $\omega, \omega' \in \Omega$ which differ only on a single bond. This is clear as the associated operators differ only by a rank 2 perturbation, see [26, Appendix]. We conclude from (3.12) and Theorem 3.7

$$\sum_{r\in\mathbb{N}} \mathbb{P}(|N_r(\lambda) - \bar{N}(\lambda)| \geq \epsilon) \leq 2\sum_{r\in\mathbb{N}} \exp\Big(-\frac{\epsilon^2|V_r|^2}{8n_r}\Big), \tag{3.13}$$

where n_r is the number of random variables entering N_r.

By construction, n_r is bounded by

$$n_r \leq |V_r||S|^{2(\rho(r)+1)} = |V_r||S|^{\frac{\ln r}{2\ln|S|}} = |V_r|\sqrt{r} \leq |V_r|^2/\sqrt{r}.$$

Use $r \leq |V_r|$ for the last step. We conclude from (3.13)

$$\sum_{r\in\mathbb{N}} \mathbb{P}(|N_r(\lambda) - \bar{N}(\lambda)| \geq \epsilon) \leq 2\sum_{r\in\mathbb{N}} \exp\big(-\epsilon^2\sqrt{r}/8\big) < \infty.$$

This is by definition almost complete convergence of $N_r(\lambda)$ to $\bar{N}(\lambda)$ and implies almost sure convergence, i.e., the existence of $\Omega_\lambda \in \mathcal{A}$ with $\mathbb{P}(\Omega_\lambda) = 1$ and

$$N_r^{(\omega)}(\lambda) \xrightarrow{r\to\infty} \bar{N}(\lambda) \qquad (\omega \in \Omega_\lambda).$$

Since $\bar{N}$ is monotone, the set of discontinuities $\mathbb{R}\backslash\mathcal{C}$ is at most countable. Let M be countable and dense subset in $\mathcal{C}$. Then, the set $\tilde{\Omega} := \bigcap_{\lambda\in M} \Omega_\lambda$ has probability 1, too. Now fix $\omega \in \tilde{\Omega}$. We know for all $\lambda \in \mathbb{R}$

$$\limsup_{r\to\infty} N_r^{(\omega)}(\lambda) \leq \inf_{\lambda'\in M\cap[\lambda,\infty)} \lim_{r\to\infty} N_r^{(\omega)}(\lambda') = \inf_{\lambda'\in M\cap[\lambda,\infty)} \bar{N}(\lambda') = \bar{N}(\lambda),$$

since $\bar{N}$ is monotone and continuous from the right, and M is dense. In the other direction, for all $\lambda \in \mathcal{C}$, we have

$$\liminf_{r\to\infty} N_r^{(\omega)}(\lambda) \geq \sup_{\lambda'\in M\cap(-\infty,\lambda]} \lim_{r\to\infty} N_r^{(\omega)}(\lambda') = \sup_{\lambda'\in M\cap(-\infty,\lambda]} \bar{N}(\lambda') = \bar{N}(\lambda).$$

Hereby, $\lim_{r\to\infty} N_r^{(\omega)}(\lambda)$ exists and equals $\bar{N}(\lambda)$ for all $\omega \in \tilde{\Omega}$ and $\lambda \in \mathcal{C}$. □

Remark 3.9. In many cases, we can allow ρ to grow much faster than permitted by (3.11). Let $m_\rho(r)$ denote the number of non-trivial random variables in $\{a(x,y) \mid x, y \in B^G_{\rho(r)}\}$. The condition

$$\sum_{r\in\mathbb{N}} q^{|V_r|/m_\rho(r)} < \infty$$

for all $q \in (0,1)$ is sufficient for (3.13) to be finite.

Assume that A has finite hopping range, i.e., there exists $R \in \mathbb{N}$ such that $a(x,y) = 0$ whenever $d(x,y) \geq R$. For instance, this is the case in the well-known Anderson model. Then $m_r \leq \binom{|S|^R}{2}$, and consequently

$$\sum_{r\in\mathbb{N}} q^{|V_r|/m_\rho(r)} \leq \sum_{r\in\mathbb{N}} q^{r/\binom{|S|^R}{2}} < \infty$$

for all $q \in (0,1)$, since $r \leq |V_r|$. Thus, for operators with finite hopping range, $\rho(r) := r/6$ suffices, as in the deterministic setting.

Complementary to Remark 3.3, Theorem 3.8 shows that the integrated density of states is deterministic. The following result sheds more light on the relation between the spectrum and the IDS.

Theorem 3.10. *The topological support of the measure ν associated to the IDS $\bar{N}$ equals the almost sure spectrum of the random operator $(A^{(\omega)})_{\omega\in\Omega}$.*

Proof. The proof is a boiled down version of [23, Corollary 5.4]. The key observation is that the measure ν is a *spectral measure* for $(A^{(\omega)})_{\omega\in\Omega}$, i. e., for any Borel set $B \in \mathcal{B}(\mathbb{R})$,

$$\nu(B) := \int_B d\bar{N} = 0 \iff E_B^{(\omega)} = 0 \quad \text{for almost all } \omega \in \Omega, \tag{3.14}$$

where $E_B^{(\omega)} := \chi_B(A^{(\omega)})$ is the spectral projection of $A^{(\omega)}$ corresponding to B. If $E_B^{(\omega)} = 0$ for almost all $\omega \in \Omega$, then, analogous to (3.8),

$$\nu(B) = \int_B d\bar{N} = \mathbb{E}[\langle \delta_{\mathrm{id}}, E_B^{(\omega)} \delta_{\mathrm{id}} \rangle] = 0.$$

In the opposite direction, we are given

$$0 = \nu(B) = \mathbb{E}[\langle \delta_{\mathrm{id}}, E_B^{(\omega)} \delta_{\mathrm{id}} \rangle] = \mathbb{E}[\langle \delta_g, E_B^{(\omega)} \delta_g \rangle]$$

for all $g \in G$. We obtain $\langle \delta_g, E_B^{(\omega)} \delta_g \rangle = 0$ for all $g \in G$ and for almost all $\omega \in \Omega$, since the integrand is never negative. The Cauchy–Schwarz inequality gives, for all $g, h \in G$ and almost all $\omega \in \Omega$,

$$|\langle \delta_g, E_B^{(\omega)} \delta_h \rangle|^2 \leq \langle \delta_g, E_B^{(\omega)} \delta_g \rangle \langle \delta_h, E_B^{(\omega)} \delta_h \rangle = 0.$$

Thus, ν is a spectral measure of $(A^{(\omega)})_\omega$.

Now, choose $B := \mathbb{R} \setminus \operatorname{spt} \nu$ to be the complement of the topological support of ν. By (3.14), $E_B^{(\omega)} = 0$ for almost all $\omega \in \Omega$. This means that $B \cap \sigma(A^{(\omega)}) = \emptyset$ for almost all $\omega \in \Omega$. So $\sigma(A^{(\omega)}) \subseteq \operatorname{spt} \nu$ for almost all $\omega \in \Omega$.

By Remark 3.3, the spectrum of the operators $A^{(\omega)}$, $\omega \in \Omega$, is almost surely equal to a non-random set Σ. We let $B := \mathbb{R} \setminus \Sigma$ be the almost sure resolvent set and conclude $\nu(B) = 0$ by (3.14). Thus, the topological support of ν does not intersect the resolvent set, and we see $\operatorname{spt} \nu \subseteq \Sigma$. $\square$

4. Examples and Applications

As an application, we show in Sect. 4.1 that the well-known percolation model on sofic groups is covered by our abstract theory. In Sect. 4.2, we describe two approaches to obtain sofic approximations for the free group.

4.1. Percolation

In this section, we apply our results to percolation models on graphs. We will study the approximability of the IDS of the corresponding Laplacian. The models in consideration will contain short-range as well as long-range percolation on sofic groups.

Let G be a finitely generated sofic group and S a finite, symmetric set of generators. Let $\Gamma_{\mathrm{co}} = (V, E_{\mathrm{co}})$ be the complete graph over the vertex set $V = G$, i.e., the edge set is

$$E_{\mathrm{co}} = \mathcal{P}_2 = \{e \subseteq G \mid |e| = 2\}.$$

Furthermore, let $p \in \ell^1(G)$ be such that

$$0 \leq p(g) = p(g^{-1}) \leq 1$$

for all $g \in G$, and consider, for distinct $g, h \in G$, independent Bernoulli random variables $X_{\{g,h\}}$ with $\mathbb{P}(X_{\{g,h\}} = 1) = p(gh^{-1})$. For each ω, we define a random subgraph $\Gamma_\omega = (V, E_\omega)$ of Γ_{co} by

$$E_\omega = \{e \in E_{\mathrm{co}} \mid X_e(\omega) = 1\}.$$

The size of the support of p distinguishes different cases. If $\mathrm{spt}(p) = S$, we deal with the (typical) percolation model of the Cayley graph $\Gamma = \Gamma(G, S)$. If $\mathrm{spt}(p)$ is a finite set, the model is referred to as *short-range percolation*. Note that under this assumption, Γ_ω is of uniformly bounded vertex degree. Otherwise, $|\mathrm{spt}(p)| = \infty$ leads to *long-range percolation*. An application of the first lemma of Borel–Cantelli shows that $p \in \ell^1(G)$ implies that Γ_ω is almost surely locally finite, see [3, Lemma 3.2]. However, there is almost surely no uniform bound for the vertex degree. Also, there exist with probability one edges of arbitrary length, measured in the word metric induced by the generating system S. This implies that the Laplacian that we define now is almost surely unbounded and not of finite hopping range.

The matrix elements of the Laplacian are given by

$$a^{(\omega)}(x, y) = \begin{cases} X_{\{x,y\}} & \text{if } x \neq y, \\ -\sum_{z \neq x} X_{\{x,z\}}(\omega) & \text{otherwise.} \end{cases} \tag{4.1}$$

Obviously, we are in the setting of (3.3). In order to apply Lemma 3.1, it remains to observe that (3.2) follows from

$$\mathbb{E}\Big[\Big(\sum_{x \in G} X_{\{\mathrm{id},x\}}\Big)^2\Big] \leq \sum_{x,y \in G} \mathbb{E}[X_{\{\mathrm{id},x\}}]\mathbb{E}[X_{\{\mathrm{id},y\}}] + \sum_{x \in G} \mathbb{E}[X_{\{\mathrm{id},x\}}] = \|p\|_1^2 + \|p\|_1.$$

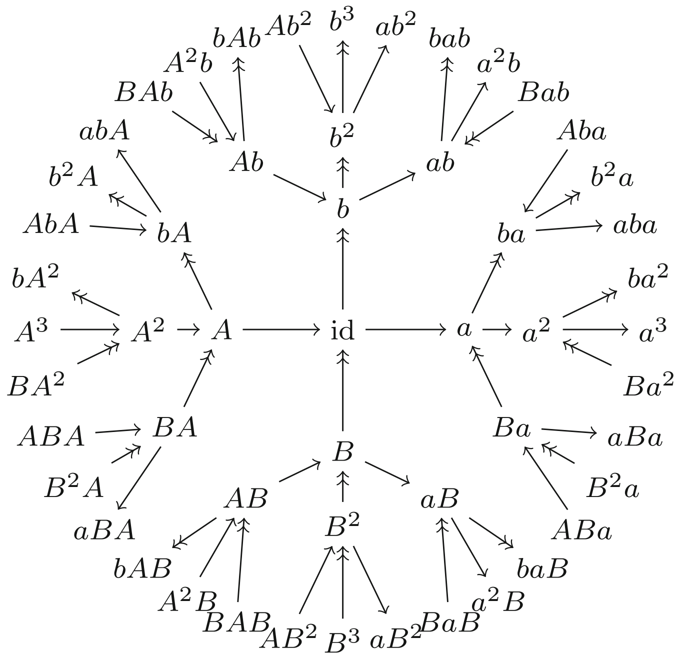


FIGURE 3. The Cayley graph of the ball B_3 of radius 3 in F_2, with $a := a_1$, $b := a_2$, $A := a^{-1}$ and $B := b^{-1}$. The *arrows* indicate the corresponding generator: $x \to y$ means $ax = y$, $x \twoheadrightarrow y$ is synonymous for $bx = y$

Therefore, there exists a random Hamiltonian $\Delta = (\Delta_\omega)_{\omega\in\Omega}$ with the properties *(i)–(v)* of Lemma 3.1 and matrix elements given by (4.1) almost surely. This operator is the Laplacian of random graph $\Gamma = (\Gamma_\omega)_{\omega\in\Omega}$. Theorem 3.8 gives almost sure weak convergence of distribution functions for almost all $\omega \in \Omega$. Of course, the same holds for Schrödinger operators.

In more restricted settings, stronger results are available. For instance, in [3,40], the authors consider long-range percolation models over amenable groups and obtain uniform convergence of the eigenvalue counting functions. However, their methods rely massively on the existence of sets with an arbitrary small boundary, which is per definition not the case for non-amenable groups.

4.2. The Free Group

The free group $F_s = \langle S_s \rangle$ with $s \in \mathbb{N}$ generators $S_s := \{a_1, \dots, a_s, a_1^{-1}, \dots, a_s^{-1}\}$, see Fig. 3, is an example of a sofic group, which is residually finite but not amenable.

We provide two explicit constructions for sofic approximations, along which the eigenvalue counting functions converge. The first approach is rather geometric and goes back to [5], while the second one uses algebraic properties of special linear groups. A relevant property of a graph is the length of the smallest circle, called *girth*. Both presented strategies construct finite Cayley graphs with arbitrarily large girth. In [20], a third approach is suggested, but it seems not as straightforward.

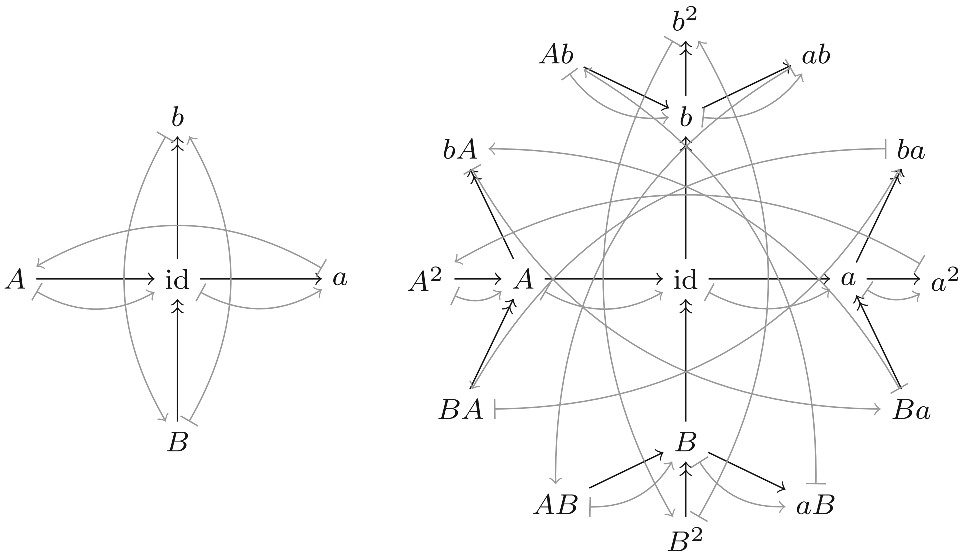


FIGURE 4. The action of the permutation $p_a^{(1)}$ on B_1 and of $p_a^{(2)}$ on B_2

4.2.1. Geometric Approach. Let B_r be the ball of radius $r \in \mathbb{N}$ in F_s centered at the identity id, with respect to the metric on the Cayley graph $\Gamma(F_s, S_s)$. A short calculation reveals

$$|B_r| = \frac{s(2s-1)^r - 1}{s-1}. \tag{4.2}$$

For each generator $x \in S_s$, we define the permutation $p_x^{(r)}$ on B_r by

$$p_x^{(r)}(w) := \begin{cases} xw & (xw \in B_r) \\ w_1^{-1}w_2^{-1}\dots w_m^{-1} & (xw \notin B_r) \end{cases}$$

for reduced words $w = w_1w_2\dots w_m \in B_r$ with $w_j \in S_s$, $j \in \{1,\dots,m\}$, see Fig. 4.

The permutation $p_x^{(r)}$ maps elements of B_{r-1} to their neighbors in direction x and members of the sphere $B_r \backslash B_{r-1}$ are sent into $B_r \backslash B_{r-2}$.

The group $H_r := \langle \{p_x^{(r)} \mid x \in S_s\}\rangle$ generated by this permutations is a subgroup of the symmetric group $\mathcal{S}_{B_r}$ on B_r.

The map $S_s \ni x \mapsto p_x^{(r)} \in \mathcal{S}_{B_r}$ has an extension to a group homomorphism $\varrho_r\colon F_s \to H_r \subseteq \mathcal{S}_{B_r}$ via $\varrho_r(w_1\dots w_m) := p_{w_1}^{(r)} \circ \dots \circ p_{w_m}^{(r)}$. We consider the normal subgroups $G_r := \ker \varrho_r$ of F_s.

Observe that a non-empty reduced word $w = w_1\dots w_m \in G_r$ is either the identity $w = \mathrm{id}$ or has at least length $m \geq 2r+1$, since the orbit of the identity id $\in B_r$ passes each sphere: $\varrho_r(w_1\dots w_j)(\mathrm{id}) \in B_j \backslash B_{j-1}$ for $j \leq r$ and $\varrho_r(w_1\dots w_j)(\mathrm{id}) \in B_r \backslash B_{2r-j}$ for $j > r$. Therefore, the girth of H_r is at least $2r+1$ and $\bigcap_r G_r = \{\mathrm{id}\}$.

We have now checked that F_s is residually finite. The corresponding sofic approximations of F_s are the Cayley graphs of H_r with respect to the generators $\{p_x^{(r)} \mid x \in S_s\}$, $r \in \mathbb{N}$. Note that since H_r is a group, all balls in H_r are isomorphic, and we set $V_r^{(0)} := H_r$, $r \in \mathbb{N}$.

The Cayley graph of (F_s, S_s) is a regular tree. For such graphs, the derivative of the spectral distribution function for the adjacency operator is explicitly calculated in [29] as

$$x \mapsto \frac{s\sqrt{4(2s-1)-x^2}}{\pi(4s^2-x^2)}\chi_{[0,2\sqrt{2s-1}]}(|x|). \tag{4.3}$$

This shows in particular that the spectral distribution function, which equals according to Theorem 2.4 the integrated density of states, is continuous. We conclude that the limit in Theorem 2.4 exists for all $\lambda \in \mathbb{R}$ and is actually uniform in λ.

A criterion for the quality of the approximation should measure the growth of the radius r in Condition (S1) relative to the growth $|V_r|$ of the approximating graphs as well as the proportion of $|V_r^{(0)}|$ in Condition (S2). In the case of the free group, we have $V_r^{(0)} = V_r$, so the growth of the girth $\gamma(H_n)$ compared with the growth of the groups H_n appears to be a reasonable choice.

The best possible situation is

$$\gamma(H_r) \geq C \ln |H_r| \tag{4.4}$$

for some $C > 0$.

This is so, because an s-regular graph with girth γ contains at least $|B_{\lfloor(\gamma-1)/2\rfloor}|$ vertices, so

$$\ln|H_r| \geq \ln|B_{\lfloor(\gamma(H_r)-1)/2\rfloor}| \approx \lfloor(\gamma(H_r)-1)/2\rfloor \ln(2s-1).$$

Equations like (4.4) have far reaching implications, see e.g. [6]. The trivial upper bound to $|H_r|$ is the number of permutations on B_r:

$$|H_r| \leq |B_r|! = \Big(\frac{s(2s-1)^r - 1}{s-1}\Big)! \ .$$

The geometric approach does not provide the optimal bound (4.4). This succeeds in the following section using an algebraic approach.

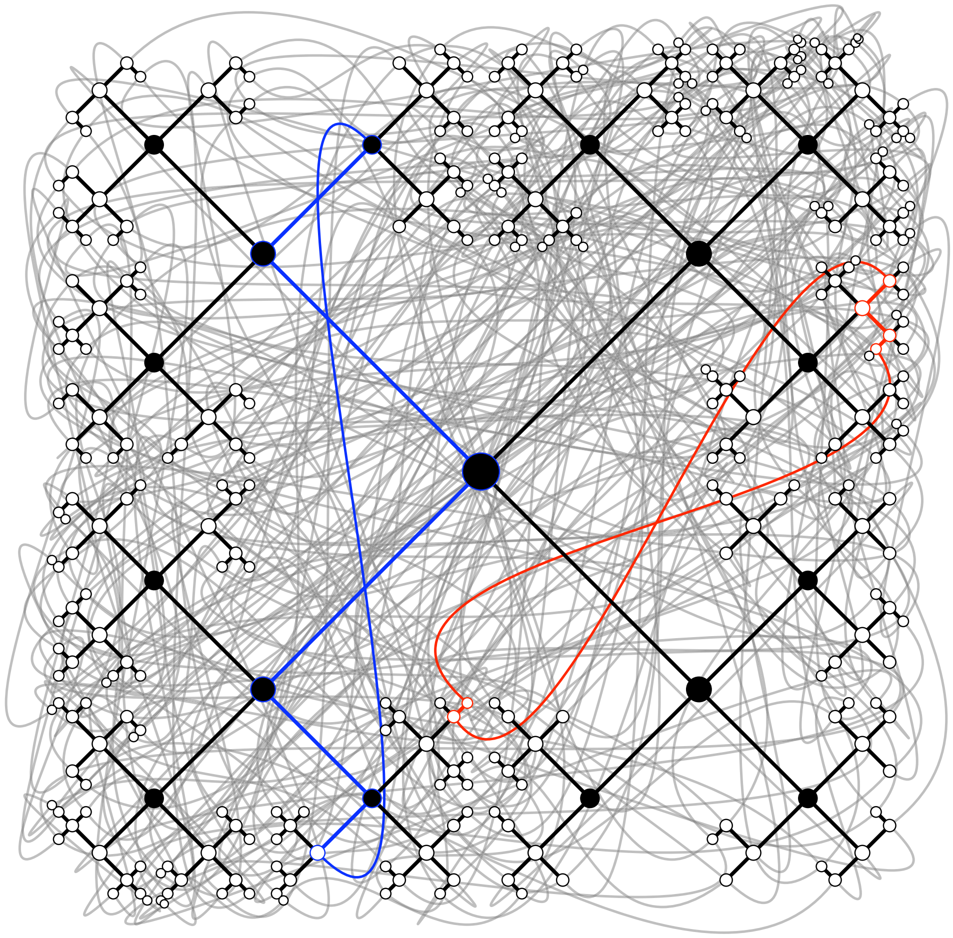
4.2.2. Algebraic Approach. According to [32, Theorem 4] and the Nielsen-Schreier theorem, the group

$$F_2 := \langle \left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right) \rangle \subseteq SL(2, \mathbb{Z})$$

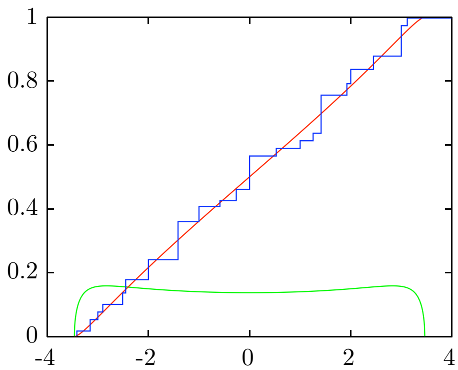
is (a faithful representation of) the free group with two generators. Note that for all such matrices, the diagonal elements are odd, while the off-diagonal entries are even. The set $S := \{\left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & -2 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ -2 & 1 \end{smallmatrix}\right)\}$ is a symmetric generator.

To see that F_2 is residually finite, we introduce the normal subgroups

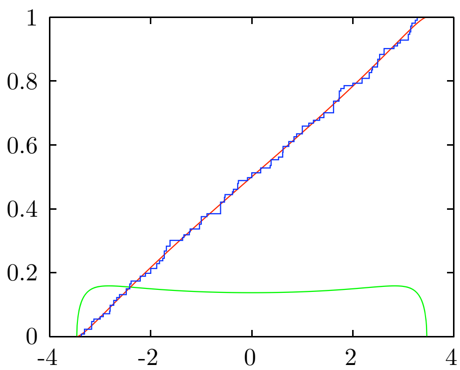
$$G_n := \{A \in F_2 \mid A \equiv \mathbf{1} \pmod{2n}\}$$



(a) The Cayley graph of H_7 has 336 nodes and girth 6. The ball of radius 2 around the identity is marked with black nodes. Each node is center of an F_2-ball of radius 2. The red and the blue lines indicate shortest circles. The high symmetry of the graph is indicated by these colored paths which are translates of each other. The nodes were found by a breadth-first search. Whenever an already found node was encountered again, a gray line was added to indicate the neighborship in the Cayley graph of H_7. H_7 is Ramanujan.



(b) $n = 7$



(c) n= 11

FIGURE 5. **a** Cayley graph of H_7, **b**, **c** numeric visualisations of the density of states (*green*), the integrated density of states (*red*) and the eigenvalue counting function of the adjacency matrix of H_n (*blue*). Colors appear only in the online version

of F_2, where $\mathbf{1} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $A \equiv \mathbf{1} \pmod{2n} \iff \exists B \in \mathrm{Mat}(2, \mathbb{Z})\colon A = 2nB + \mathbf{1}$. We further utilize the component-wise norm

$$m_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} := \max\{|a|, |b|, |c|, |d|\}.$$

Multiplication by any generator increases this norm at most by a factor 3:

$$m_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right)} = m_{\left(\begin{smallmatrix} a & 2a+b \\ c & 2c+d \end{smallmatrix}\right)} \leq 3m_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}.$$

By induction, a word $s_1 \cdots s_r$ of length $r \geq 2$, composed of generators $s_1, \dots, s_r \in S$, is bounded by

$$m_{s_1 \cdots s_r} \leq 5 \cdot 3^{r-2},$$

since $m_{s_1 s_2} \leq 5$ for all $s_1, s_2 \in S$. This means that all entries of words of length at most r are confined to $\{-5\cdot 3^{r-2}, \dots, 5\cdot 3^{r-2}\} \subseteq \{1-2\cdot 3^{r-1}, \dots, 2\cdot 3^{r-1}-1\}$. Therefore, for $2n \geq 2\cdot 3^{r-1}$, we have $B_r^{F_2} \cap G_n = \{\mathrm{id}\}$. By this, F_2 is residually finite.

The corresponding sofic approximations V_r are the Cayley graphs of

$$H_n := F_2/G_n = \{A \bmod 2n \mid A \in F_2\}$$

with respect to the generating set S, where $n \geq 3^{r-1}$. As in Sect. 4.2.1, we can choose $V_r^{(0)} := V_r$.

The size of H_n is at most $|H_n| \leq n^3$, since each matrix element is either an even or an odd number between 0 and $2n-1$, and by $\det A = 1$, the fourth entry is determined by the other three. So for $n := 3^{r-1}$, we obtain

$$|H_n| \leq 3^{3(r-1)} \leq 3^{3(\gamma(H_n)-3)/2} \leq 3^{3\gamma(H_n)/2},$$

since the girth $\gamma(H_n)$ is at least $2r+1$. Therefore, (4.4) holds with $C_1 = \frac{3}{2}\ln 3$. See Fig. 5 for numerical visualization of the eigenvalue counting function of the adjacency operator of the Cayley graphs of H_7 and H_{11} in comparison with the integrated density of states of F_2, cf. (4.3).

Acknowledgements

It is a great pleasure to thank Christian Seifert, Ivan Veselić, Marcel Hansmann, Wolfgang Spitzer, Matthias Keller, and Felix Pogorzelski for enlightening and encouraging discussions. In particular, we thank Ivan Veselić for drawing our attention to the concentration inequality by McDiarmid.

Appendix A. An Unbounded Translation Invariant Operator

We present an example of a self-adjoint, translation invariant, and unbounded operator A on $\mathbb{Z}$ for which D_0 is a core. The operator is constructed in Fourier space as multiplier $M\colon D(M) \to L^2(S^1)$ with an unbounded real-valued function $m \in L^p(S^1)$, $p \in (2, \infty)$. By this, A is self-adjoint, translation invariant, and unbounded. To prove that D_0 is a core, it suffices to show that the trigonometric polynomials $\mathcal{T}$ are a core of M. To this end, we verify

$$M \subseteq \overline{M|_{L^\infty(S^1)}}, \quad M|_{L^\infty(S^1)} \subseteq \overline{M|_{C(S^1)}} \quad \text{and} \quad M|_{C(S^1)} \subseteq \overline{M|_{\mathcal{T}}}.$$

For the third inclusion, let $f \in C(S^1)$ be given. Note that by Weierstraß' theorem, the trigonometric polynomials are dense in $C(S^1)$ with respect to supremum norm, cf. [39, Theorem 4.25]. Thus, we find a sequence (f_n) of elements in $\mathcal{T}$ with $\lim_{n\to\infty}\|f - f_n\|_\infty = 0$. This implies $\lim_{n\to\infty}\|f - f_n\|_{L^2} = 0$. Moreover, we obtain

$$\lim_{n\to\infty}\|Mf - Mf_n\|_{L^2} = \lim_{n\to\infty}\|m(f - f_n)\|_{L^2} \le \lim_{n\to\infty}\|m\|_{L^2}\|f - f_n\|_\infty = 0.$$

This proves the third inclusion.

To show the second inclusion, let $f \in L^\infty(S^1)$ and $r \in (2,\infty)$ such that $\frac{1}{2} = \frac{1}{r} + \frac{1}{p}$. We choose a sequence $f_n \in C(S^1)$ with $f_n \xrightarrow[n\to\infty]{L^r} f$. Then,

$$\|Mf_n - Mf\|_2 \le \|m\|_p\|f_n - f\|_r \xrightarrow{r\to\infty} 0.$$

Thereby, $f \in D(\overline{M|_{C(S^1)}})$ and $\overline{M|_{C(S^1)}}f = Mf$.

To prove the first inclusion, let $f \in D(M)$ be given, i.e., $f \in L^2(S^1)$ and $Mf \in L^2(S^1)$. Split

$$m = \frac{1}{m_1} + m_2 \quad \text{with} \quad m_2 := (\chi_{[0,1]} \circ |m|)\cdot m.$$

Note $m_1, m_2 \in L^\infty(S^1)$ and $\frac{1}{m_1}f \in L^2(S^1)$. Therefore, we find $g_n \in L^\infty(S^1)$ with $g_n \xrightarrow{L^2} f/m_1$. Set $f_n := m_1 g_n \in L^\infty(S^1)$. We obtain

$$f_n \xrightarrow{L^2} f \quad \text{and} \quad Mf_n = g_n + m_2 f_n \xrightarrow{L^2} \frac{1}{m_1}f + m_2 f = Mf.$$

Thus, D_0 is a core for A.

Appendix B. Weak Convergence

We thank the referee for simplifications in this proof.

Proof of Lemma 2.6. Let $N, N_1, N_2, \ldots : \mathbb{R} \to [0,1]$ be probability distribution functions. By the portmanteau theorem, *(i)* is equivalent to

$$\lim_{r\to\infty}\int f\,dN_r = \int f\,dN \quad \text{for all } f \in C_0(\mathbb{R}). \tag{B.1}$$

Thus, *(i)* implies *(ii)*. Conversely, consider the set

$$\mathcal{S} := \Big\{f : \mathbb{R} \to \mathbb{C} \;\Big|\; \lim_{r\to\infty}\int_{\mathbb{R}} f dN_r = \int_{\mathbb{R}} f dN\Big\}.$$

This is a linear space and closed with respect to $\|\cdot\|_\infty$-limits. By *(ii)*,

$$\mathcal{R} := \{\mathbb{R} \ni x \mapsto (z - x)^{-1} \mid z \in \mathbb{C}\backslash\mathbb{R}\} \subseteq \mathcal{S},$$

and since $\mathcal{S}$ is a closed linear space, it contains the $\|\cdot\|_\infty$-closure of the linear span of $\mathcal{R}$, too:

$$\overline{\operatorname{lin}\mathcal{R}} \subseteq \mathcal{S}.$$

Note now that $\mathcal{R}$ separates points and is closed under conjugation, and for any $x \in \mathbb{R}$, we find $f \in \mathcal{R}$ with $f(x) \neq 0$. Below we show that $\overline{\operatorname{lin}\mathcal{R}}$ is an algebra. These facts allow us to apply the Stone–Weierstrass theorem [12], which implies $C_0(\mathbb{R}) = \overline{\operatorname{lin}\mathcal{R}} \subseteq \mathcal{S}$, establishes (B.1), and finishes the proof.

To show that $\overline{\operatorname{lin}\mathcal{R}}$ is in fact an algebra, we first verify $\mathcal{R}^2 \subseteq \overline{\operatorname{lin}\mathcal{R}}$, i.e., that products of functions in $\mathcal{R}$ are contained in $\overline{\operatorname{lin}\mathcal{R}}$. For $f(x) = (z-x)^{-1}$ and $g(x) = (w-x)^{-1}$ with $w, z \in \mathbb{C}\backslash\mathbb{R}$, $w \neq z$, we see, by partial fraction expansion,

$$x \mapsto f(x)g(x) = \frac{1}{w-z}\Big(\frac{1}{z-x} - \frac{1}{w-x}\Big) \in \operatorname{lin}\mathcal{R} \subseteq \overline{\operatorname{lin}\mathcal{R}}.$$

For $f(x) = g(x) = (w-x)^{-1}$, $w \in \mathbb{C}\backslash\mathbb{R}$, choose a sequence $(w_n)_n$, $w_n \in \mathbb{C}\backslash\mathbb{R}$, with $w_n \neq w$ for all $n \in \mathbb{N}$, converging to w and note that

$$\sup_{x\in\mathbb{R}}\Big|\frac{1}{(w-x)^2} - \frac{1}{(w_n-x)(w-x)}\Big| \leq \frac{|w_n - w|}{|\Im w|^2|\Im w_n|} \xrightarrow{n\to\infty} 0.$$

Thus, $\mathcal{R}^2 \subseteq \overline{\operatorname{lin}\mathcal{R}}$, and consequently $(\operatorname{lin}\mathcal{R})^2 \subseteq \operatorname{lin}(\mathcal{R}^2) \subseteq \operatorname{lin}(\overline{\operatorname{lin}\mathcal{R}}) = \overline{\operatorname{lin}\mathcal{R}}$.

Now let $f, g \in \overline{\operatorname{lin}\mathcal{R}}$. By definition, we find sequences (f_n) and (g_n) in $\operatorname{lin}\mathcal{R}$ converging to f and g, respectively, with respect to $\|\cdot\|_\infty$.

We obtain

$$\begin{aligned}\|fg - f_ng_n\|_\infty &\leq \|(f-f_n)g\|_\infty + \|f_n(g-g_n)\|_\infty\\ &\leq \|f-f_n\|_\infty\|g\|_\infty + \|f_n\|_\infty\|g-g_n\|_\infty \xrightarrow{n\to\infty} 0,\end{aligned}$$

since the convergent sequence $(\|f_n\|_\infty)_n$ is bounded. For each $n \in \mathbb{N}$, we know $f_ng_n \in (\operatorname{lin}\mathcal{R})^2 \subseteq \overline{\operatorname{lin}\mathcal{R}}$, so there is a sequence $(h_k^n)_{k\in\mathbb{N}}$ in $\operatorname{lin}\mathcal{R}$ with

$$\|f_ng_n - h_k^n\|_\infty \xrightarrow{k\to\infty} 0.$$

By a diagonal argument, $fg \in \overline{\operatorname{lin}\mathcal{R}}$. More precisely, for all $\varepsilon > 0$ there are $n, k \in \mathbb{N}$ such that $\|fg - f_ng_n\|_\infty \leq \varepsilon/2$ and $\|f_ng_n - h_k^n\|_\infty \leq \varepsilon/2$. Then, we have

$$\|fg - h_k^n\|_\infty \leq \|fg - f_ng_n\|_\infty + \|f_ng_n - h_k^n\|_\infty \leq \varepsilon.$$

We conclude $(\overline{\operatorname{lin}\mathcal{R}})^2 \subseteq \overline{\operatorname{lin}\mathcal{R}}$, i.e., $\overline{\operatorname{lin}\mathcal{R}}$ is an algebra. □

Appendix C. Random Operators on Countable Groups

Let G be a countable group. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let

$$\big\{a(x,y) = \overline{a(y,x)} \colon \Omega \to \mathbb{C} \;\big|\; x, y \in G\big\}$$

be a set of random variables on $(\Omega, \mathcal{A}, \mathbb{P})$, such that for each $z \in G$

$$a(x,y) \stackrel{\mathrm{d}}{=} a(xz, yz), \tag{C.1}$$

i.e., $a(x,y)$ and $a(xz,yz)$ are identically distributed. We assume further

$$\mathbb{E}\Big[\Big(\sum_{x\in G}|a(x,\mathrm{id})|\Big)^2\Big] < \infty. \tag{C.2}$$

Lemma C.1. *For almost all $\omega \in \Omega$, the* matrix operator *$\tilde{A}^{(\omega)}$, acting on D_0 via*

$$(\tilde{A}^{(\omega)}f)(x) := \sum_{y\in G} a^{(\omega)}(x,y)f(y) \quad (x\in G), \tag{C.3}$$

is well defined. The family $\tilde{A} := (\tilde{A}^{(\omega)})_{\omega\in\Omega}$ is a random operator and satisfies for each $x \in G$

$$\mathbb{E}\big[\|\tilde{A}\delta_x\|_2^2\big] \le \mathbb{E}\big[\|\tilde{A}\delta_{\mathrm{id}}\|_1^2\big] < \infty.$$

Proof. According to [35], an operator valued function $\tilde{A}\colon \omega \mapsto \tilde{A}^{(\omega)}$ with a common core D_0 is measurable, if the functions

$$\omega \mapsto \langle v, \tilde{A}^{(\omega)}w\rangle = \lim_{r\to\infty} \sum_{x\in B_r^G(\mathrm{id})} \overline{v_x} \sum_{y\in \operatorname{spt} w} a^{(\omega)}(x,y)w_y$$

are measurable for all $v \in \ell^2(G)$ and $w \in D_0$. Since limits of sums of random variables are again measurable, $\tilde{A}$ is measurable.

From (C.1) and (C.2) follows

$$\mathbb{E}\big[\|\tilde{A}\delta_x\|_2^2\big] \le \mathbb{E}\big[\|\tilde{A}\delta_x\|_1^2\big] = \mathbb{E}\Big[\Big(\sum_{z\in G}|a(z,\mathrm{id})|\Big)^2\Big] < \infty.$$

for each $x \in G$. A direct consequence is $\mathbb{P}(\|\tilde{A}\delta_x\|_2 = \infty) = 0$ for all $x \in G$, which implies that $\tilde{A}$ is $\mathbb{P}$-a.s. well defined on D_0. □

Since our operators are up to now only defined on D_0, we need the multiplication operators $\pi_F\colon \ell^2(G) \to D_0$, $\pi_F(f) := \chi_F f$ with the indicator functions χ_F of finite sets $F \subseteq G$. The following lemma is adapted from [35, Proposition 4.1].

Lemma C.2. *Let $A, B\colon D_0 \to \ell^2(G)$ be random operators. Assume joint translation invariance of A and B and $\mathbb{E}[\|A\delta_{\mathrm{id}}\|_1], \mathbb{E}[\|B\delta_{\mathrm{id}}\|_1] < \infty$. Then, for all $x \in G$ and $r \in \mathbb{N}$,*

$$\mathbb{E}\big[\|A\pi_{B_r^G}B\delta_x\|_2^2\big] \le \|B\|_\infty^2 \mathbb{E}\big[\|A\delta_{\mathrm{id}}\|_1^2\big],$$

where $\|B\|_\infty$ is the ℓ^∞-norm of the random variable $\omega \mapsto \|B^{(\omega)}\|$ with operator norm.

We say that the operators A and B, with matrix elements $a(x,y), b(x,y)$, $x,y \in G$ respectively, are *jointly translation invariant*, if

$$\mathbb{P}\big(\big(a(x,y),b(x,y)\big)_{x,y\in F} \in E\big) = \mathbb{P}\big(\big(a(xz,yz),b(xz,yz)\big)_{x,y\in F} \in E\big)$$

for all $z \in G$, $F \subseteq G$ finite and $E \in \mathcal{B}(\mathbb{C}^F \times \mathbb{C}^F)$.

Proof. Let $a(x,y) := \langle \delta_x, A\delta_y\rangle$ and $b(x,y) := \langle \delta_x, B\delta_y\rangle$, $x,y \in G$, be the matrix elements of A and B. Then we get

$$\begin{aligned}\mathbb{E}\big[\|A\pi_{B_r^G}B\delta_x\|_2^2\big] &= \mathbb{E}\big[\langle A\pi_{B_r^G}B\delta_x, A\pi_{B_r^G}B\delta_x\rangle\big]\\ &\le \sum_{y,z\in G}\mathbb{E}\big[|\langle A\delta_y, A\delta_z\rangle b(y,x)b(z,x)|\big]\\ &= \sum\nolimits_{y,z}\mathbb{E}\big[|\langle A\delta_{yz^{-1}}, A\delta_{\mathrm{id}}\rangle b(yz^{-1}, xz^{-1})b(\mathrm{id}, xz^{-1})|\big],\end{aligned}$$

where we used joint translation invariance of A and B. We re-index the sums to rearrange the terms and use the Cauchy–Schwarz inequality:

$$\begin{aligned}\ldots &= \sum\nolimits_{y',z}\mathbb{E}\big[|\langle A\delta_{y'}, A\delta_{\mathrm{id}}\rangle b(y', xz^{-1})b(\mathrm{id}, xz^{-1})|\big]\\ &= \sum\nolimits_{y'}\mathbb{E}\big[|\langle A\delta_{y'}, A\delta_{\mathrm{id}}\rangle|\sum\nolimits_{z'}|b(y',z')b(\mathrm{id},z')|\big]\\ &\le \sum\nolimits_{y'}\mathbb{E}\big[|\langle A\delta_{y'}, A\delta_{\mathrm{id}}\rangle|\|B\delta_{y'}\|_2\|B\delta_{\mathrm{id}}\|_2\big]\\ &\le \|B\|_\infty^2\sum\nolimits_{y'}\mathbb{E}\big[|\langle A\delta_{y'}, A\delta_{\mathrm{id}}\rangle|\big].\end{aligned}$$

Now, as B is out of the way, we expand the scalar product and use translation invariance again and re-index the sums once more:

$$\begin{aligned}\ldots &= \|B\|_\infty^2\sum\nolimits_{y'}\mathbb{E}\Big[\Big|\sum\nolimits_z a(z,y')a(z,\mathrm{id})\Big|\Big]\\ &= \|B\|_\infty^2\sum\nolimits_{y'}\mathbb{E}\Big[\Big|\sum\nolimits_z a(\mathrm{id}, y'z^{-1})a(\mathrm{id}, z^{-1})\Big|\Big]\\ &\le \|B\|_\infty^2\sum\nolimits_{y',z}\mathbb{E}\big[|a(\mathrm{id}, y'z^{-1})a(\mathrm{id}, z^{-1})|\big]\\ &= \|B\|_\infty^2\mathbb{E}\Big[\sum\nolimits_{y''}|a(\mathrm{id}, y'')|\sum\nolimits_{z'}|a(\mathrm{id}, z')|\Big]\\ &= \|B\|_\infty^2\mathbb{E}\Big[\Big(\sum\nolimits_y|a(\mathrm{id}, y)|\Big)^2\Big] = \|B\|_\infty^2\mathbb{E}\big[\|A\delta_{\mathrm{id}}\|_1^2\big].\end{aligned}$$

□

Proposition C.3. *There exists $\tilde\Omega \in \mathcal{A}$ of full measure such that for all $\omega \in \tilde\Omega$ the operator $\tilde A^{(\omega)}$ defined in (C.3) is essentially self-adjoint. In particular, D_0 is a core for $\tilde A^{(\omega)}$.*

Proof. The proof of [35, (4.2) Theorem] for essential self-adjointness generalizes to our more general setting.

The operators are symmetric. By the basic criterion for self-adjointness, [38, Theorem VIII.3, p. 256f], for all $z \in \mathbb{C}\backslash\mathbb{R}$, we have to show that $(z - \tilde A^{(\omega)})D_0$ is dense in $\ell^2(G)$ with probability 1. For this, it suffices to approximate δ_g for arbitrary $g \in G$.

Define for $r \in \mathbb{N}$ the bounded random matrix operator $\tilde A_r^{(\omega)}$, acting on D_0, via its matrix elements

$$\tilde a_r^{(\omega)}(x,y) := \chi_{[0,r]}\big(|a^{(\omega)}(x,y)|\big)\chi_{B_r^G(x)}(y)a^{(\omega)}(x,y).$$

Using Lebesgue's dominated convergence theorem, continuity of squaring and monotone convergence, we see that the operators $\tilde{A}_r^{(\omega)}$ approximate $\tilde{A}^{(\omega)}$ in the following sense:

$$\lim_{r\to\infty} \mathbb{E}\big[\|(\tilde{A}-\tilde{A}_r)\delta_{\mathrm{id}}\|_1^2\big] = \lim_{r\to\infty} \mathbb{E}\Big[\Big(\sum_{h\in G}|a(h,\mathrm{id})-\tilde{a}_r(h,\mathrm{id})|\Big)^2\Big]$$
$$= \mathbb{E}\Big[\Big(\sum_{h\in G}\lim_{r\to\infty}|a(h,\mathrm{id})-\tilde{a}_r(h,\mathrm{id})|\Big)^2\Big] = 0,$$

since

$$\sum_{g\in G}|a^{(\omega)}(g,\mathrm{id})-\tilde{a}_r^{(\omega)}(g,\mathrm{id})| \le 2\sum_{g\in G}|a^{(\omega)}(g,\mathrm{id})| = 2\|\tilde{A}^{(\omega)}\delta_{\mathrm{id}}\|_1$$

uniformly in $r\in\mathbb{N}$ and $\mathbb{E}\big[\|2\tilde{A}\delta_{\mathrm{id}}\|_1^2\big]<\infty$. Consider for $r,n\in\mathbb{N}$ the element $f_{g,r,n}^{(\omega)} := \pi_{B_n^G(\mathrm{id})}(z-\tilde{A}_r^{(\omega)})^{-1}\delta_g \in D_0$. We will show that its image under $z-\tilde{A}^{(\omega)}$ converges to δ_g for a suitable limit in n and r. Therefore, we estimate

$$\begin{aligned}
&\|(z-\tilde{A}^{(\omega)})f_{g,r,n}^{(\omega)}-\delta_g\|_2\\
&\quad= \big\|(z-\tilde{A}^{(\omega)})\pi_{B_n^G}(z-\tilde{A}_r^{(\omega)})^{-1}\delta_g\\
&\qquad -(z-\tilde{A}_r^{(\omega)})(\pi_{B_n^G}+\pi_{G\setminus B_n^G})(z-\tilde{A}_r^{(\omega)})^{-1}\delta_g\big\|_2\\
&\quad\le \|(\tilde{A}_r^{(\omega)}-\tilde{A}^{(\omega)})f_{g,r,n}^{(\omega)}\|_2 + C(r)\|\pi_{G\setminus B_n^G}(z-\tilde{A}_r^{(\omega)})^{-1}\delta_g\|_2 \quad \text{a.s.}
\end{aligned}\tag{C.4}$$

with $C(r):=\sup_{\omega\in\Omega}\|z-\tilde{A}_r^{(\omega)}\|_2<\infty$. Note $\lim_{n\to\infty}\|\pi_{G\setminus B_n^G}(z-\tilde{A}_r^{(\omega)})^{-1}\delta_g\|_2 = 0$ and

$$\|\pi_{G\setminus B_n^G}(z-\tilde{A}_r^{(\omega)})^{-1}\delta_g\|_2 \le |\Im z|^{-1}$$

uniformly in $n\in\mathbb{N}$. By Lebesgue, there exists for all $r>0$ an $\tilde{n}=\tilde{n}(r,g)\in\mathbb{N}$ such that

$$C(r)\mathbb{E}\big[\|\pi_{G\setminus B_{\tilde{n}}^G}(z-\tilde{A}_r)^{-1}\delta_g\|_2\big] \le 1/r. \tag{C.5}$$

Also, in expectation, the first summand in (C.4) can be controlled by Lemma C.2 and $\|(z-\tilde{A}_r^{(\omega)})^{-1}\|\le|\Im z|^{-1}$ with the bound

$$\begin{aligned}
\big(\mathbb{E}[\|(\tilde{A}_r-\tilde{A})f_{g,r,\tilde{n}}\|_2]\big)^2 &\le \mathbb{E}\big[\|(\tilde{A}_r-\tilde{A})f_{g,r,\tilde{n}}\|_2^2\big]\\
&\le \|(z-\tilde{A}_r)^{-1}\|_\infty^2\mathbb{E}\big[\|(\tilde{A}_r-\tilde{A})\delta_{\mathrm{id}}\|_1^2\big]\\
&\le |\Im z|^{-2}\mathbb{E}\big[\|(\tilde{A}_r-\tilde{A})\delta_{\mathrm{id}}\|_1^2\big] \xrightarrow{r\to\infty} 0.
\end{aligned}$$

Together with (C.5), we infer from (C.4)

$$\mathbb{E}\big[\|(z-\tilde{A})f_{g,r,\tilde{n}}-\delta_g\|_2\big] \xrightarrow{r\to\infty} 0,$$

i.e., convergence in L^1. Thereby we find a subsequence $(r_k)_{k\in\mathbb{N}}$ and a set $\tilde{\Omega}\subseteq\Omega$ of full measure, such that for all $\omega\in\tilde{\Omega}$

$$\lim_{k\to\infty}\|(z-\tilde{A}^{(\omega)})f_{g,r_k,\tilde{n}(r_k,g)}^{(\omega)}-\delta_g\|_2 = 0,$$

and essential self-adjointness of $\tilde{A}^{(\omega)}$ is shown. $\square$

Given the set $\tilde{\Omega}$ from PropositionC.3, denote for all $\omega \in \tilde{\Omega}$ the self-adjoint extension of $\tilde{A}^{(\omega)}$ by $\bar{A}^{(\omega)}$ and define the random operator $A = (A^{(\omega)})_{\omega\in\Omega}$ by

$$A^{(\omega)} := \begin{cases} \bar{A}^{(\omega)} & \text{for } \omega \in \tilde{\Omega} \\ \mathrm{I}d & \text{for } \omega \in \Omega\backslash\tilde{\Omega}. \end{cases}$$

For the sake of completeness, we state the following well-known Corollary.

