LOCALIZATION FOR DISCRETE ONE DIMENSIONAL RANDOM WORD MODELS

DAVID DAMANIK, ROBERT SIMS, AND GÜNTER STOLZ

Abstract. We consider Schrödinger operators in $\ell^2(\mathbb{Z})$ whose potentials are obtained by randomly concatenating words from an underlying set $\mathcal{W}$ according to some probability measure ν on $\mathcal{W}$. Our assumptions allow us to consider models with local correlations, such as the random dimer model or, more generally, random polymer models. We prove spectral localization and, away from a finite set of exceptional energies, dynamical localization for such models. These results are obtained by employing scattering theoretic methods together with Furstenberg's theorem to verify the necessary input to perform a multiscale analysis.

1. Introduction

We study one-dimensional discrete random Schrödinger operators H_ω, where the potential is constructed by a random concatenation of finite words, that is, vectors in $\mathbb{R}^j$, $1 \le j \le m$. Our main goal is to extend known results on localization (spectral and dynamical) for the Anderson model and the so-called random dimer model to this more general class of random operators. The only requirement will be that at least two words $w_1 \in \mathbb{R}^{j_1}$ and $w_2 \in \mathbb{R}^{j_2}$ used in the construction do not commute in the sense that (w_1, w_2) and (w_2, w_1) are different vectors in $\mathbb{R}^{j_1+j_2}$, which avoids periodicity of the random potential.

Generalizing known results for the Anderson model (see [3] and [16] for the most general case of an arbitrary non-trivial distribution of the coupling constant), it will be shown that H_ω almost surely has pure point spectrum with exponentially decaying eigenfunctions. However, as opposed to the Anderson model, for our models the Lyapunov exponent may vanish on a finite set of exceptional energies. That this is possible was first observed by physicists in the example of the dimer model. It turns out that a proof of dynamical localization requires the exclusion of this set of energies. More precisely, we will exclude a somewhat larger, but still finite, set of energies, at which Furstenberg's theorem is not applicable.

In the case of the random dimer model, where an explicit analysis of the transfer matrices allows for an exact determination of the exceptional set, these results were proven by de Bièvre and Germinet [1]. To get finiteness of the exceptional set in our less explicit situation we employ tools from scattering theory (reflection and transmission coefficients) and a basic fact from inverse spectral theory (that reflection coefficients for non-trivial scattering cannot vanish identically). These methods were developed in our previous work [6] to prove localization for continuum

---

Date: November 20, 2002.

D. D. was supported in part by NSF grant DMS–0227289, G. S. by NSF grant DMS–0070343. R. S. is supported through an NSF Postdoctoral Fellowship.

Anderson-type models with singular distributions of the couplings, particularly for the Bernoulli case. We adapt these methods to the present situation and use them to prove positivity of the Lyapunov exponent away from the exceptional set. One can then follow the methods used in [3]: After deducing Hölder continuity of the Lyapunov exponent and the integrated density of states, one obtains a Wegner bound and an initial length scale estimate. The latter are the ingredients for a multiscale analysis, which yields the localization results.

In [11] it is shown for a subset of our models (random polymers constructed from two words) that a generic type of the appearing critical energies indeed leads to a breakdown of dynamical localization. In these cases, one actually gets superdiffusive transport. The results of [1] and [11] showed that the random dimer model provides an example of almost sure co-existence of spectral localization with dynamical delocalization. Here we generalize [1], thus allowing one to construct many other examples of this type. We note that some of these examples will give rise to non-generic types of exceptional energies, where the dynamical properties are not yet known and might well be different from the behavior obtained in [11].