Corollary C.4. *The resolvents $\omega \mapsto (z - A^{(\omega)})^{-1}$, $z \in \mathbb{C}\backslash\mathbb{R}$, are strongly measurable.*

Proof. Fix $z \in \mathbb{C}\backslash\mathbb{R}$ and denote for $\omega \in \Omega$ by $A_r^{(\omega)} \colon \ell^2(G) \to \ell^2(G)$ the self-adjoint operator with matrix elements

$$a_r^{(\omega)}(x,y) := \chi_{[0,r]}\big(|\langle\delta_x, A^{(\omega)}\delta_y\rangle|\big)\chi_{B_r^G(\mathrm{id})}(x)\chi_{B_r^G(\mathrm{id})}(y)\langle\delta_x, A^{(\omega)}\delta_y\rangle.$$

Note that this is operator is not translation invariant in distribution but has only finitely many nonzero matrix elements.

By Cramer's rule, the resolvent $\omega \mapsto (z - A_r^{(\omega)})^{-1}$ is weakly measurable. Since $\ell^2(G)$ is separable, by Pettis' measurability theorem, the resolvent is in fact strongly measurable. We will now show that the resolvents of A_r converge strongly to the corresponding resolvents of A. This will show measurability of A.

Since $A^{(\omega)}$ and $A_r^{(\omega)}$ are self-adjoint, $\|(z - A^{(\omega)})^{-1}\| \leq 1/|\Im z|$ and analogously for $A_r^{(\omega)}$. Therefore, fix some $\omega \in \Omega$, $\xi \in \ell^2(G)$ and $\kappa > 0$. By Lemma 2.5, there exists $\psi \in \ell^2(G)$ with

$$\|\xi - \psi\|_2 < \kappa \quad \text{and} \quad \phi := (z - A^{(\omega)})^{-1}\psi \in D_0.$$

Thus, by the second resolvent identity,

$$\begin{aligned} &\big\|\big((z - A^{(\omega)})^{-1} - (z - A_r^{(\omega)})^{-1}\big)\xi\big\|_2 \\ &\quad\leq \big\|\big((z - A^{(\omega)})^{-1} - (z - A_r^{(\omega)})^{-1}\big)\psi\big\|_2 + 2\|\xi - \psi\|_2/|\Im z| \\ &\quad\leq \|(z - A_r^{(\omega)})^{-1}(A^{(\omega)} - A_r^{(\omega)})(z - A^{(\omega)})^{-1}\psi\|_2 + 2\kappa/|\Im z| \\ &\quad\leq \big(\|(A^{(\omega)} - A_r^{(\omega)})\phi\|_2 + 2\kappa\big)/|\Im z|. \end{aligned}$$

Now, we use that for all $x, y \in G$

$$a_r^{(\omega)}(x,y) \xrightarrow{r\to\infty} \langle\delta_x, A^{(\omega)}\delta_y\rangle$$

to obtain, employing dominated convergence,

$$\limsup_{r\to\infty}\big\|\big((z - A^{(\omega)})^{-1} - (z - A_r^{(\omega)})^{-1}\big)\xi\big\|_2 \leq 2\kappa/|\Im z|.$$

As $\kappa > 0$ was arbitrary, this concludes the proof. $\square$

Appendix D. Conditions for (Un-)Boundedness

Proof of Lemma 3.4. Let $D=\infty$ and let $m>0$ be given. Note that condition (3.2) implies that

$$k := \mathbb{E}\Big[\sum_{z\neq \mathrm{id}} |X_{\{\mathrm{id},z\}}|\Big] < \infty.$$

Without loss of generality, we assume $m \geq 2k|\alpha|$. As D is assumed to be infinite, there exists $z\in G$ such that

$$\|X_{\{\mathrm{id},z\}}\|_\infty \geq m. \tag{D.1}$$

Let us distinguish two case. In the first case, we consider the situation where there exists $z\in G\backslash\{\mathrm{id}\}$ satisfying (D.1). Then obviously the probability $\mathbb{P}(a(\mathrm{id},z)\geq m)$ is strictly positive. In the case where there exists no $z\in G\backslash\{\mathrm{id}\}$ satisfying (D.1) the same holds true, however, we need a short calculation to see this. In this situation, we have $\|X_{\{\mathrm{id}\}}\|_\infty=\infty$. By definition of $a(\mathrm{id},\mathrm{id})$, we have by triangle inequality

$$\begin{aligned}
\mathbb{P}(|a(\mathrm{id},\mathrm{id})|\geq m) &\geq \mathbb{P}\Bigg(|X_{\{\mathrm{id}\}}| - \Big|\alpha \sum_{z\in G\backslash\{\mathrm{id}\}} X_{\{\mathrm{id},z\}}\Big| \geq m\Bigg)\\
&\geq \mathbb{P}\Bigg(|X_{\{\mathrm{id}\}}| \geq 2m,\ \Big|\alpha \sum_{z\in G\backslash\{\mathrm{id}\}} X_{\{\mathrm{id},z\}}\Big| \leq m\Bigg)\\
&= \mathbb{P}\big(|X_{\{\mathrm{id}\}}|\geq 2m\big)\mathbb{P}\Bigg(\Big|\alpha \sum_{z\in G\backslash\{\mathrm{id}\}} X_{\{\mathrm{id},z\}}\Big| \leq m\Bigg).
\end{aligned}$$

As $\|X_{\{\mathrm{id}\}}\|_\infty=\infty$ we get $\mathbb{P}(|X_{\{\mathrm{id}\}}|\geq 2m)>0$. We use the Tschebyscheff inequality to obtain

$$\mathbb{P}\Bigg(\Big|\alpha \sum_{z\in G\backslash\{\mathrm{id}\}} X_{\{\mathrm{id},z\}}\Big| \leq m\Bigg) \geq 1-\frac{|\alpha|}{m}\mathbb{E}\Bigg[\sum_{z\in G\backslash\{\mathrm{id}\}} X_{\{\mathrm{id},z\}}\Bigg] \geq \frac{1}{2}.$$

This gives $\mathbb{P}(|a(\mathrm{id},\mathrm{id})|\geq m)>0$. Together with the previous case, we showed that whenever $D=\infty$, there exists $z\in G$ such that $\mathbb{P}(|a(\mathrm{id},z)|>m)$ is positive. Furthermore, by construction we have that the random variables $a(x,zx)$, $x\in G$, are independent and identically distributed, such that we get

$$\sum_{x\in G}\mathbb{P}(|a(x,zx)|>m)=\infty.$$

Now, Borel-Cantelli gives that for almost all $\omega\in\Omega$ there are infinitely many $x\in G$ such that $|a^{(\omega)}(x,zx)|>m$. For each such ω, we choose one of these x and obtain $(A^{(\omega)}\delta_{zx})(x)=a^{(\omega)}(x,zx)$. Hence,

$$\|A^{(\omega)}\| \geq \|A^{(\omega)}\delta_{zx}\|_2 \geq m.$$

Let $D < \infty$ and A be of finite hopping range R. We set $m := (1 + |\alpha||B_R|)D$. Then, we have

$$\mathbb{P}(\exists x, y \in G \text{ with } a(x,y) \geq m) = \mathbb{P}\Big(\bigcup_{x,y \in G} \{\omega \in \Omega \mid a(x,y) \geq m\}\Big)$$
$$\leq \sum_{x,y \in G} \mathbb{P}\big(\{\omega \in \Omega \mid a(x,y) \geq m\}\big) = 0.$$

Using this we get for $f \in \ell^2(G)$ and almost all realizations $\omega \in \Omega$

$$\|A^{(\omega)} f\|_2^2 = \sum_{v \in G} \Big| \sum_{w \in B_R^G(v)} a^{(\omega)}(v,w) f(w) \Big|^2 \leq \sum_{v \in G} m^2 \Big| \sum_{w \in B_R^G(v)} |f(w)| \Big|^2$$
$$\leq \sum_{v \in G} m^2 |B_R^G| \sum_{w \in B_R^G(v)} |f(w)|^2 \leq m^2 |B_R^G|^2 \|f\|_2^2.$$

□

References

[1] Aliprantis, C.D., Border, K.C.: Infinite Dimensional Analysis: A hitchhiker's Guide, 3rd edn.. Springer, Berlin (2006)

[2] Adachi, T., Sunada, T.: Density of states in spectral geometry. Comment. Math. Helv. **68**(1), 480–493 (1993)

[3] Ayadi, S., Schwarzenberger, F., Veselić, I.: Uniform existence of the integrated density of states for randomly weighted Hamiltonians on long-range percolation graphs. Math. Phys. Anal. Geom (2012) online first. doi:10.1007/s11040-013-9133-2

[4] Aizenman, M., Warzel, S.: The canopy graph and level statistics for random operators on trees. Math. Phys. Anal. Geom. **9**(4), 291–333 (2008)

[5] Biggs, N.L.: Girth and residual finiteness. Combinatorica **8**(4) 307–312 (1988). MR 981888 (90c:05105)

[6] Brooks, S., Lindenstrauss, E.: Non-localization of eigenfunctions on large regular graphs. Israel J. Math. **193**(1), 1–14 (2013)

[7] Bowen, L.: Measure conjugacy invariants for actions of countable sofic groups. J. Am. Math. Soc. **23**(1), 217–245 (2010)

[8] Bowen, L.: Sofic entropy and amenable groups. Ergod. Theor. Dyn. Syst. **32**, 427–466 (2012)

[9] Bartholdi, L., Woess, W.: Spectral computations on lamplighter groups and Diestel-Leader graphs. J. Fourier Anal. Appl. **11**(2), 175–202 (2005). MR 2131635 (2006e:20052)

[10] Cycon, H., Froese, R., Kirsch, W., Simon, B.: Schrödinger Operators, 3rd edn. Springer, Berlin (2008)

[11] Cornulier, Y.: A sofic group away from amenable groups. Math. Ann. **350**, 269–275 (2011)

[12] de Branges, L.: The Stone-Weierstrass theorem. Proc. Amer. Math. Soc. **10**, 822–824 (1959). MR 0113131 (22 #3970)

[13] Dodziuk, J., Linnell, P., Mathai, V., Schick, T., Yates, S.: Approximating L^2-invariants, and the Atiyah conjecture. Commun. Pur. Appl. Math. **56**(7), 839–873 (2003)

[14] Dodziuk, J., Mathai, V.: Approximating l^2 invariants of amenable covering spaces: A heat kernel approach. In: Contemporary Mathematics, vol. 211, pp. 151–167. AMS (1997)

[15] Dodziuk, J., Mathai, V.: Approximating L^2 invariants of amenable covering spaces: a combinatorial approach. J. Funct. Anal. **154**(2), 359–378 (1998)

[16] Eckmann, B.: Approximating ℓ^2-Betti numbers of an amenable covering by ordinary Betti numbers. Comment. Math. Helv. **74**(1), 150–155 (1999)

[17] Elek, G., Szabó, E.: On sofic groups. J. Group Theory **9**(2), 161–171 (2006)

[18] Froese, R., Hasler, D., Spitzer, W.: A geometric approach to absolutely continuous spectrum for discrete Schrödinger operators. Prog Probab **64**, 201–226 (2011)

[19] Gromov, M.: Endomorphisms of symbolic algebraic varieties. J. Eur. Math. Soc. **1**(2), 109–197 (1999)

[20] Komarov, N., McNeill, R.T., Webster, J.T.: Normal subgroups of the free group. In: Proceedings of the Oregon State University. REU in Mathematics, pp. 61–80 (2007)

[21] Lenz, D., Müller, P., Veselić, I.: Uniform existence of the integrated density of states for models on $\mathbb{Z}^d$. Positivity **12**(4), 571–589 (2008)

[22] Lenz, D., Peyerimhoff, N., Veselić, I.: Integrated density of states for random metrics on manifolds. Proc. London Math. Soc. **88**(3), 733–752 (2004)

[23] Lenz, D., Peyerimhoff, N., Veselić, I.: Groupoids, von neumann algebras, and the integrated density of states. Math. Phys. Anal. Geom. **10**(1), 1–41 (2007)

[24] Lück, W., Schick, T.: L^2-torsion of hyperbolic manifolds of finite volume. Geom. Funct. Anal. **9**(3), 518–567 (1999). MR 1708444 (2000e:58050)

[25] Lenz, D., Stollmann, P.: An ergodic theorem for Delone dynamical systems and existence of the density of states. J. Anal. Math. **97**(1), 1–23 (2005)

[26] Lenz, D., Schwarzenberger, F., Veselić, I.: A Banach space-valued ergodic theorem and the uniform approximation of the integrated density of states. Geom. Dedicata **150**, 1–34 (2011). MR 2753695 (2012c:22010).

[27] Lück, W.: Approximating l^2-invariants by their finite-dimensional analogues. Geom. Funct. Anal. **4**(4), 455–481 (1994)

[28] McDiarmid, C.: Concentration. In: Probabilistic Methods for Algorithmic Discrete Mathematics, pp. 1–46 (1998)

[29] McKay, B.D.: The expected eigenvalue distribution of a large regular graph. Linear Algebra Appl. **40**, 203–216 (1981)

[30] Mathai, V., Schick, T., Yates, S.: Approximating spectral invariants of Harper operators on graphs II. Proc. Amer. Math. Soc. **131**(6), 1917–1923 (2003)

[31] Mathai, V., Yates, S.: Approximating spectral invariants of Harper operators on graphs. J. Funct. Anal. **188**(1), 111–136 (2002)

[32] Newman, M.: Free subgroups and normal subgroups of the modular group. Am. J. Math. **86**(1), 262–265 (1964)

[33] Pastur, L.A.: Selfaverageability of the number of states of the Schrödinger equation with a random potential. Mat. Fiz. i Funkcional. Anal. **238**(2), 111–116 (1971)

[34] Pestov, V.: Hyperliner and sofic groups: a brief guide. B. Symb. Log. **14**(4), 449–480 (2008)

[35] Pastur, L., Figotin, A.: Spectra of random and almost-periodic operators. In: Grundlehren der mathematischen Wissenschaften, vol. 297. Springer, Berlin (1992)

[36] Pogorzelski, F., Schwarzenberger, F.: A banach space-valued ergodic theorem for amenable groups and applications. J. Anal. Math. 55 (2014, to appear)

[37] Peyerimhoff, N., Veselić, I.: Integrated density of states for ergodic random Schrödinger operators on manifolds. Geometriae Dedicata **91**(1), 117–135 (2002)

[38] Reed, M., Simon, B.: Functional Analysis, vol. 1. Academic Press Inc, New York (1930)

[39] Rudin, W.: Real and Complex Analysis, 3 edn.. McGraw-Hill Book Co, Singapore (1987)

[40] Schwarzenberger, F.: Uniform approximation of the integrated density of states for long-range percolation Hamiltonians. J. Stat. Phys. **146**(6), 1156–1183 (2012)

[41] Schwarzenberger, F.: The integrated density of states for operators on groups. Ph.D. thesis, Chemnitz University of Technology (2013). http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-123241

[42] Shubin, M.A.: Spectral theory and the index of elliptic operators with almost-periodic coefficients. Russ. Math. Surv. **34**(2), 109–157 (1979)

[43] Szabó, E.: Soficgroups and direct finiteness. J. Algebra **280**(2), 426–434 (2004)

[44] Sznitman, A.-S.: Lifschitz tail and Wiener sausage on hyperbolic space. Commun. Pure Appl. Math. **42**(8), 1033–1065 (1989)

[45] Sznitman, A.-S.: Lifschitz tail on hyperbolic space: Neumann conditions. Commun. Pure Appl. Math. **43**(1), 1–30 (1990)

[46] Thom, A.: Sofic groups and diophantine approximation. Commun. Pure Appl. Math. **511**, 1155–1171 (2008)

[47] Veselić, I.: Spectral analysis of percolation Hamiltonians. Math. Ann. **331**(4), 841–865 (2005)

[48] Weiss, B.: Sofic groups and dynamical systems. Sankhyā Ser. A **62**(3), 350–359 (2000). Ergodic theory and harmonic analysis (Mumbai, 1999). MR 1803462 (2001j:37022)

Christoph Schumacher and Fabian Schwarzenberger
Fakultät für Mathematik
TU Chemnitz, Chemnitz, Germany
e-mail: `Christoph.Schumacher@mathematik.tu-chemnitz.de`;
`fschwarzenberger@web.de`

Communicated by Claude-Alain Pillet.

Received: November 4, 2013.

Accepted: March 21, 2014.

Chapter 7

The Anderson Model on the Bethe Lattice: Lifshitz tails

THE ANDERSON MODEL ON THE BETHE LATTICE: LIFSHITZ TAILS

FRANCISCO HOECKER-ESCUTI AND CHRISTOPH SCHUMACHER

ABSTRACT. This paper is devoted to the study of the (discrete) Anderson Hamiltonian on the Bethe lattice, which is an infinite tree with constant vertex degree. The Hamiltonian we study corresponds to the sum of the graph Laplacian and a diagonal operator with non-negative bounded, i. i. d. random coefficients on its diagonal. We study in particular the asymptotic behavior of the integrated density of states near the bottom of the spectrum. More precisely, under a natural condition on the random variables, we prove the conjectured double-exponential Lifshitz tail with exponent 1/2. The result is a consequence of some estimates on the Laplace transform of the density of states, which is also related to the solution of the parabolic Anderson problem on the tree. These estimates are linked to the asymptotic behavior of the ground state energy of the Anderson Hamiltonian restricted to trees of finite length. The proofs make use of Tauberian theorems, a discrete Feynman–Kac formula, a discrete IMS localization formula, the spectral theory of the free Laplacian on finite rooted trees, an uncertainty principle for low-energy states, an epsilon-net argument and standard concentration inequalities.

CONTENTS

1. Introduction

In this paper we are interested in a tight-binding, one-body Hamiltonian of a disordered alloy. This Hamiltonian is known as the Anderson model, and it was introduced in its most simple form by the American physicist Phillip W. Anderson in 1958 [And58]. Given the extensive mathematical and physical literature on the subject, see e. g. [Abr10, and references therein], we defer the discussion and review of the literature until after the rigorous statement of our results.

The underlying physical space of our model is assumed to be a *Bethe lattice*, this is, an infinite regular graph with no loops and constant coordination number $k + 1$ (see fig. 1). The Anderson model in this setting was introduced very early by Abou-Chacra, Thouless and Anderson in [ATA73]. A number of physical and numerical (e. g. [KH85; MF91; MD94; BAF04; AF05; MG09]) as well as rigorous mathematical works (e. g. [KS83; Aiz94; AK92; Kle96; Kle98; ASW06; AW11b; AW11a; War13]) in this setting have been since published.

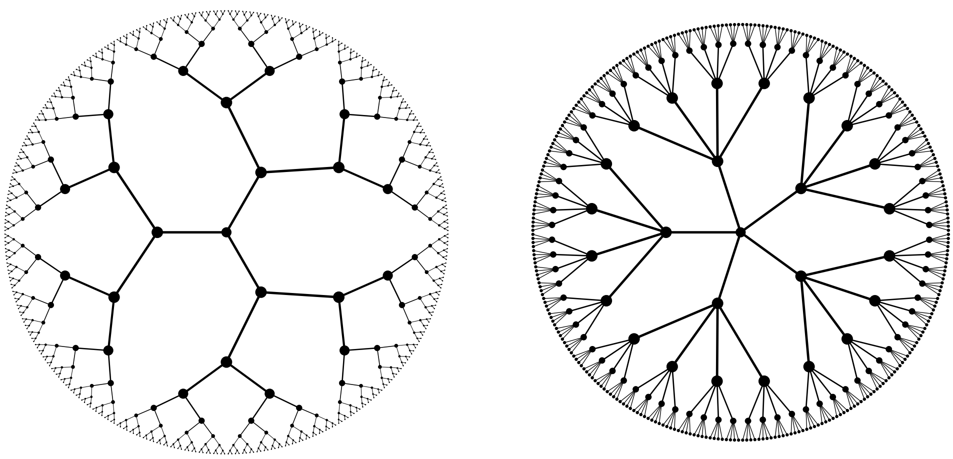


Figure 1. The Bethe lattices with coordination numbers $k + 1 \in \{3, 5\}$.

The study of transport properties of disordered models leads to the spectral theory of random Schrödinger operators. The prototypical example of these operators is the Anderson Hamiltonian. An important quantity of study is the *integrated density of states*, which is a function we denote by $\mathcal{N}$. The numerical value $\mathcal{N}(E)$ counts the available energy levels below the energy E per unit volume. Under very general assumptions, the support of the derivative of this function coincides with the spectrum of the Hamiltonian in consideration. The study of its asymptotic behavior when we approach the bottom of the spectrum E_0 has attracted a lot of attention since Lifshitz' remark [Lif65]. The physicist noted that in presence of disorder this asymptotic behavior is drastically different from the one of the free operator. Indeed, as soon as the disorder is non-trivial, this function exhibits a very fast decay at the bottom of the spectrum E_0. This behavior has also drawn the attention of many mathematicians, as it can be used as one of the main ingredients of the rigorous proofs of the occurrence of *Anderson localization*. In the setting of our paper, it was conjectured, see [KH85; BST10; BS11], that the integrated density of states exhibits a double exponential decay with exponent 1/2, i. e. that for some suitable $\epsilon > 0$

$$\exp\left(-\mathrm{e}^{\epsilon^{-1}(E-E_0)^{-1/2}}\right) \leqslant \mathcal{N}(E) \leqslant \exp\left(-\mathrm{e}^{\epsilon(E-E_0)^{-1/2}}\right) \quad \text{for } E \in (E_0, E_0+\epsilon). \tag{1.1}$$

In the literature (e. g. [War13, eq. 5]) one also finds this written in the somewhat weaker form

$$\lim_{E \searrow E_0} \frac{\log\log|\log \mathcal{N}(E)|}{\log(E-E_0)} = -\frac{1}{2}.$$

The purpose of this paper is to prove this conjecture. To do so, we study the Laplace transform $t \mapsto \tilde{\mathcal{N}}(t)$ of the measure $\mathrm{d}\mathcal{N}(\,\cdot\,)$ and we establish asymptotic bounds for large t. We will see that for a suitable $\epsilon > 0$ and all t large enough

$$\exp\left(-t\left(E_0 + \frac{\epsilon^{-1}}{(\log t)^2}\right)\right) \leqslant \tilde{\mathcal{N}}(t) \leqslant \exp\left(-t\left(E_0 + \frac{\epsilon}{(\log t)^2}\right)\right). \tag{1.2}$$

These bounds are of independent interest, as they are related to the long-time behavior of the so called *parabolic Anderson problem* in the annealed regime. This long-time behavior is in turn related to the location of the ground state energy of suitable finite-dimensional approximations of the Anderson Hamiltonian. We discuss this circle of ideas, which is well known in the literature, after stating rigorously our results.

Most of the novelty of this work lies in the proof of the bounds on the ground state energy E^L_{GS} of the Hamiltonian restricted to finite symmetric rooted trees $\mathcal{T}^L$ of length L (see fig. 2, where $L = \infty$). In absence of disorder, it behaves as

$$E_0 + CL^{-2}, \tag{1.3}$$

up to smaller terms. In presence of disorder, one expects heuristically that the ground state of the disordered Hamiltonian restricted to $\mathcal{T}^L$ lives in some smaller subtree of length $r = C' \log L$ on which the random potential is essentially zero. Hence, with good probability we should have

$$E^L_{\mathit{GS}} = E_0 + C''(\log L)^{-2}, \tag{1.4}$$

which is the order of the ground state energy of the free operator restricted to this subtree. The length scale $\log L$ appears naturally as one balances out the probability that the random potential is small in a subtree, which is exponential in the number of random variables (we find about k^r of them in a subtree of length r), and the number of trees of length r (there are about $k^{L-r} \leqslant k^L$) in

Boole's inequality. As for the shape of this region, symmetry considerations might suggest that the ground state localizes to a ball. As it turns out, symmetric rooted trees provide a tractable, good approximation for the balls of the Bethe lattice (which are also finite trees).

Let us finish this short summary by emphasizing that the usual rigorous argument does not work in our setting, the culprit being (i) the exponential growth of the trees and (ii) the spectral gap of the free Laplacian restricted to trees, which is of order L^{-3} and thus too small, compared with (1.3) and (1.4). As a consequence of (i) we are not able to use Dirichlet–Neumann bracketing and (ii) renders impossible the approximation of the ground state of the perturbed operator by the ground state of the free one. We discuss later the new ideas required to overcome these two problems.

Let us now introduce some notation and the rigorous statements of our results.

1.1. **Main results.** Let Γ be an infinite graph and denote by $\ell^2(\Gamma)$ the space of square summable functions defined on the vertices of Γ. Let Δ_Γ be the associated (negative definite) Laplacian operator, i. e.

$$\Delta_\Gamma \colon \ell^2(\Gamma) \to \ell^2(\Gamma)$$
$$(\Delta_\Gamma \varphi)(v) := \sum_{w\sim v} \big(\varphi(w) - \varphi(v)\big), \quad v \in \Gamma.$$

Here the index $w \sim v$ runs over all neighboring nodes $w \in \Gamma$ of the node $v \in \Gamma$. Let us define a *random potential* on this graph, i. e. a diagonal operator

$$V_\omega^\Gamma \colon \ell^2(\Gamma) \to \ell^2(\Gamma)$$
$$(V_\omega^\Gamma \varphi)(v) := \omega_v \varphi(v), \quad v \in \Gamma,$$

where $\omega := \{\omega_v\}_{v\in\Gamma}$ is a sequence of non-trivial, bounded, non-negative, independent and identically distributed random variables. We will also assume that

$$\operatorname{ess\,inf} \omega_0 = 0.$$

This is no additional restriction given that we can always shift the energy through a translation. We are interested in the random operator

$$H_\omega^\Gamma := -\Delta_\Gamma + \lambda V_\omega^\Gamma, \tag{1.5}$$

where λ denotes a (strictly) positive coupling constant. We will call this Hamiltonian the *Anderson model on* Γ. Choose some $0 \in \Gamma$ and let us define its associated *integrated density of states* as

$$\mathcal{N}^\Gamma(E) := \mathbb{E}[\langle \delta_0, \mathbf{1}_{(-\infty,E]}(H_\omega^\Gamma)\delta_0\rangle], \tag{1.6}$$

which is a function of the energy $E \in \mathbb{R}$. Here, and in the rest of the paper, $\mathbf{1}_S$ denotes the indicator function of the set S, the operator $\mathbf{1}_{(-\infty,E]}(H_\omega^\Gamma)$ is the spectral projector on $(-\infty, E]$, defined by functional calculus, and $\delta_0 \in \ell^2(\Gamma)$ denotes Kronecker's delta. The function $E \mapsto \mathcal{N}^\Gamma(E)$ is positive, increasing, and takes values in $[0,1]$. It is the cumulative distribution function of the *density of states measure*, which we denote by $\mathrm{d}\mathcal{N}$.

If one assumes that H_ω^Γ is *ergodic* [PF92; CL90], then (1.6) is independent of the choice of $0 \in \Gamma$, and we know that there exists some set $\Sigma \subset \mathbb{R}$ such that

$$\Sigma = \sigma(H_\omega^\Gamma) = \sigma(-\Delta_\Gamma) + \lambda \operatorname{supp} \omega_0 = \operatorname{supp} \mathrm{d}\mathcal{N}^\Gamma, \tag{1.7}$$

for almost every ω. This is the case if Γ is the graph $\mathbb{Z}^d$ or the Bethe lattice $\mathcal{B}$ defined below. We will denote by E_0 the bottom of the almost sure spectrum, i. e.

$$E_0 := \inf \Sigma = \inf \sigma(-\Delta_\Gamma).$$

It is well known that the asymptotic behavior of the integrated density of states close to the bottom of the spectrum E_0 is very different in the presence of disorder (see remark 1.2 below or [KM06] for a survey). In this work, we study this behavior on a graph known as the *Bethe lattice*, which we define as an infinite connected undirected graph, with no closed loops and degree constant and equal to $k+1$. For $k=1$, we obtain with this definition the graph $\mathbb{Z}$. From now on we fix $k \geqslant 2$ for the rest of this paper and we denote this graph by $\mathcal{B}$. Whenever we omit the index Γ it will be assumed that $\Gamma = \mathcal{B}$.

This paper is devoted to the proof of the following theorem.

Theorem 1.1. *Let $k \geqslant 2$ and H_ω be the Anderson model on the Bethe lattice of degree $k+1$. If*

$$\nu := \limsup_{\kappa \searrow 0} \sqrt{\kappa} \log\bigl|\log \mathbb{P}(\omega_0 \leqslant \kappa)\bigr| < 1, \tag{1.8}$$

then inequalities (1.1) *hold true, and thus*

$$\lim_{E \searrow E_0} \frac{\log\log|\log \mathcal{N}(E)|}{\log(E-E_0)} = -\frac{1}{2}. \tag{1.9}$$

Remark 1.2.

• The fact that the integrated density of states decays faster in presence of disorder has been known to hold rigorously since the works of Nakao [Nak77] and Pastur [Pas77] on $\mathbb{R}^d$. Analogous results have also been obtained in the discrete setting $\Gamma = \mathbb{Z}^d$. In this case, if $\lambda = 0$ in (1.5), then

$$\lim_{E \searrow E_0} \frac{\log \mathcal{N}^{\mathbb{Z}^d}(E)}{\log(E-E_0)} = \frac{d}{2}, \qquad \text{(Van Hove singularity)}$$

while as soon as $\lambda \neq 0$, (and with a restriction analogous to (1.8))

$$\lim_{E \searrow E_0} \frac{\log|\log \mathcal{N}^{\mathbb{Z}^d}(E)|}{\log(E-E_0)} = -\frac{d}{2}. \qquad \text{(Lifshitz tails)}$$

• In absence of disorder, the density of states of the free Laplacian on the Bethe lattice can be calculated explicitly, see [Kes59; McK81],

$$\begin{aligned} \mathrm{d}\mathcal{N}_0(E) &= \mathrm{d}\langle \delta_0, \mathbf{1}_{(-\infty,E]}(-\Delta_{\mathcal{B}})\delta_0\rangle \\ &= \mathbf{1}_I(E)\frac{k+1}{2\pi}\frac{\sqrt{4k-(E-k-1)^2}}{(k+1)^2-(E-k-1)^2}dE \end{aligned} \tag{1.10}$$

with $I := \sigma(-\Delta_{\mathcal{B}}) = \operatorname{supp} \mathrm{d}\mathcal{N}_0 = [(\sqrt{k}-1)^2, (\sqrt{k}+1)^2]$. In particular, we see that for any $k \geqslant 2$

$$\lim_{E \searrow E_0} \frac{\log \mathcal{N}_0(E)}{\log(E-E_0)} = \frac{3}{2}.$$

• The double exponential decay of the integrated density of states in (1.9) stems from concentration inequalities, which are exponential in the volume of shells of the Bethe lattice, and the fact that the volume of these shells grows exponentially with their radius.

• Condition (1.8) tells us that the distribution of the random variables should not decay too fast when we approach 0. It is satisfied, for example, by uniform or Bernoulli random variables. We provide in the text a slightly weaker version for which we can prove (1.9) but not (1.1). If this last condition is not satisfied, it is indeed possible to show that the lower bound fails (see lemma 2.3). Similar results are known to hold true in the Euclidean settings (see [KM06]).

To establish our main result, we will study the Laplace transform of $\mathrm{d}\mathcal{N}$. The study of the integrated density of states through the Laplace transform of its derivative goes back at least to Pastur [Pas71]. This last work together with the celebrated result of Donsker and Varadhan [DV75] on the asymptotic of the Wiener sausage were used to give the first rigorous proof of the existence Lifshitz tails for the continuous Anderson model with Poisson impurities, see [Pas77; Nak77]. Similar ideas work in the discrete setting [BK01]. The spectral theorem shows that the Laplace transform of the *density of states measure* $\mathrm{d}\mathcal{N}$ is the continuous solution $u\colon [0,+\infty) \times \mathcal{B} \to [0,+\infty)$ of a heat equation associated to H_ω evaluated at one point. Thus, the proof of our main theorem will be a consequence of our next result, which is related to the following Cauchy problem:

$$\begin{cases} \dfrac{\partial}{\partial t}u(t,v) = \Delta_{\mathcal{B}}u(t,v) - V_\omega u(t,v), & \text{for } (t,v) \in (0,\infty) \times \mathcal{B} \\ \quad u(0,v) = \delta_0(v) & \text{for } v \in \mathcal{B}. \end{cases} \tag{1.11}$$

The solution $t \mapsto u(t,\,\cdot\,)$ is the solution to the *heat equation with random coefficients* and localized initial datum δ_0. Again, $0 \in \Gamma$ is here any point of the lattice and the results are independent of this choice.

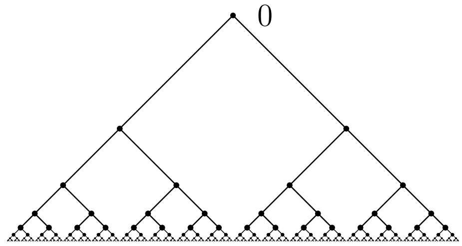


FIGURE 2. The infinite rooted tree with two children per parent.

Theorem 1.3 (Annealed regime)**.** *Assume* (1.8). *Then there exist* $\epsilon > 0$ *and* $t^* > 0$ *such that for all* $t > t^*$,

$$\exp\left(-t\left(E_0 + \frac{\epsilon^{-1}}{(\log t)^2}\right)\right) \leqslant \mathbb{E}\big[u(t,0)\big] \leqslant \exp\left(-t\left(E_0 + \frac{\epsilon}{(\log t)^2}\right)\right). \tag{1.12}$$

Remark 1.4. Obviously $\exp(-t(E_0 + |O((\log t)^{-2})|)) = \exp(-|O(t(\log t)^{-2})|)$ but the quantity $E_0 + |O((\log t)^{-2})|$ should be regarded as an energy in the spectrum Σ close to the bottom E_0.

The long term behavior (1.12) at the node $0 \in \Gamma$ of the solution to the heat equation (1.11) is well approximated by finite volume versions of the same problem (using e. g. Feynman–Kac formula). More precisely, we will look at the solution to the Cauchy problem on a ball $\mathcal{B}^L$ of radius $L \asymp t$ with *Dirichlet* boundary conditions, i. e. we require that the solution is zero outside this ball. The solution of the finite dimensional problem is then bounded above by a term of the form $\mathrm{e}^{-tE_{\mathit{GS}}(H_\omega|\mathcal{B}^L)}$, where $E_{\mathit{GS}}(H_\omega|\mathcal{B}^L)$ denotes the smallest eigenvalue of H_ω restricted to the ball $\mathcal{B}^L$. A crucial ingredient of our proof consists in replacing the balls $\mathcal{B}^L$ by *finite symmetric rooted trees* $\mathcal{T}^L$.

Let us introduce some more notation. We let $\mathcal{T}$ be a rooted tree with branching number k, this is an infinite connected graph which has no closed loops and such that the degree is constant and equal to $k+1$ on every site, except at one particular site 0, which has degree k and is called the root of the tree (see fig. 2). Note that we can embed this infinite graph into the Bethe lattice $\mathcal{B}$. In this note we consider finite versions of this tree, namely, for any integer number $L > 0$ we denote by $\mathcal{T}^L$ the subtree of $\mathcal{T}$ of finite depth L, consisting of all those sites at a distance $L-1$ or smaller from the root 0:

$$\mathcal{T}^L := \{v \in \mathcal{T} \colon d^{\mathcal{T}}(0,v) \leqslant L-1\}.$$

Here $d^\Gamma(\,\cdot\,,\,\cdot\,)$ denotes the graph distance associated to the graph Γ. By introducing the notation (which we repeatedly use later) $|v| = d^{\mathcal{T}}(0,v)+1$ for the "level" of the node v, we can also write $\mathcal{T}^L = \{v \in \mathcal{T} \colon |v| \leqslant L\}$.

These finite symmetric rooted trees look like the balls $\mathcal{B}^L \subset \mathcal{B}$, centered at 0, after removing entirely one of the branches attached to the center of the ball. Note that $\mathcal{T}^L$ is a finite connected graph which has no closed loops and such that the degree is constant at each site except at the root and at the leaves $\{|v| = L\}$. We also picture these finite subtrees as subsets of the infinite Bethe lattice $\mathcal{B}$. Let now

$$H_\omega^L := H_\omega | \mathcal{T}^L$$

be the Hamiltonian H_ω restricted to the subtree $\mathcal{T}^L$ of length L with *Dirichlet* (also called *simple*) boundary conditions. We denote by E_{GS}^L the random ground state energy of H_ω^L, i. e.

$$E_{GS}^L := \inf_{\|\varphi\|_2=1} \left\langle H_\omega^L \varphi, \varphi \right\rangle .$$

We can now state our last main result.

Theorem 1.5. *Assume* (1.8). *Then there exist* $\epsilon > 0$ *and* $L^* > 1$ *such that for all* $L > L^*$ *we have*

$$E_0 + \epsilon (\log L)^{-2} \leqslant E_{GS}^L \leqslant E_0 + \epsilon^{-1} (\log L)^{-2}$$

with probability at least

$$1 - \exp(-\epsilon L).$$

To finish the presentation of our results, let us note that an immediate corollary of theorem 1.5 is that

$$\lim_{L\to\infty} \frac{\log(E_{GS}^L - E_0)}{\log\log L} = -2 \quad \text{a. s.} \tag{1.13}$$

In fact, the Borel–Cantelli lemma implies this, since

$$\sum_{L=L^*}^{\infty} \mathbb{P}\Big(\frac{\log(E_{GS}^L - E_0)}{\log\log L} > -2 - \frac{\log \epsilon}{\log\log L} \Big) \leqslant \sum_{L=L^*}^{\infty} \exp(-\epsilon L) < \infty,$$

so that $\limsup_{L\to\infty} \frac{\log(E_{GS}^L - E_0)}{\log\log L} \leqslant -2$ a. s., and analogously for the other direction. For comparison, in absence of disorder we obtain (see section 3):

$$E_0^L := E_{GS}(-\Delta_{\mathcal{B}} | \mathcal{T}^L) = E_0 + \frac{\sqrt{k}\pi^2}{(L+1)^2} + O(L^{-4}) \tag{1.14}$$

which implies

$$\lim_{L\to\infty} \frac{\log(E_0^L - E_0)}{\log L} = -2.$$

1.2. **Discussion.** The results we presented concern an operator which appears naturally in the study of the macroscopic properties of crystals, alloys, glasses, and other materials. If one looks at the Schrödinger equation

$$\begin{cases} i\frac{d\varphi}{dt} = H_\omega^\Gamma \varphi, & i^2 = -1 \\ \varphi(0) = \varphi_0 \in \ell^2(\Gamma), & \|\varphi_0\|_2^2 = 1 \end{cases}, \tag{1.15}$$

then the Anderson model H_ω^Γ defined by (1.5) describes the Hamiltonian governing the behavior of a quantum particle having an initial state φ_0 in a disordered medium. Its integrated density of states measures the “number of energy levels per unit volume” and is a concept of fundamental importance in condensed matter physics, as it encodes various thermodynamical quantities of the material, spectral features of the operator and properties of the underlying geometry.

The Anderson model has been the subject of hundreds of physical and mathematical papers. One of the most studied mathematical features of this model is the phenomenon known as *Anderson localization*, i. e. that the spectrum of the random operator exhibits pure point spectrum with probability 1, for *any* strength of disorder, whereas the free operator has only absolutely continuous spectrum. We invite the interested reader to consult the monographs [CL90; PF92; Sto01; His08; Kir07].

One of the hallmarks of Anderson localization is the so-called *Lifshitz tails behavior*, i. e. the exponential decay of the integrated density of states at the bottom of the spectrum. It is well known that such a decay, together with additional assumptions on the regularity of the random variables, provides one of the main ingredients to start the *multi-scale analysis*, see e. g. [GK13], or to satisfy the *fractional moment criterion* ([AM93]). These strategies have been successfully applied to prove the existence of localization in a neighborhood of the spectral edges, for example when the graph is $\mathbb{Z}^d$ (the model introduced originally by P. W. Anderson in [And58]) or its continuous version on $\mathbb{R}^d$.

The Bethe lattice is of interest in statistical mechanics because of its symmetry properties and the absence of loops. It allows to obtain closed solutions for some models, e. g. in percolation theory and the non-rigorous scaling theory of Anderson localization. In our setting, the resolvent of the operator H_ω on the Bethe lattice admits a recursive representation (see e. g. [Ros12]), but in this work we make no use of these formulas. It was for these reasons that the model was studied in [ATA73], and it enjoys some renewed interest in the physical community (see e. g. [BST10; BS11]). Because of its exponential growth, it is also of interest in connection with the configuration space of many-body problems [Alt+97].

Perhaps one of the most striking features of the operator defined on the infinite tree $\mathcal{B}$ is the *absence of pure point spectrum* at weak disorder [Kle96; AW11b; AW11a]. For a survey on recent progress on the spectral properties of the Anderson model on the Bethe lattice see [War13]. At weak disorder, thus, this model exhibits no Anderson localization, even near the spectral edges where Lifshitz tails take place. For the Anderson model on the Bethe lattice, the existence of a Lifshitz tail *does not imply localization.* We remind the reader that in the Euclidean case, the absence of localization at higher energies and therefore the existence of a spectral transition is still an open problem.

The parabolic problem (1.11) is the heat equation associated with the Anderson Hamiltonian and is well studied under the name of *Parabolic Anderson model.* It describes a random particle flow in the tree $\mathcal{B}$ through a random field of soft sinks, which can also be seen as traps or obstacles via the Feynman–Kac formula. There is an additional interpretation in terms of a branching process in a field of random branching rates. We refer the reader to [GK05; KW] for a survey. In this context, one is usually interested in the behavior of the total mass $\|u(t,\,\cdot\,)\|_1$ of the solution for large $t > 0$, which is also the behavior of the solution to the heat equation with initial datum $u(\,\cdot\,, 0) \equiv 1$ at one point. Because of the exponential growth of the graph, we were not able to study this quantity. However, theorem 1.3 is a first step in this direction.

A related question of interest, which remains unanswered in this paper, is whether there is intermittency in this setting. Following the heuristics described in [KW], intermittency can be understood as a consequence of Anderson localization. On the other hand, Lifshitz tails have been proved as a by-product of the proof of intermittency in the parabolic Anderson model in [BK01]. Given that we may have absolutely continuous spectrum in spite of the existence of Lifshitz tails, the answer to this question is an interesting subject for further studies.

This discussion would not be complete without citing some previous results. The Lifshitz tails behavior for a percolation model on the Bethe lattice was studied in [Rei09], see also [MS11]. In [Sni89] similar bounds to ours are obtained for $\mathcal{N}$ and for $\tilde{\mathcal{N}}$ in the hyperbolic space, which is the continuous analog of the Bethe lattice. In our setting, Lifshitz tails were studied in [BS11; Ros12], where in particular a rigorous lower bound

$$\liminf_{E \to E_0} \frac{\log \log |\log \mathcal{N}(E - E_0)|}{\log(E - E_0)} \geqslant -\frac{1}{2}.$$

is established. A proof of any type of decay other than the trivial one has resisted several attempts to be rigorously proved. We will try to explain now why.

The first problem concerns the finite-dimensional approximation of the infinite dimensional operator. In the standard setting of the Anderson model $\Gamma = \mathbb{Z}^d$, if we let

$$\Lambda_L := \{v \in \mathbb{Z}^d \colon \|v\|_\infty \leqslant L\},$$

then the thermodynamic limit of the *normalized eigenvalue counting functions*

$$\lim_{L \to \infty} \frac{\#(\sigma(H_\omega|\Lambda_L) \cap [0, E])}{\#\Lambda_L} = \lim_{L \to \infty} \frac{\operatorname{tr} \mathbf{1}_{[0,E]}(H_\omega|\Lambda_L)}{\#\Lambda_L} \tag{1.16}$$

exists almost surely for every E where $\mathcal{N}$ is continuous. In this setting as well as in the continuous version on $\mathbb{R}^d$, the limit defined in (1.16) is *independent of the boundary conditions*. That this limit coincides with the averaged spectral function $\mathcal{N}$ at these points is known as the *Pastur–Shubin formula*. There is a wealth of results in this direction in different settings. It holds in particular on any amenable graph Γ (like $\mathbb{Z}^d$) if we choose the sequence Λ_L as a Følner sequence.

The limit (1.16) may have different limits depending on the choice of boundary conditions when the graph is not amenable [AW06]. In particular, the usual Dirichlet–Neumann bracketing is of no hope. A similar phenomenon occurs on the hyperbolic space [Sni89; Sni90]. This leaves us with the problem of finding the right finite volume approximations. In [SS14] it is proved that the Pastur–Shubin formula holds in great generality. This setting includes the Cayley graph of a free group. The finite volume approximations are the analogs of the periodic boundary conditions in the Euclidean case. Unfortunately, the rate of convergence of the approximations to the averaged spectral function (1.6) is unknown. An approach of a different vein was explored in [Gei14]. Here one looks at this problem on random regular graphs. It is indeed known, since the pioneering works of Kesten [Kes59] and McKay [McK81], that the density of eigenvalues of the free Laplacian of random regular graphs converges to the measure given by (1.10). This approach introduces another source of randomness and seems difficult to study. We avoid these difficulties altogether by approximating the integrated density of states only at low energies. This strategy works because low energy eigenfunctions have exponentially small values on the leaves and counter the exponentially growing number of leaves, see section 3.1 and [Bro91].

In this work we consider Dirichlet restrictions of the operator to finite trees $\mathcal{T}^L$. We show that these are good approximations as long as we look at the bottom of the spectrum. The problem then reduces, as usual, to that of finding good upper bounds on the ground state energy $E_{GS}(H_\omega|\mathcal{T}^L)$ with good probability. In the standard Anderson model on $\mathbb{Z}^d$, one usually uses Temple's inequality, [Sim85], or Thirring's inequality, [KM83], or perturbation theory, [Sto99]. Unfortunately, they are all based on the premise that the ground state of the perturbed operator, modulo some small error, should look like the one of the free operator. This reduces the problem to study only the effect of the random potential on the ground state of the free operator. The error term in these methods is related to the spectral gap of the free operator, i. e. the distance between the first and the second eigenvalue. In the Euclidean setting both the first and the second eigenvalue of the free operator with Dirichlet boundary conditions are of the same order L^{-2}, whereas in our setting we have (1.14) but the second eigenvalue of $-\Delta_{\mathcal{B}}|\mathcal{T}^L$ behaves as

$$E_{GS}(-\Delta_{\mathcal{B}}|\mathcal{T}^L) + O(L^{-3}).$$

See section 3 for these calculations. This behavior renders the use of the methods named above impossible. Also, the ground state of the perturbed operator is fundamentally different from the ground state of the free operator. Roughly speaking, the random potential restricts the ground state to a subtree of length on the scale of the logarithm of the radius of the ball, see section 2. Therefore, the ground state of the free operator is not a good test function to compare with the one of the random operator.

1.3. **Strategy of proof.** In order to make the reading of this paper easier, we provide here a road map for the proofs and a table of notations in page 16. This will also allow us to comment on some results and acknowledge some sources. To simplify the (not necessarily rigorous) exposition, we assume $k = 2$ and $\epsilon > 0$ a small constant which may change from line to line. We also write $A \lesssim B$, $B \gtrsim A$

if there exists a constant c such that $A < cB$; $A \asymp B$ if $A \lesssim B$ and $A \gtrsim B$. As usual, to prove theorem 1.1 we prove lower and upper bounds separately.

1.3.1. *Lower bound.* We first prove lower bounds on the integrated density of states, which is easier. This is done in section 2. Note that a rigorous proof of the lower bound was already obtained in [BS11] and [Ros12]. Our method is not very different from the one in [Ros12], but we do identify the sharper condition (1.8). Another novelty is that we also obtain a precise lower bound for the ground state energy on finite rooted trees. Indeed, as will be clear, to obtain the lower bound of the Lifshitz tails it is enough to prove that for large L

$$E^L_{\mathcal{GS}} \leqslant E_0 + \epsilon^{-1}(\log L)^{-2} \tag{1.17}$$

holds with not too small probability (we actually prove that this holds with probability $1 - \exp(-k^{\epsilon L})$, see proposition 2.1). This is, as usually, proved by finding a suitable test function.

Our proposition 2.1 corresponds to the upper bound in theorem 1.5. It is proved by localizing the test function to a subtree of length $\log L$ (see fig. 3 in page 18). This is crucial to prove that the almost sure behavior of $E^L_{\mathcal{GS}}$ is of order $(\log L)^{-2}$ for large L. We show then that this upper bound implies the lower bound in (1.2) (corollary 2.5) and a Tauberian theorem (lemma 2.6 and proposition 2.7) gives the lower bound in (1.1).

1.3.2. *Upper bound.* In section 3, we introduce some elements and tools we will need in the course of the proof. Because these calculations make no use of randomness, we decided to isolate them in their own section.

We first study the spectral theory of the free Laplacian on finite trees, calculating explicitly all the eigenvalues and an orthonormal basis of eigenfunctions (lemmas 3.1 to 3.3 and 3.5). The eigenfunctions have their support confined to disjoint subtrees. This property is crucial when we study the action of the random potential in section 4.

We then prove an analog of the Ismagilov–Morgan–Sigal (IMS) localization formula on trees (proposition 3.6). The proof is of interest on its own as it can be adapted to very general discrete settings (see remark 3.8). In this work, we have decided to prove it only for functions in $\ell^2(\mathbb{Z})$ (lemma 3.7) and then carry it over to the tree by means of the spectral theory of the free Laplacian developed earlier in this section.

We finally prove in this section an uncertainty relation for truncated eigenfunctions on the tree. Explaining the details here would make this road map too long, but see below how this truncation is used. We also prove first this property for functions in $\ell^2(\mathbb{Z})$ and then use the spectral theory to prove it on $\ell^2(\mathcal{T}^L)$.