One may also look at these models from the point of view of "subword complexity" of the potential—especially in the case where the potentials take on only finitely many values. This point of view has been discussed, for example, in [5, 7]. The subword complexity function $p : \mathbb{N} \to \mathbb{N}$ of a given potential V (taking finitely many values) is defined as follows: For every $n \in \mathbb{N}$, $p(n)$ is given by the number of distinct subwords of length n of V, where V is regarded as an infinite word. It is easy to see that for randomly generated models, the complexity function is a non-random quantity, that is, it is the same function for a full measure set of potentials. This complexity measure is popular in many disciplines since it discriminates nicely between periodic potentials, aperiodic potentials with long-range order, and random potentials. For example, a potential has a bounded complexity function p if and only if it is periodic and, on the other extremal end, (Bernoulli-type) Anderson models have maximal word complexity (since, almost surely, every possible word occurs). Heuristically, a reduction of complexity should correspond to a trend from localization to delocalization. Finite length (larger than 1) of the building blocks in random word models introduces some local correlation into the potential and thus reduces complexity compared to the Anderson model. While the trend to delocalization is not apparent on the spectral level, it shows through the result of [11] on the dynamical level.

In Section 2 we define random word models, discuss some special cases, and state our main results. Section 3 provides a formula which relates the Lyapunov exponent of H_ω to the Lyapunov exponent of products of independent unimodular matrices (the word transfer matrices), thus making Furstenberg's theorem applicable to our model. Sections 4 to 6 express the exceptional set of energies in scattering theoretic terms, show its finiteness, and prove positivity of the Lyapunov exponent away from the exceptional set. Here we closely follow arguments from [6]. The combinatorial Lemma 6.1 allows us to apply a fact from inverse spectral theory (which enters through Lemma 5.2: Potentials can be reconstructed from m-functions). In Section 7 we briefly indicate the main steps in the remaining proof of localization, which follows the method from [3] with minimal changes. Due to the varying length of words, the dynamical system underlying our random operators is a generalization of the two-sided shift in infinite product spaces. We include a detailed proof of the ergodicity properties of this dynamical system in an appendix.

Acknowledgements: The authors are grateful for an invitation to the Mittag-Leffler Institute in Fall 2002, where this work was completed. G. S. also acknowledges the hospitality of Université Paris 7 and financial support of CNRS (France). We would like to thank Elliot Lieb, Eric Rains, and Shannon Starr for useful discussions.

2. Models and Results

In this section we will present the models we consider and outline the results we obtain for them. Basically, we will study discrete Schrödinger operators in one dimension whose potentials are obtained by randomly concatenating blocks from an underlying set of words. To do so, we will fix a set of words and a probability measure on this set. This probability space will then lead to a random family of discrete Schrödinger operators which will be the object of study in subsequent sections.

To begin, we define the fundamental set of words from which we will construct our operators. Fix two parameters: $m \in \mathbb{N}$, the maximum word length, and $K \in (0,\infty)$, the maximum component value of any given word. More explicitly, let

$$\mathcal{W} := \bigcup_{j=1}^{m} \mathcal{W}_j,$$

where $\mathcal{W}_j = [-K,K]^j$. For $j = 1,\ldots,m$ let ν_j be finite Borel measures on $\mathcal{W}_j$. Assume that the ν_j are normalized such that $\sum_{j=1}^m \nu_j(\mathcal{W}_j) = 1$. We have then that ν, the direct sum of the ν_j, defined by $\nu(U) = \sum_{j=1}^m \nu_j(U_j)$, for $U = \bigcup_{j=1}^m U_j$ with $U_j \subset \mathcal{W}_j$, is a probability measure on $\mathcal{W}$.

We must further assume a non-triviality condition on the space $(\mathcal{W},\nu)$; essentially, it must contain two elements which "do not commute":

$$\text{(NC)} \quad \left\{ \begin{array}{l} \text{For } i = 0,1, \text{ there exist } w_i \in \mathcal{W}_{j_i}, \text{ both in supp}(\nu), \\ \text{such that the two vectors:} \\ (w_0(1), w_0(2), \ldots, w_0(j_0), w_1(1), w_1(2), \ldots, w_1(j_1)) \text{ and} \\ (w_1(1), w_1(2), \ldots, w_1(j_1), w_0(1), w_0(2), \ldots, w_0(j_0)) \\ \text{are distinct.} \end{array} \right.$$

Here $\operatorname{supp}(\nu)$ is the topological support of ν (where $\mathcal{W}$ carries the direct sum topology of the topologies on $\mathcal{W}_j$).