In section 4 we prove the upper bound in (1.1). It proceeds roughly as follows. The first step is a Tauberian theorem. It is an elementary consequence of an upper bound on the integrated density of states for energies close to E_0 using the large time decay of the Laplace transform $\tilde{\mathcal{N}}(t)$ of its derivative (proposition 4.2). We include it for the sake of completeness. It says the following: if for some $c > 0$

$$\tilde{\mathcal{N}}(t) \leqslant \exp(-t(E_0 + c(\log t)^{-2})) \quad \text{for } t \gg 1 \tag{1.18}$$

then

$$\mathcal{N}(E) \leqslant \exp(-\mathrm{e}^{\epsilon(E-E_0)^{-1/2}}) \quad \text{for } E - E_0 \ll 1.$$

We then do a reduction to a finite scale (proposition 4.3). We show that we can restrict the operator to a ball, as long as the ball grows linearly with the time t:

$$\tilde{\mathcal{N}}(t) = \mathbb{E}\langle \delta_0, \mathrm{e}^{-tH_\omega}\delta_0\rangle \leqslant \mathbb{E}\langle \delta_0, \mathrm{e}^{-tH_\omega|\mathcal{B}^L}\delta_0\rangle + \mathrm{e}^{-\zeta t} \quad \text{with } t = \zeta L \text{ and } \zeta \gg 1.$$

The approximations $H_\omega|\mathcal{B}^L$ considered in this step correspond to the Hamiltonian H_ω restricted to balls of radius L with *simple* (also called *Dirichlet*) boundary conditions. A proof using the Feynman–Kac formula is contained in [Ros12] (following [BK01]). We provide a somewhat elementary proof of this fact, which appears to be new in this context. The idea is to compare the series expansion of both e^{-tH_ω} and $\mathrm{e}^{-tH_\omega|\mathcal{B}^L}$ and expand the matrix products of the terms $\langle \delta_0, (-tH_\omega)^n\delta_0\rangle$ as products of paths from δ_0 to δ_0 of length at most $n/2$. This is a discrete version of the Feynman–Kac formula. Because the coefficients of both H_ω and its approximation coincide in a large ball, the error we make is easily estimated by the tail of the exponential series.

The next step consists in simply replacing the operator by its ground state energy (lemma 4.4). Using a spectral decomposition of δ_0 in terms of the eigenfunctions of H_ω leads to an upper bound of the form

$$\langle \delta_0, \mathrm{e}^{-tH_\omega|\mathcal{B}^L}\delta_0\rangle \leqslant \mathrm{e}^{-tE_{\mathit{GS}}(H_\omega|\mathcal{B}^L)}. \tag{1.19}$$

It is easy to see (lemmas 4.5 and 4.6) that every ball is embedded in a finite symmetric rooted tree and that we can replace the ball by a tree $\mathcal{T}^L$ because

$$\exp(-tE_{\mathit{GS}}(H_\omega|\mathcal{B}^L)) \leqslant \exp(-tE_{\mathit{GS}}(H_\omega^L)) \quad \text{where } H_\omega^L = H_\omega|\mathcal{T}^L.$$

This is crucial in our argument, as we are able to use the spectral theory on the tree as a makeshift "Fourier transform" in the probability estimates we describe below. Using the two last inequalities and taking the expectation in (1.19) we see that (lemma 4.7) for any $E \geqslant 0$

$$\mathbb{E}\langle \delta_0, \mathrm{e}^{-tH_\omega|\mathcal{B}^L}\delta_0\rangle \leqslant \mathrm{e}^{-tE} + \mathrm{e}^{-tE_0}\mathbb{P}(E_{\mathit{GS}}^L \leqslant E).$$

As (1.18) and the proved lower bound (1.17) suggest, one should take $E = E_0 + C(\log L)^{-2}$ in the last inequality. Note that $E_{\mathit{GS}}(-\Delta|\mathcal{T}^L) = E_0 + CL^{-2}$ is not even of the same order. We perform then a reduction to a yet smaller scale. By using the IMS localization formula we go from the scale L to $r = \epsilon^{-1}\log L$ (proposition 4.9). By doing this, we trade energy for probability: the number of subtrees of length $\log L$ is about $k^{L-\log L} \asymp k^L = k^{\exp(\epsilon r)}$ and thus

$$\mathbb{P}(E_{\mathit{GS}}^L \leqslant E_0 + C(\log L)^{-2}) \leqslant k^{\exp(\epsilon r)}\mathbb{P}(E_{\mathit{GS}}^r \leqslant E_0 + Cr^{-2}).$$

It is clear now (replacing $r \gg 1$ by $L \gg 1$ below) that we need to prove

$$\mathbb{P}(E^L_{GS} \leqslant E_0 + CL^{-2}) \leqslant \mathrm{e}^{-\exp(\epsilon' L)}, \quad L \gg 1, \tag{1.20}$$

for some $\epsilon' > 0$. After choosing $0 < \epsilon < \epsilon'$, we obtain (1.18). Proving this last bound occupies the last section of this paper.

To prove (1.20) we proceed as follows. We first note that functions with low kinetic energy average the random potential and this pushes their energy away from the bottom of the spectrum. First we show that (up to an error) we can assume that the (random) ground state φ^L of H^L_ω is a linear combination of low energy eigenfunctions. Using our uncertainty relation, we can furthermore replace these eigenfunctions by truncated versions. The rest of the proof is a careful analysis of the action of the random potential on functions of this type. The truncation is necessary to exploit the averaging properties of the eigenfunctions away from the root in the concentration inequalities.

Let us be more precise. Every eigenfunction of $\Delta|\mathcal{T}^L$ is supported in some subtree of $\mathcal{T}^L_v$ of length $L-|v|$ rooted at $v \in \mathcal{T}^L$ (see section 3). We index the eigenfunctions $\Psi_{v,m}$ by their *anchor* $v \in \mathcal{T}^L$ and their *mode* (or *frequency*) $m = 1, \dots L - |v|$. It is not hard to see (lemma 3.3) that the eigenfunctions $\Psi_{v,m}$ close to the bottom E_0 satisfy

$$\langle \Delta_{\mathcal{T}} \Psi_{v,m}, \Psi_{v,m} \rangle \approx E_0 + \Big(\frac{Cm}{L - |v| + 1}\Big)^2$$

We deduce then that the eigenfunctions $\Psi_{v,m}$ having small energy, i. e. those for which

$$\Big(\frac{Cm}{L - |v| + 1}\Big)^2 \leqslant \beta L^{-2}$$

satisfy both that their modes are bounded by β (uniformly in L), i. e.

$$m \leqslant \beta \tag{1.21}$$

and that the distance of their anchors to the root of $\mathcal{T}^L$ is bounded linearly in L, i. e.

$$|v| \leqslant C_A L, \qquad 0 < C_A < 1. \tag{1.22}$$

The subscript A here stands for "anchor". The reader may imagine that the ground state is (up to an error) "bandlimited in Fourier space".

After this projection in "Fourier space", we introduce a truncation in physical space. This is necessary because low energy functions are not "flat" in the usual sense. The reader will convince herself by looking at the ground state of $\Delta_{\mathcal{T}^L}$, which is radially symmetric and thus exponentially decaying from the root. Nevertheless, these functions distribute evenly their ℓ^2-mass in the transversal direction. We can thus throw away some of the mass close to the anchor and control precisely the error by doing so with our uncertainty principle. Let us call φ^L the (random) ground state of H^L_ω we obtain after applying the spectral projection and the truncation.

Most of the averaging now takes place away from the anchor in the radial direction. Let us look at the potential energy of φ^L. If we center the random variables

$$\langle V_\omega \varphi^L, \varphi^L \rangle = \langle (V_\omega - \mathbb{E}\omega_0)\varphi^L, \varphi^L \rangle + (\mathbb{E}\omega_0)\|\varphi^L\|_2$$

then the first term in the sum is close to zero with good probability. The second term of the sum is of order 1, which implies that we are far from the bottom of the spectrum.

We proceed now to explain the probability estimates. To show that the potential energy is concentrated around its mean we may use Hoeffding's inequality, which tells us that for fixed $\varphi \in \ell^2(\mathcal{T}^L)$ and $\kappa > 0$ we have

$$|\langle (V_\omega - \mathbb{E}\omega_0)\varphi, \varphi \rangle| \lesssim \kappa \tag{1.23}$$

with probability at least

$$1 - \exp\bigl(-O(\kappa^2 / \operatorname{Var}[\langle (V_\omega - \mathbb{E}\omega_0)\varphi, \varphi \rangle])\bigr).$$

Cauchy–Schwarz then tells us that

$$\operatorname{Var}[\langle (V_\omega - \mathbb{E}\omega_0)\varphi, \varphi \rangle] \lesssim \|\varphi\|_4^2.$$

We cannot apply this inequality with φ^L directly because it depends on the realization ω. To get rid of this problem, we exploit first the spectral theory of the free Laplacian. Because some spectral projectors have disjoint support, we are able to reduce the metric entropy in the second step, which is a classical ϵ-net argument. The problem is now reduced to estimate the probability that inequality (1.23) holds for every φ chosen from a fixed ϵ-net. The last part of the calculation is thus a uniform estimation of the ℓ^4-norm of functions both restricted in "Fourier" space and truncated in physical space, which decay exponentially fast in $k^{\epsilon L}$.

This finishes our presentation of the proofs and the introduction.

Table of Notation	
Γ	$\triangleq$ A graph. When it is missing from the notation we assume that $\Gamma = \mathcal{B}$
$-\Delta_\Gamma$	$\triangleq$ Graph Laplacian of the graph Γ
$\mathcal{B}$	$\triangleq$ Bethe lattice of degree k (infinite graph)
$\mathcal{B}^L$	$\triangleq$ Ball of radius L of the Bethe lattice $\mathcal{B}$
$\mathcal{T}^L$	$\triangleq$ Rooted tree of length L (every node has k children but the leaves)
E_0	$\triangleq$ Bottom of the spectrum, $E_0 := \inf\sigma(-\Delta_{\mathcal{B}}) = (\sqrt{k}-1)^2 \overset{\text{a.s.}}{=} \inf\sigma(H_\omega)$
$E_{\mathtt{GS}}(H)$	$\triangleq$ Ground state energy of H, i.e. $E_{\mathtt{GS}}(H) := \inf_{\Vert\varphi\Vert_2=1}\langle\varphi, H\varphi\rangle$
$\mathcal{N}$	$\triangleq$ Integrated density of states of H_ω on $\mathcal{B}$
$\mathcal{N}^L$	$\triangleq$ Expected integrated density of states of H_ω on $\mathcal{B}^L$
$\mathrm{d}\mathcal{N}$	$\triangleq$ Density of states measure
$\tilde{\mathcal{N}}$	$\triangleq$ Laplace transform of the density of states measure $\mathrm{d}\mathcal{N}$
$H\vert\Gamma$	$\triangleq$ Restriction of the operator H with *simple* b.c.
$\mathbb{P}$	$\triangleq$ Probability
$\mathbb{E}$	$\triangleq$ Expectation
δ_i	$\triangleq$ Kronecker's delta, i.e. $\delta_i(j) = 1$ for $i = j$ and zero elsewhere
$d_\Gamma(\,\cdot\,,\,\cdot\,)$	$\triangleq$ Graph distance associated to Γ
$\mathcal{B}_v^L$	$\triangleq$ Ball of the Bethe lattice of radius L centered at v
$\mathcal{T}_v^L$	$\triangleq$ The subtree of $\mathcal{T}^L$ with root in v (of length $L - \vert v\vert + 1$)
$\mathcal{T}_*^L$	$\triangleq$ Augmented finite tree, i.e. $\mathcal{T}_*^L := \mathcal{T}^L \uplus \{*\}$ and $d(*, 0) = 1$
$\vert v\vert$	$\triangleq$ Distance to $* \in \mathcal{T}_*^L$, i.e. $\vert v\vert = d(*, v) = d(0, v) + 1$.
H_ω^L	$\triangleq$ Anderson Hamiltonian restricted to $\mathcal{T}^L$ with *simple* b.c.
$\mathbf{1}_S$	$\triangleq$ Indicator function of the set S
$u \sim v$	$\triangleq$ u and v are neighbors, i.e. $d(u, v) = 1$
A_Γ	$\triangleq$ Adjacency matrix of the graph Γ
∇, ∇^*	$\triangleq$ Forward gradient on $\mathbb{Z}$ and its adjoint
η_a	$\triangleq$ Partition of unity on $\mathbb{Z}$
$\eta_{a,r}$	$\triangleq$ r-scaled, radially symmetric partition of unity on $\mathcal{T}$
$C_{\mathtt{IMS}}$	$\triangleq$ The constant in the error of the IMS formula
$E_\beta^{(L)}$	$\triangleq$ An energy defined as $E_\beta^{(L)} := 2\sqrt{k}\cos\big(\frac{(\beta+1)\pi}{L+1}\big)$, see lemma 3.3.
$\Pi_E^{(L)}$	$\triangleq$ $\Pi_E^{(L)} := \mathbf{1}_{(-\infty,E]}(k + 1 + \Delta_{\mathcal{B}}^L\vert\mathcal{T}^L) = \mathbf{1}_{[E,+\infty)}(A_{\mathcal{B}}^L\vert\mathcal{T}^L)$
$\bar{\Pi}_E^{(L)}$	$\triangleq$ $\bar{\Pi}_E^{(L)} := 1 - \Pi_E^{(L)}$
$B^{(L)}$	$\triangleq$ unit ball of $\ell^2(\mathcal{T}^L)$
$B_v^{(L)}$	$\triangleq$ unit ball of $\ell^2(\mathcal{T}_v^L)$
ω_v	$\triangleq$ One of the non-trivial, bounded, i.i.d. random variables
$\bar{\omega}$	$\triangleq$ Expectation of the random variable ω_v

p. t. o.

Table of Notation (continued)
$\tilde{\omega}_v \triangleq$ Centered random variable $\tilde{\omega}_v := \omega_v - \bar{\omega}$
$\omega_+ \triangleq$ Sup-norm of the random variables $\omega_+ := \|\omega_0\|_\infty$
$\tilde{\omega}_+ \triangleq$ Sup-norm of the centered random variables $\tilde{\omega}_+ := \|\omega_0 - \bar{\omega}\|_\infty$

2. Lifshitz tails: The lower bound

In this section we prove the upper bound in theorem 1.5, the lower bound in theorem 1.3 and the lower bound in theorem 1.1.

2.1. Locating the ground state on a finite rooted tree: The upper bound.

Denote by E_0 the infimum of the spectrum of the free Laplacian Δ on the infinite rooted tree with k children at each node, i. e. $E_0 := (\sqrt{k}-1)^2$. As we will see in section 3.1, the ground state energy of the free Laplacian restricted to the finite tree $\mathcal{T}^L$ of length L (with the root on level 1) reads

$$E_{GS}(-\Delta|\mathcal{T}^L) = E_0 + 2\sqrt{k}\big(1-\cos(\frac{\pi}{L+1})\big) > E_0.$$

The distance between these two values is thus of the order of L^{-2} as $L \to \infty$. By adding a nonnegative random potential V_ω, we increase the ground state energy by at least $\inf V_\omega(\mathcal{T}^L)$. Our first proposition gives a probabilistic upper bound on the random ground state energy E^L_{GS} of the random operator $H^L_\omega := -\Delta^L + V_\omega$ on $\mathcal{T}^L$.

Proposition 2.1. *Assume that the single-site potentials $V_\omega(v)$, $v \in \mathcal{T}$, satisfy*

$$\nu := \limsup_{\kappa \searrow 0} \sqrt{\kappa}\log\big|\log \mathbb{P}(V_\omega(v) \leqslant \kappa)\big| < 1. \tag{2.1}$$

Fix $C_1 > 1 + \pi^2\sqrt{k}(\log k)^2/(1-\nu)^2$ and $\varepsilon \in (0,1)$. Then there is a scale $L_0 = L_0(k,\nu,C_1,\varepsilon)$ such that, for all $L \geqslant L_0$, we have

$$\mathbb{P}\big(E^L_{GS} \leqslant E_0 + C_1(\log L)^{-2}\big) \geqslant 1 - \exp\big(-k^{\varepsilon L}\big).$$

Remark 2.2.

- Condition (2.1) restricts the tail behavior of the distribution function of the single site potentials at 0. This guarantees large enough probability for small single site potentials. The result shows that the ground state is shifted from the scale L^{-2} not further than $(\log L)^{-2}$ with probability exponentially close to 1 as $L \to \infty$. Without condition (2.1) the distribution of the single site potentials could have topological support bounded away from 0, which would shift the spectrum by a positive distance almost surely.
- The upper bound on the ground state provides a lower bound on the integrated density of states, see proposition 2.7. The classical assumption for a lower bound on the IDS on $\mathbb{Z}^d$ is that the cumulative distribution

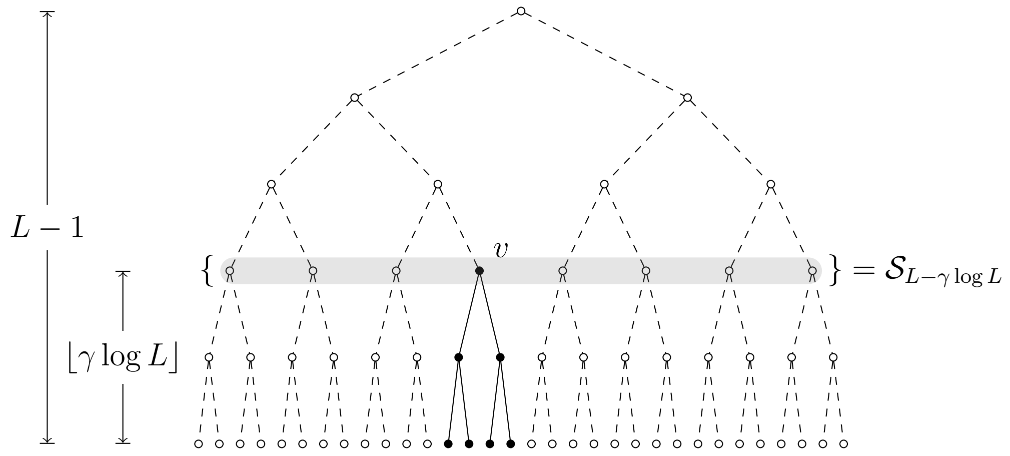


FIGURE 3. Support of test function φ_v in the tree based at $v \in \mathcal{S}_{L-\gamma \log L}$.

function of the single site potentials vanishes not faster than a polynomial at 0:

$$\exists C_2, \nu' > 0\colon \forall \kappa > 0\colon \mathbb{P}(\omega \leqslant \kappa) \geqslant (C_2\kappa)^{\nu'}. \tag{2.2}$$

This is e. g. satisfied for the uniform distribution on an interval $[0, a]$, $a > 0$, and all non-degenerate Bernoulli laws. Condition (2.2) implies $\nu = 0$ and thus (2.1).

Proof of proposition 2.1. Let $\gamma := \pi\sqrt[4]{k}/\sqrt{C_1 - 1}$ and note that

$$0 \leqslant \gamma < (1-\nu)/\log k.$$

We denote by

$$\mathcal{S}_{L-\gamma \log L} := \{v \in \mathcal{T}^L\colon |v| = \lceil L - \gamma \log L \rceil\}$$

the sphere of $\mathcal{T}^L$ at level $\lceil L - \gamma \log L\rceil$ and by $\mathcal{T}_v^L$ the subtree of $\mathcal{T}^L$ rooted at v. Define the function $\varphi_v \in \ell^2(\mathcal{T}^L)$ by

$$\varphi_v(w) := \mathbf{1}_{\mathcal{T}_v^L}(w)\sqrt{\frac{2}{(L-|v|+2)k^{|w|-|v|+1}}}\sin\Big(\pi\frac{|w|-|v|+1}{L+|v|+2}\Big), \quad (w \in \mathcal{T}^L).$$

The support of the function is then $\mathcal{T}_v^L$, see fig. 3. In section 3 we will see that φ_v is the normalized ground state of the free Laplacian restricted to $\mathcal{T}_v^L$, trivially embedded in $\mathcal{T}^L$. We also see that the corresponding eigenvalue is

$$k+1-2\sqrt{k}\cos\Big(\frac{\pi}{\lfloor\gamma\log L\rfloor+1}\Big) = (\sqrt{k}-1)^2 + 2\sqrt{k}\Big(1-\cos\Big(\frac{\pi}{\lfloor\gamma\log L\rfloor+1}\Big)\Big).$$

We will use the states φ_v, $v \in \mathcal{S}_{L-\gamma\log L}$, as test functions to probe for the ground state energy of H_ω^L. In the quadratic form $\langle -\Delta\varphi_v, \varphi_v\rangle$, we sum only over the support of φ_v. Hence, $\langle -\Delta\varphi_v, \varphi_v\rangle$ is the eigenvalue of φ_v on $\mathcal{T}_v^L$. Since $1-\cos(x) \leqslant x^2/2$ for all $x \in \mathbb{R}$, we see that

$$\langle -\Delta\varphi_v, \varphi_v\rangle \leqslant E_0 + \frac{\pi^2\sqrt{k}}{(\lfloor \gamma \log L\rfloor + 1)^2} \leqslant E_0 + \frac{\pi^2\sqrt{k}}{\gamma^2(\log L)^2}.$$

We ask the potential to be small on at least one of the subtrees $\mathcal{T}_v^L$, $v \in \mathcal{S}_{L-\gamma\log L}$. To this end, let $\kappa := (\log L)^{-2}$ and

$$\Omega'_L := \{\omega : \exists v \in \mathcal{S}_{L-\gamma\log L}\colon \max_{w\in\mathcal{T}_v^L} V_\omega(w) \leqslant \kappa\}.$$

For all $\omega \in \Omega'_L$, we have

$$\begin{aligned} E^L_{\mathit{GS}} &\leqslant \inf_{v\in\mathcal{S}_{L-\gamma\log L}} \big(\langle -\Delta\varphi_v, \varphi_v\rangle + \langle V_\omega\varphi_v, \varphi_v\rangle\big) \\ &\leqslant E_0 + \frac{\pi^2\sqrt{k}}{\gamma^2(\log L)^2} + \kappa \leqslant E_0 + C_1(\log L)^{-2}. \end{aligned}$$

For the probabilities, this implies

$$\mathbb{P}(\Omega'_L) \leqslant \mathbb{P}\big(E^L_{\mathit{GS}} \leqslant E_0 + C_1(\log L)^{-2}\big).$$

We have to estimate $\mathbb{P}(\Omega'_L)$ from below. Choose $\delta \in (0, 1-\nu-\gamma\log k)$. From $\nu < 1$ and (2.1), we get an $L'_0 > 0$ such that, for all $L \geqslant L'_0$ and all $w \in \mathcal{T}^L$,

$$|\log \mathbb{P}(V_\omega(w) \leqslant \kappa)| \leqslant \exp(\kappa^{-1/2}(\nu+\delta)) = L^{\nu+\delta}.$$

We use this to build a lower bound of $\mathbb{P}(\Omega'_L)$ in several steps. Note that for each $v \in \mathcal{S}_{L-\gamma\log L}$, the subtree $\mathcal{T}_v^L$ rooted at v has $\#\mathcal{T}_v^L = \sum_{i=0}^{\lfloor\gamma\log L\rfloor} k^i \leqslant k^{\lfloor \gamma\log L\rfloor+1} \leqslant kL^{\gamma\log k}$ nodes. Therefore, we have

$$\Big|\log \mathbb{P}\big(\max_{w\in\mathcal{T}_v^L} V_\omega(w) \leqslant \kappa\big)\Big| = \sum_{w\in\mathcal{T}_v^L} \Big|\log \mathbb{P}(V_\omega(w) \leqslant \kappa)\Big| \leqslant kL^{\gamma\log k+\nu+\delta}.$$

Together with $\#\mathcal{S}_{L-\gamma\log L} = k^{\lceil L-\gamma\log L\rceil-1} \geqslant k^{L-\gamma\log L-1}$, we get the lower bound

$$\begin{aligned} \mathbb{P}(\Omega'_L) &= 1 - \prod_{v\in\mathcal{S}_{L-\gamma\log L}} \big(1 - \mathbb{P}(\max_{w\in\mathcal{T}_v^L} V_\omega(w) \leqslant \kappa)\big) \\ &= 1 - \prod_{v\in\mathcal{S}_{L-\gamma\log L}} \big(1 - \exp(-|\log \mathbb{P}(\max_{w\in\mathcal{T}_v^L} V_\omega(w) \leqslant \kappa)|)\big) \\ &\geqslant 1 - \Big(1 - \exp\big(-kL^{\gamma\log k+\nu+\delta}\big)\Big)^{\#\mathcal{S}_{L-\gamma\log L}} \\ &\geqslant 1 - \Big(1 - \exp\big(-kL^{\gamma\log k+\nu+\delta}\big)\Big)^{k^{L-\gamma\log L-1}}. \end{aligned}$$

From $\log(1-p) = -\sum_{j=1}^{\infty} p^j/j \leqslant -p$, we see

$$(1-p)^x = \exp(x\log(1-p)) \leqslant \exp(-px)$$

for all $p \in [0,1]$ and $x > 0$. This yields

$$\begin{aligned}\mathbb{P}(\Omega'_L) &\geqslant 1 - \exp\bigl(-\exp\bigl(-kL^{\gamma \log k+\nu+\delta}\bigr)k^{L-\gamma\log L-1}\bigr)\\ &= 1 - \exp\bigl(-\exp\bigl(L\log k - kL^{\gamma\log k+\nu+\delta+1} - (\gamma\log L+1)\log k\bigr)\bigr).\end{aligned}$$

The important fact here is $\gamma \log k + \nu + \delta < 1$. Thus, for all $\varepsilon \in (0,1)$, the exponent satisfies

$$L\log k - kL^{\gamma\log k+\nu+\delta} - (\gamma\log L+1)\log k \geqslant \varepsilon L\log k$$

as soon as L is large enough, say $L \geqslant L_0 \geqslant L'_0$. The proposition readily follows. ■

We address briefly the question of the optimality of condition (2.1). Let us first note that to prove the lower bound for the Lifshitz tails with exponent $1/2$ it is enough to prove that for every $\eta \in (0,2)$ we have

$$E^L_{\mathit{GS}} \leqslant E_0 + (\log L)^{-\eta}, \quad L \gg 1$$

with good probability (compare this to the consequence of proposition 2.1). This leads us to consider the slightly weaker condition

$$\forall \eta \in (0,2)\colon\ \limsup_{\kappa\searrow 0} \kappa^{1/\eta}\log\bigl|\log\mathbb{P}(V_\omega(v)\leqslant\kappa)\bigr| = 0,$$

which is implied by condition (2.1). The following lemma shows that we can not expect to do better than this.

Lemma 2.3. *Suppose that for some $\eta > 0$*

$$\limsup_{\kappa\searrow 0} \kappa^{1/\eta}\log\bigl|\log\mathbb{P}(V_\omega(v)\leqslant\kappa)\bigr| > 0. \tag{2.3}$$

Then, if $\eta' > \eta$ and $\zeta > 0$, there is a sequence $L_j \to \infty$ for which

$$\mathbb{P}\bigl(E^{L_j}_{\mathit{GS}} \geqslant E_0 + (\log L_j)^{-\eta'}\bigr) \geqslant 1 - \exp\bigl(-\zeta L_j\bigr).$$

Proof. We start with the simple bound

$$E^L_{\mathit{GS}} \geqslant E_0 + \min_{v\in\mathcal{T}^L} V_\omega(v).$$

Then, it is enough to prove that for $\eta' > \eta$ and $\zeta > 0$, there is a sequence $L_j \to \infty$ satisfying

$$\mathbb{P}\bigl(\min_{v\in\mathcal{T}^{L_j}} V_\omega(v) \geqslant (\log L_j)^{-\eta'}\bigr) \geqslant 1 - \exp\bigl(-\zeta L_j\bigr). \tag{2.4}$$

Condition (2.3) implies for any $\eta'' > 0$ such that $\eta < \eta'' < \eta'$ there exists some sequence $\kappa_j \to 0$ satisfying

$$\bigl|\log\mathbb{P}(V_\omega(v)\leqslant\kappa_j)\bigr| \geqslant \exp(\kappa_j^{-1/\eta''}).$$

We can always assume that the κ_j are small enough by removing some elements of the sequence. Letting $L_j = \lceil \exp(\kappa_j^{-1/\eta'}) \rceil$ this implies that for any $\zeta > 0$ there exists some L^* such that for all $L_j > L^*$

$$\left|\log \mathbb{P}(V_\omega(v) \leqslant (\log L_j)^{-\eta'})\right| \geqslant \exp((\log L_j)^{\eta'/\eta''}) \geqslant \zeta L_j.$$

Using the independence of the random variables and the fact that $|\log(1-p)| \leqslant 2p$ for $0 < p \ll 1$, we see that for any $\zeta > 0$ there is some sequence $L_j \to +\infty$ so that

$$\begin{aligned}
\log \mathbb{P}\big(\min_{v \in \mathcal{T}^{L_j}} V_\omega(v) \geqslant (\log L_j)^{-\eta'}\big) &= \sum_{v \in \mathcal{T}^{L_j}} \log\big(1 - \mathbb{P}\big(V_\omega(v) \leqslant (\log L_j)^{-\eta'}\big)\big) \\
&= \#\mathcal{T}^{L_j} \log\big(1 - \mathbb{P}\big(V_\omega(0) \leqslant (\log L_j)^{-\eta'}\big)\big) \\
&\geqslant -2k^{L_j+1}\mathbb{P}\big(V_\omega(0) \leqslant (\log L_j)^{-\eta'}\big) \\
&\geqslant -2k\mathrm{e}^{(\log k)L_j - \zeta L_j} \qquad (2.5)
\end{aligned}$$

In particular, using that (2.5) is small and $\exp(-x) = 1 - x + O(x^2)$, we see that for any $\zeta > 0$ there exists some sequence $L_j \to \infty$ satisfying (2.4). This finishes the proof. ∎

If we assume condition (2.3) with $\eta < 2$, this last result and the methods we introduce later in section 4 can be used to prove that there exists some sequence $E_j' \searrow 0$ for which

$$\limsup_{j\to\infty} \frac{\log\log|\log \mathcal{N}(E_j')|}{\log(E_j' - E_0)} < -\frac{1}{2}.$$

It is thus impossible to obtain the lower bound in proposition 2.7 under this assumption.

2.2. The lower bound on $\mathcal{N}(E)$.

The upper bound on the ground state in proposition 2.1 implicates a lower bound on the integrated density of states $\mathcal{N}$, formulated in proposition 2.7. The strategy of proof is the same as in [Ros12, section 2.1]. Nonetheless, proposition 2.7 improves the prerequisites under which the lower bound holds, cf. remark 2.2.

The following lemma is taken from [Ros12, section 2.1.1] and adapted for trees $\mathcal{T}^L$ instead of balls of the Bethe lattice. Please keep in mind that the integrated density of state $\mathcal{N}$ does not depend on L. Nonetheless, the finite trees are useful to estimate the Laplace transform of $\mathrm{d}\mathcal{N}$.

Lemma 2.4. *For all $L \in \mathbb{N}$, $t > 0$ and $E' \geqslant E_0$, it holds true that*

$$\tilde{\mathcal{N}}(t) \geqslant \mathrm{e}^{-tE'}\mathbb{P}(E_{\mathcal{GS}}^L \leqslant E')/\#\mathcal{T}^L.$$

Due to the change of notation and for the convenience of the reader, we repeat and detail the proof.

Proof. Let $L \in \mathbb{N}$ and $E \in \mathbb{R}$. Denote by $\Pi_E := \mathbf{1}_{(-\infty,E]}(H_\omega)$ the spectral projection of H_ω. According to (1.6), the integrated density of states is given by

$$\mathcal{N}(E) = \mathbb{E}[\langle \delta_0, \Pi_E \delta_0 \rangle] = (\#\mathcal{T}^L)^{-1} \sum_{v \in \mathcal{T}^L} \mathbb{E}[\langle \delta_v, \Pi_E \delta_v \rangle] = (\#\mathcal{T}^L)^{-1} \mathbb{E}[\mathrm{tr}(\mathbf{1}_{\mathcal{T}^L} \Pi_E \mathbf{1}_{\mathcal{T}^L})].$$

For the Laplace transform of $\mathcal{N}$, the spectral theorem gives

$$\tilde{\mathcal{N}}(t) = \int \mathrm{e}^{-\lambda t} \mathrm{d}\mathcal{N}(\lambda) = (\#\mathcal{T}^L)^{-1} \mathbb{E}[\mathrm{tr}(\mathbf{1}_{\mathcal{T}^L} \exp(-tH_\omega) \mathbf{1}_{\mathcal{T}^L})]$$

for $t \geqslant 0$. [Sim05, Theorem 8.9] states

$$\mathrm{tr}(\mathbf{1}_{\mathcal{T}^L} \exp(-tH_\omega) \mathbf{1}_{\mathcal{T}^L}) \geqslant \mathrm{tr}(\exp(-tH_\omega^L)), \tag{2.6}$$

where $H_\omega^L := \mathbf{1}_{\mathcal{T}^L} H_\omega \mathbf{1}_{\mathcal{T}^L}$. This is easily seen with help of spectral measures. Due to the convexity of $\lambda \mapsto \mathrm{e}^{-t\lambda}$, for each $v \in \mathcal{T}^L$, Jensen's inequality gives

$$\begin{aligned} \langle \delta_v, \exp(-tH_\omega) \delta_v \rangle &= \int \exp(-t\lambda)\, \mathrm{d}\mu_{\delta_v}(\lambda) \geqslant \exp\Big(-t \int \lambda\, \mathrm{d}\mu_{\delta_v}(\lambda)\Big) \\ &= \exp(-t\langle \delta_v, H_\omega \delta_v \rangle) = \exp(-t\langle \delta_v, \mathbf{1}_{\mathcal{T}^L} H_\omega \mathbf{1}_{\mathcal{T}^L} \delta_v \rangle) \\ &= \langle \delta_v, \exp(-tH_\omega^L) \delta_v \rangle, \end{aligned}$$

where μ_{δ_v} is the spectral measure of H_ω with respect to δ_v. Summing over $v \in \mathcal{T}^L$, we obtain (2.6). The Laplace transform is thus bounded by

$$\tilde{\mathcal{N}}(t) \geqslant (\#\mathcal{T}^L)^{-1} \mathbb{E}[\mathrm{tr}(\exp(-tH_\omega^L))] \geqslant (\#\mathcal{T}^L)^{-1} \mathbb{E}[\exp(-tE_{GS}^L)].$$

The Markov inequality reduces the last expectation to a probability

$$\mathbb{P}(E_{GS}^L \leqslant E') \leqslant \mathrm{e}^{tE'} \mathbb{E}[\exp(-tE_{GS}^L)]$$

and finishes the proof. ■

The following corollary provides the inequality on the left hand side of (1.2).

Corollary 2.5. *We assume* (2.1) *and let* C_1 *be the constant in proposition 2.1. Then we have for* $t > 0$ *large enough that*

$$\tilde{\mathcal{N}}(t) \geqslant \frac{1}{2} \exp\left(-t\Big(E_0 + \frac{5C_1}{(\log t)^2}\Big)\right)$$

Proof. We choose $L = \lfloor \sqrt{t} \rfloor$. Note that for $t > 1$

$$\#\mathcal{T}^L = \sum_{i=0}^{L-1} k^i \leqslant k^L \leqslant \mathrm{e}^{\sqrt{t} \log k}.$$

From proposition 2.1 we see that $\mathbb{P}(E^L_{\mathcal{GS}} \leqslant E_0 + C_1(\log L)^{-2}) \geqslant 1/2$ for $t > 0$ large enough. Now use lemma 2.4 with the choice $E' := E_0 + C_1(\log L)^{-2} > E_0$ to obtain

$$\begin{aligned}\tilde{\mathcal{N}}(t) &\geqslant \exp\bigl(-t(E_0 + C_1(\log L)^{-2})\bigr)\mathbb{P}(E^L_{\mathcal{GS}} \leqslant E_0 + C_1(\log L)^{-2})/\#\mathcal{T}^L \\ &\geqslant \frac{1}{2}\exp\bigl(-t(E_0 + C_1(\log\lfloor\sqrt{t}\rfloor)^{-2}) - \sqrt{t}\log k\bigr) \\ &\geqslant \frac{1}{2}\exp\bigl(-t(E_0 + 5C_1(\log t)^{-2})\bigr).\end{aligned}$$

In the last line we have used that

$$\frac{t(C_1(\log\lfloor\sqrt{t}\rfloor)^{-2}) + \sqrt{t}\log k}{t(5C_1(\log t)^{-2})} \to 4/5 < 1$$

as $t \to +\infty$. ■

As known from Tauberian theorems, the behavior of $\tilde{\mathcal{N}}(t)$ as $t \to \infty$ and the behavior of $\mathcal{N}(E)$ as $E \searrow E_0$ are related. The following is taken almost verbatim from [Ros12, (2.27)].

Lemma 2.6. *For all $t > 0$ and $E \geqslant E_0$, it holds true that*

$$\mathcal{N}(E) \geqslant \mathrm{e}^{tE_0}\tilde{\mathcal{N}}(t) - \mathrm{e}^{-t(E-E_0)}.$$

For completeness, we give the short proof.

Proof. Integration by parts, with vanishing boundary terms since $\mathcal{N}(E_0) = 0$, gives

$$\begin{aligned}\tilde{\mathcal{N}}(t) &= \int_{E_0}^{\infty} \mathrm{e}^{-t\lambda}\mathrm{d}\mathcal{N}(\lambda) = \int_{E_0}^{\infty} t\mathrm{e}^{-t\lambda}\mathcal{N}(\lambda)\,\mathrm{d}\lambda \\ &\leqslant \mathcal{N}(E)\int_{E_0}^{E} t\mathrm{e}^{-t\lambda}\,\mathrm{d}\lambda + \int_{E}^{\infty} t\mathrm{e}^{-t\lambda}\,\mathrm{d}\lambda \leqslant \mathrm{e}^{-tE_0}\mathcal{N}(E) + \mathrm{e}^{-tE}.\end{aligned}$$

This is equivalent to the claim. ■

Together with proposition 2.1, lemmas 2.4 and 2.6 are all that is needed to prove the lower bound of the Lifshitz tails. More precisely, we obtain the following.

Proposition 2.7. *Assume (2.1) and fix C_1 as in proposition 2.1. Then there exists $\lambda > E_0$ such that, for all $E \in (E_0, \lambda)$, it holds true that*

$$\mathcal{N}(E) \geqslant k^{-2-2\exp\left(\sqrt{2C_1/(E-E_0)}\right)}/16.$$

In particular,

$$\liminf_{E\searrow E_0} \frac{\log\log|\log\mathcal{N}(E)|}{\log(E-E_0)} \geqslant -\frac{1}{2}.$$

Proof. Lemmas 2.4 and 2.6 concatenate to

$$\begin{aligned}\mathcal{N}(E) &\geqslant (\#\mathcal{T}^L)^{-1}\mathrm{e}^{-t(E'-E_0)}\mathbb{P}(E^L_{\mathcal{GS}} \leqslant E') - \mathrm{e}^{-t(E-E_0)} \\ &= \bigl((\#\mathcal{T}^L)^{-1}\mathrm{e}^{t(E-E')}\mathbb{P}(E^L_{\mathcal{GS}} \leqslant E') - 1\bigr)\mathrm{e}^{-t(E-E_0)},\end{aligned}$$

which is true for all $t > 0$ and $E, E' > E_0$. We choose $E' := (E_0 + E)/2$. This ensures $E - E' > 0$ and will enable us to choose t large enough to make the lower bound positive.

But first we have to deal with the probability. In order to apply proposition 2.1, we let $L := \lceil \exp\big(\sqrt{2C_1/(E - E_0)}\big) \rceil$. That way, $E' \geqslant E_0 + C_1/(\log L)^2$ and, provided $E - E_0$ is small enough so that L is large enough,

$$\mathbb{P}\big(E^L_{GS} \leqslant E'\big) \geqslant \mathbb{P}\big(E^L_{GS} \leqslant E_0 + C_1(\log L)^{-2}\big) \geqslant 1 - \exp(-k^{L/2}) \geqslant 1/2.$$

Up to now we know, for all $t > 0$ and L large enough,

$$\mathcal{N}(E) \geqslant \big((2\#\mathcal{T}^L)^{-1} \mathrm{e}^{t(E-E_0)/2} - 1\big) \mathrm{e}^{-t(E-E_0)}.$$

It is time to choose $t := 2\log(4\#\mathcal{T}^L)/(E - E_0)$, that is, $(2\#\mathcal{T}^L)^{-1} \mathrm{e}^{t(E-E_0)/2} = 2$ and

$$\mathcal{N}(E) \geqslant \exp\big(-2\log(4\#\mathcal{T}^L)\big) = \frac{(\#\mathcal{T}^L)^{-2}}{16} \geqslant \frac{k^{-2L}}{16} \geqslant \frac{1}{16} k^{-2-2\exp\left(\sqrt{2C_1/(E-E_0)}\right)}.$$

It is now easy to read the exponent of $E - E_0$ from the limit inferior of this lower bound on the Lifshitz tail behavior, i. e. $-1/2$. This finishes the proof. ■

3. Deterministic preparations

We develop the spectral theory of finite rooted trees. The spectrum was already calculated in [RR07], but we need the eigenfunctions, too. The radially symmetric generalized eigenfunctions for the (infinite) Bethe lattice were calculated in [Bro91].

Recall that we denote by $\mathcal{T}^L$ the (nodes of the) rooted tree of length L with k children at each node except the leaves, by 0 the root of the tree and by $|v| = d(0, v) + 1$ the "level" of the node v. For indexing reasons, we introduce the notation $\mathcal{T}^L_* := \mathcal{T}^L \uplus \{*\}$ for the (nodes of the) rooted tree of length L augmented by a vertex $*$ with $|*| = 0$, such that $*$ is a parent of the root. Any function in $\ell^2(\mathcal{T}^L)$ is understood as an element of $\ell^2(\mathcal{T}^L_*)$, too, with the value 0 on $*$.

3.1. The spectrum of the adjacency matrix on a finite rooted tree.

Lemma 3.1. *For each $m \in \{1, \ldots, L\}$, the radially symmetric function defined by*

$$\mathcal{T}^L \ni v \mapsto \psi^L_m(v) = \sqrt{\frac{2}{(L+1)k^{|v|-1}}} \sin\Big(\frac{m\pi}{L+1}|v|\Big)$$

is a normalized eigenfunction of the adjacency matrix $A^{(L)}$ of the rooted tree $\mathcal{T}^L$ with eigenvalue

$$\lambda^L_m := 2\sqrt{k} \cos\Big(\frac{m\pi}{L+1}\Big).$$

Proof. Let $\theta := \frac{m\pi}{L+1} \in \mathbb{R}$. We check first the eigenvalue equation for $v \in \mathcal{T}^L$:

$$\begin{aligned} A^{(L)}\psi_m^L(v) &= \sum_{w\in\mathcal{T}^L, w\sim v} \psi_m^L(w) \\ &= \sqrt{\tfrac{2}{(L+1)k^{|v|-2}}}\sin((|v|-1)\theta) + k\sqrt{\tfrac{2}{(L+1)k^{|v|}}}\sin((|v|+1)\theta) \\ &= \sqrt{\tfrac{2}{(L+1)k^{|v|-2}}}\cdot 2\sin(|v|\theta)\cos(\theta) = \lambda_m^L\psi_m^L(v). \end{aligned}$$

The third equation employs $\sin(\alpha+\beta) = \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta)$, $\alpha, \beta \in \mathbb{R}$. We check now that they are normalized. This is seen via

$$\begin{aligned} \|\psi_m^L\|_2^2 = \sum_{v\in\mathcal{T}^L} |\psi_m^L(v)|^2 &= \sum_{\ell=1}^{L} \frac{2\sin^2(\theta\ell)}{L+1} = \frac{1}{2(L+1)}\sum_{\ell=1}^{L}\big(2 - \mathrm{e}^{2\mathrm{i}\theta\ell} - \mathrm{e}^{-2\mathrm{i}\theta\ell}\big) \\ &= \frac{1}{2(L+1)}\Big(2L - \frac{\mathrm{e}^{2\mathrm{i}\theta} - \mathrm{e}^{2\mathrm{i}(L+1)\theta}}{1-\mathrm{e}^{2\mathrm{i}\theta}} - \frac{\mathrm{e}^{-2\mathrm{i}\theta} - \mathrm{e}^{-2\mathrm{i}(L+1)\theta}}{1-\mathrm{e}^{-2\mathrm{i}\theta}}\Big) = 1, \end{aligned}$$

where we used $\mathrm{e}^{\pm 2\mathrm{i}(L+1)\theta} = \mathrm{e}^{\pm 2\pi\mathrm{i}m} = 1$ in the last step. ∎

Since the radially symmetric functions on $\mathcal{T}^L$ form a linear subspace of $\ell^2(\mathcal{T}^L)$ of dimension L, lemma 3.1 lists all radially symmetric eigenfunctions of $A^{(L)}$. We now construct the remaining non-radially symmetric eigenfunctions on $\mathcal{T}^L$. Recall that, for each $v \in \mathcal{T}^L$, we denote by $\mathcal{T}_v^L$ the subtree of $\mathcal{T}^L$ rooted at v and of length $L - |v| + 1$.

Let $v \in \mathcal{T}^{L-1} \subseteq \mathcal{T}^L$ and $u \in \mathcal{T}_v^L$, $u \sim v$. The node u is the root of a subtree $\mathcal{T}_u^L$ isomorphic to $\mathcal{T}^{L-|v|}$. According to lemma 3.1, we have $L - |v|$ radially symmetric eigenfunctions $\psi_{u,m}^{L-|v|}$, $m \in \{1, \dots, L-|v|\}$, of the adjacency matrix of $\mathcal{T}_u^L$, given by

$$\psi_{u,m}^{L-|v|}(w) = \sqrt{\frac{2}{(L+1-|v|)k^{|w|-|v|-1}}}\sin\Big(\frac{m\pi}{L+1-|v|}(|w|-|v|)\Big) \tag{3.1}$$

for $w \in \mathcal{T}_u^L$. We trivially extend $\psi_{u,m}^{L-|v|}$ to a function on $\mathcal{T}^L$ by assigning 0 to the complement of $\mathcal{T}_u^L$. For a given $v \in \mathcal{T}^L$, we will agglutinate below the functions $\psi_{u,m}^{L-|v|}$, $u \in \mathcal{T}_v^L$, $u \sim v$, at v, see (3.2).

Note that $\mathcal{T}_v^{|v|+1} = \{v\} \cup \{u \in \mathcal{T}_v^L : u \sim v\}$ is isomorphic to $\mathcal{T}^2$ as a graph. The matrix representation of $A^{(2)}$ with respect to a basis $(\delta_v; v \in \mathcal{T}^2)$ with the root as the first entry is

$$\begin{pmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix}$$

with dimensions $(k+1)\times(k+1)$. The kernel of $A^{(2)}$ on $\mathcal{T}^2$ has dimension $k-1$, so we can find $k-1$ normalized and orthogonal real eigenvectors $\psi^{\perp}_{v,j}$, $j \in \{1,\dots,k-1\}$, of $A^{(2)}$ associated to the eigenvalue 0 on $\mathcal{T}_v^{|v|+1}$. These eigenvectors assign the value 0 to v, since for any $u \in \mathcal{T}_v^{|v|+1}$, $u \neq v$, we have

$$\psi^{\perp}_{v,j}(v) = A^{(2)}\psi^{\perp}_{v,j}(u) = 0 \cdot \psi^{\perp}_{v,j}(u) = 0.$$

We set

$$\Psi^L_{v,j,m} := \sum_{u \in \mathcal{T}_v^L, u \sim v} \psi^{\perp}_{v,j}(u)\psi^{L-|v|}_{u,m}. \tag{3.2}$$

To unify notation, we define $\Psi^L_{*,1,m} := \psi^L_m$ and $\psi^{\perp}_{*,1}(v) := 1$ for the root v of $\mathcal{T}^L$, too, as well as

$$J_v := \begin{cases} \{1\} & \text{for } v = * \text{ and} \\ \{1,\dots,k-1\} & \text{if } v \in \mathcal{T}^{L-1}. \end{cases}$$

We call a triple (v,j,m) *L-admissible* if $v \in \mathcal{T}_*^{L-1}$, $j \in J_v$, $m \in \{1,\dots,L-|v|\}$.

Lemma 3.2. *The vectors $\Psi^L_{v,j,m}$ with (v,j,m) L-admissible are normalized eigenvectors of $A^{(L)}$ with eigenvalues $\lambda^L_{v,j,m} := \lambda^{L-|v|}_m = 2\sqrt{k}\cos\left(\frac{m\pi}{L+1-|v|}\right)$, respectively, and form an orthonormal basis of $\ell^2(\mathcal{T}^L)$.*

Proof. Let (v,j,m) be a L-admissible. In the case $v = *$, lemma 3.1 tells us that $\Psi^L_{v,j,m} = \psi^L_m$ is a normalized eigenfunction of $A^{(L)}$. From now on, we let $v \in \mathcal{T}^{L-1}$. Note that

$$\sum_{u \in \mathcal{T}_v^L, u \sim v} \psi^{\perp}_{v,j}(u) = A^{(2)}\psi^{\perp}_{v,j}(v) = 0 \cdot \psi^{\perp}_{v,j}(v) = 0. \tag{3.3}$$

Since $\Psi^L_{v,j,m}$ is pieced together from eigenfunctions on trees with the same eigenvalue, the only node we need to check is v itself. We use (3.3) to see that

$$A^{(L)}\Psi^L_{v,j,m}(v) = \sum_{u \in \mathcal{T}_v^L, w \sim v} \psi^{\perp}_{v,j}(u)\psi^{L-|v|}_{u,m}(u) = 0 = \lambda^L_{v,j,m}\Psi^L_{v,j,m}(v).$$

Thus, all $\Psi^L_{v,j,m}$ are eigenfunctions of $A^{(L)}$.

Orthonormality is our next goal. For $v \in \mathcal{T}^{L-1}$, $m \in \{1,\dots,L-|v|\}$, $m' \in \{1,\dots,L\}$ and $j \in \{1,\dots,k-1\}$, we have

$$\langle \Psi^L_{*,1,m'}, \Psi^L_{v,j,m}\rangle = \sum_{u \in \mathcal{T}_v^L, u \sim v} \psi^{\perp}_{v,j}(u)\langle \psi^L_{m'}, \psi^{L-|v|}_{u,m}\rangle = 0,$$

since $\langle \psi^L_{m'}, \psi^{L-|v|}_{u,m}\rangle$ is constant in u and (3.3). For (v,j,m) and (v',j',m') L-admissible with $v, v' \in \mathcal{T}^{L-1}$ we distinguish the following cases.

- If $\lambda^L_{v,j,m} \neq \lambda^L_{v',m',j'}$, then $\langle \Psi^L_{v,j,m}, \Psi^L_{v',m',j'}\rangle = 0$, since $A^{(L)}$ is symmetric.

- Let $v \neq v'$. If $v \in \mathcal{T}_{v'}^L$ or $v' \in \mathcal{T}_v^L$, then the argument from above for $v' = *$ applies. If v and v' have disjoint subtrees, then the supports of $\Psi_{v,j,m}^L$ and $\Psi_{v',m',j'}^L$ are disjoint. Either way we reach $\langle \Psi_{v,j,m}^L, \Psi_{v',m',j'}^L \rangle = 0$.
- Assume $v = v'$, $\lambda_{v,j,m}^L = \lambda_{v',m',j'}^L$. We thus have $\cos(\frac{m\pi}{L+1-|v|}) = \cos(\frac{m'\pi}{L+1-|v|})$. Since $\frac{m\pi}{L-|v|+1} \in (0,\pi)$ and $\cos|_{(0,\pi)}$ is injective, we deduce $m = m'$. Consequently,
$$\begin{aligned}\langle \Psi_{v,j,m}^L, \Psi_{v,j',m}^L \rangle &= \sum_{u,u' \in \mathcal{T}_v^L, u,u' \sim v} \overline{\psi_{v,j}^\perp(u)} \psi_{v,j'}^\perp(u') \langle \psi_{u,m}^{L-|v|}, \psi_{u',m}^{L-|v|} \rangle \\ &= \sum_{u \in \mathcal{T}_v^L, u \sim v} \overline{\psi_{v,j}^\perp(u)} \psi_{v,j'}^\perp(u) = \delta_{j,j'},\end{aligned}$$
since $\psi_{v,j}^\perp$ and $\psi_{v,j'}^\perp$ are orthonormal and $\psi_{v,j}^\perp(v) = 0$.

We now know that the set of all $\Psi_{v,j,m}^L$ with (v,j,m) L-admissible is orthonormal.

To identify this orthonormal set as a basis, we simply count all L-admissible triples:
$$\begin{aligned}&\sum_{v \in \mathcal{T}_*^{L-1}} \sum_{j \in J_v} \sum_{m=1}^{L-|v|} 1 = \sum_{m=1}^{L} 1 + \sum_{v \in \mathcal{T}^{L-1}} \sum_{j=1}^{k-1} \sum_{m=1}^{L-|v|} 1 \\ &= L + \sum_{\ell=1}^{L-1} k^{\ell-1}(k-1)(L-\ell) = L + (k-1)\Big(L \sum_{\ell=1}^{L-1} k^{\ell-1} - \sum_{\ell=1}^{L-1} \ell k^{\ell-1}\Big) \\ &= L + (k-1)\Big(L\frac{k^{L-1}-1}{k-1} - \frac{Lk^{L-1}(k-1) - (k^L-1)}{(k-1)^2}\Big) = \frac{k^L-1}{k-1}.\end{aligned}$$
This is exactly the dimension $\#\mathcal{T}^L = \sum_{\ell=1}^L k^{\ell-1} = \frac{k^L-1}{k-1}$ of $\ell^2(\mathcal{T}^L)$. ■

We study the behavior of the principal eigenvalue $\lambda_{*,1,1}^L$ of $A^{(L)}$ as a function of L and identify the states in its vicinity. This will be used in section 5, and it is a crucial part of our argument.

Lemma 3.3. *Let $L \in \mathbb{N}$. For $\beta \in \mathbb{R}$ we define $E_\beta^{(L)} := 2\sqrt{k}\cos\big(\frac{(\beta+1)\pi}{L+1}\big)$. For L-admissible (v,j,m) and $\beta \in [0,L]$, we have*
$$\lambda_{v,j,m}^L \in [E_\beta^{(L)}, \lambda_{*,1,1}^L] \iff |v| \leqslant (L+1)\big(1 - \tfrac{m}{\beta+1}\big) \implies m \in \{1, \dots, \lfloor \beta+1 \rfloor\}.$$

Remark 3.4. Note that $E_0^{(L)} = E_\beta^{(L)}|_{\beta=0} = \lambda_{*,1,1}^L$.

Proof. Remember that $1 - \frac{x^2}{2} \leqslant \cos x \leqslant 1 - \frac{x^2}{2} + \frac{x^4}{24}$ for all $x \in \mathbb{R}$. This reveals
$$-\frac{1}{24}\Big(\frac{(\beta+1)\pi}{L+1}\Big)^4 \leqslant \cos\Big(\frac{\pi}{L+1}\Big) - \cos\Big(\frac{(\beta+1)\pi}{L+1}\Big) - \frac{\pi^2\beta(\beta+2)}{2(L+1)^2} \leqslant \frac{1}{24}\Big(\frac{\pi}{L+1}\Big)^4.$$

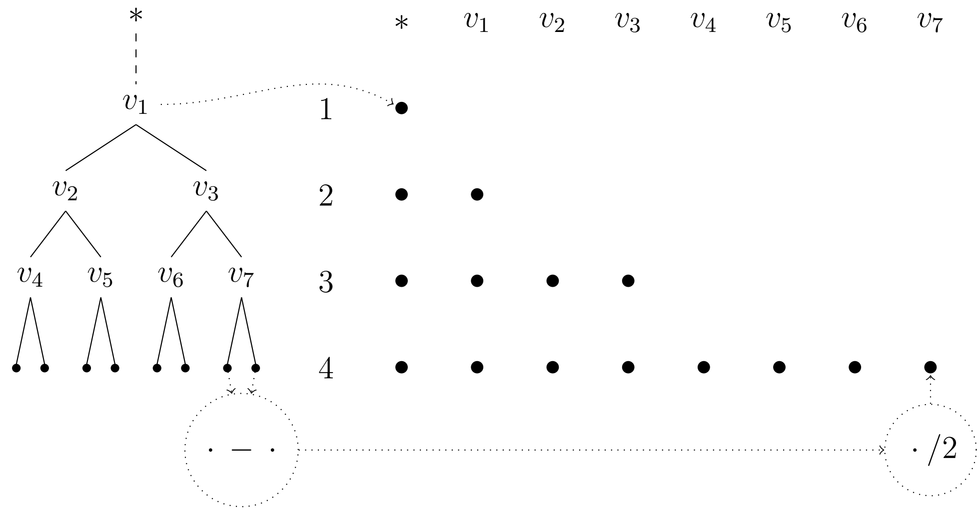


FIGURE 4. The action of the map $\widehat{\cdot}$ is indicated with the dotted arrows.

For $\beta \leqslant L$, we use that $\cos|_{[0,\pi]}$ is strictly decreasing to obtain

$$\lambda^L_{v,j,m} \geqslant E^{(L)}_\beta \iff \frac{m}{L+1-|v|} \leqslant \frac{\beta+1}{L+1} \iff \frac{m}{\beta+1} + \frac{|v|}{L+1} \leqslant 1$$
$$\iff |v| \leqslant (L+1)\Big(1 - \frac{m}{\beta+1}\Big) \implies m \leqslant \lfloor \beta + 1 \rfloor. \qquad \blacksquare$$

Next, we study the spectral projections

$$P_{v,j}\colon \ell^2(\mathcal{T}^L) \to \ell^2(\mathcal{T}^L), \quad P_{v,j}\varphi := \sum_{m=1}^{L-|v|} \langle \varphi, \Psi^L_{v,j,m}\rangle \Psi^L_{v,j,m} \tag{3.4}$$

of $A^{(L)}$ for $v \in \mathcal{T}^{L-1}_*$ and $j \in J_v$. We introduce the map

$$\widehat{\cdot}\colon \ell^2(\mathcal{T}^L) \to \bigoplus_{v\in\mathcal{T}^{L-1}_*} \bigoplus_{j\in J_v} \ell^2(\{|v|+1,\dots,L\}),$$
$$\widehat{\psi}_{v,j}(z) := k^{-(z-|v|-1)/2} \sum_{u\in\mathcal{T}^L_v, u\sim v} \overline{\psi^\perp_{v,j}(u)} \sum_{w\in\mathcal{T}^L_u, |w|=z} (P_{v,j}\psi)(w) \tag{3.5}$$

for $z \in \{|v|+1,\dots,L\}$. For a rough illustration see fig. 4. The map $\widehat{\cdot}$ has been sketched in [AW06, Proposition A.2] and it is similar to an infinite dimensional version in [AF00].

The action of $\widehat{\cdot}$ is best illustrated on radially symmetric eigenfunctions ψ^L_m of $A^{(L)}$. As we will see in lemma 3.5, they are mapped to functions supported on $\{1,\dots,L\}$, and in the process, the exponential weights $k^{(|v|-1)/2}$ are removed:

$$(\widehat{\psi^L_m})_{*,1}(z) = \sqrt{\frac{2}{L+1}} \sin\Big(\frac{m\pi}{L+1} z\Big)$$

for $z \in \{1, \dots, L\}$ and $m \in \{1, \dots, L\}$. The result is an eigenfunction of the adjacency matrix of $\mathbb{Z}$ restricted to $\{1, \dots, L\}$. Given a non-radially symmetric eigenfunction $\Psi^L_{v,j,m}$ of $A^{(L)}$, $\widehat{\cdot}$ reconstructs the underlying radially symmetric eigenfunction, removes the exponential weight and presents the result as a function on the copy of $\{|v|+1, \dots, L\}$ which is indexed by (v, j):

$$(\widehat{\Psi^L_{v,j,m}})_{v,j}(z) = \sqrt{\frac{2}{L+1-|v|}} \sin\Big(\frac{m\pi}{L+1-|v|}(z-|v|)\Big) \tag{3.6}$$

for $z \in \{|v|+1, \dots, L\}$ and (v, j, m) L-admissible. This is again an eigenfunction of the adjacency matrix of $\mathbb{Z}$ restricted to $\{|v|+1, \dots, L\}$.

We define the adjacency matrix on the image of $\widehat{\cdot}$, which is the Hilbert sum of the ℓ^2-spaces of segments of $\mathbb{Z}$. The direct sum of the adjacency matrices of the segments of $\mathbb{Z}$ is the natural choice. For $\varphi \in \bigoplus_{v,j} \ell^2(\{|v|+1, \dots, L\})$, it is given by

$$(\widehat{A}\varphi)_{v,j}(z) := (A_{\mathbb{Z}}\varphi_{v,j})(z) = \varphi_{v,j}(z-1) + \varphi_{v,j}(z+1)$$

for $v \in \mathcal{T}^{L-1}$, $j \in J_v$, $z \in \{|v|+1, \dots, L\}$, and with the boundary values $\varphi_{v,j}(|v|) := \varphi_{v,j}(L+1) := 0$.

Lemma 3.5. *For all $\psi \in \ell^2(\mathcal{T}^L)$, we have the following.*

(i) *The map $\widehat{\cdot}$ conjugates $A^{(L)}$ and $\sqrt{k}\widehat{A}$: $\widehat{A^{(L)}\psi} = \sqrt{k}\widehat{A}\widehat{\psi}$.*

(ii) *The map $\widehat{\cdot}$ is unitary: $\|\psi\|_2 = \|\widehat{\psi}\|_2$. In particular, $\sigma(A^{(L)}) = \sqrt{k}\sigma(\widehat{A})$.*

(iii) *Let $v \in \mathcal{T}^{L-1}_*$ and $j \in J_v$. The subspace $P_{v,j}\ell^2(\mathcal{T}^L)$ contains ψ if and only if* $\operatorname{supp}(\psi) \subseteq \mathcal{T}^L_v \setminus \{v\}$ *and*

$$\psi(w)\psi^{\perp}_{j,v}(u') = \psi(w')\psi^{\perp}_{j,v}(u)$$

for all $u, u' \in \mathcal{T}^L_v$ with $u, u' \sim v$ and all $w \in \mathcal{T}^L_u, w' \in \mathcal{T}^L_{u'}$ such that $|w| = |w'|$.

(iv) *For all radially symmetric functions $\eta\colon \mathcal{T}^L \to \mathbb{C}$, i. e. $\eta(w) = \eta_{\mathbb{Z}}(|w|)$ for all $w \in \mathcal{T}^L$ and an $\eta_{\mathbb{Z}}\colon \{1, \dots, L\} \to \mathbb{C}$, we have $P_{v,j}\eta = \eta P_{v,j}$ and $(\widehat{\eta\psi})_{v,j} = \eta_{\mathbb{Z}}\widehat{\psi}_{v,j}$ for all $v \in \mathcal{T}^{L-1}_*$, $j \in J_v$. Here, $\eta_{\mathbb{Z}}$ denotes the multiplication with the function $\eta_{\mathbb{Z}}|_{\{|v|+1,\dots,L\}}$.*

Proof. Ad (i). We study the linear map $\widehat{\cdot}$ on the orthonormal basis $\Psi^L_{v,j,m}$. To this end, let (v, j, m) be admissible, $v' \in \mathcal{T}^{L-1}_*$, $j' \in J_{v'}$ and $z \in \{|v'|+1, \dots, L\}$. For $v' \neq v$ or $j' \neq j$, we have $P_{v',j'}\Psi^L_{v,j,m} = 0$ and ergo $(\widehat{\Psi^L_{v,j,m}})_{v',j'}(z) = 0$, too. So from now on, we assume $v' = v$ and $j' = j$. We then have $P_{v,j}\Psi^L_{v,j,m} = \Psi^L_{v,j,m}$. For $u \in \mathcal{T}^L_v$, $u \sim v$ and $w \in \mathcal{T}^L_u$, we find

$$\Psi^L_{v,j,m}(w) = \sum_{u' \in \mathcal{T}^L_v, u' \sim v} \psi^{\perp}_{v,j}(u')\psi^{L-|v|}_{u',m}(w) = \psi^{\perp}_{v,j}(u)x_{|w|,|v|,m,L}$$

with $x_{|w|,|v|,m,L} := \sqrt{\frac{2}{(L+1-|v|)k^{|w|-|v|-1}}}\sin\big(\frac{m\pi}{L+1-|v|}(|w|-|v|)\big)$. We now see

$$\begin{aligned}
(\widehat{\Psi^L_{v,j,m}})_{v,j}(z) &= k^{-(z-|v|-1)/2}\sum_{u\in\mathcal{T}^L_v, u\sim v}\overline{\psi^\perp_{v,j}(u)}\sum_{w\in\mathcal{T}^L_u,|w|=z}\Psi^L_{v,j,m}(w)\\
&= k^{-(z-|v|-1)/2}\sum_{u\in\mathcal{T}^L_v, u\sim v}\overline{\psi^\perp_{v,j}(u)}\psi^\perp_{v,j}(u)\sum_{w\in\mathcal{T}^L_u,|w|=z}x_{z,|v|,m,L}\\
&= k^{-(z-|v|-1)/2}k^{z-|v|-1}x_{z,|v|,m,L}\\
&= \sqrt{\frac{2}{L+1-|v|}}\sin\Big(\frac{m\pi}{L+1-|v|}(z-|v|)\Big).
\end{aligned}$$

We now identify $\widehat{\Psi^L_{v,j,m}}$ as an eigenfunction of $\widehat{A}$. Let $\varphi := \frac{m\pi}{L+1-|v|}$ and note that, for $z\in\{|v|_1,\dots,L\}$, by the angle sum and difference identities,

$$\begin{aligned}
&\sin(\varphi(z-1-|v|))+\sin(\varphi(z+1-|v|))\\
&=\sin(\varphi(z-|v|))\cos(\varphi)+\cos(\varphi(z-|v|))\sin(\varphi)\\
&\quad+\sin(\varphi(z-|v|))\cos(\varphi)-\cos(\varphi(z-|v|))\sin(\varphi)\\
&=2\cos(\varphi)\sin(\varphi(z-|v|)).
\end{aligned}$$

The boundary values $\sin(\varphi(|v|-|v|))=0$ and $\sin(\varphi(L+1-|v|))=0$ are satisfied, too. Thus, $\sqrt{k}\widehat{A}\widehat{\Psi^L_{v,j,m}} = \lambda^L_{v,j,m}\widehat{\Psi^L_{v,j,m}} = \widehat{\lambda^L_{v,j,m}\Psi^L_{v,j,m}} = \widehat{A^{(L)}\Psi^L_{v,j,m}}$ for all L-admissible (v,j,m).

Ad (ii). We have to check that the image of an orthonormal basis is again an orthonormal basis. Let (v,j,m) be admissible. The fact that $\|\widehat{\Psi^L_{v,j,m}}\|_2^2=1$ is seen exactly as the normalization part in lemma 3.1. Let (v',j',m') be another admissible triple. For $(v,j)\neq(v',j')$, $\widehat{\Psi^L_{v,j,m}}$ and $\widehat{\Psi^L_{v,j,m}}$ have disjoint support and are thus orthogonal. In case $(v,j)=(v',j')$ and $m\neq m'$, $\widehat{\Psi^L_{v,j,m}}$ and $\widehat{\Psi^L_{v,j,m}}$ are orthogonal, too, since the corresponding eigenvalues $\lambda^L_{v,j,m}\neq\lambda^L_{v',j',m'}$ with respect to the symmetric operator $\sqrt{k}\widehat{A}$ are not equal. Finally, $\widehat{\cdot}$ is surjective, since the dimensions of its preimage and its image agree.

Ad (iii). Fix $v\in\mathcal{T}^{L-1}_*$ and $j\in J_v$. We denote the linear subspace defined by the condition in (iii) by $\mathcal{D}_{v,j}$. By construction, $\Psi^L_{v,j,m}\in\mathcal{D}_{v,j}$, so $P_{v,j}\ell^2(\mathcal{T}^L)\subseteq\mathcal{D}_{v,j}$. Furthermore, $\dim(\mathcal{D}_{v,j})=L-|v|$, since the condition allows one degree of freedom per sphere of $\mathcal{T}^L_v\setminus\{v\}$. On the other hand, $\dim(P_{v,j}\ell^2\mathcal{T}^L_v)=L-|v|$, because the vectors $\Psi^L_{v,j,m}$, $m\in\{|v|+1,\dots,L\}$, are a basis of $P_{v,j}\ell^2\mathcal{T}^L_v$. The statement follows.

Ad (iv). Let $\psi\in\ell^2(\mathcal{T}^L)$ and $\eta,\eta_{\mathbb{Z}}$ be given as in the statement. Because of (iii), $\eta P_{v,j}\psi\in P_{v,j}\ell^2(\mathcal{T}^L)$. This and the fact that the spectral projectors are orthogonal implies that $P_{v,j}$ and multiplication with $\eta_{\mathbb{Z}}$ commute:

$$P_{v,j}(\eta\psi) = P_{v,j}\Big(\eta \sum_{v',j'} P_{v',j'}\psi\Big) = \sum_{v',j'} P_{v,j}(\eta P_{v',j'}\psi) = \eta P_{v,j}\psi.$$

We use this in

$$\begin{aligned}\widehat{\eta\psi}_{v,j}(z) &= k^{-(z-|v|-1)/2} \sum_{u\in\mathcal{T}_v^L, u\sim v} \overline{\psi_{v,j}^{\perp}(u)} \sum_{w\in\mathcal{T}_u^L, |w|=z} (P_{v,j}(\eta\psi))(w) \\ &= \eta_{\mathbb{Z}}(z) k^{-(z-|v|-1)/2} \sum_{u\in\mathcal{T}_v^L, u\sim v} \overline{\psi_{v,j}^{\perp}(u)} \sum_{w\in\mathcal{T}_u^L, |w|=z} (P_{v,j}\psi)(w) \\ &= \eta_{\mathbb{Z}}(z)\widehat{\psi}_{v,j}(z). \qquad \blacksquare\end{aligned}$$

3.2. **The IMS localization formula.** In this subsection we provide a proof of the following proposition. It will be needed in section 4 (proposition 4.9).

Proposition 3.6 (IMS localization formula)**.** *There is a constant $C_{\mathit{IMS}} > 0$ such that for each $r > 2$, we have a partition of unity $\{\eta_{a,r}\}_{a\geqslant 0} \subseteq \ell^2(\mathcal{T}^L)$, consisting of radially symmetric functions normalized to $\sum_{a\geqslant 0}\eta_{a,r}^2 = 1$, such that for all $\psi \in \ell^2(\mathcal{T}^L)$ we have*

$$\big\langle A^{(L)}\psi, \psi\big\rangle \geqslant \sum_{a\geqslant 0}\big\langle A^{(L)}(\eta_{a,r}\psi), \eta_{a,r}\psi\big\rangle - \frac{C_{\mathit{IMS}}}{r^2}\|\psi\|_2^2.$$

Furthermore, the support of $\eta_{a,r}$ is a union of disjoint trees of length at most r.

The proof of proposition proposition 3.6 is made in two steps. We first prove this formula for the discrete, one-dimensional Laplacian. Then, we carry this formula onto the tree by means of the spectral theory of the rooted tree.

3.2.1. *The IMS localization formula on $\mathbb{Z}$.* In this subsection we consider the discrete Laplacian

$$\Delta_{\mathbb{Z}} := \tau^{-1} - 2 + \tau$$

on $\ell^2(\mathbb{Z})$, where τ is the translation operator, i. e. given by $(\tau f)(x) = f(x+1)$ for $f\colon \mathbb{Z} \to \mathbb{C}$ and $x \in \mathbb{Z}$. Note that on $\ell^2(\mathbb{Z})$ we have $\tau^{-1} = \tau^*$. We will also employ the discrete gradient

$$\nabla := \tau - 1.$$

Lemma 3.7. *Let $f \in \ell^2(\mathbb{Z})$. For any partition of unity $\{\eta_a\}$, normalized so that $\sum_a \eta_a^2 = 1$, we have*

$$\big\langle -\Delta_{\mathbb{Z}} f, f\big\rangle \leqslant \sum\nolimits_a \big\langle -\Delta_{\mathbb{Z}}(\eta_a f), \eta_a f\big\rangle + \Big\|\sum\nolimits_a (\nabla\eta_a)^2\Big\|_\infty \|f\|_2^2.$$

Remark 3.8. (1) In the proof, we actually show the operator equality

$$-\Delta_{\mathbb{Z}} - \sum\nolimits_a \eta_a(-\Delta_{\mathbb{Z}})\eta_a = -\tfrac{1}{2}\sum\nolimits_a\big((\nabla\eta_a)^2\,\tau + (\nabla^*\eta_a)^2\,\tau^*\big).$$

Thus, the reverse inequality holds, too.

(2) On $\mathbb{Z}^d$, the Laplacian decomposes: $\Delta_{\mathbb{Z}^d} = \sum_{j=1}^d \Delta_{\mathbb{Z}_j}$. Thus, we immediately get the d-dimensional IMS formula

$$-\Delta_{\mathbb{Z}^d} - \sum\nolimits_a \eta_a(-\Delta_{\mathbb{Z}^d})\eta_a = -\tfrac{1}{2}\sum\nolimits_a \sum\nolimits_{j=1}^d \big((\nabla_j\eta_a)^2\,\tau_j + (\nabla_j^*\eta_a)^2\,\tau_j^*\big),$$

where $(\tau_j f)(z) = f(z+e_j) - f(z)$, and $\nabla_j = \tau_j - 1$ is a discrete partial derivative.

(3) Actually, the above formula holds on the Cayley graph of any finitely generated group, as long as the generator does not contain an idempotent element. This is proved basically with the exact same proof as given below for $\mathbb{Z}$, except that one has to read the notation higher dimensional. To be more precise, let $\mathcal{S}$ be the generator corresponding to the Cayley graph. Since the group acts on itself, we get for each $s \in \mathcal{S}$ a translation $\tau_s f(z) := f(zs)$. We treat

$$\tau := (\tau_s)_{s\in\mathcal{S}}, \quad \nabla := (\nabla_s)_{s\in\mathcal{S}}$$

as columns and ∇^* as row and use matrix multiplication when interpreting

$$-\nabla^*\nabla = \sum_{s\in\mathcal{S}} \nabla_s^*\nabla_s = \Delta.$$

We also have to write sums whenever appropriate.

(4) The formulation of lemma 3.7 with the quadratic form instead of the operators has the advantage, that it is easily restricted to subgraphs, e. g. $G = \{1,\dots,L\}$. All we have to do is to note that $\ell^2(G)$ is embedded trivially into $\ell^2(\mathbb{Z})$. The corresponding operator to the restricted quadratic form is the restriction with simple boundary conditions.

(5) Thanks to the simple boundary conditions, the adjacency operator $A_{\mathbb{Z}} := \tau^{-1} + \tau = \Delta_{\mathbb{Z}} + 2$ is only a shift of the Laplacian $\Delta_{\mathbb{Z}}$. Lemma 3.7 transfers to $A_{\mathbb{Z}}$:

$$\begin{aligned}\big\langle A_{\mathbb{Z}} f, f\big\rangle &= \big\langle \Delta_{\mathbb{Z}} f, f\big\rangle + 2\|f\|_2^2 \\ &\geqslant \sum\nolimits_a \big\langle \Delta_{\mathbb{Z}}\eta_a f, \eta_a f\big\rangle - \Big\|\sum\nolimits_a (\nabla\eta_a)^2\Big\|_\infty \|f\|_2^2 + 2\|f\|_2^2 \\ &\geqslant \sum\nolimits_a \big\langle A_{\mathbb{Z}}\eta_a f, \eta_a f\big\rangle - \Big\|\sum\nolimits_a (\nabla\eta_a)^2\Big\|_\infty \|f\|_2^2,\end{aligned}$$

since $\sum_a \eta_a^2 = 1$.

(6) Another noteworthy generalization of lemma 3.7 is the following. Note that any multiplication operator commutes with the multiplication of η_a. Thus, lemma 3.7 holds for Schrödinger operators, i. e. $-\Delta + V$ with a potential $V\colon \mathbb{Z} \to \mathbb{R}$ acting via multiplication.

Proof of lemma 3.7. We follow the proof of [Sim83, Lemma 3.1], the analogous statement on $\mathbb{R}^d$. With the above definitions,

$$\Delta_{\mathbb{Z}} = -\nabla^*\nabla.$$

For $f, g \in \ell^2(\mathbb{Z})$, it is easy to check that

$$\nabla(fg) = (\nabla f)\,\tau g + f\,(\nabla g) \quad \text{and} \quad \nabla^*(fg) = (\nabla^* f)\,\tau^* g + f\,(\nabla^* g).$$

Using this and $-\nabla^*\tau = \nabla$ as well as $-\tau^*\nabla = \nabla^*$, we immediately calculate

$$\begin{aligned}\Delta_{\mathbb{Z}}(fg) &= -\nabla^*\nabla(fg) = -\nabla^*\big(\nabla f\,\tau g + f\,\nabla g\big)\\ &= \Delta_{\mathbb{Z}} f\, g - \nabla f\,\nabla^*\tau g - \nabla^* f\,\tau^*\nabla g + f\,\Delta_{\mathbb{Z}} g\\ &= \Delta_{\mathbb{Z}} f\, g + \nabla f\,\nabla g + \nabla^* f\,\nabla^* g + f\,\Delta_{\mathbb{Z}} g.\end{aligned}$$

Consequently,

$$[f, -\Delta_{\mathbb{Z}}] = (\Delta_{\mathbb{Z}} f) + (\nabla f)\,\nabla + (\nabla^* f)\,\nabla^*.$$

To compute $[f,[f,-\Delta_{\mathbb{Z}}]]$, consider, for $g \in \ell^2(\mathbb{Z})$,

$$\begin{aligned}[f,-\Delta_{\mathbb{Z}}](fg) &= (\Delta_{\mathbb{Z}} f) fg + \nabla f\,\nabla(fg) + \nabla^* f\,\nabla^*(fg)\\ &= f(\Delta_{\mathbb{Z}} f) g + \nabla f\,(\nabla f\,\tau g + f\,\nabla g) + \nabla^* f\,(\nabla^* f\,\tau^* g + f\,\nabla^* g)\\ &= f\big(\Delta_{\mathbb{Z}} f + (\nabla f)\,\nabla + (\nabla^* f)\,\nabla^*\big) g + (\nabla f)^2\,\tau g + (\nabla^* f)^2\,\tau^* g\\ &= f[f,-\Delta_{\mathbb{Z}}] g + (\nabla f)^2\,\tau g + (\nabla^* f)^2\,\tau^* g.\end{aligned}$$

Thus,

$$[f,[f,-\Delta_{\mathbb{Z}}]] = -(\nabla f)^2\,\tau - (\nabla^* f)^2\,\tau^*.$$

On the other hand, expanding the commutators yields

$$[f,[f,-\Delta_{\mathbb{Z}}]] = [f, f(-\Delta_{\mathbb{Z}}) + \Delta_{\mathbb{Z}} f] = -f^2\Delta_{\mathbb{Z}} + 2f\Delta_{\mathbb{Z}} f - \Delta_{\mathbb{Z}} f^2.$$

We combine the last two formulas for $f := \eta_a$, sum over a and use $\sum_a \eta_a^2 = 1$ to derive

$$-\Delta_{\mathbb{Z}} - \sum\nolimits_a \eta_a(-\Delta_{\mathbb{Z}})\eta_a = -\tfrac{1}{2}\sum\nolimits_a \big((\nabla\eta_a)^2\,\tau + (\nabla^*\eta_a)^2\,\tau^*\big).$$

For $f \in \ell^2(\mathbb{Z})$, we see

$$\begin{aligned}&\tfrac{1}{2}\Big|\Big\langle -\tfrac{1}{2}\sum\nolimits_a \big((\nabla\eta_a)^2\,\tau + (\nabla^*\eta_a)^2\,\tau^*\big) f, f\Big\rangle\Big|\\ &\qquad \leqslant \tfrac{1}{2}\Big(\Big\|\sum\nolimits_a (\nabla\eta_a)^2\Big\|_\infty + \Big\|\sum\nolimits_a (\nabla^*\eta_a)^2\Big\|_\infty\Big)\|f\|_2^2 = \Big\|\sum\nolimits_a (\nabla\eta_a)^2\Big\|_\infty \|f\|_2^2.\end{aligned}$$

Thus,

$$\Big|\langle -\Delta_{\mathbb{Z}} f, f\rangle - \sum\nolimits_a \langle -\Delta_{\mathbb{Z}}(\eta_a f), \eta_a f\rangle\Big| \leqslant \Big\|\sum\nolimits_a (\nabla\eta_a)^2\Big\|_\infty \|f\|_2^2.$$

The triangle inequality finishes the proof. ■

3.2.2. *The IMS localization formula on the tree.* The discrete IMS formula is also valid on trees, in a very general setting. Indeed, points 3 and 4 of Remark 3.8 hint at the following way of proving the IMS localization formula on a tree of bounded degree. First note that the Cayley graph of the free group with s generators is a tree of degree $2s$. Then we can embed the bounded degree tree into the Cayley graph of a free group, and restrict to the subgraph again. It is enough for our purposes to consider a radially symmetric partition of unity, so that instead we will use in this section the spectral theory of the rooted trees to extend the IMS formula on $\mathbb{Z}$ to trees.

Proof of proposition 3.6. Step I. Fix $\eta_{\mathbb{R}} \in C^1(\mathbb{R}, [0,1])$ with support $\operatorname{supp}(\eta_{\mathbb{R}}) \subseteq [-1,1]$ such that, for any $x \in \mathbb{R}$,

$$\sum_{a\in\mathbb{Z}} \big(\eta_{\mathbb{R}}(x-a)\big)^2 = 1.$$

We define a partition of unity on $\mathbb{Z}$ as follows. For $r > 2$, let

$$\eta_{\mathbb{Z},r,a} \colon \mathbb{Z} \to \mathbb{R}, \quad \eta_{\mathbb{Z},r,a}(x) := \eta_{\mathbb{R}}\left(\frac{2x}{r} - a\right). \tag{3.7}$$

This gives a partition on $\mathbb{Z}$ satisfying $\#\operatorname{supp}\eta_{\mathbb{Z},r,a} \leqslant r$. Furthermore, by the mean value theorem and $\operatorname{supp}\eta_{\mathbb{R}} \subseteq [-1,1]$, we get

$$\begin{aligned}
|\nabla\eta_{\mathbb{Z},r,a}(x)| &= \big|\eta_{\mathbb{R}}\big(2r^{-1}(x+1)-a\big) - \eta_{\mathbb{R}}\big(2r^{-1}x-a\big)\big| \\
&\leqslant 2r^{-1} \sup_{\xi\in[0,1]} \big|\eta_{\mathbb{R}}'\big(2r^{-1}(x+\xi)-a\big)\big| \\
&\leqslant 2r^{-1} \sup|\eta_{\mathbb{R}}'(\mathbb{R})| \cdot \mathbf{1}_{[2r^{-1}x-1, 2r^{-1}(x+1)+1]}(a).
\end{aligned}$$

There are at most two values of $a \in \mathbb{Z}$ where the gradient is nonzero, since $r > 2$ and $2r^{-1}(x+1) + 1 - (2r^{-1}x - 1) = 2 + 2/r$. We can thus bound the following sum by

$$\sum_{a\in\mathbb{Z}} (\nabla\eta_{\mathbb{Z},r,a}(x))^2 \leqslant 4\sup|\eta_{\mathbb{R}}'(\mathbb{R})|^2 r^{-2} \sum_{a\in\mathbb{Z}} \mathbf{1}_{[2r^{-1}x-1, 2r^{-1}(x+1)+1]}(a) \leqslant C_3 r^{-2}$$

with $C_3 := 8\sup|\tilde{\eta}'(\mathbb{R})|^2$.

Step II. We now define the partition on the tree. Let

$$\eta_{r,a} \colon \mathcal{T} \to [0,1], \quad \eta_{r,a}(v) := \eta_{\mathbb{Z},r,a}(|v|). \tag{3.8}$$

With this definition we have

$$\sum_{a\in\mathbb{N}} (\eta_{r,a})^2 = 1$$

on $\mathcal{T}$. The support of each $\eta_{r,a}$ is a disjoint union of rooted trees of length at most r, see fig. 5. For $\psi \in \ell^2(\mathcal{T}^L)$, we employ remark 3.8 and learn that

$$\begin{aligned}
\langle A^{(L)}\psi, \psi\rangle &= \big\langle \widehat{A^{(L)}\psi}, \widehat{\psi}\big\rangle = \sqrt{k}\big\langle \widehat{A}\widehat{\psi}, \widehat{\psi}\big\rangle \\
&= \sqrt{k}\sum\nolimits_{v,j} \big\langle (\widehat{A}\widehat{\psi})_{v,j}, \widehat{\psi}_{v,j}\big\rangle = \sqrt{k}\sum\nolimits_{v,j} \big\langle A_{\mathbb{Z}}\widehat{\psi}_{v,j}, \widehat{\psi}_{v,j}\big\rangle \\
&\geqslant \sqrt{k}\sum\nolimits_{v,j} \Big(\sum\nolimits_a \big\langle A_{\mathbb{Z}}(\eta_{\mathbb{Z},r,a}\widehat{\psi}_{v,j}), \eta_{\mathbb{Z},r,a}\widehat{\psi}_{v,j}\big\rangle - C_3\|\widehat{\psi}_{v,j}\|_2^2/r^2\Big) \\
&= \sum\nolimits_a \sqrt{k}\sum\nolimits_{v,j} \big\langle A_{\mathbb{Z}}\widehat{(\eta_{r,a}\psi)}_{v,j}, \widehat{\eta_{r,a}\psi}_{v,j}\big\rangle - C_{\mathtt{IMS}}\|\widehat{\psi}\|_2^2/r^2 \\
&= \sum\nolimits_a \big\langle \widehat{A^{(L)}(\eta_{r,a}\psi)}, \widehat{\eta_{r,a}\psi}\big\rangle - C_{\mathtt{IMS}}\|\widehat{\psi}\|_2^2/r^2 \\
&= \sum\nolimits_a \big\langle A^{(L)}(\eta_{r,a}\psi), \eta_{r,a}\psi\big\rangle - C_{\mathtt{IMS}}\|\widehat{\psi}\|_2^2/r^2,
\end{aligned}$$

where $C_{\mathtt{IMS}} := C_3\sqrt{k}$. ∎

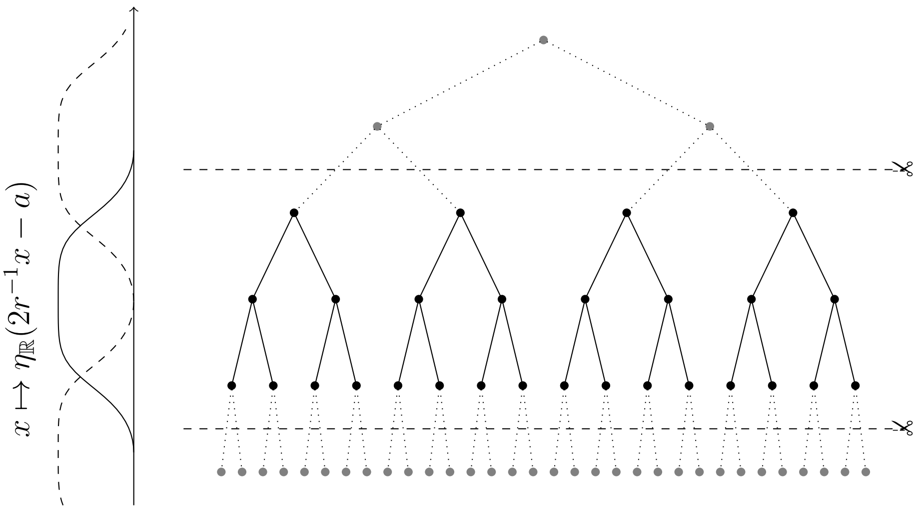


FIGURE 5. Shells of a tree split in trees

3.3. **An uncertainty principle.**

3.3.1. *On a finite segment of* $\mathbb{Z}$. Let us first prove a one-dimensional version of proposition 3.10. Afterwards, we transfer the result to the tree with lemma 3.5. To this end, let $L \in \mathbb{N}$ and $v \in \mathcal{T}_*^{L-1}$. Consider a function $\varphi \in \ell^2(\{|v|+1,\dots,L\})$, which can be written in the orthonormal basis of eigenfunctions of $-\Delta_{\mathbb{Z}}|\{|v|+1,\dots,L\}$ as

$$\varphi(z) = \sum_{1\leqslant m\leqslant L-|v|} \alpha_m \sqrt{\frac{2}{L+1-|v|}} \sin\Big(\frac{m\pi(z-|v|)}{L-|v|+1}\Big).$$

Given $\beta > 0$, we define the spectral projector $\hat{P}_\beta^{|v|,L}$ on $\ell^2(\{|v|+1,\dots,L\})$ via

$$\hat{P}_\beta^{|v|,L}\varphi(z) = \sum_{1\leqslant m\leqslant \beta+1} \alpha_m \sqrt{\frac{2}{L+1-|v|}} \sin\Big(\frac{m\pi(z-|v|)}{L-|v|+1}\Big),$$

$z \in \{|v|+1,\dots,L\}$.

Lemma 3.9. *Let* $L \in \mathbb{N}$, $\beta > 0$, $0 < \delta < 1$ *and* $v \in \mathcal{T}_*^L$ *be fixed, such that* $|v| + 1 + \delta L \leqslant L$. *Define, for* $\varphi \in \ell^2(|v|+1,\dots,L\})$, *the truncation*

$$T'_{|v|,\delta}\varphi := \mathbf{1}_{\{|v|+1+\lceil\delta L\rceil,\dots,L\}}\varphi.$$

Then, for $|v| \leqslant \big(1 - \frac{1}{\beta+1}\big)(L+1)$, *we have*

$$\|(1 - T'_{|v|,\delta})\hat{P}_\beta^{|v|,L}\varphi\|_2 \leqslant \sqrt{2}\pi\delta^{3/2}(\beta+1)^3\|\hat{P}_\beta^{|v|,L}\varphi\|_2.$$

Proof. We fix $\varphi \in \ell^2(|v|+1,\dots,L\})$ and calculate, using Cauchy–Schwarz,

$$\|(1-T'_{|v|,\delta})\hat{P}_\beta^{|v|,L}\varphi\|_2^2 = \frac{2}{L+1-|v|}\sum_{z=|v|+1}^{|v|+\lceil\delta L\rceil}\Big|\sum_{m=1}^{\lfloor\beta+1\rfloor}\alpha_m\sin\Big(\frac{m\pi(z-|v|)}{L+1-|v|}\Big)\Big|^2$$
$$\leqslant \frac{2}{L+1-|v|}\sum_{z=1}^{\lceil\delta L\rceil}\Big(\sum_{m=1}^{\lfloor\beta+1\rfloor}|\alpha_m|^2\Big)\sum_{m=1}^{\lfloor\beta+1\rfloor}\Big(\sin\Big(\frac{m\pi z}{L+1-|v|}\Big)\Big)^2.$$

Now using $\|\hat{P}_\beta^{|v|,L}\varphi\|_2^2 = \sum_{m=1}^{\lfloor\beta+1\rfloor}|\alpha_m|^2$, and $\sin(t)\leqslant|t|$, valid for all $t\in\mathbb{R}$, the last line is smaller than

$$\frac{2\|\hat{P}_\beta^{|v|,L}\varphi\|_2^2}{L+1-|v|}\sum_{z=1}^{\lceil\delta L\rceil}\sum_{m=1}^{\lfloor\beta+1\rfloor}\Big(\frac{m\pi z}{L+1-|v|}\Big)^2 = \frac{2\pi^2\|\hat{P}_\beta^{|v|,L}\varphi\|_2^2}{(L+1-|v|)^3}\sum_{z=1}^{\lceil\delta L\rceil}z^2\sum_{m=1}^{\lfloor\beta+1\rfloor}m^2$$
$$\leqslant \frac{2\pi^2(\delta L)^3(\beta+1)^3}{(L+1-|v|)^3}\|\hat{P}_\beta^{|v|,L}\varphi\|_2^2.$$

Note that the assumption $|v|\leqslant\big(1-\frac{1}{\beta+1}\big)(L+1)$ implies

$$\frac{L}{L+1-|v|}\leqslant\frac{\beta+1}{L+1}L\leqslant\beta+1.$$

This bound and taking the square root yields the result. ∎

3.3.2. *On a finite rooted tree.* For any $\beta>0$, we recall the definition of

$$E_\beta^{(L)} := 2\sqrt{k}\cos\Big(\frac{\beta+1}{L+1}\pi\Big) \tag{3.9}$$

from lemma 3.3. We want to study the neighborhood $[E_\beta^{(L)},E_0^{(L)}]$ of the principal eigenvalue $E_0^{(L)}$ of the adjacency matrix on the rooted tree $\mathcal{T}^L$. We define the spectral projector of $A^{(L)}$ on the energy interval $[E_\beta^{(L)},\infty)$ as

$$\Pi_{E_\beta^{(L)}}^{(L)}\colon \ell^2(\mathcal{T}^L)\to\ell^2(\mathcal{T}^L),\quad \Pi_{E_\beta^{(L)}}^{(L)}\varphi := \sum_{v,j,m\colon\lambda_{v,j,m}^L\geqslant E_\beta^{(L)}}\langle\varphi,\Psi_{v,j,m}^L\rangle\Psi_{v,j,m}^L.$$

We also define the space truncations

$$T_{|v|,\delta}\colon \ell^2(\mathcal{T}^L)\to\ell^2(\mathcal{T}^L),\quad T_{|v|,\delta}\varphi := \varphi\mathbf{1}_{\{x\in\mathcal{T}^L:|x|>|v|+1+\delta L\}}$$

and a truncated version of $\Pi_{E_\beta^{(L)}}^{(L)}$

$$\tilde{\Pi}_{E_\beta^{(L)}}^{(L)}\colon \ell^2(\mathcal{T}^L)\to\ell^2(\mathcal{T}^L),\quad \tilde{\Pi}_{E_\beta^{(L)}}^{(L)}\varphi := \sum_{v,j,m\colon\lambda_{v,j,m}^L\geqslant E_\beta^{(L)}}\langle\varphi,\Psi_{v,j,m}^L\rangle T_{|v|,\delta}\Psi_{v,j,m}^L.$$

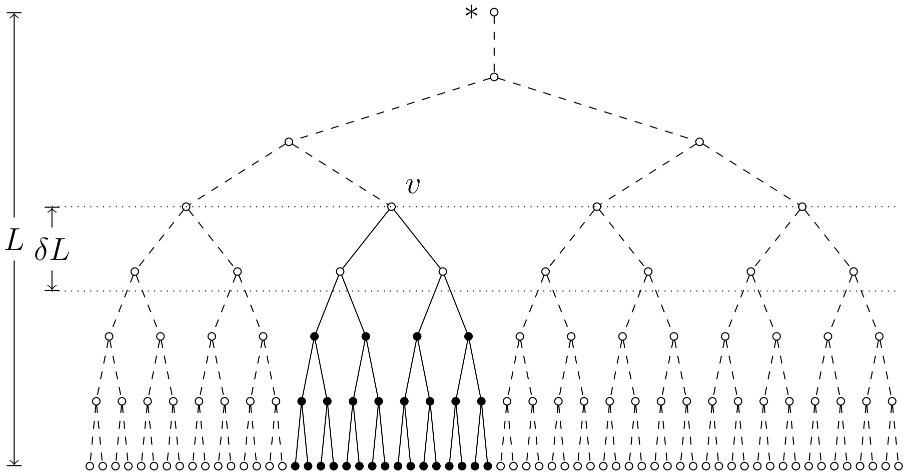


FIGURE 6. Illustration of (3.10). The subtree $\mathcal{T}_v^L$ is indicated with solid edges. Nodes in the support of functions truncated with $T_{|v|,\delta}$ in $\mathcal{T}_v^L$ are filled black.

It is straightforward to check on the eigenbasis that $\tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\mathbb{P}_{v,j} = T_{|v|,\delta}P_{v,j}\Pi^{(L)}_{E^{(L)}_\beta}$, so

$$\tilde{\Pi}^{(L)}_{E^{(L)}_\beta} = \sum_{v,j} T_{|v|,\delta}P_{v,j}\Pi^{(L)}_{E^{(L)}_\beta}. \tag{3.10}$$

Proposition 3.10. *Let $L \in \mathbb{N}$, $\beta > 0$ and $0 < \delta < 1$. Then, for any $\varphi \in \ell^2(\mathcal{T}^L)$,*

$$\|\Pi^{(L)}_{E^{(L)}_\beta}\varphi - \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi\|_2 \leqslant \sqrt{2}\pi\delta^{3/2}(\beta+1)^3\|\Pi^{(L)}_{E^{(L)}_\beta}\varphi\|_2.$$

Proof. We will show equivalently that

$$\forall\varphi \in \Pi^{(L)}_{E^{(L)}_\beta}\ell^2(\mathcal{T}^L)\colon \|\varphi - \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi\|_2 \leqslant \sqrt{2}\delta^{3/2}(\beta+1)^2\|\varphi\|_2.$$

Indeed, it follows from (3.10) that if $\Pi^{(L)}_{E^{(L)}_\beta}\varphi = 0$ then $\tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi = 0$. We assume thus from now on that $\varphi = \Pi^{(L)}_{E^{(L)}_\beta}\varphi$. For such φ, we see, by (3.10),

$$\big(\Pi^{(L)}_{E^{(L)}_\beta} - \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\big)\varphi = \sum_{v,j}(\mathbf{1}_{\mathcal{T}^L} - T_{|v|,\delta})P_{v,j}\varphi.$$

By lemma 3.5, we know that $P_{v,\delta}$ commutes with the radially symmetric truncation. Thus, by the orthogonality of the projections $P_{v,j}$,

$$\big\|\big(\Pi^{(L)}_{E^{(L)}_\beta}\varphi - \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\big)\varphi\big\|_2^2 = \Big\|\sum_{v,j} P_{v,j}(\mathbf{1}_{\mathcal{T}^L} - T_{|v|,\delta})\varphi\Big\|_2^2 = \sum_{v,j}\big\|P_{v,j}(\mathbf{1}_{\mathcal{T}^L} - T_{|v|,\delta})\varphi\big\|_2^2.$$

We study this norms via the unitary $\widehat{\cdot}$, see (3.5) and lemma 3.5. For each v, j, we have

$$\begin{aligned}\big\|P_{v,j}(\mathbf{1}_{\mathcal{T}^L} - T_{|v|,\delta})\varphi\big\|_2 &= \big\|\big(P_{v,j}(\mathbf{1}_{\mathcal{T}^L} - T_{|v|,\delta})\varphi\big)^\wedge\big\|_2 = \big\|\big((\mathbf{1}_{\mathcal{T}^L} - T_{|v|,\delta})\varphi\big)^\wedge_{v,j}\big\|_2 \\ &= \big\|\big(\mathbf{1}_{\{|v|+1,\dots,L\}} - T'_{|v|,\delta}\big)\widehat{\varphi}_{v,j}\big\|_2.\end{aligned}$$

We learn from lemma 3.3 that the coefficients $(\alpha_{v,j,m})_{v,j,m}$ of

$$\varphi = \Pi^{(L)}_{E^{(L)}_\beta}\varphi = \sum_{v,j,m} \alpha_{v,j,m}\Psi^L_{v,j,m}$$

vanish as soon as $|v| > (L+1)(1 - \frac{m}{\beta+1})$ or $m > \beta + 1$. Therefore, we have $\widehat{\varphi}_{v,j} = \hat{P}^{|v|,L}_\beta \widehat{\varphi}_{v,j}$, and we can thus invoke lemma 3.9 to conclude

$$\big\|\big(\Pi^{(L)}_{E^{(L)}_\beta}\varphi - \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\big)\varphi\big\|_2^2 \leqslant 2\pi^2\delta^3(\beta+1)^6 \sum_{v,j} \|\widehat{\varphi}_{v,j}\|_2^2 = 2\pi^2\delta^3(\beta+1)^6\|\varphi\|_2^2.$$

Since $\varphi = \Pi^{(L)}_{E^{(L)}_\beta}\varphi$, this is what we set out to prove. ∎

4. Lifshitz tails: The upper bound

This section is devoted to the proof of the following theorem.

Theorem 4.1. *Let $E_0 := (\sqrt{k} - 1)^2$. Then,*

$$\limsup_{E \to E_0} \frac{\log\log|\log \mathcal{N}(E)|}{\log(E - E_0)} \leqslant -\frac{1}{2}.$$

This theorem provides the converse to proposition 2.7. Note that no condition on the random variables is needed for the upper bound.