If $w \in \mathcal{W}$ belongs to $\mathcal{W}_j$, we say that w has length j and write $|w| = j$. Denote the expectation of $|w|$ by

$$\langle L \rangle = \sum_{j=1}^{m} j\nu(\mathcal{W}_j).$$

Moreover, set

$$\Omega_0 = \mathcal{W}^{\mathbb{Z}}, \text{ take } \mathbb{P}_0 = \bigotimes_{\mathbb{Z}} \nu$$

on the σ-algebra generated by the cylinder sets in Ω_0, and

$$\Omega = \bigcup_{j=1}^{m} \Omega_j \subset \Omega_0 \times \{1,\ldots,m\}, \text{ where } \Omega_j = \{\omega \in \Omega_0 : |\omega_0| = j\} \times \{1,2,\ldots,j\}.$$

We also define a probability measure $\mathbb{P}$ on Ω as follows: For any $\mathbb{P}_0$-measurable $A \subset \Omega_0$ such that there is $1 \le j \le m$ with $|\omega_0| = j$ for every $\omega \in A$, we let for $1 \le k \le j$,

$$\mathbb{P}\left(A \times \{k\}\right) = \frac{\mathbb{P}_0(A)}{\langle L \rangle}. \tag{2.1}$$

This determines $\mathbb{P}$ uniquely on the σ-algebra in Ω generated by sets of the type $A \times \{k\}$, see the Appendix for details. Finally, we define shifts $T_0 : \Omega_0 \to \Omega_0$ and $T : \Omega \to \Omega$ by

$$(T_0\omega)_n = \omega_{n+1}$$

and

$$T(\omega, k) = \begin{cases} (\omega, k+1) & \text{if } k < |\omega_0| \\ (T_0\omega, 1) & \text{if } k = |\omega_0|. \end{cases} \tag{2.2}$$

It is well known that (Ω_0, T_0) is ergodic (e.g., [15, p. 49]). It can also be shown that (Ω, T) is ergodic. More precisely, we have the following

Proposition 2.1. *Let $J := \{j : \nu(\mathcal{W}_j) > 0\}$.*
(a) *If J is relatively prime, then (Ω, T) is strongly mixing.*
(b) *If J is not relatively prime, then (Ω, T) is ergodic, but not weakly mixing.*

As this result seems to be of some interest in its own right, and we could not find it in the literature, we include a detailed proof in the appendix. A special case of Proposition 2.1 (see example (v) below) was used in [11] without proof. If all the words have equal length, then the proof is simple and reduces to a discrete version of the suspension procedure described by Kirsch for continuum random operators in [13].

We define a family of discrete Schrödinger operators as follows. For $(\omega, k) \in \Omega$, we consider the operator

$$(H_{(\omega,k)}u)(n) = u(n+1) + u(n-1) + V_{(\omega,k)}(n)u(n)$$

in $\ell^2(\mathbb{Z})$, where the potential $V_{(\omega,k)}$ results from the concatenation of

$$\ldots, \omega_{-1}, \omega_0, \omega_1, \omega_2, \ldots$$

such that the origin, $n = 0$, coincides with the k-th position in ω_0. The operator $H_{(\omega,k)}$ is $\mathbb{Z}$-ergodic, that is, $V_{(\omega,k)}(n)$ is $\mathbb{P}$-measurable for every n, and

$$H_{T(\omega,k)} = U H_{(\omega,k)} U^{-1},$$

where U is the shift on $\ell^2(\mathbb{Z})$.