4.1. **Bound by a probability.** We remind that $E \mapsto \mathcal{N}(E)$ denotes the integrated density of states given by (1.6), which is monotone and thus the cumulative distribution function of the measure $\mathrm{d}\mathcal{N}$ called the density of states measure, and $0 \leqslant t \mapsto \tilde{\mathcal{N}}(t)$ is the Laplace transform of $\mathrm{d}\mathcal{N}$. We start by proving the following Tauberian theorem, which links the long time behavior of $\tilde{\mathcal{N}}$ to the low energy asymptotic of $\mathcal{N}$.

Proposition 4.2. *Let $\tilde{\mathcal{N}}$ be the Laplace transform of the density of states measure $\mathrm{d}\mathcal{N}$. Suppose that for some $\eta > 0$,*

$$\limsup_{t\to\infty} \mathrm{e}^{t(E_0 + (\log t)^{-\eta})}\tilde{\mathcal{N}}(t) \leqslant 1 \tag{4.1}$$

with $E_0 := (\sqrt{k} - 1)^2$. Then,

$$\limsup_{E \to E_0} \frac{\log\log|\log \mathcal{N}(E)|}{\log(E - E_0)} \leqslant -\frac{1}{\eta}.$$

Proof. Assume that inequality (4.1) holds. Then, there exists some t^* such that for all $t > t^*$,

$$\tilde{\mathcal{N}}(t) \leqslant 2\exp(-t(E_0 + (\log t)^{-\eta})). \tag{4.2}$$

Clearly, for $E \geqslant E_0$,

$$\mathcal{N}(E) = \int_{E_0}^{E} \mathrm{d}\mathcal{N}(\lambda) \leqslant \mathrm{e}^{tE} \int_{E_0}^{E} \mathrm{e}^{-t\lambda} \mathrm{d}\mathcal{N}(\lambda) \leqslant \mathrm{e}^{tE} \int_{E_0}^{\infty} \mathrm{e}^{-t\lambda} \mathrm{d}\mathcal{N}(\lambda) = \mathrm{e}^{tE} \tilde{\mathcal{N}}(t),$$

and by (4.2), for large t,

$$\mathcal{N}(E) \leqslant 2\exp(t(E - E_0) - t(\log t)^{-\eta}).$$

Now we choose t as follows

$$t = t(E) := \exp\Big((2(E - E_0))^{-1/\eta}\Big).$$

We see that, for small $E - E_0$,

$$\begin{aligned} \mathcal{N}(E) &\leqslant 2\exp\left(-(E - E_0)\exp\Big((2(E - E_0))^{-1/\eta}\Big)\right) \\ &\leqslant \exp\left(-\frac{1}{2}(E - E_0)\exp\Big((2(E - E_0))^{-1/\eta}\Big)\right), \end{aligned}$$

and

$$\log|\log \mathcal{N}(E)| \geqslant \log((E - E_0)/2) + (2(E - E_0))^{-1/\eta}.$$

Now we take another logarithm and divide by $\log(E - E_0) < 0$:

$$\begin{aligned} \frac{\log\log|\log \mathcal{N}(E)|}{\log(E - E_0)} &\leqslant \frac{\log(\log((E - E_0)/2) + (2(E - E_0))^{-1/\eta})}{\log(E - E_0)} \\ &= \frac{\log\big((E - E_0)^{1/\eta}\log((E - E_0)/2) + 2^{-\frac{1}{\eta}}\big)}{\log(E - E_0)} - \frac{1}{\eta} \xrightarrow{E \searrow E_0} -\frac{1}{\eta}. \end{aligned}$$

Taking the lim sup proves the proposition. ∎

The rest of this section will be devoted to prove that, as a consequence of theorem 4.10, condition (4.1) holds for any $\eta > 2$. This proves theorem 4.1.

Our next proposition compares $\tilde{\mathcal{N}}$ to a finite dimensional analog $\tilde{\mathcal{N}}^L$. For any $\Gamma \subset \mathcal{B}$, we denote by $H_\omega|\Gamma$ the operator H_ω with *simple* (sometimes called *Dirichlet*) boundary conditions, i. e. the operator defined by

$$H_\omega|\Gamma := \mathbf{1}_\Gamma H_\omega \mathbf{1}_\Gamma,$$

or equivalently, writing $H_\omega(v, w)$, $v, w \in \mathcal{B}$, for the matrix coefficients, it can be defined by

$$(H_\omega|\Gamma)(v, w) := \begin{cases} H_\omega(v, w) & \text{if } v, w \in \Gamma \\ 0 & \text{elsewhere.} \end{cases} \tag{4.3}$$

Remember that $\mathcal{B}^L$ denotes the ball of radius L of the Bethe lattice. Let us define the averaged spectral density $\mathcal{N}^L$ of $H_\omega|\mathcal{B}^L$ by

$$\mathcal{N}^L(E) := \mathbb{E}\langle \delta_0, \mathbf{1}_{(-\infty,E]}(H_\omega|\mathcal{B}^L)\delta_0\rangle.$$

In particular, its Laplace transform can be written

$$\tilde{\mathcal{N}}^L(t) = \mathbb{E}\left[\langle \delta_0, \mathrm{e}^{-tH_\omega|\mathcal{B}^L}\delta_0\rangle\right].$$

Note also that, using functional calculus, we have

$$\tilde{\mathcal{N}}(t) = \mathbb{E}\left[\langle \delta_0, \mathrm{e}^{-tH_\omega}\delta_0\rangle\right].$$

In the following proposition we compare these two quantities. We define $\omega_+ := \|\omega_0\|_\infty$ for further use.

Proposition 4.3. *Let $\tilde{\mathcal{N}}^L$ be the Laplace transform of $\mathrm{d}\mathcal{N}^L$. Pick some positive constant $\zeta > \mathrm{e}^2\|H_\omega\| = \mathrm{e}^2((\sqrt{k}+1)^2 + \omega_+)$ and let $L = \lceil \zeta t \rceil$. Then, for any $t \geqslant 1$ the following holds:*

$$|\tilde{\mathcal{N}}(t) - \tilde{\mathcal{N}}^L(t)| \leqslant \mathrm{e}^{-\zeta t}.$$

Here, $\|H_\omega\| = \sup \Sigma$.

Proof. Assume $\zeta > \mathrm{e}^2\|H_\omega\|$ and $t \geqslant 1$. First let us note that H_ω is a bounded operator and (we actually have $\|H_\omega\| = (k+1+2\sqrt{k}) + \|V_\omega\|_\infty$). This allows us to expand the exponential as a sum like

$$\langle \delta_0, \mathrm{e}^{-tH_\omega}\delta_0\rangle = \sum_{n=0}^{L} \frac{(-t)^n}{n!}\langle \delta_0, H_\omega^n\delta_0\rangle + \sum_{n>L} \frac{(-t)^n}{n!}\langle \delta_0, H_\omega^n\delta_0\rangle,$$

which is also valid if we replace H_ω by $H_\omega|\mathcal{B}^L$. It is easy to see that the two first terms of this sum are 1 and $-tH_\omega(0,0) = -t(H_\omega|\mathcal{B}^L)(0,0)$ respectively. Expanding the matrix product, we see that, for $n \in \mathbb{N}$, $2 \leqslant n \leqslant L$

$$\langle \delta_0, H_\omega^n\delta_0\rangle = \sum_{x_1,\ldots,x_{n-1}\in\mathcal{B}} H_\omega(0,x_1)H_\omega(x_2,x_3)\cdots H_\omega(x_{n-2},x_{n-1})H_\omega(x_{n-1},0).$$

Now, using that $H_\omega(v,w) = 0$ for $v, w \in \mathcal{B}^L$ satisfying $d(v,w) > 1$, the last sum reduces to

$$\sum_{(p_0,\ldots,p_n):0\rightsquigarrow 0} H_\omega(p_0,p_1)H_\omega(p_2,p_3)\cdots H_\omega(p_{n-2},p_{n-1})H_\omega(p_{n-1},p_n), \tag{4.4}$$

where we have written $(p_0,\ldots,p_n) : 0 \rightsquigarrow 0$ to denote a path $(p_0,\ldots,p_n) \in (\mathcal{B})^{n+1}$ (which may include loops) starting at 0 and ending at 0. In particular, if $(p_0,\ldots,p_n) : 0 \rightsquigarrow 0$,

$$d(p_0,p_i) \leqslant \sum_{0\leqslant j\leqslant n-1} d(p_j,p_{j+1}) \leqslant n \leqslant L \qquad \text{for any } 0 \leqslant i \leqslant n,$$

i. e. the paths in the sum (4.4) are entirely contained in $\mathcal{B}^L$. Using (4.3), we see that for $2 \leqslant n \leqslant L$

$$\langle \delta_0, H_\omega^n \delta_0 \rangle = \sum_{(p_0,\dots,p_n):0\rightsquigarrow 0} \prod_{1\leqslant i\leqslant n} (H_\omega|\mathcal{B}^L)(p_{i-1}, p_i) = \langle \delta_0, (H_\omega|\mathcal{B}^L)^n \delta_0 \rangle.$$

Thus, the first $L+1$ terms of the expansions of $\langle \delta_0, \mathrm{e}^{-tH_\omega}\delta_0\rangle$ and $\langle \delta_0, \mathrm{e}^{-tH_\omega|\mathcal{B}^L}\delta_0\rangle$ coincide, and

$$\begin{aligned}\left|\langle \delta_0, \mathrm{e}^{-tH_\omega}\delta_0\rangle - \langle \delta_0, \mathrm{e}^{-tH_\omega|\mathcal{B}^L}\delta_0\rangle\right| &\leqslant \sum_{n>L} \frac{t^n}{n!} |\langle \delta_0, H_\omega^n \delta_0\rangle| + \sum_{n>L} \frac{t^n}{n!} |\langle \delta_0, (H_\omega|\mathcal{B}^L)^n \delta_0\rangle| \\ &\leqslant 2\sum_{n>L} \frac{t^n}{n!} \|H_\omega\|^n.\end{aligned}$$

Here, we used the Cauchy–Schwarz inequality and $\|H_\omega|\mathcal{B}^L\| \leqslant \|H_\omega\|$. Let us estimate this error with

$$\begin{aligned}2\sum_{n>L}\frac{t^n}{n!}\|H_\omega\|^n = 2\sum_{n\geqslant 1}\frac{t^{L+n}}{(L+n)!}\|H_\omega\|^{L+n} &\leqslant \sum_{n\geqslant 1}\frac{t^{L+n}}{(L+n)^{(L+n)}\mathrm{e}^{-(L+n)}}\|H_\omega\|^{L+n} \\ &\leqslant \sum_{n\geqslant 1}\left(\frac{\mathrm{e}t\|H_\omega\|}{L}\right)^{L+n},\end{aligned}$$

where we have used $n! \geqslant \sqrt{2\pi n}(n/\mathrm{e})^n \geqslant 2n^n\mathrm{e}^{-n}$ and $(L+n)^{-1} \leqslant L^{-1}$. In particular, if $L = \lceil \zeta t \rceil$, we see that

$$\left(\frac{\mathrm{e}t\|H_\omega\|}{L}\right)^{L+n} \leqslant (\mathrm{e}\|H_\omega\|/\zeta)^{\lceil \zeta t\rceil + n},$$

and as $\zeta > \mathrm{e}^2\|H_\omega\|$ and $t \geqslant 1$, we can bound the error as

$$\left|\langle \delta_0, \mathrm{e}^{-tH_\omega}\delta_0\rangle - \langle \delta_0, \mathrm{e}^{-tH_\omega|\mathcal{B}^L}\delta_0\rangle\right| \leqslant \sum_{n>L}\frac{t^n}{n!}\|H_\omega\|^n \leqslant \sum_{n\geqslant 1}\mathrm{e}^{-(\lceil\zeta t\rceil+n)} \leqslant \mathrm{e}^{-\zeta t}.$$

The geometric series $\sum_{n\geqslant 1}\mathrm{e}^{-n} = (\mathrm{e}-1)^{-1} \leqslant 1$ enters in the last inequality. Taking the expectation ends the proof. ∎

We will now study the large time behavior of $\tilde{\mathcal{N}}^L$.

Lemma 4.4. *Let $\Gamma \subset \mathcal{B}$ be finite and $H_\omega|\Gamma$ the restriction of H_ω with simple boundary conditions. Then*

$$\mathbb{E}\left[\langle \delta_0, \mathrm{e}^{-tH_\omega|\Gamma}\delta_0\rangle\right] \leqslant \mathbb{E}\left[\mathrm{e}^{-tE_{GS}(H_\omega|\Gamma)}\right].$$

Proof. Fix a realization ω and let $\{\lambda_i;\psi_i\}_{i\in\Gamma} = \{\lambda_i(\omega);\psi_i(\omega)\}_{i\in\Gamma}$ be a complete set of (eigenvalues, eigenfunctions) of $H_\omega|\Gamma$. Then, writing

$$\delta_0 = \sum \alpha_i(\omega)\psi_i(\omega)$$

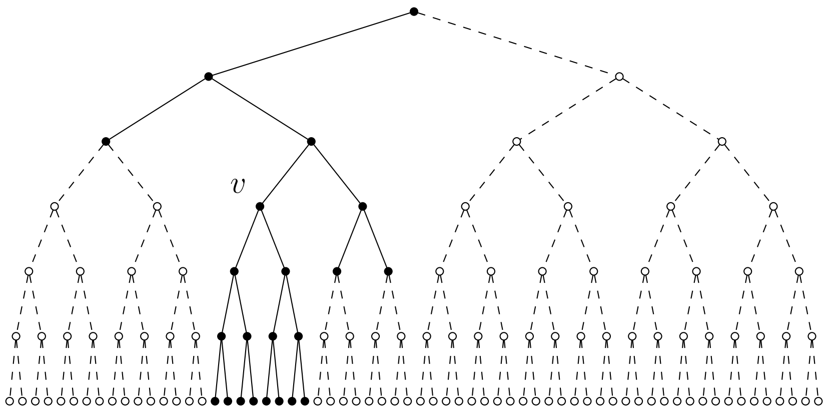


FIGURE 7. A ball $\mathcal{B}_v^3$ centered at v, indicated with solid edges and filled nodes, is contained in the tree $\mathcal{T}^7$.

we see that, as $E_{\mathit{GS}}(H_\omega|\Gamma) = \min_{i\in\Gamma} \lambda_i(\omega)$ and $\sum|\alpha_i|^2 = 1$,

$$\langle \delta_0, \mathrm{e}^{-tH_\omega|\Gamma}\delta_0\rangle = \sum_{i\in\Gamma} |\alpha_i(\omega)|^2 \mathrm{e}^{-t\lambda_i(\omega)} \leqslant \sum_{i\in\Gamma} |\alpha_i(\omega)|^2 \mathrm{e}^{-tE_{\mathit{GS}}(H_\omega|\Gamma)} = \mathrm{e}^{-tE_{\mathit{GS}}(H_\omega|\Gamma)}.$$

Taking the expectation in this inequality yields the desired result. ■

The following two lemmas link the behavior of the ground state energy of the Hamiltonian on a ball to the one on a finite rooted tree. This is needed in order to use the spectral theory developed in section 3. The trick is to embedd a ball in the Bethe lattice in a finite symmetric rooted tree, cf. fig. 7. Of course, the ball is not centered at the root of the tree, but taking advantage of the translation invariance of the Anderson Hamiltonian on the Bethe lattice, as soon as we take expectations, the location of the ball is arbitrary.

Lemma 4.5. *Let $L \geqslant 1$ and $\mathcal{B}_v^L$ a ball of radius L centered at $v \in \mathcal{B}$, i. e.*

$$\mathcal{B}_v^L := \{w \in \mathcal{B} : d_{\mathcal{B}}(w, v) \leqslant L\}.$$

Then, for every $v \in \mathcal{B}$ with $|v| = L + 2$ there exists a rooted tree $\mathcal{T}^{3L} \subset \mathcal{B}$, of length $3L$, which contains $\mathcal{B}_v^L$.

Proof. Label the k branches of the Bethe lattice by the nodes $x \in \mathcal{T}^L$ satisfying $|x| = 1$ and assume that $d(0, v) = L + 1$. Then, there exists a unique minimal path $[0, v] = (0, v_1, v_2, \ldots, v)$ of length $L + 1$. Because $d(v_1, v) = L$, we know that $v_1 \in \mathcal{B}^L$. In particular the whole ball is contained in the branch of the Bethe lattice v_1. Now choose $k - 1$ other branches to form the infinite rooted tree $\mathcal{T}$. The result is now clear because by definition $\mathcal{T}^L := \{v \in \mathcal{T} : |v| \leqslant L\}$ and for any $x \in \mathcal{B}_v^L$ we have $|x| \leqslant 2L + 1 \leqslant 3L$. ■

Conversely, it is easy to see that $\mathcal{T}^L \subset \mathcal{B}^L$, for all $L > 1$. This leads to the following lemma.

Lemma 4.6. *For any $L \geqslant 1$ and $|v| = L+2$,*

$$E_{GS}(H_\omega|\mathcal{B}_v^L) \leqslant E_{GS}(H_\omega|\mathcal{T}^{3L}) \leqslant E_{GS}(H_\omega|\mathcal{B}^{3L})$$

Here $\mathcal{T}^{3L}$ is the tree satisfying $\mathcal{B}_v^L \subseteq \mathcal{T}^{3L} \subseteq \mathcal{B}^{3L}$.

Proof. Let $v \in \mathcal{B}$ with $|v| = L+2$ and $\mathcal{T}^{3L}$ be the rooted tree containing $\mathcal{B}_v^L$. Then

$$E_{GS}(H_\omega|\mathcal{B}_v^L) = \inf_{\substack{\varphi\in\ell^2(\mathcal{B}_v^L)\\ \|\varphi\|_2=1}} \langle H_\omega\varphi,\varphi\rangle \leqslant \inf_{\substack{\varphi\in\ell^2(\mathcal{T}^{3L})\\ \|\varphi\|_2=1}} \langle H_\omega\varphi,\varphi\rangle$$
$$= E_{GS}(H_\omega|\mathcal{T}^{3L}) \leqslant E_{GS}(H_\omega|\mathcal{T}^{3L}).$$

The second inequality is proved analogously. ■

Using translation invariance, we can translate the point where we calculate the integrated densities of states $\mathcal{N}$ and $\mathcal{N}^L$. Proposition 4.3 tells us then that it is enough to study, for some $v \in \mathcal{B}$ with $|v| = L+2$,

$$\tilde{\mathcal{N}}^L(t) = \mathbb{E}\left[\langle\delta_v, \mathrm{e}^{-tH_\omega|\mathcal{B}_v^L}\delta_v\rangle\right].$$

We remind that $\mathcal{B}_v^L$ is the ball centered at v.

From now on we write $H_\omega^L := H_\omega|\mathcal{T}^L$. The next lemma is a simple bound on the expectation by a probability.

Lemma 4.7. *For any $\epsilon > 0$, $L > 1$ and $t > 1$, we have*

$$\mathbb{E}\big[\mathrm{e}^{-tE_{GS}(H_\omega^L)}\big] \leqslant \mathrm{e}^{-t(E_0+2\epsilon(\log t)^{-2})} + \mathrm{e}^{-tE_0}\mathbb{P}\big(E_{GS}(H_\omega^L) < E_0 + 2\epsilon(\log t)^{-2}\big). \tag{4.5}$$

Proof. We have indeed for all $E \geqslant E_0$

$$\mathbb{E}\big[\mathrm{e}^{-tE_{GS}(H_\omega^L)}\big] = \mathbb{E}\big[(\mathbf{1}_{\{E_{GS}(H_\omega^L)\geqslant E\}} + \mathbf{1}_{\{E_{GS}(H_\omega^L)<E\}})\mathrm{e}^{-tE_{GS}(H_\omega^L)}\big]$$
$$\leqslant \mathrm{e}^{-tE} + \mathrm{e}^{-tE_0}\mathbb{P}(E_{GS}(H_\omega^L) < E).$$
■

We summarize the results of this section in the following proposition.

Proposition 4.8. *Assume that $\epsilon > 0$ and $\zeta > \mathrm{e}^2\|H_\omega\|_2 = \mathrm{e}^2((\sqrt{k}+1)^2 + \omega_+)$ satisfy*

$$\limsup_{L\to\infty} \mathrm{e}^{\epsilon L/\zeta}\mathbb{P}\left(E_{GS}(H_\omega^L) < E_0 + 4\epsilon(\log L)^{-2}\right) \leqslant 1. \tag{4.6}$$

Then, we have

$$\limsup_{t\to\infty} \mathrm{e}^{t(E_0+\epsilon(\log t)^{-2})}\tilde{\mathcal{N}}(t) \leqslant 1. \tag{4.7}$$

Proof. Let $t > 1$ and $L = \lceil\zeta t\rceil$. Then,

$$\begin{aligned}
&\exp(t(E_0+\epsilon(\log t)^{-2}))\tilde{\mathcal{N}}(t)\\
&\leqslant \exp(t(E_0+\epsilon(\log t)^{-2}))(\tilde{\mathcal{N}}^L(t) + \mathrm{e}^{-\zeta t}) \quad \text{by proposition 4.3}\\
&\leqslant \exp(t(E_0+\epsilon(\log t)^{-2}))(\mathbb{E}[\mathrm{e}^{-tE_{GS}(H_\omega^{3L})}] + \mathrm{e}^{-\zeta t}) \quad \text{using lemmas 4.4 to 4.6}\\
&\leqslant \mathrm{e}^{-\epsilon t(\log t)^{-2}} + \mathrm{e}^{-t(\zeta-E_0-\epsilon(\log t)^{-2})} + \mathrm{e}^{\epsilon t(\log t)^{-2}}\mathbb{P}\left(E_{GS}(H_\omega^{3L}) < E_0 + 2\epsilon(\log t)^{-2}\right),
\end{aligned}$$

using lemma 4.7. Note that $\zeta > E_0$, so for the first two terms in this sum we have

$$\mathrm{e}^{-\epsilon t(\log t)^{-2}} + \mathrm{e}^{-t(\zeta - E_0 - (\log t)^{-2}))} \xrightarrow{t\to\infty} 0.$$

For the third term, noting that $\mathrm{e}^{\epsilon t(\log t)^{-2}} \leqslant \mathrm{e}^{\epsilon(3L)/\zeta}$ and that

$$\frac{2\epsilon(\log t)^{-2}}{4\epsilon(\log(3L))^{-2}} = \frac{1}{2}\left(\frac{\log(3\lceil\zeta t\rceil)}{\log t}\right)^2 \xrightarrow{t\to\infty} \frac{1}{2} < 1$$

yields the result. ∎

It is not hard to see that (4.7) implies (4.1) for every $\eta > 2$, so that theorem 4.1 is a consequence of condition (4.6).

4.2. **Reduction to a smaller scale.** In the following lemma we trade energy for probability. The *IMS localization formula* (proposition 3.6) furnishes a crucial ingredient of the proof.

Proposition 4.9. *For every $\epsilon > 0$ there exists $L^* > 1$ so that for any $L > L^*$ and $r = \lfloor \epsilon^{-1/2} \log L \rfloor$,*

$$\mathbb{P}\left(E_{\mathit{GS}}(H^L_\omega) \leqslant E_0 + \frac{4\epsilon}{(\log L)^2}\right) \leqslant k^{\exp((r+1)\sqrt{\epsilon})}\mathbb{P}\left(E_{\mathit{GS}}(H^r_\omega) \leqslant E_0 + \frac{4 + C_{\mathit{IMS}}}{r^2}\right).$$

Proof. Assume both

$$r = \lfloor \epsilon^{-1/2} \log L \rfloor = r > 2 \tag{4.8}$$

and

$$E_{\mathit{GS}}(H^L_\omega) \leqslant E_0 + \frac{4\epsilon}{(\log L)^2} \leqslant E_0 + \frac{4}{r^2}. \tag{4.9}$$

Let $\{\eta_{a,r}\}_a$ be the family of spherically symmetric functions on $\mathcal{T}^L$ given by proposition 3.6. They satisfy

$$\sum_a (\eta_{a,r}(v))^2 = 1$$

for all $v \in \mathcal{T}^L$, and

$$\mathcal{S}_{a,r} := \operatorname{supp} \eta_{a,r} \subseteq \mathcal{T}^L.$$

If φ^L_{GS} is the normalized the ground state of H^L_ω, then the IMS formula and the normalization $\sum_a \|\varphi^L_{GS}\eta_{a,r}\|^2_2 = 1$ yield

$$\begin{aligned}
E_{\mathit{GS}}(H^L_\omega) = \langle \varphi^L_{GS}, H^L_\omega \varphi^L_{GS}\rangle &\geqslant \sum_a E_{\mathit{GS}}(H^L_\omega | \mathcal{S}_{a,r}) \|\varphi^L_{GS}\eta_{a,r}\|^2_2 - \frac{C_{\mathit{IMS}}}{r^2}\|\varphi^L_{GS}\|^2_2 \\
&\geqslant \min_a E_{\mathit{GS}}(H^L_\omega | \mathcal{S}_{a,r}) - \frac{C_{\mathit{IMS}}}{r^2}\|\varphi^L_{GS}\|^2_2 \\
&\geqslant \min_{v\in\mathcal{T}^L :\, \mathcal{T}^r_v \subseteq \mathcal{T}^L} E_{\mathit{GS}}(H^L_\omega | \mathcal{T}^r_v) - \frac{C_{\mathit{IMS}}}{r^2}\|\varphi^L_{GS}\|^2_2.
\end{aligned}$$

The last estimate is true by proposition 3.6, which states that $\mathcal{S}_{a,r}$ is the disjoint union of finite subtrees of length at most r. From (4.9) we deduce then

$$\min_{v\in\mathcal{T}^L:\,\mathcal{T}_v^r\subseteq\mathcal{T}^L} E_{\mathit{GS}}(H_\omega^L|\mathcal{T}_v^r) \leqslant E_0 + \frac{4+C_{\mathit{IMS}}}{r^2}$$

and thus

$$\begin{aligned}&\mathbb{P}\left(E_{\mathit{GS}}(H_\omega^L) \leqslant E_0 + \frac{4\epsilon}{(\log L)^2}\right)\\ \leqslant& \sum_{v\in\mathcal{T}^L:\,\mathcal{T}_v^r\subseteq\mathcal{T}^L} \mathbb{P}\left(E_{\mathit{GS}}(H_\omega^L|\mathcal{T}_v^r) \leqslant E_0 + \frac{4+C_{\mathit{IMS}}}{r^2}\right)\\ \leqslant& k^L\mathbb{P}\left(E_{\mathit{GS}}(H_\omega^r) \leqslant E_0 + \frac{4+C_{\mathit{IMS}}}{r^2}\right).\end{aligned}$$

To end the proof, note from (4.8) that $L \leqslant \mathrm{e}^{(r+1)\sqrt{\epsilon}}$. ∎

We state the main probability estimate, which we will prove in the next section.

Theorem 4.10. *For every $\beta' > 0$ there exists some $\epsilon_{\beta'} > 0$ and $L^* > 1$ so that for any $L > L^*$,*

$$\mathbb{P}\left(E_{\mathit{GS}}(H_\omega^L) \leqslant E_0 + \beta' L^{-2}\right) \leqslant \exp(-\exp(\epsilon_{\beta'}L)). \tag{4.10}$$

We first state and prove the following important corollary.

Corollary 4.11. *For any $\epsilon > 0$ small enough and any $\zeta > 1$ there exists some $L^* > 1$ such that for all $L > L^*$*

$$\mathbb{P}\left(E_{\mathit{GS}}(H_\omega^L) \leqslant E_0 + \frac{4\epsilon}{(\log L)^2}\right) \leqslant \mathrm{e}^{-\zeta L}.$$

In particular, condition (4.6) of proposition 4.8 holds.

Proof. Let $\beta' \geqslant 4(4 + C_{\mathit{IMS}})$. Then, by theorem 4.10, which we assume to hold true, we get $\epsilon_{\beta'} > 0$ and $r^* > 1$ such that for all $r > r^*$, we have

$$\mathbb{P}\left(E_{\mathit{GS}}(H_\omega^r) \leqslant E_0 + \frac{4+C_{\mathit{IMS}}}{r^2}\right) \leqslant \exp(-\exp(\epsilon_{\beta'}r)). \tag{4.11}$$

Now fix $0 < \epsilon < \epsilon_{\beta'}^2$ and let $r := \lfloor \epsilon^{-1/2}\log L\rfloor$. In order to make sure that $r > r^*$, we need $L > \exp((r^*+1)\sqrt{\epsilon})$. We estimate, using proposition 4.9, (4.11), $\sqrt{\epsilon} < \epsilon_{\beta'}$, and $L > L^*$

$$\begin{aligned}\mathbb{P}\left(E_{\mathit{GS}}(H_\omega^L) \leqslant E_0 + \frac{4\epsilon}{(\log L)^2}\right) &\leqslant \exp\big(\exp((r+1)\sqrt{\epsilon})\log k - \exp(\epsilon_{\beta'}r)\big)\\ &\leqslant \exp\big(-\exp(\epsilon_{\beta'}r)/2\big) \leqslant \exp(-\mathrm{e}^{-\epsilon_{\beta'}}L^{\epsilon_{\beta'}/\sqrt{\epsilon}}/2) \leqslant \exp(-\zeta L)\end{aligned}$$

with $L^* := \max\{\exp((r^*+1)\sqrt{\epsilon}), \exp\big((\frac{\log(2\mathrm{e}^{\sqrt{\epsilon}}\log k)}{\epsilon_{\beta'}-\sqrt{\epsilon}}+1)\sqrt{\epsilon}\big), (2\mathrm{e}^{\epsilon_{\beta'}}\zeta)^{\sqrt{\epsilon}/(\epsilon_{\beta'}-\sqrt{\epsilon})}\}$. ∎

5. Main probability estimate

We remind the reader that

$$-\Delta_{\mathcal{B}} := k + 1 - A_{\mathcal{B}}$$

where $A_{\mathcal{B}}$ is the adjacency matrix of the infinite Bethe lattice $\mathcal{B}$ with symmetric spectrum $\sigma(A_{\mathcal{B}}) = [-2\sqrt{k}, 2\sqrt{k}]$. Thus, the Anderson Hamiltonian H_ω defined by (1.5) satisfies

$$H_\omega = k + 1 - A_{\mathcal{B}} + V_\omega.$$

We introduce the restriction of $A_{\mathcal{B}}$ to the finite rooted tree $\mathcal{T}^L$, which we denote by $A^{(L)}$. Note that $A^{(L)}$ is also the adjacency matrix of $\mathcal{T}^L$. Any property of the ground state energy $E_{\mathit{GS}}(H^L_\omega)$ can be restated in terms of the principal eigenvalue $\Lambda^{(L)}_\omega$ of the operator $A^{(L)}_\omega = A^{(L)} - V^{(L)}_\omega$, which we define as

$$\Lambda^{(L)}_\omega := \sup_{\|\varphi\|_2=1} \langle \varphi, A^{(L)}_\omega \varphi \rangle = k + 1 - E_{\mathit{GS}}(H^L_\omega).$$

We have indeed for $L \in \mathbb{N}$ and $\beta > 0$ the equivalence

$$E_{\mathit{GS}}(H^L_\omega) \leqslant E_0 + \beta L^{-2} \iff \Lambda^{(L)}_\omega \geqslant 2\sqrt{k} - \beta L^{-2}.$$

If we take $\beta < \sqrt{k}\pi^2$, then this inequality almost surely does not hold (trivial and obviously not very useful for our purposes). We restate theorem 4.10 as follows.

Theorem 5.1. *For every $\beta > 0$ there exists some $\epsilon_\beta > 0$, $L^* > 1$ so that for any $L > L^*$,*

$$\Lambda^{(L)}_\omega < 2\sqrt{k} - \beta L^{-2}$$

with probability at least

$$1 - \exp(-\mathrm{e}^{\epsilon_\beta L}).$$

This section will be devoted to the proof of theorem 5.1. Note that it furnishes the lower bound of theorem 1.5.

5.1. **Cutoffs in energy and space.** We claim first that, in order to attain an energy $E_0 + O(L^{-2})$ close to the bottom of the spectrum of H^L_ω (i. e. the top of the spectrum of $A^{(L)}_\omega$), a state must have both low kinetic energy and its potential energy close to the bottom of the spectrum. This will force the potential energy to deviate considerably from its mean, see proposition 5.4, which happens only with double exponentially small probability, see proposition 5.5.

To exploit the low energy of the states considered, we cut off all energies above a threshold. We implement this with the spectral projectors

$$\Pi^{(L)}_E \colon \ell^2(\mathcal{T}^L) \to \ell^2(\mathcal{T}^L)$$
$$\Pi^{(L)}_E \varphi = \mathbf{1}_{[E,+\infty)}(A^{(L)})\varphi = \sum_{\lambda^L_{v,j,m} \geqslant E} \langle \Psi^L_{v,j,m}, \varphi \rangle \Psi^L_{v,j,m},$$

where $E \in \mathbb{R}$ and the sum is taken over L-admissible indexes (v, j, m) with eigenvalue bounded below by E, see lemma 3.2.

Recall that at the beginning of section 3 we introduced a vertex $*$ and the notation $\mathcal{T}_*^L$. We used them to index the eigenvalues and eigenfunctions on the tree, see lemma 3.2.

Definition 5.2. For every $v \in \mathcal{T}_*^{L-1}$, define the orthogonal spectral projectors

$$P_v := \sum_{j \in J_v} P_{v,j} \tag{5.1}$$

using the notation from (3.4).

Remark 5.3. Here are some properties of these projectors. Let $v \in \mathcal{T}_*^{L-1}$. Then

- If $\chi_v = \mathbf{1}_{\mathcal{T}_v^L}$ is the characteristic function of the subtree $\mathcal{T}_v^L$, then for any $w \in \mathcal{T}_v^{L-1}$,
$$P_w = P_w \chi_v = \chi_v P_w.$$
In particular $P_v = P_v \chi_v = \chi_v P_v$.
- If we denote by $\operatorname{supp}\varphi$ the support of $\varphi \in \ell^2(\mathcal{T}^L)$, then for any $w \in \mathcal{T}_*^{L-1} \setminus \mathcal{T}_v^{L-1}$
$$\operatorname{supp}\bigl(P_v\varphi\bigr) \cap \operatorname{supp}\bigl(P_w\varphi\bigr) = \varnothing.$$

Given $\delta \in (0,1)$, the truncated spectral projector $\tilde{\Pi}^{(L)}_{E_\beta^{(L)}}$, see (3.10), can be written with this notation as

$$\tilde{\Pi}^{(L)}_E = \sum_{v \in \mathcal{T}_*^{L-1}} T_{|v|,\delta} P_v \Pi^{(L)}_E. \tag{5.2}$$

We note, for further use, that for any $v \in \mathcal{T}_*^{L-1}$,

$$\tilde{\Pi}^{(L)}_E P_v = T_{|v|,\delta} P_v \Pi^{(L)}_E. \tag{5.3}$$

This is easily seen using the commutativity and orthogonality of the spectral projectors. Using lemma 3.3, we also note that if $|v| > (1 - \frac{1}{\beta+1})(L+1)$ then

$$P_v \Pi^{(L)}_E = 0. \tag{5.4}$$

We finally introduce a notation for the centered potential:

$$\overline{V}^{(L)}_\omega := V^{(L)}_\omega - \bar{\omega}\mathbf{1}_{\mathcal{T}^L}, \tag{5.5}$$

where $\bar{\omega}$ is the expected value of the potential. We remind that the quantity $E_\beta^{(L)} := 2\sqrt{k}\cos\bigl(\frac{(\beta+1)\pi}{L+1}\bigr)$ was introduced in lemma 3.3. Let us now state the proposition.

Proposition 5.4. *Let $\beta' > 0$. For every $\beta \gg \beta'$ large enough, there exists some $\delta = \delta_\beta > 0$ and $L^* > 1$, so that for any $L > L^*$, then, the following inequality holds:*

$$\mathbb{P}\big(\Lambda_\omega^{(L)} \geqslant 2\sqrt{k} - \beta' L^{-2}\big) \leqslant \mathbb{P}\Big(\sup_{\|\varphi\|_2 \leqslant 1} \big|\langle \tilde{V}_\omega^{(L)} \varphi, \varphi\rangle\big| \geqslant \frac{\bar{\omega}}{16}\Big),$$

where we have introduced the notation

$$\tilde{V}_\omega^{(L)} := (\tilde{\Pi}^{(L)}_{E_\beta^{(L)}})^* \overline{V}_\omega^{(L)} \tilde{\Pi}^{(L)}_{E_\beta^{(L)}},$$

with $\tilde{\Pi}_E^{(L)}$ defined as in (5.2).

The key estimate is then given by the following proposition.

Proposition 5.5. *For any $\beta > 0$ large enough, let $\delta = \delta_\beta > 0$ given by proposition 5.4. Then, for L large enough,*

$$\mathbb{P}\Big(\sup_{\|\varphi\|_2 \leqslant 1} \big|\langle \tilde{V}_\omega^{(L)} \varphi, \varphi\rangle\big| \geqslant \frac{\bar{\omega}}{16}\Big) \leqslant \exp\big(-C_{k,\tilde{\omega}_+,\bar{\omega},\beta} k^{\delta_\beta L}\big).$$

Let us first prove proposition 5.4. We thereby reduce theorem 5.1 to proposition 5.5. The proof of proposition 5.5 is at the very end of this section. It hinges upon a series of lemmas and propositions which occupy the rest of this paper.

Proof of proposition 5.4. Fix a realization ω of the random potential with the property

$$\Lambda_\omega^{(L)} \geqslant 2\sqrt{k} - \beta' L^{-2}.$$

Then, there exists a $\varphi \in \ell^2(\mathcal{T}^L)$ with $\|\varphi\|_2 = 1$ such that

$$\langle A_\omega^{(L)} \varphi, \varphi\rangle \geqslant 2\sqrt{k} - \beta' L^{-2},$$

or, equivalently,

$$\langle (2\sqrt{k} - A^{(L)})\varphi, \varphi\rangle + \langle V_\omega^{(L)} \varphi, \varphi\rangle \leqslant \beta' L^{-2},$$

using $\|\varphi\|_2 = 1$. Note that the principal eigenvalue of $A^{(L)}$ is smaller than $2\sqrt{k}$. Thus, both $2\sqrt{k} - A^{(L)}$ and $V_\omega^{(L)}$ are non-negative operators. This implies that we have both

$$\langle (2\sqrt{k} - A^{(L)})\varphi, \varphi\rangle \leqslant \beta' L^{-2} \tag{5.6}$$

and

$$\langle V_\omega^{(L)} \varphi, \varphi\rangle \leqslant \beta' L^{-2}. \tag{5.7}$$

We now proceed as follows. In a first step, we introduce the energy cutoff $\Pi^{(L)}_{E_\beta^{(L)}}$ into (5.7). Here, (5.6) tells us how to choose β in order to keep the truncated version of (5.7) powerful enough. In a second step, we bring the spatial cutoff in $\tilde{\Pi}^{(L)}_{E_\beta^{(L)}}$ into play. This time, we have to choose $\delta > 0$ small enough, depending on β.

For the first step, let us write

$$\amalg^{(L)}_{E^{(L)}_\beta} := \mathbf{1}_{\ell^2(\mathcal{T}^L)} - \Pi^{(L)}_{E^{(L)}_\beta} \tag{5.8}$$

and $\omega_+ := \|V_\omega\|_\infty$. Then, we find that

$$\begin{aligned}\langle V^{(L)}_\omega \varphi, \varphi\rangle &= \langle V^{(L)}_\omega \Pi^{(L)}_{E^{(L)}_\beta}\varphi, \Pi^{(L)}_{E^{(L)}_\beta}\varphi\rangle + 2\Re\langle V^{(L)}_\omega \Pi^{(L)}_{E^{(L)}_\beta}\varphi, \amalg^{(L)}_{E^{(L)}_\beta}\varphi\rangle + \\ &\quad + \langle V^{(L)}_\omega \amalg^{(L)}_{E^{(L)}_\beta}\varphi, \amalg^{(L)}_{E^{(L)}_\beta}\varphi\rangle \\ &\geqslant \langle V^{(L)}_\omega \Pi^{(L)}_{E^{(L)}_\beta}\varphi, \Pi^{(L)}_{E^{(L)}_\beta}\varphi\rangle - 2\omega_+\|\Pi^{(L)}_{E^{(L)}_\beta}\varphi\|_2 \|\amalg^{(L)}_{E^{(L)}_\beta}\varphi\|_2.\end{aligned}$$

This, (5.7) and $\|\Pi^{(L)}_{E^{(L)}_\beta}\varphi\|_2 \leqslant 1$ imply that

$$\langle V^{(L)}_\omega \Pi^{(L)}_{E^{(L)}_\beta}\varphi, \Pi^{(L)}_{E^{(L)}_\beta}\varphi\rangle \leqslant \beta' L^{-2} + 2\omega_+\|\amalg^{(L)}_{E^{(L)}_\beta}\varphi\|_2.$$

We use now def. (5.5) in order to center the random variables so that their mean is zero. This gives,

$$\langle \overline{V}^{(L)}_\omega \Pi^{(L)}_{E^{(L)}_\beta}\varphi, \Pi^{(L)}_{E^{(L)}_\beta}\varphi\rangle \leqslant \beta' L^{-2} + 2\omega_+\|\amalg^{(L)}_{E^{(L)}_\beta}\varphi\|_2 - \bar{\omega}\|\Pi^{(L)}_{E^{(L)}_\beta}\varphi\|_2^2. \tag{5.9}$$

Using the non-negativity of the operator $2\sqrt{k} - A^{(L)}$, we see that

$$\begin{aligned}\left\langle (2\sqrt{k} - A^{(L)})\, \amalg^{(L)}_{E}\, \varphi, \amalg^{(L)}_{E}\varphi\right\rangle &= \sum_{\lambda^L_{v,j,m} \geqslant E} (2\sqrt{k} - \lambda^L_{v,j,m})|\langle \Psi^L_{v,j,m}, \varphi\rangle|^2 \\ &\leqslant \sum_{\lambda^L_{v,j,m} \geqslant -\infty} (2\sqrt{k} - \lambda^L_{v,j,m})|\langle \Psi^L_{v,j,m}, \varphi\rangle|^2 \\ &= \langle (2\sqrt{k} - A^{(L)})\varphi, \varphi\rangle.\end{aligned}$$

We use this with (5.6) to deduce that

$$\left\langle (2\sqrt{k} - A^{(L)})\, \amalg^{(L)}_{E^{(L)}_\beta}\, \varphi, \amalg^{(L)}_{E^{(L)}_\beta}\varphi\right\rangle \leqslant \beta' L^{-2}$$

and thus, using the definitions (5.8) and (3.9), this implies that

$$\left(2\sqrt{k} - E^{(L)}_\beta\right)\left\|\amalg^{(L)}_{E^{(L)}_\beta}\varphi\right\|^2 \leqslant \beta' L^{-2}.$$

Hence, using $\cos(x) \geqslant 1 - x^2/2$,

$$2\sqrt{k}\frac{(\beta+1)^2\pi^2}{(L+1)^2}\left\|\amalg^{(L)}_{E^{(L)}_\beta}\varphi\right\|^2 \leqslant \beta' L^{-2},$$

and thus

$$\left\|\amalg^{(L)}_{E^{(L)}_\beta}\varphi\right\| \leqslant \frac{L+1}{L(\beta+1)\pi}\sqrt{\frac{\beta'}{2\sqrt{k}}} \leqslant \frac{2}{(\beta+1)\pi}\sqrt{\frac{\beta'}{2\sqrt{k}}}.$$

From now on we assume we have chosen β so large that

$$\frac{1}{(\beta+1)\pi}\sqrt{\frac{2\beta'}{\sqrt{k}}} < \min\{1/\sqrt{2}, \bar{\omega}/(8\omega_+)\}.$$

This choice implies

$$1/2 \leqslant \|\Pi^{(L)}_{E^{(L)}_\beta}\varphi\|_2^2 \leqslant 1 \quad \text{and} \quad 2\omega_+\|\Pi^{(L)}_{E^{(L)}_\beta}\varphi\|_2\|\Pi^{(L)}_{E^{(L)}_\beta}\varphi\|_2 \leqslant \bar{\omega}/4.$$

We deduce from (5.9) that, for $L^2 > 8\beta'/\bar{\omega}$,

$$\langle \overline{V}^{(L)}_\omega \Pi^{(L)}_{E^{(L)}_\beta}\varphi, \Pi^{(L)}_{E^{(L)}_\beta}\varphi\rangle \leqslant -\frac{\bar{\omega}}{8}.$$

For the second step, let us now replace $\Pi^{(L)}_{E^{(L)}_\beta}$ by $\tilde{\Pi}^{(L)}_{E^{(L)}_\beta}$. We denote $\tilde{\omega}_+ := \|V_\omega - \bar{\omega}\|_\infty$. Choose $0 < \delta < 1$ satisfying

$$\sqrt{2}\pi\delta^{3/2}(\beta+1)^3 \leqslant \bar{\omega}/(32\tilde{\omega}_+).$$

Then, proposition 3.10 tells us that

$$2\tilde{\omega}_+\|\Pi^{(L)}_{E^{(L)}_\beta}\varphi - \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi\|_2 \leqslant \frac{\bar{\omega}}{16}\|\Pi^{(L)}_{E^{(L)}_\beta}\varphi\|_2 \leqslant \frac{\bar{\omega}}{16}.$$

Using this and $\|\tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi\|_2 \leqslant 1$, we deduce

$$\begin{aligned}\langle \overline{V}^{(L)}_\omega \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi, \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi\rangle &= \langle \overline{V}^{(L)}_\omega \Pi^{(L)}_{E^{(L)}_\beta}\varphi, \Pi^{(L)}_{E^{(L)}_\beta}\varphi\rangle + \langle \overline{V}^{(L)}_\omega (\tilde{\Pi}^{(L)}_{E^{(L)}_\beta} - \Pi^{(L)}_{E^{(L)}_\beta}\varphi, \Pi^{(L)}_{E^{(L)}_\beta}\varphi\rangle + \\ &\quad + \langle \overline{V}^{(L)}_\omega \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi, (\tilde{\Pi}^{(L)}_{E^{(L)}_\beta} - \Pi^{(L)}_{E^{(L)}_\beta}\varphi\rangle \\ &\leqslant -\frac{\bar{\omega}}{8} + 2\tilde{\omega}_+\|\Pi^{(L)}_{E^{(L)}_\beta}\varphi - \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi\|_2 \leqslant -\frac{\bar{\omega}}{16}.\end{aligned}$$

We have thereby proved that, for L large enough,

$$\{\omega : \Lambda^{(L)}_\omega \geqslant 2\sqrt{k} - \beta' L^{-2}\} \subseteq \Big\{\omega : \sup_{\|\varphi\|_2\leqslant 1}\big|\langle \overline{V}^{(L)}_\omega \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi, \tilde{\Pi}^{(L)}_{E^{(L)}_\beta}\varphi\rangle\big| \geqslant \frac{\bar{\omega}}{16}\Big\}.$$

This proves proposition 5.4. ∎

The spatial truncation we introduced into $\tilde{V}^{(L)}_\omega$ is adapted to the energy decomposition of the argument. More specifically, eigenfunctions with different anchors are treated differently. We therefore split the probability into different components depending on the anchors, see lemma 5.7.

We now prove a simple lemma.

Lemma 5.6. *Let $L \geqslant 1$ and $\varphi \in \ell^2(\mathcal{T}^L)$. Then:*

$$\sum_{v\in\mathcal{T}^{L-1}_*}\|\chi_v\varphi\|_2^2 \leqslant (L+1)\|\varphi\|_2^2.$$

Proof. We have $v \in \mathcal{T}_*^L$ and $w \in \mathcal{T}_v^L$ if and only if v lies in the shortest path from $*$ to w, which we write $v \in [*, w]$. Thus,

$$\sum_{v\in\mathcal{T}_*^{L-1}} \|\chi_v\varphi\|_2^2 \leqslant \sum_{v\in\mathcal{T}_*^{L}} \|\chi_v\varphi\|_2^2 = \sum_{v\in\mathcal{T}_*^{L}} \sum_{w\in\mathcal{T}_*^{L}} \chi_v(w)|\varphi(w)|^2 = \sum_{w\in\mathcal{T}_*^{L}} |\varphi(w)|^2 \sum_{v\in[*,w]} 1.$$

Now it suffices to remark that the maximum length of any shortest path from $*$ to any point of the tree is smaller or equal to $L+1$. ■

We introduce the following quantity. For any given $L \geqslant 1$, $v \in \mathcal{T}^L$, and $w \in \mathcal{T}_v^L$, define

$$\Xi(L, v, w) := \frac{1}{2}(L+1)^{-1} k^{-(|w|-|v|)/2}. \tag{5.10}$$

We also adopt the convention $0/0 = 0$.

Lemma 5.7. *Let $L \geqslant 1$, $\kappa > 0$, $B^{(L)}$ the unit ball of $\ell^2(\mathcal{T}^L)$ and $\mathcal{E}, \mathcal{F} \subseteq B^{(L)}$. Then the following inequality holds true*

$$\begin{aligned} &\mathbb{P}\Big(\sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}} |\langle \tilde V_\omega^{(L)}\varphi, \psi\rangle| > \kappa\Big) \\ &\qquad \leqslant \sum_{v\in\mathcal{T}_*^{L-1}} \sum_{w\in\mathcal{T}_v^{L-1}} \mathbb{P}\Big(\sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}} \frac{|\langle \tilde V_\omega^{(L)} P_v\varphi, P_w\psi\rangle|}{\|P_v\varphi\|_2 \|P_w\psi\|_2} > \kappa\Xi(L,v,w)\Big). \end{aligned}$$

Proof. First let us remark that for any $\varphi \in \ell^2(\mathcal{T}^L)$ we have $\varphi = \sum_{v\in\mathcal{T}_*^{L-1}} P_v\varphi$ and thus, using remark 5.3, we see that

$$\begin{aligned} \sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}} |\langle \tilde V_\omega^{(L)}\varphi, \psi\rangle| &\leqslant \sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}} \Big|\Big\langle \tilde V_\omega^{(L)} \sum_{v\in\mathcal{T}_*^{L-1}} P_v\varphi, \sum_{w\in\mathcal{T}_v^{L-1}} P_w\psi\Big\rangle\Big| \\ &\leqslant 2 \sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}} \sum_{v\in\mathcal{T}_*^{L-1}} \sum_{w\in\mathcal{T}_v^{L-1}} |\langle \tilde V_\omega^{(L)} P_v\varphi, P_w\psi\rangle|. \end{aligned} \tag{5.11}$$

The proof now proceeds as follows. In order to prove the inequality $\mathbb{P}(A) \leqslant \sum_j \mathbb{P}(B_j)$, we will show that $\bigcap_j B_j^c \subseteq A^c$. To do so, fix ω with the following property: for all $\varphi \in \mathcal{E}, \psi \in \mathcal{F}, v \in \mathcal{T}_*^{L-1}, w \in \mathcal{T}_v^{L-1}$,

$$|\langle \tilde V_\omega^{(L)} P_v\varphi, P_w\psi\rangle| \leqslant \kappa\Xi(L,v,w)\|P_v\varphi\|_2 \|P_w\psi\|_2. \tag{5.12}$$

Then, for any $\varphi \in \mathcal{E}, \psi \in \mathcal{F}$, we can use assumption (5.12) to bound

$$\begin{aligned}
&2 \sum_{v \in \mathcal{T}_*^{L-1}} \sum_{w \in \mathcal{T}_v^{L-1}} |\langle \tilde{V}_\omega^{(L)} P_v \varphi, P_w \psi \rangle| \\
&\leqslant \kappa (L+1)^{-1} \sum_{v \in \mathcal{T}_*^{L-1}} \|P_v \varphi\|_2 \sum_{w \in \mathcal{T}_v^{L-1}} k^{-(|w|-|v|)/2} \|P_w \psi\|_2 \\
&\leqslant \kappa (L+1)^{-1} \Big(\sum_{v \in \mathcal{T}_*^{L-1}} \|P_v \varphi\|_2^2 \Big)^{1/2} \Big(\sum_{v \in \mathcal{T}_*^{L-1}} B_v^2 \Big)^{1/2}, \qquad (5.13)
\end{aligned}$$

where we have used Cauchy–Schwarz in the last line and furthermore defined

$$B_v := \sum_{w \in \mathcal{T}_v^{L-1}} k^{-(|w|-|v|)/2} \|P_w \psi\|_2.$$

Using Cauchy–Schwarz and $P_w = P_w \chi_v$ (remark 5.3) in this last quantity, we see that

$$\sum_{v \in \mathcal{T}_*^{L-1}} B_v^2 \leqslant \sum_{v \in \mathcal{T}_*^{L-1}} \sum_{w \in \mathcal{T}_v^{L-1}} k^{-(|w|-|v|)} \sum_{w \in \mathcal{T}_v^{L-1}} \|P_w \chi_v \psi\|_2^2.$$

We can use polar coordinates to estimate the first sum over w. Indeed, note that, for $n \geqslant |v|$, the number of elements of the sphere $\{w \in \mathcal{T}_v^{L-1} : |w| = n\}$ is bounded by $k^{n-|v|}$. Thus, $\sum_{w \in \mathcal{T}_v^{L-1}} k^{-(|w|-|v|)} \leqslant \sum_{n=|v|}^{L} 1 \leqslant L+1$. With the orthogonality of the P_w and lemma 5.6, we see that

$$\sum_{v \in \mathcal{T}_*^{L-1}} B_v^2 \leqslant (L+1) \sum_{v \in \mathcal{T}_*^{L-1}} \|\chi_v \psi\|_2^2 \leqslant (L+1)^2 \|\psi\|_2^2.$$

We insert this bound into (5.13), apply $\sum_{v \in \mathcal{T}_*^{L-1}} \|P_v \varphi\|_2^2 = \|\varphi\|_2^2$ once more, and plug the result into (5.11), to see that assuming (5.12) for all $\varphi \in \mathcal{E}, \psi \in \mathcal{F}, v \in \mathcal{T}_*^{L-1}, w \in \mathcal{T}_v^{L-1}$ leads to

$$\sup_{\varphi \in \mathcal{E}, \psi \in \mathcal{F}} |\langle \tilde{V}_\omega^{(L)} \varphi, \psi \rangle| \leqslant \kappa.$$

This finishes the proof. ∎

5.2. **The epsilon-net argument.** The next problem we deal with is the fact that the ground state of H_ω is random. This is reflected in proposition 5.4 as follows. The supremum is inside the probability, so that φ and ψ are adapted to ω. In order to remove the supremum, we approximate the ball with a finite ϵ-net and show, with a union bound, that it suffices to consider only the elements of the net. The following two lemmas implement a classical ϵ-net argument.

Lemma 5.8. *Let $v \in \mathcal{T}_*^{L-1}$ and $B_v^{(L)}$ be the unit ball of $\operatorname{Im} P_v = P_v\big(\ell^2(\mathcal{T}^L)\big)$. Then there exists a finite set $\mathcal{M}_v \subseteq B_v^{(L)}$ so that for any $\varphi \in B_v^{(L)}$ there exists some $\tilde{\varphi} \in \mathcal{M}_v$ so that*

$$\|\varphi - \tilde{\varphi}\|_2 \leqslant 1/8$$

and furthermore

$$\#\mathcal{M}_v \leqslant 32^{k(L-|v|)}.$$

Proof. The existence of an ϵ-covering of the unit ball of a finite dimensional space having a cardinality smaller than $(4/\epsilon)^d$, where d is the dimension of the space, is a well-known fact, which can be established by scaling and volume counting, see for example [Pis99, formula (4.22)]. It suffices now to remark from the definition (5.1) that

$$\dim \operatorname{Im} P_v \leqslant k(L-|v|)$$

to establish the result. ■

Lemma 5.9. *Let a scale $L \geqslant 1$, a constant $\kappa > 0$, $B^{(L)}$ the unit ball of $\ell^2(\mathcal{T}^L)$ and sets $\mathcal{E}, \mathcal{F} \subseteq B^{(L)}$ be given. Further, we fix, for each $v \in \mathcal{T}_*^{L-1}$, some set $\mathcal{M}_v$ given by lemma 5.8. Then, the following inequality*

$$\begin{aligned}\mathbb{P}\Big(&\sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}} \frac{|\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\psi\rangle|}{\|P_v\varphi\|_2\|P_w\psi\|_2} > \kappa\Xi(L,v,w)\Big)\\ &\leqslant \sum_{i\in\mathbb{N}} \sum_{\tilde\varphi\in\mathcal{M}_v,\tilde\psi\in\mathcal{M}_w} \mathbb{P}\big(|\langle \tilde{V}_\omega^{(L)} P_v\tilde\varphi, P_w\tilde\psi\rangle| > 2^i\kappa\Xi(L,v,w)\|P_v\tilde\varphi\|_2\|P_w\tilde\psi\|_2\big)\end{aligned}$$

holds for all $v \in \mathcal{T}_*^{L-1}$, $w \in \mathcal{T}_v^{L-1}$.

Proof. Fix $v \in \mathcal{T}_*^{L-1}$, $w \in \mathcal{T}_v^{L-1}$. Using the fact that $P_v^2 = P_v$, we see that

$$\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\psi\rangle = \Big\langle \tilde{V}_\omega^{(L)} P_v \frac{P_v\varphi}{\|P_v\varphi\|_2}, P_w\frac{P_w\psi}{\|P_w\psi\|_2}\Big\rangle \|P_v\varphi\|_2\|P_w\psi\|_2.$$

We deduce that

$$\sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}} \frac{|\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\psi\rangle|}{\|P_v\varphi\|_2\|P_w\psi\|_2} \leqslant \sup_{\varphi\in B_v^{(L)},\psi\in B_w^{(L)}} |\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\psi\rangle|.$$

We remind the reader that $0/0 = 0$.

Let $\mathcal{M}_v, \mathcal{M}_w$ be the $\frac{1}{8}$-coverings of $B_v^{(L)}, B_w^{(L)}$ given by lemma 5.8, respectively. Suppose that $\kappa' > 0$ and that

$$|\langle \tilde{V}_\omega^{(L)} P_v\tilde\varphi, P_w\tilde\psi\rangle| \leqslant \frac{\kappa'}{2}\|P_v\tilde\varphi\|_2\|P_w\tilde\psi\|_2 \tag{5.14}$$

for all $\tilde\varphi \in \mathcal{M}_v$, $\tilde\psi \in \mathcal{M}_w$. Assume furthermore that

$$|\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\psi\rangle| \leqslant 2\kappa'\|P_v\varphi\|_2\|P_w\psi\|_2 \tag{5.15}$$

for every $\varphi \in B_v^{(L)}$, $\psi \in B_w^{(L)}$. Using the definition of $\mathcal{M}_v$, $\mathcal{M}_w$, we see that for any $\varphi \in B_v^{(L)}$, $\psi \in B_w^{(L)}$, there exists some $\tilde{\varphi} \in \mathcal{M}_v$, $\tilde{\psi} \in \mathcal{M}_w$ such that

$$\begin{aligned}|\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\psi\rangle| &\leqslant |\langle \tilde{V}_\omega^{(L)} P_v\tilde{\varphi}, P_w\tilde{\psi}\rangle| + |\langle \tilde{V}_\omega^{(L)} P_v\big(\tilde{\varphi}-\varphi\big), P_w\tilde{\psi}\rangle| \\ &\quad + |\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\big(\tilde{\psi}-\psi\big)\rangle| \\ &\leqslant \kappa'/2 + 2\kappa'\|P_v\big(\tilde{\varphi}-\varphi\big)\|_2 + 2\kappa'\|P_w(\tilde{\psi}-\psi)\|_2 \\ &\leqslant \kappa'/2 + \kappa'/4 + \kappa'/4 = \kappa'.\end{aligned}$$

We deduce that if $|\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\psi\rangle| > \kappa'$ then we cannot have both (5.14) and (5.15). We use this below with $\kappa', 2\kappa', 4\kappa', \ldots$. Thus,

$$\begin{aligned}\mathbb{P}\Big(\sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}} \frac{|\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\psi\rangle|}{\|P_v\varphi\|_2\|P_w\psi\|_2} > \kappa'\Big) &\leqslant \mathbb{P}\Big(\sup_{\varphi\in B_v^{(L)},\psi\in B_w^{(L)}} |\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\psi\rangle| > \kappa'\Big) \\ &\leqslant \sum_{\tilde{\varphi}\in\mathcal{M}_v,\psi\in\mathcal{M}_w} \mathbb{P}\big(|\langle \tilde{V}_\omega^{(L)} P_v\tilde{\varphi}, P_w\tilde{\psi}\rangle| > \kappa'\|P_v\tilde{\varphi}\|_2\|P_w\tilde{\psi}\|_2\big) \\ &\qquad + \mathbb{P}\Big(\sup_{\varphi\in B_v^{(L)},\psi\in B_w^{(L)}} |\langle \tilde{V}_\omega^{(L)} P_v\varphi, P_w\psi\rangle| > 2\kappa'\Big) \\ &\leqslant \sum_{i=1}^{\infty}\sum_{\tilde{\varphi}\in\mathcal{M}_v,\tilde{\psi}\in\mathcal{M}_w} \mathbb{P}\big(|\langle \tilde{V}_\omega^{(L)} P_v\tilde{\varphi}, P_w\tilde{\psi}\rangle| > 2^i\kappa'\|P_v\tilde{\varphi}\|_2\|P_w\tilde{\psi}\|_2\big).\end{aligned}$$

Now we choose $\kappa' := \kappa\Xi(L, v, w)$, and lemma 5.9 is proved. ∎

5.3. **Concentration inequalities.** With the lemmas we have up to now, we can attack the probability in proposition 5.5, but we will accumulate sums over $v \in \mathcal{T}_*^{L-1}$, $w \in \mathcal{T}_v^{L-1}$, $i \geqslant 1$, $\tilde{\phi} \in \mathcal{M}_v$ and $\tilde{\psi} \in \mathcal{M}_w$. The probabilities we sum over in the end should be very small in order to get a meaningful upper bound. We estimate these probabilities in proposition 5.10, which is the main probability estimate.

We remind the reader that $\Xi(L, v, w)$ was defined in (5.10) just before lemma 5.7, and that $\tilde{V}_\omega^{(L)} = (\Pi_{E_\beta^{(L)}}^{(L)})^*\overline{V}_\omega^{(L)}\Pi_{E_\beta^{(L)}}^{(L)}$. We further recall that $\overline{V}_\omega^{(L)}$ is the centered potential, see (5.5), and that the random variables ω_v are bounded almost surely, so $|\overline{V}_\omega^{(L)}| \leqslant \tilde{\omega}_+ := \|\omega_0 - \bar{\omega}\|_\infty$ almost surely.

Proposition 5.10. *Let $L \in \mathbb{N}$, $\beta \in (0, L]$, $v \in \mathcal{T}_*^{L-1}$, $w \in \mathcal{T}_v^{L-1}$, $\delta \in (0,1)$, and $\tilde{\varphi} \in \mathcal{M}_v$, $\tilde{\psi} \in \mathcal{M}_w$. Then*

$$|\langle \tilde{V}_\omega^{(L)} P_v\tilde{\varphi}, P_w\tilde{\psi}\rangle| > \kappa\Xi(L, v, w)\big\|P_v\Pi_{E_\beta^{(L)}}^{(L)}\tilde{\varphi}\big\|_2\big\|P_w\Pi_{E_\beta^{(L)}}^{(L)}\tilde{\psi}\big\|_2$$

holds true with probability smaller than

$$2\exp\big(-C_{k,\tilde{\omega}_+,\beta}\kappa^2 k^{\delta L}\big).$$

Here,

$$C_{k,\tilde{\omega}_+,\beta} := (64\tilde{\omega}_+^2 k^6(\beta+1)^4)^{-1} > 0.$$

To prove proposition 5.10 we will need the following two lemmas, the proofs of which are just below the proof of proposition 5.10. The first one is just an application of a well-known sub-Gaussian estimate.

Lemma 5.11. *For all $L \geqslant 1$, $\kappa > 0$ and any $\varphi, \psi \in \ell^2(\mathcal{T}_L)$, we have*

$$\mathbb{P}\big(|\langle \overline{V}_\omega^{(L)}\varphi, \psi\rangle| \geqslant \kappa\big) \leqslant 2\exp\Big(-\frac{\kappa^2}{2\tilde{\omega}_+^2\|\varphi\|_4^2\|\psi\|_4^2}\Big).$$

After applying lemma 5.11, we will be interested in certain ℓ^4-norms. The following estimate is tailored to our needs.

Lemma 5.12. *For all $L \in \mathbb{N}$, $\beta \in (0, L]$, $v \in \mathcal{T}_*^{L-1}$, $x \in \mathcal{T}_v^{L-1}$ satisfying $|x| > |v|$, and $\varphi \in \ell^2(\mathcal{T}^L)$,*

$$\big\|\chi_x P_v \Pi^{(L)}_{E_\beta^{(L)}}\varphi\big\|_4^4 \leqslant \frac{8k^6(\beta+1)^4}{(L+1)^2} k^{-2(|x|-|v|)}\big\|P_v\Pi^{(L)}_{E_\beta^{(L)}}\varphi\big\|_2^4$$

holds true.

We now prove proposition 5.10.

Proof of proposition 5.10. First, recall (5.3). This allows us to write

$$\begin{aligned} P_w\tilde{V}_\omega^{(L)}P_v &= P_w(\tilde{\Pi}^{(L)}_{E_\beta^{(L)}})^*\overline{V}_\omega^{(L)}\tilde{\Pi}^{(L)}_{E_\beta^{(L)}}P_v \\ &= \Pi^{(L)}_{E_\beta^{(L)}}P_w T_{|w|,\delta}\overline{V}_\omega^{(L)}T_{|v|,\delta}P_v\Pi^{(L)}_{E_\beta^{(L)}}, \end{aligned}$$

since the operators $T_{|\bullet|,\delta}$, $P_\bullet$, and $\chi_\bullet$ are self-adjoint. Furthermore, note that $T_{|w|,\delta} = T_{|w|,\delta}^2$ and recall from remark 5.3 that $P_w = P_w\chi_w$. The diagonal operators $T_{|w|,\delta}$, χ_w and $\overline{V}_\omega^{(L)}$ commute, so

$$P_w\tilde{V}_\omega^{(L)}P_v = \Pi^{(L)}_{E_\beta^{(L)}}P_w T_{|w|,\delta}\overline{V}_\omega^{(L)}\chi_w T_{|w|,\delta}T_{|v|,\delta}P_v\Pi^{(L)}_{E_\beta^{(L)}}.$$

Finally, compute $T_{|w|,\delta}T_{|v|,\delta} = T_{|w|,\delta}$. This leads us to study the quantity

$$\begin{aligned} X(x,v,w) &:= \langle \tilde{V}_\omega^{(L)}P_v\tilde{\varphi}, P_w\tilde{\psi}\rangle \\ &= \langle \overline{V}_\omega^{(L)}\chi_w T_{|w|,\delta}P_v\Pi^{(L)}_{E_\beta^{(L)}}\tilde{\varphi}, T_{|w|,\delta}P_w\Pi^{(L)}_{E_\beta^{(L)}}\tilde{\psi}\rangle, \end{aligned}$$

which is a sum of independent, bounded random variables. Note that the number of nodes in $\{x \in \mathcal{T}_w^{L-1} : |x| = |w| + \lceil \delta L\rceil\}$ is smaller than or equal to $k^{\lceil \delta L\rceil}$. Use this and lemma 5.12 to calculate

$$\begin{aligned}\big\|T_{|w|,\delta}P_w\Pi^{(L)}_{E^{(L)}_\beta}\tilde\psi\big\|_4^4 &= \sum_{\substack{x\in\mathcal{T}_w^{L-1}\\|x|=|w|+\lceil\delta L\rceil}}\big\|\chi_xP_w\Pi^{(L)}_{E^{(L)}_\beta}\tilde\psi\big\|_4^4 \leqslant k^{\lceil\delta L\rceil}\max_{\substack{x\in\mathcal{T}_w^{L-1}\\|x|=l}}\big\|\chi_xP_w\Pi^{(L)}_{E^{(L)}_\beta}\tilde\psi\big\|_4^4\\&\leqslant 8k^6\frac{(\beta+1)^4}{(L+1)^2}k^{-\lceil\delta L\rceil}\big\|P_w\Pi^{(L)}_{E^{(L)}_\beta}\tilde\psi\big\|_2^4\end{aligned}$$

and

$$\begin{aligned}\big\|\chi_wT_{|w|,\delta}P_v\Pi^{(L)}_{E^{(L)}_\beta}\tilde\varphi\big\|_4^4 &= \sum_{\substack{x\in\mathcal{T}_w^{L-1}\\|x|=|w|+\lceil\delta L\rceil}}\big\|\chi_xP_v\Pi^{(L)}_{E^{(L)}_\beta}\tilde\varphi\big\|_4^4 \leqslant k^{\lceil\delta L\rceil}\max_{\substack{x\in\mathcal{T}_w^{L-1}\\|x|=l}}\big\|\chi_xP_v\tilde\varphi\big\|_4^4\\&\leqslant 8k^6\frac{(\beta+1)^4}{(L+1)^2}k^{-2|w|-\lceil\delta L\rceil+2|v|}\big\|P_v\Pi^{(L)}_{E^{(L)}_\beta}\tilde\varphi\big\|_2^4.\end{aligned}$$

With these estimations, lemma 5.11, tells us that, if $\kappa'>0$,

$$\begin{aligned}\log\Big(\mathbb{P}\big(|X(x,v,w)|\geqslant\kappa'\big)/2\Big) &\leqslant -\frac{\kappa'^2}{2\tilde\omega_+^2\big\|\chi_xP_v\Pi^{(L)}_{E^{(L)}_\beta}\tilde\varphi\big\|_4^2\big\|\chi_xP_w\Pi^{(L)}_{E^{(L)}_\beta}\tilde\psi\big\|_4^2}\\&\leqslant -\frac{\kappa'^2(L+1)^2k^{|w|-|v|+\lceil\delta L\rceil}}{16\tilde\omega_+^2k^6(\beta+1)^4\big\|P_v\Pi^{(L)}_{E^{(L)}_\beta}\tilde\varphi\big\|_2^2\big\|P_w\Pi^{(L)}_{E^{(L)}_\beta}\tilde\psi\big\|_2^2}.\end{aligned}$$

We plug in

$$\kappa'=\kappa\Xi(L,v,w)\big\|P_v\Pi^{(L)}_{E^{(L)}_\beta}\tilde\varphi\big\|_2\big\|P_w\Pi^{(L)}_{E^{(L)}_\beta}\tilde\psi\big\|_2$$

and get

$$\begin{aligned}\log\Big(\frac12\mathbb{P}\big(|X(l,v,w)|\geqslant\kappa\Xi(L,v,w)\big\|P_v\Pi^{(L)}_{E^{(L)}_\beta}\tilde\varphi\big\|_2\big\|P_w\Pi^{(L)}_{E^{(L)}_\beta}\tilde\psi\big\|_2\big)\Big)&\\ \leqslant -\frac{\kappa^2k^{\lceil\delta L\rceil}}{64\tilde\omega_+^2k^6(\beta+1)^4}&.\end{aligned}$$

This finishes the proof. ■

Proof of lemma 5.11. Fix $\varphi,\psi\in\ell^2(\mathcal{T}^L)$. The expression

$$\langle\overline{V}^{(L)}_\omega\varphi,\psi\rangle=\sum_{v\in\mathcal{T}^L}(\omega_v-\bar\omega)\varphi(v)\psi(v)$$

is a sum of $\#\mathcal{T}^L$ independent random variables, namely $\{(\omega_v-\bar\omega)\varphi(v)\psi(v)\}_{v\in\mathcal{T}^L}$, all of them having mean zero. For every $v\in\mathcal{T}^L$, we have almost surely

$$|(\omega_v-\bar\omega)\varphi(v)\psi(v)|\leqslant\tilde\omega_+|\varphi(v)\psi(v)|.$$

To bound the probability in question, we use Hoeffding's inequality ([Hoe63]) and Cauchy–Schwarz:

$$\begin{aligned}\mathbb{P}\Big(|\langle (V_\omega^{(L)}-\bar\omega)\varphi,\psi\rangle|\geqslant\kappa\Big)&\leqslant 2\exp\Big(\frac{-2\kappa^2}{\sum_{v\in\mathcal{T}^L}(2\tilde\omega_+|\varphi(v)||\psi(v)|)^2}\Big)\\&\leqslant 2\exp\Big(\frac{-\kappa^2}{2\tilde\omega_+^2\|\varphi\|_4^2\|\psi\|_4^2}\Big).\end{aligned}$$ ■

Proof of lemma 5.12. Let $\varphi\in\ell^2(\mathcal{T}^L)$. To simplify notation, we let $\varphi\in P_v\Pi^{(L)}_{E^{(L)}_\beta}\ell^2(\mathcal{T}^L)$ For any L-admissible (v,j,m), let $\alpha_{v,j,m}$ be defined by

$$\varphi=\sum_{m=1}^{\lfloor\beta+1\rfloor}\sum_{j\in J_v}\alpha_{v,j,m}\Psi^L_{v,j,m},$$

and thus $\sum_{j,m}|\alpha_{v,j,m}|^2=\|\varphi\|_2^2$.

Using Cauchy–Schwarz,

$$\begin{aligned}\|\chi_x\varphi\|_4^4&=\sum_{w\in\mathcal{T}_x^L}\Big|\sum_{m=1}^{\lfloor\beta+1\rfloor}\sum_{j\in J_v}\alpha_{v,j,m}\Psi^L_{v,j,m}(w)\Big|^4\\&\leqslant\Big(\sum_{m=1}^{\lfloor\beta+1\rfloor}\sum_{j\in J_v}|\alpha_{v,j,m}|^2\Big)^2\sum_{w\in\mathcal{T}_x^L}\Big(\sum_{m=1}^{\lfloor\beta+1\rfloor}\sum_{j\in J_v}\big|\Psi^L_{v,j,m}(w)\big|^2\Big)^2\\&=\|\varphi\|_2^4\sum_{w\in\mathcal{T}_x^L}\Big(\sum_{m=1}^{\lfloor\beta+1\rfloor}\sum_{j\in J_v}\Big|\sum_{u\in\mathcal{T}_v^L,u\sim v}\psi^\perp_{v,j}(u)\psi^{L-|v|}_{u,m}(w)\Big|^2\Big)^2.\end{aligned}$$

Again with Cauchy–Schwarz and then with the definition (3.1) of $\psi^{L-|v|}_{u,m}$ we see

$$\begin{aligned}\sum_{j\in J_v}\Big|\sum_{u\in\mathcal{T}_v^L,u\sim v}\psi^\perp_{v,j}(u)\psi^{L-|v|}_{u,m}(w)\Big|^2&\leqslant\sum_{j\in J_v}\sum_{u\in\mathcal{T}_v^L,u\sim v}|\psi^\perp_{v,j}(u)|^2\sum_{u\in\mathcal{T}_v^L,u\sim v}|\psi^{L-|v|}_{u,m}(w)|^2\\&\leqslant\frac{2k^2}{(L+|v|-1)k^{|w|-|v|-1}}.\end{aligned}$$

We use all this in the estimate above and derive

$$\begin{aligned}\|\chi_x\varphi\|_4^4&\leqslant\frac{4k^4}{(L+|v|-1)^2}\|\varphi\|_2^4\sum_{w\in\mathcal{T}_x^L}\Big(\sum_{m=1}^{\lfloor\beta+1\rfloor}k^{-(|w|-|v|-1)}\Big)^2\\&\leqslant\frac{4k^4\lfloor\beta+1\rfloor^2}{(L+|v|-1)^2}\|\varphi\|_2^4\sum_{w\in\mathcal{T}_x^L}k^{-2(|w|-|v|-1)}\end{aligned}$$

The remaining sum can be treated with radial coordinates:

$$\sum_{w\in\mathcal{T}_x^L} k^{-2(|w|-|v|-1)} \leqslant \sum_{l=|x|}^{L} k^{l-|x|}k^{-2(l-|v|-1)} = k^{-|x|-2|v|+2}\sum_{l=|x|}^{L} k^{-l}$$
$$\leqslant k^{-|x|-2|v|+2}\frac{k^{-|x|}}{1-k^{-1}} = \frac{k^3}{k-1}k^{-2(|x|-|v|)}.$$

Since $k\geqslant 2$, we can beautify $k^3/(k-1)\leqslant 2k^2$ and get

$$\|\chi_x\varphi\|_4^4 \leqslant \frac{8k^6\lfloor\beta+1\rfloor^2}{(L+|v|-1)^2}\|\varphi\|_2^4 k^{-2(|x|-|v|)}.$$

Because of (5.4), we can assume $|v|\leqslant(1-\frac{1}{\beta+1})(L+1)$. This is equivalent to $\frac{1}{L+1-|v|}\leqslant\frac{\beta+1}{L+1}$, and the claim follows. ■

We finally are in position to finish the proof of the key probability estimate.

Proof of proposition 5.5. We need to bound

$$p := \mathbb{P}\Big(\sup_{\|\varphi\|_2\leqslant 1}\big|\langle\tilde V_\omega^{(L)}\varphi,\varphi\rangle\big| \geqslant \frac{\bar\omega}{16}\Big) \leqslant \mathbb{P}\Big(\sup_{\|\varphi\|_2\leqslant 1}\big|\langle\tilde V_\omega^{(L)}\varphi,\varphi\rangle\big| > \frac{\bar\omega}{32}\Big)$$

from above. Let us define

$$\mathcal{E} := \mathcal{F} := \big\{\varphi\in\Pi^{(L)}_{E_\beta^{(L)}}\big(\ell^2(\mathcal{T}^L)\big) : \|\varphi\|_2\leqslant 1\big\}.$$

By definition of $\tilde V_\omega^{(L)}$, see proposition 5.4, and by (5.3), we have

$$\sup_{\|\varphi\|_2\leqslant 1}\big|\langle\tilde V_\omega^{(L)}\varphi,\varphi\rangle\big| \leqslant \sup_{\varphi\in\mathcal{E}}\sup_{\psi\in\mathcal{F}}\big|\langle\tilde V_\omega^{(L)}\varphi,\psi\rangle\big|.$$

Using this, we can estimate with the help of lemma 5.7 and see

$$p \leqslant \sum_{v\in\mathcal{T}_*^{L-1}}\sum_{w\in\mathcal{T}_v^{L-1}}\mathbb{P}\bigg(\sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}}\frac{|\langle\tilde V_\omega^{(L)}P_v\varphi,P_w\psi\rangle|}{\|P_v\varphi\|_2\|P_w\psi\|_2} > \frac{\bar\omega}{32}\Xi(L,v,w)\bigg). \tag{5.16}$$

The terms in the sum (5.16) can be bounded using lemma 5.9,

$$\mathbb{P}\bigg(\sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}}\frac{|\langle\tilde V_\omega^{(L)}P_v\varphi,P_w\psi\rangle|}{\|P_v\varphi\|_2\|P_w\psi\|_2} > \frac{\bar\omega}{32}\Xi(L,v,w)\bigg)$$
$$\leqslant \sum_{i\in\mathbb{N}}\sum_{\tilde\varphi\in\mathcal{M}_v,\tilde\psi\in\mathcal{M}_w}\mathbb{P}\Big(|\langle\tilde V_\omega^{(L)}P_v\tilde\varphi,P_w\tilde\psi\rangle| > 2^i\frac{\bar\omega}{32}\Xi(L,v,w)\|P_v\tilde\varphi\|_2\|P_w\tilde\psi\|_2\Big), \tag{5.17}$$

where, for $v \in \mathcal{T}^L$, $\mathcal{M}_v \subseteq \ell^2(\mathcal{T}_v^L)$ with $\#\mathcal{M}_v \leqslant 32^{kL}$, see lemma 5.8. Using $\|P_v\tilde{\varphi}\|_2 \geqslant \big\|P_v\Pi^{(L)}_{E^{(L)}_\beta}\tilde{\varphi}\big\|_2$, valid for all $\tilde{\varphi} \in \ell^2(\mathcal{T}^L)$, proposition 5.10 tells us that

$$\begin{aligned}
&\mathbb{P}\Big(|\langle \tilde{V}^{(L)}_\omega P_v\tilde{\varphi}, P_w\tilde{\psi}\rangle| > 2^i\frac{\bar{\omega}}{32}\Xi(L,v,w)\|P_v\tilde{\varphi}\|_2\|P_w\tilde{\psi}\|_2\Big)\\
&\leqslant \mathbb{P}\Big(|\langle \tilde{V}^{(L)}_\omega P_v\tilde{\varphi}, P_w\tilde{\psi}\rangle| > 2^i\frac{\bar{\omega}}{32}\Xi(L,v,w)\big\|P_v\Pi^{(L)}_{E^{(L)}_\beta}\tilde{\varphi}\big\|_2\big\|P_w\Pi^{(L)}_{E^{(L)}_\beta}\tilde{\psi}\big\|_2\Big)\\
&\leqslant 2\exp\Big(-C_{\kappa,\tilde{\omega}_+,\beta}2^{2i}\frac{\bar{\omega}^2}{1024}k^{\delta L}\Big).
\end{aligned}$$

Plugging back into (5.17) and using $\#\mathcal{M}_v \leqslant 32^{kL}$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\varphi\in\mathcal{E},\psi\in\mathcal{F}}\frac{|\langle \tilde{V}^{(L)}_\omega P_v\varphi, P_w\psi\rangle|}{\|P_v\varphi\|_2\|P_w\psi\|_2} > \frac{\bar{\omega}}{16}\Xi(L,v,w)\Big)\\
&\leqslant \sum_{i\in\mathbb{N}}\sum_{\tilde{\varphi}\in\mathcal{M}_v,\tilde{\psi}\in\mathcal{M}_w} 2\exp\Big(-C_{\kappa,\tilde{\omega}_+,\beta}\frac{\bar{\omega}^2}{1024}2^{2i}k^{\delta L}\Big)\\
&\leqslant 2\cdot 32^{2kL}\sum_{i\in\mathbb{N}}\exp\Big(-C_{\kappa,\tilde{\omega}_+,\beta}2^{2i}\frac{\bar{\omega}^2}{1024}k^{\delta L}\Big).
\end{aligned}$$

The remaining sum can be bounded with a geometric series, since for all $x > \log 2$, we have

$$\sum_{i\in\mathbb{N}}\exp(-x2^{2i}) \leqslant \sum_{i=1}^{\infty}(\mathrm{e}^{-x})^i = \frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}} \leqslant 2\mathrm{e}^{-x}.$$

Finally, put this back into (5.16), to get, for all $L \in \mathbb{N}$ large enough,

$$\begin{aligned}
p &\leqslant \sum_{v\in\mathcal{T}_*^{L-1}}\sum_{w\in\mathcal{T}_v^{L-1}} 4\cdot 32^{2kL}\exp\Big(-C_{\kappa,\tilde{\omega}_+,\beta}\frac{\bar{\omega}^2}{1024}k^{\delta L}\Big)\\
&\leqslant 4k^{2L}32^{2kL}\exp\Big(-C_{\kappa,\tilde{\omega}_+,\beta}\frac{\bar{\omega}^2}{1024}k^{\delta L}\Big).
\end{aligned}$$

Taking L large enough, we get

$$p \leqslant \exp\Big(-C_{\kappa,\tilde{\omega}_+,\beta}\frac{\bar{\omega}^2}{2024}k^{\delta L}\Big).$$

The end. ■

References

[Abr10] E. Abrahams. *50 Years of Anderson Localization.* World Scientific Publishing Company, 2010. 612 pp. ISBN: 978-9-8142-9906-0 (cit. on p. 2).

[AF00] C. Allard and R. Froese. “A Mourre Estimate for a Schroedinger Operator on a Binary Tree.” *Reviews in Mathematical Physics* 12.12 (2000), pp. 1655–1667. DOI: 10.1142/S0129055X00000575 (cit. on p. 28).

[AF05] A. Alvermann and H. Fehske. "Local distribution approach to disordered binary alloys." *The European Physical Journal B. Condensed Matter and Complex Systems* 48.3 (2005), pp. 295–303. ISSN: 1434-6028. DOI: 10.1140/epjb/e2005-00408-8 (cit. on p. 2).

[Aiz94] M. Aizenman. "Localization at weak disorder: Some elementary bounds." *Reviews in Mathematical Physics* 6 (1994), pp. 1163–1182. DOI: 10.1142/S0129055X94000419 (cit. on p. 2).

[AK92] V. Acosta and A. Klein. "Analyticity of the density of states in the Anderson model on the Bethe lattice." *Journal of Statistical Physics* 69.1-2 (1992), pp. 277–305. DOI: 10.1007/BF01053794 (cit. on p. 2).

[Alt+97] B. Altshuler et al. "Quasiparticle Lifetime in a Finite System: A Nonperturbative Approach." *Physical Review Letters* 78.14 (1997), pp. 2803–2806. DOI: 10.1103/PhysRevLett.78.2803 (cit. on p. 9).

[AM93] M. Aizenman and S. Molchanov. "Localization at large disorder and at extreme energies: An elementary derivations." *Communications in Mathematical Physics* 157.2 (1993), pp. 245–278. ISSN: 1432-0916. DOI: 10.1007/bf02099760 (cit. on p. 9).

[And58] P. W. Anderson. "Absence of Diffusion in Certain Random Lattices." *Physical Review* 109.5 (1958), pp. 1492–1505. ISSN: 1536-6065. DOI: 10.1103/physrev.109.1492 (cit. on pp. 2, 9).

[ASW06] M. Aizenman, R. Sims, and S. Warzel. "Stability of the Absolutely Continuous Spectrum of Random Schrödinger Operators on Tree Graphs." *Probability Theory and Related Fields* 136.3 (2006), pp. 363–394. ISSN: 0178-8051. DOI: 10.1007/s00440-005-0486-8 (cit. on p. 2).

[ATA73] R. Abou-Chacra, D. J. Thouless, and P. W. Anderson. "A selfconsistent theory of localization." *Journal of Physics C: Solid State Physics* 6.10 (1973), pp. 1734–1752. DOI: 10.1088/0022-3719/6/10/009 (cit. on pp. 2, 9).

[AV09a] T. Antunović and I. Veselić. "Equality of Lifshitz and van Hove exponents on amenable Cayley graphs." *Journal de Mathématiques Pures et Appliquées* 92.4 (2009), pp. 342–362. DOI: 10.1016/j.matpur.2009.05.014. arXiv: 0706.2844.

[AV09b] T. Antunović and I. Veselić. "Spectral Asymptotics of Percolation Hamiltonians on Amenable Cayley Graphs." In: *Methods of Spectral Analysis in Mathematical Physics. Conference on Operator Theory, Analysis and Mathematical Physics (OTAMP) 2006, Lund, Sweden.* Vol. 186. Operator Theory: Advances and Applications. Birkhäuser Basel, 2009, pp. 1–29. ISBN: 978-3-7643-8754-9. DOI: 10.1007/978-3-7643-8755-6_1. arXiv: 0707.4292.

[AW06] M. Aizenman and S. Warzel. "The Canopy Graph and Level Statistics for Tandom Operators on Trees." *Mathematical Physics, Analysis and*

Geometry 9.4 (2006), pp. 291–333. DOI: 10.1007/s11040-007-9018-3 (cit. on pp. 10, 28).

[AW11a] M. Aizenman and S. Warzel. "Absence of mobility edge for the Anderson random potential on tree graphs at weak disorder." *EPL (Europhysics Letters)* 96.3 (2011), pp. 1–6. DOI: 10.1209/0295-5075/96/37004 (cit. on pp. 2, 9).

[AW11b] M. Aizenman and S. Warzel. "Extended States in a Lishitz Tail Regime for Random Schrödinger Operators on Trees." *Physical Review Letters* 106 (2011), pp. 1–6. DOI: 10.1103/PhysRevLett.106.136804 (cit. on pp. 2, 9).

[BAF04] F. X. Bronold, A. Alvermann, and H. Fehske. "Anderson localization in strongly coupled disordered electron-phonon systems." *Philosophical Magazine* 84.7 (2004), pp. 673–704. DOI: 10.1080/14786430310001624884 (cit. on p. 2).

[BK01] M. Biskup and W. König. "Long-Time Tails in the Parabolic Anderson Model with Bounded Potential." *The Annals of Probability* 29.2 (2001), pp. 636–682. URL: http://projecteuclid.org/euclid.aop/1008956688 (cit. on pp. 6, 10, 13).

[BLM89] J. Bourgain, J. Lindenstrauss, and V. Milman. "Approximation of zonoids by zonotopes." *Acta Mathematica* 162.1 (1989), pp. 73–141. DOI: 10.1007/BF02392835.

[Bro91] R. Brooks. "The Spectral Geometry of k-regular Graphs." *Journal d'Analyse Mathématique* 57 (1991), pp. 120–151. ISSN: 0021-7670. DOI: 10.1007/BF03041067 (cit. on pp. 11, 24).

[BS11] V. Bapst and G. Semerjian. "Lifshitz Tails on the Bethe Lattice: A Combinatorial Approach." *Journal of Statistical Physics* 145.1 (2011), pp. 51–92. DOI: 10.1007/s10955-011-0319-3 (cit. on pp. 3, 9, 10, 12).

[BST10] G. Biroli, G. Semerjian, and M. Tarzia. "Anderson model on Bethe lattices: density of states, localization properties and isolated eigenvalue." *Progress of Theoretical Physsics Supplements* 184 (2010), pp. 187–199. DOI: 10.1143/PTPS.184.187. arXiv: 1005.0342 (cit. on pp. 3, 9).

[CL90] R. Carmona and J. Lacroix. *Spectral Theory of Random Schrödinger Operators*. Probability and Its Applications. Boston: Birkhäuser, 1990. ISBN: 978-1-4612-8841-1. DOI: 10.1007/978-1-4612-4488-2 (cit. on pp. 5, 9).

[DV75] M. D. Donsker and S. R. S. Varadhan. "Asymptotics for the Wiener sausage." *Communications in Pure and Applied Mathematics* 28.4 (1975), pp. 525–565. ISSN: 1097-0312. DOI: 10.1002/cpa.3160280406 (cit. on p. 6).

[Gei14] L. Geisinger. "Convergence of the density of states and delocalization of eigenvectors on random regular graphs" (2014). arXiv: 1305.1039v2 (cit. on p. 11).

[GK05] J. Gärtner and W. König. "The Parabolic Anderson Model." In: *Interacting Stochastic Systems*. .I. Springer Berlin Heidelberg, 2005, pp. 153–179. ISBN: 978-3-540-23033-5. DOI: 10.1007/3-540-27110-4_8 (cit. on p. 10).

[GK13] F. Germinet and A. Klein. "A comprehensive proof of localization for continuous Anderson models with singular random potentials." *Journal of the European Mathematical Society* 15.1 (2013), pp. 53–143. ISSN: 1435-9855. DOI: 10.4171/jems/356 (cit. on p. 9).

[His08] P. D. Hislop. "Lectures on random Schrödinger operators." In: *Fourth Summer School in Analysis and Mathematical Physics. Topics in Spectral Theory and Quantum Mechanics*. Ed. by C. Villegas-Blas. Vol. 476. American Mathematical Society, 2008, pp. 41–131. ISBN: 978-0-8218-4064-1. DOI: 10.1090/conm/476 (cit. on p. 9).

[Hoe63] W. Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables." *Journal of the American Statistical Association* 58.301 (1963), pp. 13–30. DOI: 10.2307/2282952. JSTOR: 2282952 (cit. on p. 57).

[HS82] B. Hughes and M. Sahimi. "Random walks on the Bethe lattice." *Journal of Statistical Physics* 29.4 (1982), pp. 781–794. DOI: 10.1007/BF01011791.

[Kes59] H. Kesten. "Symmetric Random Walks on Groups." *Transactions of the American Mathematical Society* 92.2 (1959), pp. 336–354. DOI: 10.2307/1993160 (cit. on pp. 6, 11).

[KH85] Y. Kim and A. B. Harris. "Density of states of the random-hopping model on a Cayley tree." *Physical Review B* 31.11 (1985), pp. 7393–7407. DOI: 10.1103/PhysRevB.31.7393 (cit. on pp. 2, 3).

[Kir07] W. Kirsch. "An Invitation to Random Schroedinger operators" (2007). arXiv: 0709.3707 (cit. on p. 9).

[Kle96] A. Klein. "Spreading of wave packets in the Anderson model on the Bethe lattice." *Communication in Mathematical Physics* 177.3 (1996), pp. 755–773. ISSN: 0010-3616. DOI: 10.1007/BF02099546 (cit. on pp. 2, 9).

[Kle98] A. Klein. "Extended States in the Anderson Model on the Bethe Lattice." *Advances in Mathematics* 133.1 (1998), pp. 163–184. DOI: 10.1006/aima.1997.1688 (cit. on p. 2).

[KM06] W. Kirsch and B. Metzger. "The Integrated Density of States for Random Schroedinger Operators" (2006). arXiv: 0608066 (cit. on pp. 5, 6).

[KM83] W. Kirsch and F. Martinelli. "Large deviations and Lifshitz singularity of the integrated density of states of random Hamiltonians." *Communications in Mathematical Physics* 89.1 (1983), pp. 27–40. DOI: 10.1007/BF01219524 (cit. on p. 11).

[KN03] F. Klopp and S. Nakamura. "A note on Anderson localization for the random hopping model." *Journal of Mathematical Physics* 44.11 (2003), pp. 4975–4980. DOI: 10.1063/1.1616998.

[KS83] H. Kunz and B. Souillard. "The localization transition on the Bethe lattice." *Journal de Physique Lettres* 44.11 (1983), pp. 411–414. DOI: 10.1051/jphyslet:019830044011041100 (cit. on p. 2).

[KW] W. König and T. Wolff. "The Parabolic Anderson Model." Non published draft. URL: http://www.wias-berlin.de/people/koenig/www/PAMsurvey.pdf (visited on 10/10/2014) (cit. on p. 10).

[Lif65] I. M. Lifshitz. "Energy spectrum structure and quantum states of disordered condensed systems." *Soviet Physics Uspekhi* (1965), pp. 549–573. DOI: 10.1070/PU1965v007n04ABEH003634 (cit. on p. 3).

[McK81] B. D. McKay. "The expected eigenvalue distribution of a large regular graph." *Linear Algebra and its Applications* 40 (1981), pp. 203–216. DOI: 10.1016/0024-3795(81)90150-6 (cit. on pp. 6, 11).

[MD94] J. D. Miller and B. Derrida. "Weak disorder expansion for the Anderson model on a tree." *Journal of Statistical Physics* 75 (1994), pp. 357–388. ISSN: 0022-4715. DOI: 10.1007/BF02186867. arXiv: cond-mat/9312023 (cit. on p. 2).

[MF91] A. D. Mirlin and Y. V. Fyodorov. "Localization transition in the Anderson model on the Bethe lattice. Spontaneous symmetry breaking and correlation functions." *Nuclear Physics B* 366.3 (1991), pp. 507–532. ISSN: 0550-3213. DOI: 10.1016/0550-3213(91)90028-V (cit. on p. 2).

[MG09] C. Monthus and T. Garel. "Anderson transition on the Cayley tree as a traveling wave critical point for various probability distributions." *Journal of Physics A. Mathematical and Theoretical* 42.7 (2009), pp. 1–34. DOI: 10.1088/1751-8113/42/7/075002 (cit. on p. 2).

[MS11] P. Müller and P. Stollmann. "Percolation Hamiltonians." In: *Random Walks, Boundaries and Spectra*. Ed. by D. Lenz, F. Sobieczky, and W. Woess. Vol. 64. Progress in Probability. Springer Basel, 2011, pp. 235–258. ISBN: 978-3-0346-0243-3. DOI: 10.1007/978-3-0346-0244-0_13 (cit. on p. 10).

[Nak77] S. Nakao. "On the spectral distribution of the Schrödinger operator with random potential." *Japanese Journal of Mathematics. New Series* 3.1 (1977), pp. 111–139. URL: https://www.jstage.jst.go.jp/article/math1924/3/1/3_1_111/_article (visited on 10/10/2014) (cit. on pp. 5, 6).

[Pas71] L. A. Pastur. "On the Schrödinger equation with a random potential." *Theoretical and Mathematical Physics* 6.3 (1971), pp. 299–306. ISSN: 1573-9333. DOI: 10.1007/bf01030111 (cit. on p. 6).

[Pas77] L. A. Pastur. "Behavior of some Wiener integrals as $t \to \infty$ and the density of states of Schrödinger equations with random potential."

Theoretical and Mathematical Physics 32.1 (1977), pp. 615–620. ISSN: 1573-9333. DOI: 10.1007/bf01041435 (cit. on pp. 5, 6).

[PF92] L. A. Pastur and A. Figotin. *Spectra of random and almost-periodic operators*. Vol. 297. Grundlehren der mathematischen Wissenschaften. Berlin: Springer, 1992. 593 pp. ISBN: 3-540-50622-5 (cit. on pp. 5, 9).

[Pis99] G. Pisier. *The Volume of Convex Bodies and Banach Space Geometry*. Cambridge Tracts in Mathematics 94. Cambridge University Press, 1999. ISBN: 978-0-521-66635-0 (cit. on p. 53).

[Rei09] T. Reinold. "Über die integrierte Zustandsdichte des Laplace-Operators auf Bond-Perkolationsgraphen des Bethe-Gitters." German. Diploma thesis. Institut für Theoretische Physik der Georg-August-Universität Göttingen, 2009. 63 pp. (cit. on p. 10).

[Ros12] V. Ros. "On the Lifshitz tails phenomenon for the Density of States of Anderson Hamiltonians on trees." PhD thesis. LMU and TU München, 2012. 71 pp. (cit. on pp. 9, 10, 12, 13, 21, 23).

[RR07] O. Rojo and M. Robbiano. "An explicit formula for eigenvalues of Bethe trees and upper bounds on the largest eigenvalue of any tree." *Linear Algebra and its Applications* 427.1 (2007), pp. 138–150. DOI: 10.1016/j.laa.2007.06.024 (cit. on p. 24).

[SCW02] W. Schlag, S. C., and T. Wolff. "Frequency concentration and location lengths for the Anderson model at small disorders." *Journal d'Analyse Mathématique* 88 (2002), pp. 173–220. DOI: 10.1007/BF02786577.

[Sim05] B. Simon. *Trace Ideals and Their Applications*. 2nd ed. Vol. 120. Mathematical Surveys and Monographs. American Mathematical Society, 2005. 152 pp. ISBN: 978-0-8218-3581-4 (cit. on p. 22).

[Sim83] B. Simon. "Semiclassical analysis of low lying eigenvalues. I. Non-degenerate minima: asymptotic expansions." *Annales de l'Institut Henri Poincaré. Section A. Physique Théorique. Nouvelle Série* 38.3 (1983), pp. 295–308. ISSN: 0246-0211 (cit. on p. 32).

[Sim85] B. Simon. "Lifschitz tails for the Anderson model." *Journal of Statistical Physics* 38.1-2 (1985), pp. 65–76. ISSN: 0022-4715. DOI: 10.1007/BF01017848 (cit. on p. 11).

[Sni89] A. Snitzman. "Lifschitz Tail and Wiener Sausage on hyperbolic space." *Communications on Pure and Applied Mathematics* 42.8 (1989), pp. 1033–1065. DOI: 10.1002/cpa.3160420802 (cit. on p. 10).

[Sni90] A. Snitzman. "Lifschitz tail on hyperbolic space: Neumann conditions." *Communications on Pure and Applied Mathematics* 43.1 (1990), pp. 1–30. DOI: 10.1002/cpa.3160430102 (cit. on p. 10).

[SS14] C. Schumacher and F. Schwarzenberger. "Approximation of the Integrated Density of States on Sofic Groups." *Annales Henri Poincaré* (2014), 35 pp. ISSN: 1424-0661. DOI: 10.1007/s00023-014-0342-4. arXiv: 1212.5207 (cit. on p. 11).

[Sto01] P. Stollmann. *Caught by disorder. Bound States in Random Media*. Vol. 20. Progress in Mathematical Physics. Boston: Birkhäuser, 2001. 183 pp. ISBN: 978-1-4612-6644-0. DOI: 10.1007/978-1-4612-0169-4 (cit. on p. 9).

[Sto99] P. Stollmann. "Lifshitz Asymptotics via Linear Coupling of Disorder." *Mathematical Physics, Analysis and Geometry* 2.3 (1999), pp. 279–289. ISSN: 1385-0172. DOI: 10.1023/a:1009845206542 (cit. on p. 11).