We note here that the reason for having to use the probability space Ω rather than the more trivial product space Ω_0 is the fact that words have varying length. In essence, sequences in Ω have the additional property that the position of their zeroth component has also been "randomized" relative to the origin. If all words have the same length, that is, $\nu(\mathcal{W}_\ell) = 1$ for some ℓ, then in all our considerations we can directly work with Ω_0, (always choose $V_\omega(0) = \omega_0(1)$ and get $H_{T_0\omega} = U_\ell H_\omega U_\ell^{-1}$, where U_ℓ is the shift by ℓ). In this case (NC) holds whenever $\mathcal{W}$ contains at least two words.

Let us discuss a few examples that can be studied within this framework:

(i) *Standard Anderson model.* If we set $m = 1$, we get $\Omega = \Omega_0$, $\mathbb{P} = \mathbb{P}_0$, and the potentials are just given by independent, identically distributed random variables. This is the one-dimensional Anderson model whose localization properties have been studied in many papers, most generally for arbitrary non-trivial distribution by Carmona et al. [3] and Shubin et al. [16].

(ii) *Generalized Anderson model.* One can generalize this model by taking again a sequence of independent, identically distributed random variables with distribution supported in $A \subset [-U, U]$ and using them as random coupling constants of a fixed single site potential. That is, the potential is of the form

$$V_\omega(n) = \sum_{k\in\mathbb{Z}} q_k(\omega) f(n - k\ell),$$

where the single site potential $f : \mathbb{Z} \to \mathbb{R}$ is supported on $\{0, \dots, \ell - 1\}$. In our notation, this means that $\mathcal{W} = \{\lambda w : \lambda \in A\}$, where $w = (f(0), \dots, f(\ell - 1)) \in \mathbb{R}^\ell$.

(iii) *Discrete displacement model.* Fix integers $0 < m < \ell$ and $f : \mathbb{Z} \to \mathbb{R}$ supported in $\{0, \dots, m - 1\}$. Set

$$V_\omega(n) = \sum_{k\in\mathbb{Z}} f(n - k\ell - d_k(\omega)),$$

with discrete i.i.d. random variables d_k taking values in $\{0, \dots, \ell - m\}$. This corresponds to $\mathcal{W} = \{w_0, \dots, w_{\ell-m}\} \subset \mathbb{R}^\ell$ with $w_0 = (f(0), \dots, f(m - 1), 0, \dots, 0)$, $\dots$, $w_{\ell-m} = (0, \dots, 0, f(0), \dots, f(m - 1))$ and $\nu(\{w_j\}) = \mathbb{P}(d_k = j)$. (NC) holds whenever $f \neq 0$.

(iv) *Random dimer model.* A special case of the class of examples in (ii) is given by the random dimer model, where one sets $\ell = 2$, $f = \chi_{\{0,1\}}$. This model has been studied by de Bièvre and Germinet in [1]. These authors were particularly interested in the Bernoulli case, that is, they considered the case $\mathcal{W} = \{(\lambda, \lambda), (-\lambda, -\lambda)\}$ for $\lambda > 0$. This set clearly satisfies (NC).

(v) *Random polymer model.* Building on [1], Jitomirskaya et al. [11] studied random polymer models, where one considers random concatenations of two words. Thus, $\#(\mathcal{W}) = 2$. Rather than studying localization properties for this model, these authors focused on proving non-trivial lower bounds on transport in the presence of so-called critical energies, that is, energies at which the transfer matrices associated with the two words are both elliptic and commute. This may very well happen even if (NC) holds, for example at the energies $E = \pm\lambda$ in the dimer model if $\lambda < 1$.