[War13] S. Warzel. "Surprises in the phase diagram of the Anderson model on the Bethe lattice." *XVIIth International Congress on Mathematical Physics* (2013), pp. 239–253. DOI: 10.1142/9789814449243_0014. arXiv: 1212.4367v1 (cit. on pp. 2, 3, 9).

Acknowledgment

The research of the authors was supported by DFG projects *Zufällige und periodische Quantengraphen* and *Eindeutige Fortsetzungsprinzipien und Gleichverteilungseigenschaften von Eigenfunktionen*. Most of this paper was written while both authors were members of the Stochastics Chair at the TU Chemnitz. The authors thank Constanza Rojas-Molina and Prof. Ivan Veselić for reading an early version of the introduction.

(F. Hoecker-Escuti) Hamburg University of Technology, Institute of Mathematics, Am Schwarzenberg-Campus 3, D-21073 Hamburg, Germany

E-mail address: hoecker@math.univ-paris13.fr

(C. Schumacher) Technische Universität Dortmund, Fakultät für Mathematik, Vogelpothsweg 87, D-44227 Dortmund, Germany

E-mail address: christoph.schumacher@mathematik.tu-dortmund.de

Chapter 8

Lifshitz Asymptotics in case of monotone random breather potentials

LIFSHITZ ASYMPTOTICS IN CASE OF MONOTONE RANDOM BREATHER POTENTIALS

IVAN VESELIĆ AND CHRISTOPH SCHUMACHER

Abstract. We recall the construction of the integrated density of states (IDS) of random Schrödinger operators on $\mathbb{R}^d$ with periodic background potential. For all non–negative random potentials, we prove Lifshitz behavior at the bottom of the spectrum, which is that for low energies, the IDS is exponentially small. The theory is developed for the breather potential and generalized to all non–negative random potentials in a second step. We use the Lifshitz tails result to prove an initial length scale estimate which may in turn be useful in a proof of Anderson localization via multiscale analysis. Finally, we use complement the Lifshitz behavior with a lower bound.

Contents

1. Introduction

1.1. **Random Schrödinger operators and the IDS.** We consider Schrödinger operators on $L^2(\mathbb{R}^d)$ with a random, $\mathbb{Z}^d$-ergodic potential. More precisely, we fix a measurable space $(\Omega_0, \mathcal{A}_0)$ and a jointly measurable *single site potential* $u\colon \Omega_0 \times \mathbb{R}^d \to \mathbb{R}$. With the notiation $u_\lambda := u(\lambda, \,\cdot\,)\colon \mathbb{R}^d \to \mathbb{R}$, $\lambda \in \Omega_0$, it becomes obvious that Ω_0 serves as an index set for a whole family of single site potentials.

The single site potentials are combined randomly on $\mathbb{R}^d$. To this end, we use the canonical probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $\Omega := \bigotimes_{\mathbb{Z}^d} \Omega_0$ and an i. i. d.

family of random variables $\lambda_k\colon \Omega \to \Omega_0$, indexed by $k \in \mathbb{Z}^d$, via

$$W_\omega\colon \mathbb{R}^d \to \mathbb{R}, \quad W_\omega(x) = \sum_{k\in\mathbb{Z}^d} u_{\lambda_k(\omega)}(x-k) \qquad (\omega \in \Omega). \tag{1.1}$$

We assume that there exists $p > \max\{2, d/2\}$ such that

$$\operatorname{ess\,sup}_{\mathbb{P}} |u_{\lambda_0}| \in L^p(\ell^1),$$

where $L^p(\ell^1)$ is the set of all (equivalence classes of) functions $f\colon \mathbb{R}^d \to \mathbb{R}$ such that $\left\|\chi_{\mathcal{D}} \sum_{k\in\mathbb{Z}^d} |f(\,\cdot\, - k)|\right\|_{L^p} < \infty$ with the indicator function $\chi_{\mathcal{D}}$ of the fundamental domain $\mathcal{D} := [0,1)^d$ of $\mathbb{Z}^d$. Under this assumption, the random potential (1.1) is in $L^p_{\mathrm{loc,unif}}(\mathbb{R}^d)$, uniformly in $\omega \in \Omega$. An application of the Kato–Rellich theorem, see e. g. [RS78, Theorem XIII.96], for a $\mathbb{Z}^d$-periodic potential $V_{\mathrm{per}} \in L^\infty(\mathbb{R}^d, \mathbb{R})$, the operators

$$H_{\mathrm{per}} := -\Delta + V_{\mathrm{per}} \quad \text{and} \quad H_\omega := H_{\mathrm{per}} + W_\omega \tag{1.2}$$

are self-adjoint on the domain $\operatorname{dom}\Delta$ of Δ and lower bounded uniformly in $\omega \in \Omega$.

The joint measurability of the single site potential implies the measurablilty of the family $(H_\omega)_{\omega\in\Omega}$, cf. [KM82b]. Moreover, $(H_\omega)_{\omega\in\Omega}$ forms an *ergodic family of operators* in the following sense. Let $U\colon \mathbb{Z}^d \to \mathcal{B}(L^2(\mathbb{R}^d))$, $(U(z)f)(x) = f(x+z)$ be the unitary representation of $\mathbb{Z}^d$, acting by translation. There is an ergodic $\mathbb{Z}^d$-action $\vartheta\colon \mathbb{Z}^d \times \Omega \to \Omega$ on $(\Omega, \mathcal{A}, \mathbb{P})$, which satisfies

$$H_{\vartheta(x,\omega)} = U(x)^{-1} H_\omega U(x) \qquad (x \in \mathbb{Z}^d, \omega \in \Omega).$$

Hence there exists a closed set $\Sigma \subseteq \mathbb{R}$ and an event $\Omega' \in \mathcal{A}$ of full probability, such that for all $\omega \in \Omega'$, the spectrum of H_ω coincides with Σ, cf. [KM82b].

For the definition of the integrated density of states (IDS) $N\colon \mathbb{R} \to \mathbb{R}$ for $(H_\omega)_{\omega\in\Omega}$ we follow [Pas80; KM82a]. Denote for $L \in \mathbb{N}$ and $x \in \mathbb{R}^d$

$$\Lambda_L := [-L, L)^d, \quad \Lambda_L(x) := \Lambda_L + x, \quad \mathcal{F} := \{\Lambda_L(x) \mid x \in \mathbb{Z}^d, L \in \mathbb{N}\}.$$

Neumann (N), Dirichlet (D), periodic (P) and Mezincescu (M) boundary conditions, a specific choice of Robin boundary conditions, see Section 1.2, give rise to self-adjoint restrictions $H^{\Lambda,\sharp}$, $\sharp \in \{N, D, P, M\}$, of an operator H to the box $\Lambda \in \mathcal{F}$. We defer detailed definitions to Section 1.2. We write in short $H^{L,\sharp} := H^{\Lambda_L,\sharp}$.

It is well known, see [KM82a; Ves08], that the finite volume restrictions of H_ω, $\omega \in \Omega$, have compact resolvents, so that their spectrum is purely discrete. The eigenvalue counting functions

$$n^\sharp(E, H_\omega, \Lambda) := \operatorname{Tr}\bigl(\chi_{(-\infty,E]}(H_\omega^{\Lambda,\sharp})\bigr)$$

and its normalized versions

$$N^\sharp(E, H_\omega, \Lambda) := |\Lambda|^{-1} n^\sharp(E, H_\omega, \Lambda),$$

are thereby well defined for $\omega \in \Omega$, $E \in \mathbb{R}$, $\Lambda \in \mathcal{F}$ and $\sharp \in \{N, D, P, M\}$. Again we write briefly $n^\sharp_L(E, H_\omega) := n^\sharp(E, H_\omega, \Lambda_L)$, $L \in \mathbb{N}$, and analogously $N^\sharp_L$.

The eigenvalue counting functions are *equivariant*, i. e. for all $k \in \mathbb{Z}^d$, $E \in \mathbb{R}$, $\Lambda \in \mathcal{F}$, $\omega \in \Omega$ and $\sharp \in \{D, N, P, M\}$ we have

$$N^\sharp(E, H_{\vartheta(k,\omega)}, \Lambda) = N^\sharp(E, H_\omega, \Lambda + k).$$

Moreover, as we will see in Section 1.2, n^N and n^M are *subadditive*, i. e. for any $\Lambda \in \mathcal{F}$ given as a finite disjoint union $\Lambda = \bigcup_j \Lambda_j$ of cubes $\Lambda_j \in \mathcal{F}$,

$$n^\sharp(E, H_\omega, \Lambda) \leq \sum\nolimits_j n^\sharp(E, H_\omega, \Lambda_j) \qquad (\sharp \in \{N, M\})$$

holds. Together with the ergodicity of $(H_\omega)_{\omega\in\Omega}$, it follows that there exists for each $E \in \mathbb{R}$ an event $\Omega_E \in \mathcal{A}$ of probability 1, such that for all $\omega \in \Omega_E$

$$\lim_{L\to\infty} N_L^\sharp(E, H_\omega) = \inf_{L\in\mathbb{N}} \mathbb{E}[N_L^\sharp(E, H_\bullet)] \qquad (\sharp \in \{N, M\}).$$

Analogously, n^D is superadditive, and, w. l. o. g., for the same event Ω_E we have

$$\lim_{L\to\infty} N_L^D(E, H_\omega) = \sup_{L\in\mathbb{N}} \mathbb{E}[N_L^D(E, H_\bullet)] \qquad (\omega \in \Omega_E, E \in \mathbb{R}),$$

cf. [KM82a].

We are now in position to define the IDS of $(H_\omega)_{\omega\in\Omega}$ in two steps. For $\sharp \in \{D, N, M\}$ and $\omega \in \tilde{\Omega} := \bigcap_{E\in\mathbb{Q}} \Omega_E$, the function

$$\tilde{N}^\sharp\colon \mathbb{Q} \to \mathbb{R}, \quad \tilde{N}^\sharp(E') := \lim_{L\to\infty} N_L^\sharp(E', H_\omega)$$

is well-defined, non-decreasing and does not depend on ω almost surely. The IDS $N\colon \mathbb{R} \to \mathbb{R}$ of $(H_\omega)_{\omega\in\Omega}$ is the right continuous version of $\tilde{N}^\sharp$:

$$N(E) := \lim_{\mathbb{Q}\ni E'\searrow E} \tilde{N}^\sharp(E') = \inf\bigl\{\tilde{N}^\sharp(E') \mid E' \in \mathbb{Q} \cap (E, \infty)\bigr\}.$$

As indicated by the notation, the IDS N is independent of the choice of boundary conditions. This can be infered from [KM82a; HS04] as follows. See also [Hup+01; DIM01]. Since Neumann and Dirichlet boundary conditions bracket Mezincescu boundary conditions, it suffices to show

$$N_L^N(E, H_\omega) - N_L^D(E, H_\omega) \xrightarrow{L\to\infty} 0 \qquad (\text{for a. a. } \omega \text{ and all } E \in \mathbb{R}).$$

This is proved in [KM82a, Theorem 3.3] under the additional assumption that the Laplace transform

$$\mathbb{E}\bigl[\operatorname{Tr}\exp(-t_0(-\Delta^{N,\mathcal{D}} + q(V_{\mathrm{per}} + W_\bullet)))\bigr] < \infty$$

for some $q > 1$, $t_0 > 0$. Here, Tr denotes the trace of operators. If we split $V_{\mathrm{per}} = V_+ - V_-$, $V_+, V_- \geq 0$, by [HS04], the above condition is satisfied as soon as V_- is relatively form bounded with respect to $-\Delta^{N,\mathcal{D}}$, $V_+ + W_\omega \in L^1_{\mathrm{loc}}(\mathcal{D})$, and

$$\operatorname{Tr}\exp(\Delta^{N,\mathcal{D}} - 2V_-) < \infty.$$

But all this follows from $V_{\mathrm{per}} \in L^\infty(\mathbb{R}^d)$ and [KM82a, Proposition 2.1].

The right continuity of N implies

$$N(E) \geq \inf_{L\in\mathbb{N}} \mathbb{E}[N_L^\sharp(E, H_\bullet)] \qquad (\sharp \in \{N, M\})$$

for all $E \in \mathbb{R}$. Indeed, for all $\varepsilon > 0$ we find an $E' \in \mathbb{Q} \cap (E, \infty)$ and an $L \in \mathbb{N}$ with

$$N(E) + 2\varepsilon \geq \tilde{N}^\sharp(E') + \varepsilon \geq \mathbb{E}[N_L^\sharp(E', H_\bullet)] \geq \mathbb{E}[N_L^\sharp(E, H_\bullet)].$$

Actually, for every continuity point $E \in \mathbb{R}$ of N we have

$$N(E) = \inf_{L \in \mathbb{N}} \mathbb{E}[N_L^\sharp(E, H_\bullet)] \qquad (\sharp \in \{N, M\}). \tag{1.3}$$

Indeed, by continuity of N in E, there exists for all $\varepsilon > 0$ a $\delta > 0$ such that for all $L \in \mathbb{N}$ and $E', E'' \in \mathbb{Q}$ such that $E - \delta < E' < E'' < E$, and $\sharp \in \{N, M\}$,

$$N(E) - \varepsilon \leq N(E') \leq \tilde{N}^\sharp(E'') \leq \mathbb{E}[N_L^\sharp(E'', H_\bullet)] \leq \mathbb{E}[N_L^\sharp(E, H_\bullet)].$$

By an analogous argument we have for continuity points $E \in \mathbb{R}$ of N

$$N(E) = \sup_{L \in \mathbb{N}} \mathbb{E}[N_L^D(E, H_\bullet)]. \tag{1.4}$$

Note, that in [BK11] the continuity of N is proved for $d \in \{1, 2, 3\}$ and bounded W_ω. For specific types of W_ω, Wegner estimates are available, implying the continuity of the IDS, see e. g. [Ves08; Nak+15].

1.2. **Boundary Conditions.** We have used above that the eigenvalue counting functions are equivariant and sub- resp. superadditive. These properties are inherited from the corresponding properties of the respective operator family $(H_\omega^{\Lambda,\sharp})_{\omega \in \Omega, \Lambda \in \mathcal{F}}$. The latter is *equivariant*, if $U(x)^{-1} H_\omega^{\Lambda + x, \sharp} U(x) = H_{\vartheta(x,\omega)}^{\Lambda,\sharp}$, for all $x \in \mathbb{Z}^d$, all $\Lambda \in \mathcal{F}$ and almost all $\omega \in \Omega$. All boundary conditions we consider lead to equivariant operator families.

With Dirichlet boundary conditions, the family $(H_\omega^{\Lambda,D})_{\omega \in \Omega, \Lambda \in \mathcal{F}}$ is *subadditive*, meaning that for all disjoint unions $\Lambda = \bigcup \Lambda_j \in \mathcal{F}$ of cubes $\Lambda_j \in \mathcal{F}$, we have

$$H_\omega^{\Lambda,D} \leq \bigoplus H_\omega^{\Lambda_j,D},$$

see [RS78, p. 270, Proposition 4]. *Superadditivity* is defined with the opposite inequality and applies for Neumann boundary conditions:

$$\bigoplus H_\omega^{\Lambda_j,N} \leq H_\omega^{\Lambda,N}.$$

Covariance of $N^\sharp$ is implied by equivariance of $(H_\omega^{\Lambda,\sharp})_{\omega \in \Omega, \Lambda \in \mathcal{F}}$. Superadditivity of $(H_\omega^{\Lambda,\sharp})$ implies subadditivity of $n^\sharp$, and subadditivity of $(H_\omega^{\Lambda,\sharp})$ implies superadditivity of $n^\sharp$. This is because the smaller operator has more eigenvalues below a given threshold.

In the case $V_{\mathrm{per}} = 0$, Neumann boundary conditions work well for our purposes. Otherwise we have to resort to Mezincescu boundary conditions. Like Neumann boundary conditions they lead to equivariance and superadditivity. But they additionally preserve the ground state energy of the periodic operator: $\inf \sigma(H_{\mathrm{per}}^{\Lambda,M}) = \inf \sigma(H_{\mathrm{per}})$, $\Lambda \in \mathcal{F}$.

Following [Mez87; KW05; KW06; KV10], we define Mezincescu boundary conditions as Robin boundary conditions with a specific function $\rho_\Lambda \in L^\infty(\partial\Lambda, \sigma)$, where σ is the surface measure on the boundary $\partial\Lambda$. This means, $-\Delta^{\Lambda,M}$ is the operator associated with the sesquilinear form

$$(\varphi,\psi)\mapsto\int_\Lambda\overline{\nabla\varphi(x)}\nabla\psi(x)\,\mathrm{d}x+\int_{\partial\Lambda}\overline{\varphi(x)}\psi(x)\rho_\Lambda(x)\,\sigma(\mathrm{d}x)\tag{1.5}$$

with the Sobolev space $H^1(\Lambda)$ as its form domain. Here and in the following, we use the same name for a function on Λ and for its trace on $\partial\Lambda$. The domain of $-\Delta^{\Lambda,M}$ turns out to be the set of $\varphi\in H^2(\Lambda)$ which satisfy $\rho_\Lambda\varphi+\frac{\partial\varphi}{\partial n}=0$ on $\partial\Lambda$, where $\frac{\partial}{\partial n}$ is the outer normal derivative. Also, on its domain, $-\Delta^{\Lambda,M}$ acts as usual as the negative of the sum of the second derivatives. The details for case of von Neumann boundary conditions $\rho_\Lambda=0$ can be found found in e. g. [Sch12a, section 10.6.2]. The general case $\rho_\Lambda\in L^\infty(\Lambda)$ can be established analogously.

Note that for all $\rho_\Lambda\in L^\infty(\partial\Lambda)$, we have

$$H^{\Lambda,D}_\omega\geq H^{\Lambda,M}_\omega\tag{1.6}$$

in the form sense, where $H^{\Lambda,M}_\omega:=-\Delta^{\Lambda,M}+V_{\mathrm{per}}+V_\omega$. In fact, the form which defines the Dirichlet Laplace operator is the restriction of (1.5) to $H^1_0(\Lambda)$.

Mezincescu's choice for the function ρ_Λ is constructed as follows. Note that the restriction $H^{\mathcal{D},P}_{\mathrm{per}}$ of H_{per} to $\mathcal{D}=[0,1)^d$ with periodic boundary conditions has a positive and normed ground state $\Psi_{\mathcal{D}}\in H^1(\mathcal{D})$, $\|\Psi_{\mathcal{D}}\|_2=1$. If we extend $\Psi_{\mathcal{D}}$ periodically to $\mathbb{R}^d$, we obtain

$$\Psi\in L^\infty(\mathbb{R}^d)\cap H^1_{\mathrm{loc}}(\mathbb{R}^d).\tag{1.7}$$

In fact, by [Sim82, Theorem B.3.5], Ψ is continuously differentiable with a Hölder continuous gradient. Since $-\Delta$ is elliptic, Harnack's inequality, see [Sim82, Theorem C.1.3], applies:

$$0<\Psi_-:=\min\Psi(\mathbb{R}^d)\leq\Psi_+:=\max\Psi(\mathbb{R}^d)<\infty.\tag{1.8}$$

We now define $\rho_\Lambda:=-\frac{1}{\Psi}\frac{\partial\Psi}{\partial n}$ on $\partial\Lambda$. Equivariance of $(H^{\Lambda,M}_\omega)_{\omega\in\Omega,\Lambda\in\mathcal{F}}$ is clear from construction, and superadditivity is shown in [Mez87, Proposition 1].

Next, we argue that, for all $\Lambda\in\mathcal{F}$,

$$E_1(H^{\Lambda,P}_{\mathrm{per}})=E_1(H^{\Lambda,M}_{\mathrm{per}})=\inf\sigma(H_{\mathrm{per}}).\tag{1.9}$$

By construction, Ψ satisfies the eigenvalue equation $H_{\mathrm{per}}\Psi=E_1(H^{\mathcal{D},P}_{\mathrm{per}})\Psi$. Since Ψ is bounded, [Sim82, Theorem C.4.1] implies $E_1(H^{\mathcal{D},P}_{\mathrm{per}})\in\sigma(H_{\mathrm{per}})$. On the other hand, Ψ is positive, so that $E_1(H^{\mathcal{D},P}_{\mathrm{per}})\leq\inf\sigma(H_{\mathrm{per}})$ by [Sim82, Theorem C.8.1]. Thus we conclude $E_1(H^{\Lambda,P}_{\mathrm{per}})=\inf\sigma(H_{\mathrm{per}})$. Furthermore, for all $\Lambda\in\mathcal{F}$, the function $\Psi_\Lambda:=|\Lambda|^{-1/2}\chi_\Lambda\Psi$ is in the domains of $H^{\Lambda,P}_{\mathrm{per}}$ and of $H^{\Lambda,M}_{\mathrm{per}}$, and an eigenvector with the eigenvalue $E_1(H^{\mathcal{D},P}_{\mathrm{per}})$. Again by positivity and [Sim82, Theorem C.8.1], Ψ_Λ the ground state of $H^{\Lambda,M}_{\mathrm{per}}$ and of $H^{\Lambda,M}_{\mathrm{per}}$. That proves (1.9)

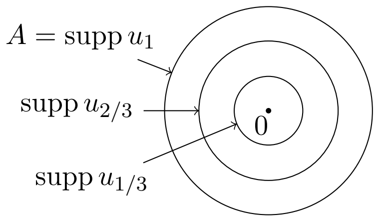


FIGURE 1. Support of single site potential u_λ for different values of λ with circular base set A

For $V_{\mathrm{per}} = 0$ we have $E_2(-\Delta^{L,N}) = \frac{\pi^2}{4L^2}$. The spectral gap between the two lowest eigenvalues of $H_{\mathrm{per}}^{L,M}$ satisfies, for some $C_1 > 0$ and all $L \in \mathbb{N}$,

$$E_2(H_{\mathrm{per}}^{L,M}) - E_1(H_{\mathrm{per}}^{L,M}) \geq C_1/L^2, \tag{1.10}$$

cf. [KS87].

2. The breather model

For the breather model, we consider a specific measurable space $\Omega_0 := [0,1]$ and a specific single site potential $u\colon [0,1] \times \mathbb{R}^d \to \mathbb{R}$. Choose a measurable set $A \subseteq \mathcal{D} := [-\frac{1}{2}, \frac{1}{2})^d \subseteq \mathbb{R}^d$ of positive Lebesgue measure and a coupling strength $\mu > 0$. The single site potential is defined via

$$u(\lambda, x) := \mu \chi_{\lambda A}(x), \tag{2.1}$$

see Figures 1 and 2 for illustrations. As in Section 1.1, the random variables $\lambda_k\colon \Omega \to [0,1]$, $k \in \mathbb{Z}$, shall be independent and identically distributed. We require 0 to be in the support of the distribution of λ_0:

$$\forall \varepsilon > 0\colon \mathbb{P}\{\lambda_0 \leq \varepsilon\} > 0, \tag{2.2}$$

but of cause we want randomness, so we assume

$$\mathbb{P}\{\lambda_0 = 0\} < 1. \tag{2.3}$$

Note that the distribution of λ_0 may but does not have to have an atom at $0 \in \mathbb{R}$. Choosing this specific type of single site potential (2.1) in the random potential in (1.1) and the corresponding random Schrödinger operator $(H_\omega)_{\omega\in\Omega}$ gives rise to the random breather model.

The basic main result of this chapter is the following.

Theorem 1. *The IDS N of the random breather model with single site potentials given by* (2.1) *satisfies a Lifshitz bound, i. e.* $\exists C_2, C_3 > 0, E' > E_0\colon \forall E \in (E_0, E']$:

$$N(E) \leq C_2(E - E_0)^{d/2} \exp\bigl(-C_3(E - E_0)^{-d/2}\bigr). \tag{2.4}$$

Before we prove Theorem 1 in Section 2.2, we show that the support condition (2.2) ensures the following well known lemma.

Lemma 2. *For the random breather model with* (2.2)*, we have*

$$E_0 = \inf \sigma(H_\omega) \qquad \text{for } \mathbb{P}\text{-a. a. } \omega \in \Omega. \tag{2.5}$$

Proof. We have $W_\omega \geq 0$. Hence by the min-max-principle $\inf \sigma(H_\omega) \geq E_0$.

For the other inequality, namely $\inf \sigma(H_\omega) \leq E_0$ for a. a. $\omega \in \Omega$, we construct for a. a. $\omega \in \Omega$ an approximate eigenfunction for H_ω. We use a mollyfier $\tilde{\chi} \in C^\infty(\mathbb{R}^d, [0,1])$ with $\chi_{\mathcal{D}} \leq \tilde{\chi} \leq \chi_{\Lambda_1}$, or rather $\tilde{\chi}_{x,L}(y) := \tilde{\chi}((y-x)/L)$, $x \in \mathbb{Z}^d$, $L \in \mathbb{N}$. Our ansatz for the approximate eigenfunction is

$$\Psi_{x,L} := \Psi \cdot \tilde{\chi}_{x,L} / \|\Psi \cdot \tilde{\chi}_{x,L}\|_2 \in \mathrm{dom}(\Delta),$$

with Ψ from (1.7), for suitable $x \in \mathbb{Z}^d$ and $L \in \mathbb{N}$ to be choosen later. We have to bound

$$\|(H_\omega - E_0)\psi_{x,L}\|_2 \leq \|(H_{\mathrm{per}} - E_0)\psi_{x,L}\|_2 + \|W_\omega \psi_{x,L}\|_2.$$

Lemmas 3 and 4 provide for $\mathbb{P}$-almost all $\omega \in \Omega$ and all $L \in \mathbb{N}$ an $x \in \mathbb{Z}^d$ such that

$$\|(H_\omega - E_0)\psi_{x,L}\|_2 \leq (C_4 + 1)/L.$$

This suffices to conclude $\inf \sigma(H_\omega) \leq E_0$. □

The following lemma provides the non–random estimate used in Lemma 2.

Lemma 3. *There exists a constant $C_4 > 0$ such that all $L \in \mathbb{N}$*

$$\|(H_{\mathrm{per}} - E_0)\Psi_{x,L}\|_2 \leq C_4/L.$$

Proof. We use $(-\Delta + V_{\mathrm{per}})\Psi = E_0\Psi$ for

$$\begin{aligned}\|(H_{\mathrm{per}} - E_0)(\Psi\tilde{\chi}_{x,L})\|_2 &= \|2\nabla\Psi \cdot \nabla\tilde{\chi}_{x,L} + \Psi\Delta\tilde{\chi}_{x,L}\|_2 \\ &\leq 2\|\chi_{\Lambda_L(x)}\nabla\Psi\|_2\|\nabla\tilde{\chi}_{x,L}\|_\infty + \|\Psi\|_\infty\|\Delta\tilde{\chi}_{x,L}\|_2.\end{aligned}$$

A short calculation shows

$$\|\nabla\tilde{\chi}_{x,L}\|_\infty = L^{-1}\|\nabla\tilde{\chi}\|_\infty \quad \text{and} \quad \|\Delta\tilde{\chi}_{x,L}\|_2 = L^{d/2-2}\|\Delta\tilde{\chi}\|_2.$$

Furthermore,

$$\|\chi_{\Lambda_L(x)}\nabla\Psi\|_2 = L^{d/2}\|\chi_{\Lambda_1}\nabla\Psi\|_2 \quad \text{and} \quad \|\Psi\tilde{\chi}_{x,L}\|_2 \geq L^{d/2}\|\Psi\chi_{\mathcal{D}}\|_2 = L^{d/2}.$$

We combine this to get

$$\|(H_{\mathrm{per}} - E_0)\Psi_{x,L}\|_2 \leq \frac{2\|\chi_{\Lambda_1}\nabla\Psi\|_2\|\nabla\tilde{\chi}\|_\infty + \Psi_+\|\Delta\tilde{\chi}\|_2/L}{L}$$

and choose $C_4 := 2\|\chi_{\Lambda_1}\nabla\Psi\|_2\|\nabla\tilde{\chi}\|_\infty + \Psi_+\|\Delta\tilde{\chi}\|_2$. □

Lemma 4 deals with the random part in the estimate from Lemma 2.

Lemma 4. *There exists a set $\Omega_{BC} \in \mathcal{A}$ of full probability $\mathbb{P}(\Omega_{BC}) = 1$, such that for all $\omega \in \Omega_{BC}$, $\varepsilon > 0$ and $L \in \mathbb{N}$, there exist $x \in \mathbb{Z}^d$ satisfying*

$$\|W_\omega\Psi_{x,L}\|_2 \leq \varepsilon.$$

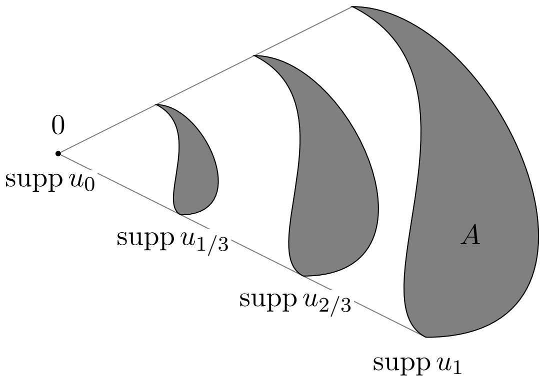


FIGURE 2. Support of single site potential u_λ for different values of λ with arbitrary base set A

Proof. Let $I_L := \mathbb{Z}^d \cap \Lambda_L$ and $I_L(x) := x + I_L$. For all $\alpha > 0$ and $L \in \mathbb{N}$, the Borel–Cantelli lemma provides us with a set $\Omega_{\alpha,L} \in \mathcal{A}$ of full measure, such that for all $\omega \in \Omega_{\alpha,L}$, there exists $x = x_{L,\alpha,\omega} \in \mathbb{Z}^d$ such that

$$\sup_{k \in I_L(x)} \lambda_k(\omega) \leq \alpha. \tag{2.6}$$

We let $\Omega_{BC} := \bigcap_{\alpha \in \mathbb{Q}, \alpha > 0} \bigcap_{L \in \mathbb{N}} \Omega_{\alpha,L} \in \mathcal{A}$ and note that $\mathbb{P}(\Omega_{BC}) = 1$, and that (2.6) holds for all $\alpha > 0$, $L \in \mathbb{N}$, and $\omega \in \Omega_{BC}$. Furthermore, (2.6) implies $\sup_{k \in I_L(x)} \|u_{\lambda_k(\omega)}\|_2 \leq \mu\sqrt{\alpha^d |A|}$. Hence,

$$\|W_\omega \chi_{\Lambda_L(x)}\|_2 \leq \sum_{k \in I_L(x)} \|u_{\lambda_k(\omega)}(\,\cdot\, - k)\|_2 \leq \mu\sqrt{\alpha^d |A|}\, \# I_L$$

and

$$\|W_\omega \Psi_{x,L}\|_2 \leq \frac{\|\Psi\|_\infty}{\|\Psi \tilde{\chi}_{x,L}\|_2} \|W_\omega \chi_{\Lambda_L(x)}\|_2 \leq \frac{\Psi_+ \mu\sqrt{\alpha^d |A|}\, \# I_L}{\|\Psi \tilde{\chi}_{x,L}\|_2}.$$

Now choose $\alpha := \bigl(\varepsilon \|\Psi \tilde{\chi}_{x,L}\|_2 / (\Psi_+ \mu \sqrt{|A|}\, \# I_L)\bigr)^{2/d}$ to finish the proof. $\square$

2.1. **Temple vs. Thirring.** The method of [KV10] is not applicable to the potential (2.1), at least not directly, because of the use of Temple's inequality [Tem28]. Provided that ψ is a vector in the domain of a lower bounded self-adjoint operator A with $E_1(A) = \inf \sigma(A)$ being a simple eigenvalue and $E_2(A) = \inf(\sigma(A) \setminus \{E_1(A)\}$ being an eigenvalue, and $\nu \in \mathbb{R}$ a number such that $\inf \sigma(A) \leq \langle \psi, A\psi \rangle < \nu \leq E_2(A)$, then Temple's inequality states that the lowest eigenvalue $E_1(A)$ of A is lower bounded by

$$E_1(A) \geq \langle \psi, A\psi \rangle - \frac{\|A\psi\|^2 - (\langle \psi, A\psi \rangle)^2}{\nu - \langle \psi, A\psi \rangle}.$$

Here $\langle \psi, A\psi \rangle$ can be considered as the first order approximation to $E_1(A)$. Indeed, it is a true *upper bound*. To obtain a lower bound one has to subtract a normalization of the second order correction $\|A\psi\|^2 - (\langle\psi, A\psi\rangle)^2$. This expression is non-negative, since it is a variance. In our application, considering eigenvalues close to zero, $\langle\psi, A\psi\rangle$ is small, consequently is $(\langle\psi, A\psi\rangle)^2$ quadratically small and thus negligible. If the single site potential $u(\lambda, x) = \chi_{\lambda A}(x)$ is a characteristic function of λA, we have

$$\langle\psi, u(\lambda, \cdot)^2\psi\rangle = \langle\psi, u(\lambda, \cdot)\psi\rangle$$

and if the translates $k + \lambda A, k \in \mathbb{Z}^d$, do not overlap for any allowed value of λ, e. g. is A is small, then we have for the resulting breather potential W_ω in (1.1) again

$$\langle\psi, W_\omega^2\psi\rangle = \langle\psi, W_\omega\psi\rangle$$

The natural choice of a test function ψ is the ground state of $H_{\mathrm{per}}^{L,M}$ with eigenvalue $E_1(H_{\mathrm{per}}^{L,M}) = 0$. Then we have

$$\|H_\omega^{L,M}\psi\| = \langle\psi, H_\omega^{L,M}\psi\rangle$$

hence the second moment is equal to the first one and cannot be considered as small correction. Note that the difference $\nu - \langle\psi, H_\omega^{L,M}\psi\rangle$ is bounded by the gap between the first two eigenvalues of $H_\omega^{L,M}$, typically of order L^{-2}. Thus, dividing by this number actually makes the correction term even larger.

It turns out that Thirring's inequality [Thi94, p. 3.5.32] is better adapted to the model under consideration. It was used before in [KM83] in a similar context. For the readers convenience we reprove Thirring's inequality here.

Lemma 5 (Thirring). *Let V be an invertible, positive operator on the Hilbert space $\mathcal{H}$, $P\colon \mathcal{H} \to P(\mathcal{H}) \subseteq \mathcal{H}$ an orthogonal projection and suppose, that the operator $PV^{-1}P^* \in \mathcal{B}(P\mathcal{H})$ is invertible. Then*

$$P^*(PV^{-1}P^*)^{-1}P \leq V.$$

In consequence, if H is a self-adjoint operator on $\mathcal{H}$, bounded from below, and

$$E_1(H) \leq E_2(H) \leq E_3(H) \leq \cdots \leq E_\infty(H) := \inf \sigma_{\mathrm{ess}}(H)$$

is the sequence of eigenvalues of H below $E_\infty(H)$, counted with multiplicity, then for all $n \in \mathbb{N}$

$$E_n\big(H + P^*(PV^{-1}P^*)^{-1}P\big) \leq E_n(H + V).$$

Proof. Let $Q := V^{-\frac{1}{2}}P^*(PV^{-1}P^*)^{-1}PV^{-\frac{1}{2}}$. Note that because of

$$Q^2 = V^{-\frac{1}{2}}P^*(PV^{-1}P^*)^{-1}PV^{-1}P^*(PV^{-1}P^*)^{-1}PV^{-\frac{1}{2}} = Q = Q^*,$$

Q is itself an orthogonal projection. Therefore $Q \leq \mathrm{Id}_{\mathcal{H}}$, i. e. $\langle\psi, Q\psi\rangle \leq \langle\psi, \psi\rangle$ for all $\psi \in \mathcal{H}$. This directly implies

$$P^*(PV^{-1}P^*)^{-1}P = V^{\frac{1}{2}}QV^{\frac{1}{2}} \leq V.$$

By the min-max-principle, see e. g. [Thi94, 3.5.21], for all $n \in \mathbb{N}$

$$\begin{aligned} &E_n\big(H + P^*(PV^{-1}P^*)^{-1}P\big) \\ &\quad = \inf_{\substack{\mathcal{H}' \subseteq \mathcal{H} \\ \dim(\mathcal{H}')=n}} \sup_{\substack{\psi \in \mathcal{H}' \\ \|\psi\|=1}} \langle \psi, (H + P^*(PV^{-1}P^*)^{-1}P)\psi \rangle \\ &\quad \le \inf_{\substack{\mathcal{H}' \subseteq \mathcal{H} \\ \dim(\mathcal{H}')=n}} \sup_{\substack{\psi \in \mathcal{H}' \\ \|\psi\|=1}} \langle \psi, (H + V)\psi \rangle = E_n(H+V) \end{aligned}$$

□

We actually need only the following special case of Lemma 5.

Corollary 6. *Let $\psi \in \mathcal{H}$ be a (normalised) ground state of H, i. e. $H\psi = E_1(H)\psi$ and $\|\psi\| = 1$, and $P = P_\psi$ the orthogonal projection onto $\operatorname{span}\{\psi\} \subseteq \mathcal{H}$. Then*

$$\min\{E_1(H) + \langle \psi, V^{-1}\psi \rangle^{-1}, E_2(H)\} \le E_1(H+V).$$

Proof. Since $P\mathcal{H}$ is one dimensional,

$$PV^{-1}P^* = \langle P\psi, PV^{-1}P^*P\psi \rangle = \langle P^*P\psi, V^{-1}P^*P\psi \rangle = \langle \psi, V^{-1}\psi \rangle,$$

where the scalar on the right hand side is interpreted as multiplication operator on $P\mathcal{H}$. We use the min-max-principle to show

$$E_1(H + P^*(PV^{-1}P^*)^{-1}P) \ge \min\{E_1(H) + \langle \psi, V^{-1}\psi \rangle^{-1}, E_2(H)\}.$$

Then Corollary 6 follows from Lemma 5.

In order to apply the min-max-principle, we decompose the arbitrary vector $\varphi \in \mathcal{H}$ of unit length $\|\varphi\| = 1$ into $\varphi = \alpha\psi + \psi^\perp$ with $\psi^\perp$ orthogonal to ψ. Self-adjointness of H gives us $\langle \psi, H\psi^\perp \rangle = 0$. In addition we know $\langle \psi^\perp, H\psi^\perp \rangle \ge E_2(H)\|\psi^\perp\|^2 = E_2(H)(1 - |\alpha|^2)$. We now see

$$\begin{aligned} &\langle \varphi, (H + P^*(PV^{-1}P^*)^{-1}P)\varphi \rangle \\ &\quad = |\alpha|^2\langle \psi, H\psi \rangle + |\alpha|^2\langle \psi, V^{-1}\psi \rangle^{-1} + \langle \psi^\perp, H\psi^\perp \rangle \\ &\quad \ge |\alpha|^2\big(E_1(H) + \langle \psi, V^{-1}\psi \rangle^{-1}\big) + (1 - |\alpha|^2)E_2(H). \end{aligned}$$

The last expression is affine linear in $|\alpha|^2 \in [0,1]$, so the minimum is realised for $|\alpha|^2 \in \{0,1\}$. □

2.2. **Proof for the breather model.**

Proof of Theorem 1. We adapt the notation of the introduction to the present situation. For this, we shift the fundamental domain $\mathcal{D}_c := [-\frac{1}{2}, \frac{1}{2})^d$ of $\mathbb{Z}^d$. The advantage is that $\operatorname{supp} u_\lambda \subseteq \mathcal{D}_c$ for all $\lambda \in [0,1]$. We define a relevant index set $I_L := [-L, L)^d \cap \mathbb{Z}^d$ for $L \in \mathbb{N}$ and shift the box $\Lambda_L := I_L + \mathcal{D}_c$, too. With $|\cdot|$ for Lebesgue measure and $\#$ for cardinality of sets we have in particular $\# I_L \cdot |\mathcal{D}_c| = |\Lambda_L|$. Since we will only use Mezincescu boundary conditions in this proof, we drop the M in the notation for finite volume restrictions of operators: $H^L := H^{L,M} := H^{\Lambda_L,M}$.

By adding a constant to the periodic potential V_{per}, we can without loss of generality assume

$$E_0 = 0. \tag{2.7}$$

We also need to show (2.4) only for points E of continuity of N, since N is monotone and the right hand side is continuous in E.

We single out the following properties of the single site potential:

$$u_{\lambda_k} = \mu\chi_{\operatorname{supp} u_{\lambda_k}} \leq \mu\chi_{\mathcal{D}_c} \quad \text{Lebesgue-a. e., and} \quad \mathbb{E}[|\operatorname{supp} u_{\lambda_k}|] > 0, \tag{2.8}$$

where the first conditions holds $\mathbb{P}$-almost surely. These are the properties of u we will actually use. This will be useful in Theorem 10, where we can recycle large parts from here. Note that we do not require precise information about the level set $\{x \in \mathbb{R}^d \mid u_\lambda(x) = \mu\}$. In particular, the breathing structure is actually irrelevant.

For non–negative single site potentials it is well known that, using (1.3),

$$\begin{aligned} N(E) &\leq \mathbb{E}[N_L(E, H_\bullet)] = \mathbb{E}\big[\chi_{\{E_1(H_\bullet^L) \leq E\}} N_L(E, H_\bullet)\big] \\ &\leq \mathbb{E}\big[\chi_{\{E_1(H_\bullet^L) \leq E\}} N_L(E, H_{\mathrm{per}})\big] \\ &\leq N_L(E, H_{\mathrm{per}}) \mathbb{P}\{\omega \colon E_1(H_\omega^L) \leq E\} \end{aligned}$$

for $L \in \mathbb{N}$ and points $E \in \mathbb{R}$ of continuity of N. Since $N_L\big(E, H_{\mathrm{per}}\big) \sim C_2 E^{d/2}$ as $E \searrow E_0 = 0$ uniformly in $L \in \mathbb{N}$, see [Sto01, p. 4.1.8], it is sufficient to derive an exponential bound on the probability that the first eigenvalue $E_1(H_\omega^L)$ of H_ω^L does not exceed E for a suitably chosen $L = L_E$.

In order to apply Thirring's Corollary 6, we need the potential to be strictly positive. We therefore regularise the potential by letting

$$H_0^L := -\Delta^L + V_{\mathrm{per}} - \gamma_L \quad \text{and} \quad V_\omega := W_\omega + \gamma_L$$

with $\gamma_L := C_1/(2L^2)$ and C_1 from (1.10). This shift by γ_L scales like the gap between the first and the second eigenvalue of $(-\Delta + V_{\mathrm{per}})^L$, cf. (1.10).

Recall from (1.7) the $\mathbb{Z}^d$-periodic function $\Psi \colon \mathbb{R}^d \to (0, \infty)$. As pointed out in Section 1.2, the normalised ground state Ψ_L of H_0^L is given by $\Psi_L = |\Lambda_L|^{-1/2}\chi_{\Lambda_L}\Psi$. Due to the normalisation (2.7), the ground state energy of H_0^L is $E_1(H_0^L) = -\gamma_L$. Furthermore, by (1.10),

$$E_2(H_0^L) \geq \frac{C_1}{L^2} - \gamma_L = \gamma_L \qquad (L \in \mathbb{N}). \tag{2.9}$$

We define for all $k \in \mathbb{Z}^d$, $L \in \mathbb{N}$ the random variables $X_k, S_L \colon \Omega \to [0,1]$,

$$X_k(\omega) := \frac{1}{|\mathcal{D}_c|} \int_{\operatorname{supp} u_{\lambda_k(\omega)}} |\Psi(x)|^2 \,\mathrm{d}x \quad \text{and} \quad S_L := \frac{1}{\# I_L} \sum_{k \in I_L} X_k. \tag{2.10}$$

The values of X_k and S_L are in $[0,1]$ because of $\|\Psi\chi_{\mathcal{D}_c}\|_2 = 1$. The next lemma employs the properties (2.8) of the single site potential. We postpone the proof to page 13.

Lemma 7. *For all $L \in \mathbb{N}$, $L \geq L_0 := \left\lceil \sqrt{\frac{C_1}{2\mu}} \right\rceil$,*

$$\frac{\gamma_L}{2} S_L(\omega) \leq E_1(H_0^L) + \langle \Psi_L, V_\omega^{-1} \Psi_L \rangle^{-1}.$$

From $\Psi(x) \geq \Psi_- > 0$ and (2.8) we have $\mathbb{E}[X_0] > 0$. We define the crucial length $\hat{L}_E$:

$$\hat{L}_E := \left\lfloor \sqrt{C_1/(2E)} \right\rfloor$$

and see, using (2.9), that for all $L \in \mathbb{N}$, $L \leq \hat{L}_E$,

$$E \leq \frac{C_1}{2\hat{L}_E^2} = \gamma_{\hat{L}_E} \leq \gamma_L \leq E_2(H_0^L).$$

On the event $\{\omega \colon E_1(H_\omega^{\hat{L}_E}) \leq E\}$, Thirring's inequality Corollary 6 implies for all $L \in \mathbb{N}$, $L_0 \leq L \leq \hat{L}_E$,

$$\frac{\gamma_L}{2} S_L(\omega) \leq E_1(H_0^L) + \langle \Psi_L, V_\omega^{-1} \Psi_L \rangle^{-1} \leq E_1(H_\omega^L) \leq E < E_2(H_0^L).$$

Let $L_E := \left\lceil \sqrt{C_1 \mathbb{E}[X_0]/(8E)} \right\rceil$. We use $X_0 \leq 1$ to check that for E small enough, $L_0 \leq L_E \leq \hat{L}_E$. Hence, since $\mathbb{E}[X_0] = \mathbb{E}[S_L]$ ist constant in L,

$$\mathbb{P}\{\omega \colon E_1(H_\omega^{L_E}) \leq E\} \leq \mathbb{P}\left\{\tfrac{\gamma_{L_E}}{2} S_{L_E} \leq E\right\} \leq \mathbb{P}\{S_{L_E} \leq \tfrac{1}{2}\mathbb{E}[S_{L_E}]\}.$$

Finally, observe that the random variables X_k, $k \in \mathbb{Z}$, are independent by (2.8). The standard large deviation estimate in Lemma 8 bounds this probability by $\mathrm{e}^{-2^d C_5 L_E^d}$ with some positive constant C_5, since $\# I_{L_E} = (2L_E)^d$. We see, from the definition of L_E,

$$N(E) \leq C_2 E^{d/2} \exp\bigl(-2^d C_5 L_E^d\bigr) \leq C_2 E^{d/2} \exp\bigl(-C_3 E^{-d/2}\bigr)$$

with $C_3 = C_5 \bigl(2\sqrt{C_1 \mathbb{E}[X_0]/8}\bigr)^d$, which is (2.4). ☐

Lemma 8. *Given a sequence X_k, $k \in \mathbb{N}$, of non–negative i. i. d. random variables with $0 < \mathbb{E}[X_1] < \infty$. Let $S_n := \frac{1}{n}\sum_{k=1}^n X_k$. Then there exists $C_5 > 0$ with*

$$\mathbb{P}\{S_n \leq \mathbb{E}[S_n]/2\} \leq \mathrm{e}^{-C_5 n}.$$

Proof. Observe that for all non–negative numbers $t \geq 0$ by independence

$$\begin{aligned}
\mathbb{P}\bigl\{S_n \leq \tfrac{1}{2}\mathbb{E}[S_n]\bigr\} &= \mathbb{E}\bigl[\chi_{\{\exp(nt(\mathbb{E}[S_n]-2S_n))\geq 1\}}\bigr] \\
&\leq \mathbb{E}\bigl[\exp\bigl(nt(\mathbb{E}[S_n]-2S_n)\bigr)\bigr] \\
&= \prod_{k=1}^n \mathbb{E}\bigl[\exp\bigl(t(\mathbb{E}[X_k]-2X_k)\bigr)\bigr].
\end{aligned}$$

The identical distribution of the random variables X_k shows

$$\mathbb{P}\bigl\{S_n \leq \tfrac{1}{2}\mathbb{E}[S_n]\bigr\} \leq \bigl(\mathbb{E}\bigl[\exp\bigl(t(\mathbb{E}[X_1]-2X_1)\bigr)\bigr]\bigr)^n = \exp\bigl(n \log M(t)\bigr),$$

employing the moment generating function

$$M(t) := \mathbb{E}\bigl[\exp\bigl(t(\mathbb{E}[X_0]-2X_0)\bigr)\bigr] \quad (t \in \mathbb{R})$$

of the random variable $\mathbb{E}[X_0] - 2X_0$.

Note, that $M(0) = 1$ and $M'(0) = \mathbb{E}\big[\mathbb{E}[X_0] - 2X_0\big] = -\mathbb{E}[X_0] < 0$. Therefore we find $s > 0$ with $M(s) < 1$, which proves the claim with $C_5 := |\log M(s)|$. □

Proof of Lemma 7. As V_ω does not vanish, V_ω^{-1} is well-defined as a multiplication operator. By construction, we have

$$\begin{aligned}\langle \Psi_L, V_\omega^{-1}\Psi_L\rangle &= \int_{\Lambda_L} \frac{|\Psi_L(x)|^2}{V_\omega(x)}\,\mathrm{d}x = \frac{1}{|\Lambda_L|}\int_{\Lambda_L}\frac{|\Psi(x)|^2}{V_\omega(x)}\,\mathrm{d}x\\ &= \frac{1}{\# I_L}\sum_{k\in I_L}\frac{1}{|\mathcal{D}_c|}\int_{\mathcal{D}_c+k}\frac{|\Psi(x)|^2}{V_\omega(x)}\,\mathrm{d}x.\end{aligned}$$

Using $\|\Psi\chi_{\mathcal{D}_c}\|_2^2 = 1$, we treat the summand, introducing X_k:

$$\begin{aligned}\frac{1}{|\mathcal{D}_c|}\int_{\mathcal{D}_c+k}\frac{|\Psi(x)|^2}{V_\omega(x)}\,\mathrm{d}x &= \frac{1}{|\mathcal{D}_c|}\int_{\mathcal{D}_c}\frac{|\Psi(x)|^2}{u_{k,\omega}(x)+\gamma_L}\,\mathrm{d}x\\ &= \left(\frac{X_k(\omega)}{\mu+\gamma_L} + \frac{1-X_k(\omega)}{\gamma_L}\right) = \frac{1}{\gamma_L}\frac{\mu+\gamma_L-\mu X_k(\omega)}{\mu+\gamma_L}.\end{aligned}$$

The average over $k \in I_L$ now reads

$$\langle \Psi_L, V_\omega^{-1}\Psi_L\rangle = \frac{\mu+\gamma_L-\mu S_L(\omega)}{(\mu+\gamma_L)\gamma_L}.$$

The inequality $L \geq L_0$ implies $\gamma_L \leq \mu$. Hence, using $S_L \geq 0$, too,

$$E_1(H_0^L) + \langle \Psi_L, V_\omega^{-1}\Psi_L\rangle^{-1} = \gamma_L\frac{\mu S_L(\omega)}{\mu+\gamma_L-\mu S_L(\omega)} \geq \frac{\gamma_L}{2}S_L(\omega).$$ □

3. Reduction to the breather model

In the present section we reduce a far more general situation to the setting (2.8). This shows Lifshitz tails for a broad family of random potentials. In the generalization we keep the independence of the single site potentials, but we do not require them to be identically distributed any more. For well–definedness of the random Schrödinger operator and to be able to perform the mentioned reduction we need to impose some uniformity conditions on the single site potential.

We start with some notions capturing a kind of uniformness of the single site potentials.

Definition 9. *Let $\lambda_k\colon \Omega \to \Omega_0$, $k \in \mathbb{Z}^d$, be a family of random variables on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $u\colon \Omega_0 \times \mathbb{R}^d \to [0,\infty)$ measurable.*

- *A measurable function $U\colon \mathbb{R}^d \to \mathbb{R}$ is an* (almost sure) bound *on u, if*

$$\sup_{k\in\mathbb{Z}^d} u_{\lambda_k} \leq U \quad \mathbb{P}\text{-a. s.}$$

- *A bound $U \in L^p(\ell^1)$ on u with $p > \max\{2, d/2\}$ is a* localizer *for u. As usual, $U \in L^p(\ell^1)$ means*

$$\|U\|_{L^p(\ell^1)} := \Big\| \sum_{k\in\mathbb{Z}^d} \big|U(\,\cdot\, - k)\big| \Big\|_{L^p(\mathcal{D}_c)} < \infty,$$

where $\mathcal{D}_c := [-\frac{1}{2}, \frac{1}{2})^d$ is a fundamental domain of the lattice $\mathbb{Z}^d$.
- *The function u is* μ-non–degenerate *with $\mu > 0$, if*

$$\inf_{k\in\mathbb{Z}^d} \mathbb{P}\big\{|\{x \in \mathcal{D}_c \mid u_{\lambda_k}(x) \geq \mu\}| \geq \mu\big\} \geq \mu.$$

We say that u is non–degenerate, if there exists a $\mu > 0$ such that u is μ-non–degenerate.

Note that u is non–degenerate if and only if

$$\inf_{k\in\mathbb{Z}^d} \mathbb{E} \int_{\mathcal{D}_c} \min\{u_{\lambda_k}(x), 1\}\,\mathrm{d}x > 0.$$

Under the additional assumption that u_{λ_k}, $k \in \mathbb{Z}^d$, are identically distributed, this is furthermore equivalent to $u_{\lambda_k} \neq 0$ on $\mathcal{D}_c$ with positive probability. Note also that the single site potential of the breather model (2.1) is μ-non–degenerate for some $\mu > 0$ and has as a localizer $U := \mu\chi_{\mathcal{D}_c}$.

Let $u\colon \Omega_0 \times \mathbb{R}^d \to [0,\infty)$ be a measurable, non–degenerate single site potential with a localizer. Via (1.1) we form the random potential $(W_\omega)_{\omega\in\Omega}$, and by (1.2) we get the operator H_{per} with periodic background potential and the random operator $(H_\omega)_{\omega\in\Omega}$. Both operators are well defined on the domain of $-\Delta$, see [Kir81, p. 19, Satz 1].

Since we dropped the assumption of identical distribution, translational invariance is lost, too, and the IDS may cease to exist. We therefore state the conclusion in the following theorem directly on the probabilities of low eigenvalues for finite volume restrictions of a suited scale. We will again consider only Mezincescu boundary conditions and suppress the superscript ^{M}.

Theorem 10. *Let $u\colon \Omega_0 \times \mathbb{R}^d \to [0,\infty)$ be measurable and non–degenerate with a localizer, and let $\lambda_k\colon \Omega \to \Omega_0$, $k \in \mathbb{Z}^d$, be independent random variables. For (H_ω) as in* (1.1) *and* (1.2)*, there exist $C_7, \delta \geq 0$ and $E' > E_0$ such that for all $E \in (E_0, E']$*

$$\mathbb{P}\big\{\omega \in \Omega : E_1(H_\omega^{L_E}) \leq E\big\} \leq \exp\big(-C_7(E-E_0)^{-\frac{d}{2}}\big), \tag{3.1}$$

where $L_E = \big\lceil \sqrt{\delta/(E-E_0)}\big\rceil$ and $E_0 := \inf\sigma(H_{\mathrm{per}})$.

Remark 11. *In the case of an ergodic operator family $(H_\omega)_{\omega\in\Omega}$, the IDS N exists, and a direct consequence of Theorem 10 is Lifshitz behavior at E_0:*

$$\exists C_2, C_3 > 0, E' > E_0\colon \forall E \in (E_0, E']\colon$$
$$N(E) \leq C_2(E-E_0)^{d/2}\exp(-C_3(E-E_0)^{-\frac{d}{2}}) \leq \exp(-C_3(E-E_0)^{-\frac{d}{2}}).$$

Remark 12. *Of course,* (3.1) *is only interesting if* $E_0 \in \sigma(H_\omega)$ *almost surely. This holds true, if additionally*

$$\forall \varepsilon > 0\colon \inf_{k\in\mathbb{Z}^d} \mathbb{P}\big\{\|u_{\lambda_k}\|_{L^2(\ell^1)} \le \varepsilon\big\} > 0. \tag{3.2}$$

This is a generalization of (2.2), *where we assumed translational invariance.*

To prove $E_0 \in \sigma(H_\omega)$, *proceed as in the proof of Lemma 2. All we have to change is the proof of Lemma 4.*

Proof of (3.2) $\implies$ *Lemma 4.* Let $L \in \mathbb{N}$ and denote the localizer of u by U. We estimate

$$\begin{aligned}
\|W_\omega \tilde\chi_{x,L}\|_2 &\le \|W_\omega \chi_{\Lambda_L+x}\|_2 \le \sum_{j\in I_L(x)} \Big\|\sum_{k\in\mathbb{Z}^d} u_{\lambda_k(\omega)}(\,\cdot\, - k)\chi_{\mathcal{D}_c+j}\Big\|_2 \\
&= \sum_{j\in I_L(x)} \Big\|\sum_{k\in\mathbb{Z}^d} u_{\lambda_k(\omega)}(\,\cdot\, - j - k)\chi_{\mathcal{D}_c}\Big\|_2 = \sum_{j\in I_L(x)} \Big\|\sum_{\ell\in\mathbb{Z}^d} u_{\lambda_{\ell-j}(\omega)}(\,\cdot\, - \ell)\chi_{\mathcal{D}_c}\Big\|_2 \\
&\le \sum_{j\in I_L(x)} \Big(\Big\|\sum_{\ell\in I_R} u_{\lambda_{\ell-j}(\omega)}(\,\cdot\, - \ell)\chi_{\mathcal{D}_c}\Big\|_2 + \Big\|\sum_{\ell\in\mathbb{Z}^d\setminus I_R} u_{\lambda_{\ell-j}(\omega)}(\,\cdot\, - \ell)\chi_{\mathcal{D}_c}\Big\|_2\Big) \\
&\le \sum_{j\in I_{L+R}(x)} \Big\|\sum_{\ell\in I_R} u_{\lambda_j(\omega)}(\,\cdot\, - \ell)\chi_{\mathcal{D}_c}\Big\|_2 + \#\, I_L \cdot \Big\|\sum_{\ell\in\mathbb{Z}^d\setminus I_R} U(\,\cdot\, - \ell)\chi_{\mathcal{D}_c}\Big\|_2 \\
&\le \sum_{j\in I_{R+L}} \|u_{\lambda_j(\omega)}\|_{L^2(\ell^1)} + \#\, I_L \cdot \Big\|\sum_{k\in\mathbb{Z}^d\setminus I_R} U(\,\cdot\, - k)\chi_{\mathcal{D}_c}\Big\|_2.
\end{aligned}$$

Because of $U \in L^p(\ell^1) \subseteq L^2(\ell^1)$, the last term vanishes for $R \to \infty$. We fix $\varepsilon > 0$ and choose $R \in \mathbb{N}$ so large that

$$\Big\|\sum_{\ell\in\mathbb{Z}^d\setminus I_R} U(\,\cdot\, - \ell)\chi_{\mathcal{D}_c}\Big\|_2 \le \frac{\varepsilon\|\Psi\tilde\chi_{x,L}\|_2}{2\Psi_+\,\#\, I_L}.$$

By (3.2), the Borel-Cantelli lemma provides us again with a set $\Omega_{BC} \in \mathcal{A}$ of full probability with the following property. For all $\omega \in \Omega_{BC}$ and $L \in \mathbb{N}$ we find $x = x_{\omega,L} \in \mathbb{Z}^d$ such that for all $j \in I_{L+R}(x)$

$$\|u_{\lambda_j(\omega)}\|_{L^2(\ell^1)} \le \frac{\varepsilon\|\Psi\tilde\chi_{x,L}\|_2}{2\Psi_+\,\#\, I_{L+R}}.$$

Combined this reads:

$$\|W_\omega \Psi_{x,L}\|_2 \le \Psi_+\|W_\omega\tilde\chi_{x,L}\|_2/\|\Psi\tilde\chi_{x,L}\|_2 \le \varepsilon. \qquad \square$$

Example 13 (general breather model). *Let* $\Omega_0 := [0,\infty)$ *and* $u_1 \in L^\infty(\mathbb{R}^d, \Omega_0)$ *be non–vanishing and compactly supported, and define for* $\lambda \ge 0$ *the breather type single site potential*

$$u\colon [0,\infty)\times\mathbb{R}^d \to [0,\infty)\,, \quad u(\lambda,x) := \begin{cases} u_1(x/\lambda) & (\lambda > 0) \\ 0 & (\lambda = 0)\,. \end{cases}$$

Further let λ_k, $k \in \mathbb{Z}^d$, *be independent and identically distributed non–negative random variables, such that* $\mathbb{P}\{\lambda_0 < \varepsilon\} > 0$ *for all* $\varepsilon > 0$ *and* $\mathbb{P}\{\lambda_0 = 0\} < 1$.

Then, the family $(H_\omega)_{\omega\in\Omega}$ is ergodic, and consequently the IDS N is well-defined and shows Lifshitz behavior as in Theorem 1.

Proof of Theorem 10. As in Section 2.2 we use $\mathcal{D}_c := [-\frac{1}{2}, \frac{1}{2}]^d$ for the fundamental domain, $I_L := [-L, L)^d \cap \mathbb{Z}^d$ for a relevant index set with $L \in \mathbb{N}$ and $\Lambda_L := I_L + \mathcal{D}_c$ for the closed cube of side length $2L$. The superscript ^{L} denotes the restriction of operators to Λ_L with Mezincescu boundary conditions, cf. Section 1.2.

Without loss of generality we assume that $E_0 = \inf \sigma(H_{\mathrm{per}}) = 0$.

We will reduce the model to a random Schrödinger operator with simplified single site potential with properties described in (2.8). For this purpose we employ for $\mu > 0$, such that u is μ-non–degenerate, the cut–off operator

$$A_\mu \colon L^p(\mathbb{R}^d, [0,\infty)) \to L^p(\mathbb{R}^d, \{0,\mu\}), \quad A_\mu u := \mu \chi_{\{q\in\mathcal{D}_c | u(q) \geq \mu\}}. \tag{3.3}$$

The non–linear operator A_μ is weakly measurable, since it maps measurable functions u on measurable images $\tilde{u} := A_\mu u = \mu\chi_{\mathcal{D}_c} \cdot (\chi_{[\mu,\infty)} \circ u)$. The random Schrödinger operator

$$\tilde{H}_\omega := -\Delta + V_{\mathrm{per}} + \tilde{V}_\omega \quad \text{with} \quad \tilde{V}_\omega := \sum_{k\in\mathbb{Z}^d} \tilde{u}_{\lambda_k(\omega)}(\,\cdot\, - k)$$

has bounded potential and is thus well defined on the domain of $-\Delta$.

Since $\tilde{u} \leq u$ and thereby $\tilde{V}_\omega \leq V_\omega$, we have $\tilde{H}_\omega \leq H_\omega$ for $\mathbb{P}$-almost all $\omega \in \Omega$. And since Mezincescu boundary conditions depend on V_{per} and the periodic background is the same for $(\tilde{H}_\omega)$ and (H_ω), for all the finite volume restrictions

$$\tilde{H}^L_\omega \leq H^L_\omega$$

for $\mathbb{P}$-a. a. $\omega \in \Omega$ and all $L \in \mathbb{N}$. It therefore suffices to show the upper bound of the theorem for $\tilde{H}_\omega$ instead of H_ω. By the non–degeneracy of the single site potential, we henceforth reduce ourselves without loss of generality to the situation (2.8) and skip from now on the ˜ in the notation.

The random variables X_k, $k \in \mathbb{Z}$, defined in (2.10), are now no longer identically distributed. But their expectations still share a positive infimum:

$$\mathbb{E}[X_k] = \mathbb{E}\int_{\operatorname{supp} u_{\lambda_k}} |\Psi(x)|^2 \,\mathrm{d}x \geq \Psi_-^2 \mathbb{E}|\operatorname{supp} u_{\lambda_k}| \geq \Psi_-^2 \mu^2 =: \beta > 0.$$

Of course, $\mathbb{E}[S_L] \geq \beta$, too. We adapt the definition of L_E, substituting β for $\mathbb{E}[X_0]$:

$$L_E := \left\lceil \sqrt{\frac{C_1\beta}{8E}} \right\rceil, \text{ i. e.} \quad \delta := \frac{C_1\beta}{8}.$$

With this change we can inherit a large part of the proof of Theorem 1. In particular, we get for all $E > 0$ small enough

$$\mathbb{P}\{\omega \colon E_1(H^{L_E}_\omega) \leq E\} \leq \mathbb{P}\{S_{L_E} \leq \beta/2\}.$$

Lemma 8 is replaced by the Bernstein inequality Lemma 14, cf. [Sch12b], and gives

$$\begin{aligned}\mathbb{P}\{\omega\colon E_1(H_\omega^{L_E}) \le E\} &\le \mathbb{P}\{S_{L_E} \le \tfrac{\beta}{2}\}\\ &\le \exp\bigl(-\tfrac{\beta^2}{16}\,\#\,I_{L_E}\bigr) \le \exp\bigl(-C_7 E^{-d/2}\bigr),\end{aligned}$$

with $C_7 := \beta^2\delta^{d/2}/16$. □

Lemma 14. *Given independent and non–negative random variables* $X_k\colon \Omega \to [0,\alpha]$, $k \in \mathbb{N}$, *with* $0 < \beta \le \mathbb{E}[X_k] \le 1$. *Let* $S_n := \frac{1}{n}\sum_{k=1}^n X_k$. *Then*

$$\mathbb{P}\bigl\{S_n \le \tfrac{\beta}{2}\bigr\} \le \exp\bigl(-\tfrac{\beta^2}{16\alpha^2}n\bigr).$$

Proof. Let

$$Y_k := \frac{X_k - \mathbb{E}[X_k]}{\alpha} \in [-1,1] \quad (k \in \mathbb{N}).$$

The exponential moments of Y_k are, for $|h| \le 1/2$, bounded by

$$\mathbb{E}[\exp(hY_k)] = \sum_{m=0}^{\infty} \frac{h^m \mathbb{E}[Y_k^m]}{m!} \le 1 + \frac{h^2}{2}\sum_{m=0}^{\infty} |h|^m \le 1 + h^2 \le \mathrm{e}^{h^2}.$$

Therefore we have

$$\mathbb{E}\bigl[\exp\bigl(t(S_n - \mathbb{E}S_n)\bigr)\bigr] = \prod_{k=1}^{n} \mathbb{E}\bigl[\exp\bigl(\tfrac{\alpha t}{n}Y_k\bigr)\bigr] \le \exp\bigl((\alpha t)^2/n\bigr)$$

for all $|t| \le \frac{n}{2\alpha}$. Now we employ Markov's inequality with a parameter $t \in [0, \frac{n}{2\alpha}]$:

$$\begin{aligned}\mathbb{P}\{S_n \le \tfrac{\beta}{2}\} &\le \mathbb{P}\{S_n - \mathbb{E}S_n \le -\tfrac{\beta}{2}\}\\ &= \mathbb{P}\{\exp(-t(S_n - \mathbb{E}S_n)) \ge \exp(\beta t/2)\}\\ &\le \exp(-\beta t/2)\mathbb{E}[\exp(-t(S_n - \mathbb{E}S_n))]\\ &\le \exp((\alpha t)^2/n - \beta t/2).\end{aligned}$$

The minimum is achieved for $t = \frac{\beta}{4\alpha^2}n \le \frac{n}{2\alpha}$ and establishes the Lemma. □

4. Initial Length Scale Estimate

We state and prove an initial length scale estimate. Such estimates serve as base in an induction scheme called multiscale analysis to proof localization, cf. [KSS98; Sto01]. As for other random Schrödinger operators, initial length scale estimates follows from our main result, low probability for low eigenvalues, by a Combes–Thomas estimate, Lemma 15. We therefore state the initial length scale estimate as a Corollary to Theorem 10.

In this section, we have to deal with Dirichlet and Mezincescu boundary conditions, and we denote them explicitly again. We further denote the distance between two sets $A, B \subseteq \mathbb{R}^d$ by

$$\operatorname{dist}(A,B) := \inf\bigl\{|x-y| \bigm| x \in A, y \in B\bigr\}.$$

We will use the following Combes–Thomas estimate, as found e. g. in [Sto01, Theorem 2.4.1 and Remark 2.4.3], with Dirichlet boundary conditions and adapted to our needs:

Lemma 15. *Let $\omega \in \Omega$. Let further p, W_ω and H_ω be as in* (1.1) *and* (1.2), *and let $M_W \geq \sup_{x\in\mathbb{R}^d} \|W_\omega \chi_{[x,x+1]^d}\|_p$. Then there exist $C_8 = C_8(M_W)$ and $C_9 = C_9(M_W)$ such that the conditions*

(i) $\Lambda \subseteq \mathbb{R}^d$ an open cube, $A, B \in \mathcal{B}(\Lambda)$ s. t. $\delta := \operatorname{dist}(A,B) > 0$, and

(ii) $E < E_1(H_\omega^{\Lambda,D})$,

imply the estimate

$$\|\chi_A(E - H_\omega^{\Lambda,D})^{-1}\chi_B\| \leq \frac{C_8}{E_1(H_\omega^{\Lambda,D}) - E} \exp\bigl(-C_9(E_1(H_\omega^{\Lambda,D}) - E)\delta\bigr).$$

With this tool we prove the following corollary to Theorem 10.

Corollary 16 (Initial Length Scale Estimate). *Let $(H_\omega)_{\omega\in\Omega}$ and $E_0 := \inf\sigma(H_{\mathrm{per}})$ be as in Theorem 10, $\ell, \kappa \in \mathbb{N}$, and $L := \ell^\kappa$. Let further $A, B \in \mathcal{B}(\Lambda_L)$ be as in Lemma 15 with distance $\delta := \operatorname{dist}(A,B) > 0$.*

Then there exists $C_{10} > 0$ such that with C_8, C_9 from Lemma 15:

$$\begin{aligned}\mathbb{P}\bigl\{\omega\colon \|\chi_A(E_0 + L^{-2/\kappa} - H_\omega^{L,D})^{-1}\chi_B\| \leq C_8 L^{2/\kappa}\exp\bigl(-C_9\delta/L^{2/\kappa}\bigr)\bigr\}&\\ \geq 1 - 2^d L^{(1-1/\kappa)d}\exp\bigl(-C_{10}L^{d/\kappa}\bigr)&.\end{aligned}$$

Remark 17. *Usually one arranges $\delta \geq L/3$. In this case, the upper bound in the event*

$$C_8 L^{2/\kappa}\exp\bigl(-C_9\delta/L^{2/\kappa}\bigr) \leq C_8 L^{2/\kappa}\exp\bigl(-C_9 L^{1-2/\kappa}/3\bigr)$$

vanishes exponentially, given $\kappa > 2$.

Proof of Corollary 16. Let $E := E_0 + \ell^{-2}$. For all $\omega \in \Omega$ with $E_1(H_\omega^{L,D}) \geq E_0 + 2\ell^{-2}$, we have by Lemma 15

$$\|\chi_A(E_0 + \ell^{-2} - H_\omega^{L,D})^{-1}\chi_B\| \leq C_8\ell^2\exp\bigl(-C_9\delta/\ell^2\bigr).$$

Therefore we estimate

$$\begin{aligned}q := \mathbb{P}\bigl\{\omega\colon \|\chi_A(E_0 + L^{-2/\kappa} - H_\omega^{L,D})^{-1}\chi_B\| > C_8 L^{2/\kappa}\exp\bigl(-C_9\delta/L^{2/\kappa}\bigr)\bigr\}&\\ \leq \mathbb{P}\bigl\{\omega\colon E_1(H_\omega^{L,D}) \leq E_0 + 2\ell^{-2}\bigr\}&\end{aligned}$$

from above. By (1.6), $E_1(H_\omega^{L,M}) \leq E_1(H_\omega^{L,D})$, so

$$q \leq \mathbb{P}\bigl\{\omega\colon E_1(H_\omega^{L,M}) \leq E_0 + 2\ell^{-2}\bigr\}.$$

We now introduce more Mezincescu boundary conditions and lower the eigenvalues further: $H_\omega^{L,M} \geq \bigoplus_{k\in I_{L,\ell}} H_\omega^{\Lambda_\ell + k,M}$, where $I_{L,\ell} := \Lambda_L \cap \ell\mathbb{Z}^d$. Thus:

$$\begin{aligned}q &\leq \mathbb{P}\bigl\{\omega\colon E_1\bigl(\bigoplus_{k\in I_{L,\ell}} H_\omega^{\Lambda_\ell+k,M}\bigr) \leq E_0 + 2\ell^{-2}\bigr\}\\ &\leq \sum_{k\in I_{L,\ell}} \mathbb{P}\{\omega\colon E_1(H_\omega^{\ell,M}) \leq E_0 + 2\ell^{-2}\}.\end{aligned}$$

Now we invoke Theorem 10 and conclude

$$q \le (2\ell^{(\kappa-1)})^d \exp\bigl(-c(2\ell^{-2})^{-d/2}\bigr) = 2^d L^{(1-1/\kappa)d} \exp\bigl(-C_{10} L^{d/\kappa}\bigr). \qquad \square$$

5. The lower bound

For an ergodic operator family $(H_\omega)_{\omega\in\Omega}$ the IDS N exists and in the setting of Theorem 10 shows Lifshitz behavior at E_0. A lower bound on N is more involved. If we strive only for the exponent in the exponent, we have by l'Hôpital's rule

$$\begin{aligned}\limsup_{E\searrow E_0} \frac{\ln\bigl(-\ln(N(E))\bigr)}{\ln(E-E_0)} &\le \lim_{E\searrow E_0} \frac{\ln\bigl(C_3(E-E_0)^{-\frac{d}{2}}\bigr)}{\ln(E-E_0)} \\ &= \lim_{E\searrow E_0} \frac{(E-E_0)(-\frac{d}{2}C_3(E-E_0)^{-\frac{d}{2}-1})}{C_3(E-E_0)^{-\frac{d}{2}}} = -\frac{d}{2}.\end{aligned}$$

Under moderate additional conditions, Theorem 18 shows that the limit of the logarithms actually exists and equals $-d/2$. We use the norm $\|f\|_{\ell^1(L^p)} := \sum_{k\in\mathbb{Z}^d} \|f\|_{p,\mathcal{D}+k}$.

Theorem 18. *Let $(H_\omega)_{\omega\in\Omega}$ be an ergodic random operator with measurable single site potential $u\colon \Omega_0 \times \mathbb{R}^d \to [0,\infty)$ and IDS $N\colon \mathbb{R} \to [0,1]$. We assume that the random variables λ_k are i. i. d., again. The single site potential has a summable decay:*

$$\exists C, \epsilon > 0\colon \forall k \in \mathbb{Z}^d\colon \|u_{\lambda_0}\|_{p,\mathcal{D}+k} \le C\|k\|^{-(d+\epsilon)} \quad \textit{a. s.,} \tag{5.1}$$

and low values of the single site potential are not too improbable:

$$\exists \alpha_0, \eta > 0\colon \forall \alpha \in [0,\alpha_0]\colon \mathbb{P}\bigl\{\|u_{\lambda_0}\|_{\ell^1(L^p)} \le \alpha\bigr\} \ge \alpha^\eta. \tag{5.2}$$

Then we have

$$\liminf_{E\searrow E_0} \frac{\ln(-\ln(N(E)))}{\ln(E-E_0)} \ge -\frac{d}{\min\{2,\epsilon\}}. \tag{5.3}$$

Remark 19. *To understand the decay condition* (5.1) *better, note that it is a quantitative version of an upper bound, see Definition 9. Also, consider its following consequence. There exists $C_{11} \ge 0$ such that for almost all $\omega \in \Omega$ and $R > 0$*

$$\begin{aligned}\|W_\omega\|_{p,\mathcal{D}} = \Bigl\|\sum_{k\in\mathbb{Z}^d} u_{\lambda_k(\omega)}(\,\cdot\, - k)\Bigr\|_{p,\mathcal{D}} &\le \sum_{k\in\mathbb{Z}^d} \|u_{\lambda_k(\omega)}\|_{p,\mathcal{D}+k} \\ &\le \frac{C_{11}}{\epsilon R^\epsilon} + \sum_{k\in I_R} \|u_{\lambda_k(\omega)}\|_{\ell^1(L^p)}.\end{aligned}$$

We can hereby controll the norm of W_ω on $\mathcal{D}$ with the norm of the single site potential in a box of side length $2R$, where R is determined by the allowed error.

Remark 20. *The requirement* (5.2) *is a quantitative version of* (3.2). *Note also, that in the case $\epsilon \ge 2$ the limit exists and is equal to $-d/2$. In specific models (alloy type with long range single site potentials) one can derive upper bounds on the IDS which match the bound* (5.3) *also in the case $\epsilon \in (0,2)$.*

Proof of Theorem 18. The proof follows the line of [KS86]. We denote the restriction of H_ω to Λ_L with Dirichlet boundary condition by $H_\omega^{L,D}$. By (1.4) and Čebyšev's inequality we see

$$
\begin{aligned}
N(E) &\geq L^{-d}\mathbb{E}\big[N_L^D(E,H_\bullet)\big] \\
&\geq L^{-d}\mathbb{P}\big\{\omega\in\Omega\colon N_L^D(E,H_\omega)\geq 1\big\} \qquad (5.4)\\
&= L^{-d}\mathbb{P}\big\{\omega\in\Omega\colon \inf\sigma(H_\omega^{L,D})\leq E\big\}.
\end{aligned}
$$

For all $\varphi\in D(H_\omega^{L,D})\setminus\{0\}$ holds by the min-max-principle for a. a. $\omega\in\Omega$

$$\inf\sigma(H_\omega^{L,D})\leq\frac{\langle\varphi,H_\omega^{L,D}\varphi\rangle}{\|\varphi\|^2}.$$

To continue estimate (5.4), we use a smoothly truncated Version $\varphi=\tilde\chi_L\Psi$ of the periodic solution $\Psi\colon\mathbb{R}^d\to\mathbb{R}$ of $(-\Delta+V_{\mathrm{per}})\Psi=0$ with $\|\Psi\|_{2,\mathcal{D}}=1$, see (1.7). Here, $\tilde\chi_L=\tilde\chi(\,\cdot\,/L)$ is a properly scaled mollifier function, i. e. $\tilde\chi\in C^\infty(\mathbb{R}^d,[0,1])$ such that $\tilde\chi|_{\Lambda_{1/2}}=1$ and $\operatorname{supp}(\tilde\chi)\subseteq\operatorname{interior}\Lambda_1$.

Lemma 21. *There exists a constant $C_{12}>0$ such that for a. a. $\omega\in\Omega$ and all $L\in\mathbb{N}$*

$$\frac{\langle\tilde\chi_L\Psi,H_\omega^{L,D}(\tilde\chi_L\Psi)\rangle}{\|\tilde\chi_L\Psi\|_2^2}-E_0\leq\frac{\langle\tilde\chi_L\Psi,W_\omega\tilde\chi_L\Psi\rangle}{\|\tilde\chi_L\Psi\|_2^2}+\frac{C_{12}}{L^2}.$$

Proof. It suffices to show

$$\frac{\langle\tilde\chi_L\Psi,(H_{\mathrm{per}}^{L,D}-E_0)(\tilde\chi_L\Psi)\rangle}{\|\tilde\chi_L\Psi\|_2^2}\leq\frac{C_{12}}{L^2}.$$

We use $(-\Delta+V_{\mathrm{per}})\Psi=E_0\Psi$, the fact that Ψ is real-valued and $\Psi(\mathbb{R}^d)=[\Psi_-,\Psi_+]\subseteq(0,\infty)$, cf. (1.8):

$$
\begin{aligned}
D &:= \langle\tilde\chi_L\Psi,(-\Delta+V_{\mathrm{per}})(\tilde\chi_L\Psi)\rangle-E_0\|\tilde\chi_L\Psi\|_2^2\\
&= -\langle\tilde\chi_L\Psi,(\Delta\tilde\chi_L)\Psi+2(\nabla\tilde\chi_L)\nabla\Psi\rangle\\
&= -\langle\tilde\chi_L\Psi^2,\Delta\tilde\chi_L\rangle-2\langle\tilde\chi_L\Psi\nabla\Psi,\nabla\tilde\chi_L\rangle.
\end{aligned}
$$

Now partial integration gives

$$
\begin{aligned}
D &= \langle\nabla(\tilde\chi_L\Psi^2),\nabla\tilde\chi_L\rangle-2\langle\tilde\chi_L\Psi\nabla\Psi,\nabla\tilde\chi_L\rangle\\
&= \langle(\nabla\tilde\chi_L)\Psi^2,\nabla\tilde\chi_L\rangle=\|\Psi\nabla\tilde\chi_L\|_2^2.
\end{aligned}
$$

Finally, the rescaling produces the needed factor:

$$D=L^{-2}\|\big((\nabla\tilde\chi)(\,\cdot\,/L)\big)\Psi\|_2^2=L^{d-2}\|\Psi(L\,\cdot\,)\nabla\tilde\chi\|_2^2\leq L^{d-2}\|\nabla\tilde\chi\|_2^2\Psi_+^2.$$

Combined with

$$\|\tilde\chi_L\Psi\|_2^2=L^d\|\Psi(L\,\cdot\,)\tilde\chi\|_2^2\geq L^d\|\tilde\chi\|_2^2\Psi_-^2,$$

we see that $C_{12}:=\big(\frac{\|\nabla\tilde\chi\|_2\Psi_+}{\|\tilde\chi\|_2\Psi_-}\big)^2$ is a valid choice. □

Remark 22. *Note, that Lemma 21 does neither follow from nor imply Lemma 3, since*
$$\frac{\langle \tilde\chi_L \Psi, (H_{\mathrm{per}}^{L,D} - E_0)(\tilde\chi_L\Psi)\rangle}{\|\tilde\chi_L\Psi\|_2^2} \leq \frac{\|(H_{\mathrm{per}} - E_0)(\tilde\chi_L\Psi)\|_2}{\|\tilde\chi_L\Psi\|_2} \leq \frac{C_4}{L}$$
is not strong enough, while $\|(H_{\mathrm{per}} - E_0)(\tilde\chi_L\Psi)\|_2$ *cannot be controlled by* $\langle \tilde\chi_L \Psi, (H_{\mathrm{per}}^{L,D} - E_0)(\tilde\chi_L\Psi)\rangle$.

Lemma 21 and the choice $L_E := \lceil \sqrt{2C_{12}/(E-E_0)}\rceil$ allow us to continue the estimate (5.4):

$$\begin{aligned} N(E) &\geq L_E^{-d}\mathbb{P}\{\omega\colon \inf\sigma(H_\omega^{L_E,D}) \leq E\} \\ &\geq L_E^{-d}\mathbb{P}\Big\{\omega\colon \frac{\langle \tilde\chi_{L_E}\Psi, W_\omega\tilde\chi_{L_E}\Psi\rangle}{\|\tilde\chi_{L_E}\Psi\|_2^2} + \frac{C_{12}}{L_E^2} \leq E - E_0\Big\} \\ &\geq L_E^{-d}\mathbb{P}\Big\{\omega\colon \frac{\|W_\omega\Psi^2\|_{1,\Lambda_{L_E}}}{\|\tilde\chi_{L_E}\Psi\|_2^2} \leq \frac{E-E_0}{2}\Big\}. \end{aligned} \tag{5.5}$$

Next we break Λ_{L_E} into copies of the fundamental domain $\mathcal{D} := [0,1)^d$, using $p, q \in [1,\infty]$, $\frac{1}{p} + \frac{1}{q} = 1$:

$$\begin{aligned} \frac{\|W_\omega\Psi^2\|_{1,\Lambda_{L_E}}}{\|\tilde\chi_{L_E}\Psi\|_2^2} &\leq \frac{\|\Psi^2\|_{q,\Lambda_{L_E}}\|W_\omega\|_{p,\Lambda_{L_E}}}{L_E^d\Psi_-^2} \\ &= \frac{2^d}{\Psi_-^2}\Big(\frac{1}{(2L_E)^d}\int_{\Lambda_{L_E}}\Psi(x)^{2q}\,\mathrm{d}x\Big)^{\frac{1}{q}}\Big(\frac{1}{(2L_E)^d}\int_{\Lambda_{L_E}} W_\omega(x)^p\,\mathrm{d}x\Big)^{\frac{1}{p}} \\ &= \frac{2^d\|\Psi^2\|_{q,\mathcal{D}}}{\Psi_-^2}\Big(\frac{1}{\# I_{L_E}}\sum_{k\in I_{L_E}}\int_{\mathcal{D}+k} W_\omega(x)^p\,\mathrm{d}x\Big)^{\frac{1}{p}}, \end{aligned}$$

for $\mathbb{P}$-a. a. $\omega \in \Omega$, $L \in \mathbb{N}$. We return to inequality (5.5) and use $C_{13} := \frac{\Psi_-^2}{2^{d+1}\|\Psi^2\|_{q,\mathcal{D}}}$:

$$\begin{aligned} N(E) &\geq L_E^{-d}\mathbb{P}\Big\{\omega\colon \Big(\frac{1}{\# I_{L_E}}\sum_{k\in I_{L_E}}\|W_\omega(x)\|_{\mathcal{D}+k}^p\Big)^{\frac{1}{p}} \leq (E-E_0)C_{13}\Big\} \\ &\geq L_E^{-d}\mathbb{P}\{\omega\colon \forall k\in I_{L_E}\colon \|W_\omega\|_{p,\mathcal{D}+k} \leq (E-E_0)C_{13}\}. \end{aligned} \tag{5.6}$$

The next step is to reduce the condition on W_ω to a condition the single site potentials $u_{\lambda_k(\omega)}$, using the decay estimate from Remark 19. To garantee $\|W_\omega\|_{p,\mathcal{D}+k} \leq (E-E_0)C_{13}$ for all $k \in I_{L_E}$, it suffices to establish

$$\frac{C_{11}}{\epsilon R^\epsilon} \leq \frac{(E-E_0)C_{13}}{2} \quad \text{and} \quad \|u_{\lambda_k(\omega)}\|_{\ell^1(L^p)} \leq \frac{(E-E_0)C_{13}}{2\ \# I_{L_E+R}}$$

for all $k \in I_{L_E+R}$. The first condition is met for

$$R = R_E := \Big\lceil\Big(\frac{2C_{11}}{(E-E_0)C_{13}\epsilon}\Big)^{1/\epsilon}\Big\rceil.$$

To establish the second condition, we use the equivalence relation "$a_E \sim b_E$ as $E \searrow E_0$" defined as $\lim_{E\searrow E_0} a_E/b_E \in (0,\infty)$. Observe that, as $E \searrow E_0$,

$$L_E + R_E \sim (E-E_0)^{-1/2} + (E-E_0)^{-1/\epsilon} \sim (E-E_0)^{-1/\bar\epsilon},$$

where we let $\bar\epsilon := \min\{2,\epsilon\}$. Since $\# I_{L_E+R_E} = 2^d(L_E+R_E)^d$ and $(E-E_0)^{1+(d/\bar\epsilon)} \sim \frac{(E-E_0)C_{13}}{2^d(L_E+R_E)^d}$ as $E \searrow E_0$, the second condition is implied by

$$\|u_{\lambda_k(\omega)}\|_{\ell^1(L^p)} \le C_{14}(E-E_0)^{1+(d/\bar\epsilon)}$$

with a suitable $C_{14} > 0$ and $E - E_0$ small enough.

We continue (5.6), using the independence of λ_k and (5.2):

$$\begin{aligned} N(E) &\ge L_E^{-d}\mathbb{P}\bigl\{\omega\colon \forall k \in I_{L_E}\colon \|W_\omega\|_{p,\mathcal{D}+k} \le (E-E_0)C_{13}\bigr\} \\ &\ge L_E^{-d}\prod\nolimits_{k\in I_{L_E+R_E}} \mathbb{P}\bigl\{\|u_{\lambda_k}\|_{\ell^1(L^p)} \le C_{14}(E-E_0)^{1+(d/\bar\epsilon)}\bigr\} \\ &\ge L_E^{-d}\bigl(C_{14}(E-E_0)^{1+(d/\bar\epsilon)}\bigr)^{\eta\,\# I_{L_E+R_E}}, \end{aligned}$$

By assumption, the estimate for the probability works for $E-E_0 < \alpha_0$. With $C_{15}, C_5 > 0$ such that $\# I_{L_E+R_E} \le C_{15}(E-E_0)^{-d/\bar\epsilon}$ and $L_E^{-d} \ge C_5(E-E_0)^{d/2}$ we see

$$N(E) \ge C_5(E-E_0)^{d/2}\bigl(C_{14}(E-E_0)^{1+\frac{d}{\bar\epsilon}}\bigr)^{\eta C_{15}(E-E_0)^{-d/\bar\epsilon}}.$$

We isolate the topmost exponent by

$$\begin{aligned} &\frac{\ln\bigl(-\ln(N(E))\bigr)}{\ln(E-E_0)} \ge \\ &\quad \frac{\ln\bigl(-\ln(C_5(E-E_0)^{d/2}) - \eta C_{15}(E-E_0)^{d/\bar\epsilon}\ln\bigl(C_{14}(E-E_0)^{1+\frac{d}{\bar\epsilon}}\bigr)\bigr)}{\ln(E-E_0)} \\ &\qquad \xrightarrow{E\searrow E_0} -\frac{d}{\bar\epsilon}, \end{aligned}$$

as l'Hôpital's rule shows. Note $\ln(E-E_0) < 0$ for $E - E_0 < 1$. □

References

[BK11] J. Bourgain and A. Klein. "Bounds on the Density of States for Schrödinger Operators" (2011). arXiv: 1112.1716v1 (cit. on p. 4).

[DIM01] S.-I. Doi, A. Iwatsuka, and T. Mine. "The uniqueness of the integrated density of states for the Schrödinger operators with magnetic field". *Math. Z.* 237 (2001), pp. 335–371 (cit. on p. 3).

[HS04] D. Hundertmark and B. Simon. "A diamagnetic inequality for semigroup differences". *J. reine angew. Math.* 571 (2004), pp. 107–130 (cit. on p. 3).

[Hup+01] T. Hupfer et al. "Existence and uniqueness of the integrated density of states for Schrödinger operators with magnetic fields and unbounded random potentials". *Reviews in Mathematical Physics* 13.1547 (12 Dec. 2001), pp. 1547–1581. DOI: 10.1142/S0129055X01001083 (cit. on p. 3).

[Kir81] W. Kirsch. "Über Spektren Stochastischer Schrödingeroperatoren". German. PhD thesis. Bochum: Ruhr-Universität Bochum, 1981 (cit. on p. 14).

[KM82a] W. Kirsch and F. Martinelli. "On the density of states of Schrödinger operators with a random potential". *J. Phys. A: Math. Gen.* 15 (1982), pp. 2139–2156. URL: http://iopscience.iop.org/0305-4470/15/7/025 (visited on 01/17/2012) (cit. on pp. 2, 3).

[KM82b] W. Kirsch and F. Martinelli. "On the ergodic properties of the spectrum of general random operators". *Journal für reine und angewandte Mathematik* 334 (1982), pp. 141–156. ISSN: 1435-5345. DOI: 10.1515/crll.1982.334.141. URL: http://www.degruyter.com/view/j/crll.1982.issue-334/crll.1982.334.141/crll.1982.334.141.xml (visited on 06/15/2012) (cit. on p. 2).

[KM83] W. Kirsch and F. Martinelli. "Large deviations and Lifshitz singularity of the integrated density of states of random Hamitonians". *Commun. Math. Phys.* 89 (1983), pp. 27–40 (cit. on p. 9).

[KS86] W. Kirsch and B. Simon. "Lifshitz Tails for Periodic Plus Random Potentials". *Journal of Statistical Physics* 42 (1986), pp. 799–808 (cit. on p. 20).

[KS87] W. Kirsch and B. Simon. "Comparison theorems for the gap of Schrödinger operators". *J. Funct. Anal.* 75 (1987), pp. 396–410 (cit. on p. 6).

[KSS98] W. Kirsch, P. Stollmann, and G. Stolz. "Localization for random perturbations of periodic Schrödinger operators". *Random Oper. and Stoch. Equ.* 6.3 (1998), pp. 241–268 (cit. on p. 17).

[KV10] W. Kirsch and I. Veselić. "Lifshitz Tails for a Class of Schrödinger Operators with Random Breather-Type Potential". *Letters in Mathematical Physics* 94 (1 2010), pp. 27–39. DOI: 10.1007/s11005-010-0417-1. arXiv: math-ph/0609014v2 (cit. on pp. 4, 8).

[KW05] W. Kirsch and S. Warzel. "Lifshitz tails caused by anisotropic decay: the emergence of a quantum-classical regime". *Math. Phys. Anal. Geom.* 8 (3 2005), pp. 257–285. DOI: 10.1007/s11040-005-0582-0. arXiv: math-ph/0310033v2 (cit. on p. 4).

[KW06] W. Kirsch and S. Warzel. "Anderson localization and Lifshits tails for random surface potentials". *J. Funct. Anal.* 230 (2006), pp. 222–250. DOI: 10.1016/j.jfa.2005.05.009. arXiv: math-ph/0412079 (cit. on p. 4).

[Mez87] G. A. Mezincescu. "Lifschitz Singularities for Periodic Operators Plus Random Potentials". *Journal of Statistical Physics* 49 (5-6 1987), pp. 1181–1190 (cit. on pp. 4, 5).

[Nak+15] I. Nakić et al. "Scale-free uncertainty principles and Wegner estimates for random breather potentials". *Comptes Rendus Mathematique* 353.10 (2015), pp. 919–923. ISSN: 1631-073X. DOI: 10.1016/j.crma.2015.08.005. arXiv: 1410.5273 [math.AP] (cit. on p. 4).

[Pas80] L. A. Pastur. "Spectral Properties of Disordered Systems in the One-Body Approximation". *Commun. Math. Phys.* 75 (2 1980), pp. 179–196. DOI: 10.1007/BF01222516. Project Euclid: cmp/1103908097 (cit. on p. 2).

[RS78] M. Reed and B. Simon. *Methods of Modern Mathematical Physics*. Vol. 4: *Analysis of Operators*. Academic Press, Inc., 1978. ISBN: 0-12-585004-2 (cit. on pp. 2, 4).

[Sch12a] K. Schmüdgen. *Unbounded Self-adjoint Operators on Hilbert Space*. Vol. 265. Graduate Texts in Mathematics. Springer, 2012. ISBN: 978-94-007-4752-4. DOI: 10.1007/978-94-007-4753-1 (cit. on p. 5).

[Sch12b] F. Schwarzenberger. "Uniform Approximation of the Integrated Density of States for Long-range Percolation Hamiltonians". *J. Stat. Phys.* 146 (6 2012), pp. 1156–1183. arXiv: 1011.4192v1 (cit. on p. 17).

[Sim82] B. Simon. "Schrödinger Semigroups". *Bulletin of the American Mathematical Society* 7.3 (1982), pp. 447–526. Project Euclid: bams/1183549767 (cit. on p. 5).

[Sto01] P. Stollmann. *Caught by disorder. A Course on Bound States in Random Media*. Vol. 20. Progress in Mathematical Physics. Birkhäuser, 2001 (cit. on pp. 11, 17, 18).

[Tem28] G. Temple. "The Theory of Rayleigh's Principle as Applied to Continuous Systems". *Proc. R. Soc. Lond. A* 119 (1928), pp. 276–293. DOI: 10.1098/rspa.1928.0098. URL: http://rspa.royalsocietypublishing.org/content/119/782/276.full.pdf (visited on 07/06/2012) (cit. on p. 8).

[Thi94] W. Thirring. *Lehrbuch der mathematischen Physik*. Vol. 3: *Quantenmechanik von Atomen und Molekülen*. German. 2nd ed. Wien: Springer-Verlag, 1994 (cit. on pp. 9, 10).

[Ves08] I. Veselić. *Existence and Regularity Properties of the Integrated Density of States of Random Schrödinger Operators*. 1st ed. Vol. 1917.

Lecture Notes in Mathematics. Springer-Verlag, 2008. X, 147. ISBN: 978-3-540-72691-3. DOI: 10.1007/978-3-540-72691-3 (cit. on pp. 2, 4).

TU Dortmund, Fakultät für Mathematik, 44227 Dortmund

Preprints ab 2015/03

2020-01 **Christoph Schumacher**
Habilitationsschrift: Concentration inequalities in random Schrödinger operators

2019-06 **Matthias Röger and Ben Schweizer**
Relaxation analysis in a data driven problem with a single outlier

2019-05 **Klaas Hendrik Poelstra, Ben Schweizer and Maik Urban**
The geometric average of curl-free fields in periodic geometries

2019-04 **Agnes Lamacz and Ben Schweizer**
Representation of solutions to wave equations with profile functions

2019-03 **Ben Schweizer**
Existence results for the Helmholtz equation in periodic wave-guides with energy methods

2019-02 **Ben Schweizer and Maik Urban**
On a limiting absorption principle for sesquilinear forms with an application to the Helmholtz equation in a waveguide

2019-01 **Michael Voit and Jeannette H.C. Woerner**
Functional central limit theorems for multivariate Bessel processes in the freezing regime

2018-08 **Ben Schweizer**
Effective Helmholtz problem in a domain with a Neumann sieve perforation

2018-07 **Sergio Andraus and Michael Voit**
Limit theorems for multivariate Bessel processes in the freezing regime

2018-06 **Michael Voit**
Central limit theorems for multivariate Bessel processes in the freezing regime

2018-05 **Elena El Behi-Gornostaeva, Koondanibha Mitra and Ben Schweizer**
Traveling wave solutions for the Richards equation with hysteresis

2018-04 **Mario Ohlberger, Ben Schweizer, Maik Urban and Barbara Verfürth**
Mathematical analysis of transmission properties of electromagnetic meta-materials

2018-03 **Margit Rösler and Michael Voit**
Beta distributions and Sonine integrals for Bessel functions on symmetric cones

2018-02 **Michael Voit**
Continuous Association Schemes and Hypergroups

2018-01 **Merdan Artykov and Michael Voit**
Some central limit theorems for random walks associated with hypergeometric functions of type BC

2017-05 **Ben Schweizer and Florian Theil**
Lattice dynamics on large time scales and dispersive effective equations

2017-04 **Frank Klinker and Christoph Reineke**
A note on the regularity of matrices with uniform polynomial entries

2017-03 **Tomáš Dohnal and Ben Schweizer**
A Bloch wave numerical scheme for scattering problems in periodic wave-guides

2017-02 **Matthias Röger and Ben Schweizer**
Strain gradient visco-plasticity with dislocation densities contributing to the energy

2017-01 **Ben Schweizer and Maik Urban**
Effective Maxwell´s equations in general periodic microstructures

2016-05 **Robert Lipton and Ben Schweizer**
Effective Maxwell´s equations for perfectly conducting split ring resonators

2016-04 **Ben Schweizer**
Resonance meets homogenization - Construction of meta-materials with astonishing properties

2016-03 **Ben Schweizer**
On Friedrichs inequality, Helmholtz decomposition, vector potentials, and the div-curl lemma

2016-02 **Michael Voit**
Generalized commutative association schemes, hypergroups, and positive product formulas

2016-01 **Agnes Lamacz and Ben Schweizer**
Effective acoustic properties of a meta-material consisting of small Helmholtz resonators

2015-13 **Christian Eggert, Ralf Gäer, Frank Klinker**
The general treatment of non-symmetric, non-balanced star circuits: On the geometrization of problems in electrical metrology

2015-12 **Daniel Kobe and Jeannette H.C. Woerner**
Oscillating Ornstein-Uhlenbeck processes and modelling electricity prices

2015-11 **Sven Glaser**
A distributional limit theorem for the realized power variation of linear fractional stable motions

2015-10 **Herold Dehling, Brice Franke and Jeannette H.C. Woerner**
Estimating drift parameters in a fractional Ornstein Uhlenbeck process with periodic mean

2015-09 **Harald Garcke, Johannes Kampmann, Andreas Rätz and Matthias Röger**
A coupled surface-Cahn-Hilliard bulk-diffusion system modeling lipid raft formation in cell membrans

2015-08 **Agnes Lamacz and Ben Schweizer**
Outgoing wave conditions in photonic crystals and transmission properties at interfaces

2015-07 **Manh Hong Duong, Agnes Lamacz, Mark A. Peletier and Upanshu Sharma**
Variational approach to coarse-graining of generalized gradient flows

2015-06 **Agnes Lamacz and Ben Schweizer**
A negative index meta-material for Maxwell´s equations

2015-05 **Michael Voit**
Dispersion and limit theorems for random walks associated with hypergeometric functions of type *BC*

2015-04 **Andreas Rätz**
Diffuse-interface approximations of osmosis free boundary problems

2015-03 **Margit Rösler and Michael Voit**
A multivariate version of the disk convolution