Our goal is to prove, in the general context introduced above, spectral localization at all energies and dynamical localization away from a finite set of exceptional energies. This recovers known results for the examples in (i) and (iv) (cf. [1, 3, 16]) and, more importantly, establishes new results for the examples in (ii), (iii) and (v). Namely, assuming the non-triviality condition (NC), we will prove the following pair of theorems:

Theorem 2.2 (Exponential Localization). *For $\mathbb{P}$-almost every $(\omega, k) \in \Omega$, the operator $H_{(\omega,k)}$ has pure point spectrum and all eigenfunctions decay exponentially at $\pm\infty$.*

Theorem 2.3 (Strong Dynamical Localization)**.** *There exists a finite set $M \subset \mathbb{R}$ such that for every compact interval $I \subset \mathbb{R} \setminus M$, and every finitely supported $\phi \in \ell^2(\mathbb{Z})$, and every $p > 0$,*

$$\mathbb{E}\left\{\sup_{t>0} \left\| |X|^p e^{-itH} P_I(H)\phi \right\| \right\} < \infty, \tag{2.3}$$

where P_I is the spectral projection onto I.

These theorems will be proven by adapting the scattering theoretic approach to localization developed in [6] to the discrete setting. More precisely, we will show that, away from a finite set of energies, one can apply Furstenberg's theorem to yield positivity of the Lyapunov exponent and subsequently establish the necessary ingredients to start a multiscale analysis. It is by now well known that the above theorems "follow from" multiscale analysis.

Let us briefly compare the results stated above with those found in [11], for the models in (v). While we have to exclude certain other types of exceptional energies as well, the critical energies studied in [11] are a special case of the energies included in the set M. Thus the results in [11] show that in Theorem 2.3 the restriction to an interval I outside M is generally necessary. Also, by Theorem 2.2 every two-word random polymer model with at least one critical energy in the sense of [11] provides an example of a random Schrödinger operator with almost sure coexistence of exponential localization and superdiffusive transport.

As becomes clear from the methods used in this work, in particular Furstenberg's theorem, the number of exceptional energies decreases if the set of words, that is, the support of the measure ν, increases. One might conjecture that for suitably rich word spaces there are no exceptional energies. The following example shows that non-discreteness or even connectedness of $\operatorname{supp}\nu$ is not sufficient to guarantee this: Choose $\mathcal{W} = \{\lambda w : \lambda \in A\}$ as in example (ii) above with $w = (-1, 0, 1)$. At $E = 0$ this yields the word transfer matrix (see Section 3)

$$M(\lambda w, 0) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

independent of λ. This implies that the Lyapunov exponent vanishes at 0 for any choice of A.

3. The Lyapunov Exponent

In this section we discuss the Lyapunov exponent, which is a quantity that measures the growth of transfer matrix norms. This growth is also related to growth/decay properties of generalized eigenfunctions associated with the operators $H_{(\omega,k)}$. Due to the structure of the underlying word space, it is of some technical advantage to define two families of transfer matrices and two Lyapunov exponents. We shall, however, show that these two quantities are essentially the same.

Let us first work in the abstract setting of random products of unimodular matrices. Given $w = (w(1), w(2), \dots, w(j)) \in \mathcal{W}$ and $z \in \mathbb{C}$, we define

$$M(w, z) = \mathcal{T}(w(j), z) \times \cdots \times \mathcal{T}(w(1), z),$$

where for $a \in \mathbb{R}$,

$$\mathcal{T}(a, z) = \begin{pmatrix} z - a & -1 \\ 1 & 0 \end{pmatrix}.$$

For every $z \in \mathbb{C}$, there is a number $\gamma_0(z) \in [0, \infty)$, called the Lyapunov exponent (of the family of random products of matrices $M(w, z)$, $w \in \mathcal{W}$) such that we have

$$\gamma_0(z) = \lim_{N\to\infty} \frac{1}{N} \ln \|M(\omega_N, z) \times \cdots \times M(\omega_1, z)\| \quad \text{for } \mathbb{P}_0\text{-a.e. } \omega \in \Omega_0. \tag{3.4}$$

Let us now turn to the transfer matrices associated with the operators $H_{(\omega,k)}$. If $(\omega, k) \in \Omega$, we define

$$M_{(\omega,k)}(n, z) = \mathcal{T}(V_{(\omega,k)}(n), z) \times \cdots \times \mathcal{T}(V_{(\omega,k)}(1), z)$$

with the $\mathcal{T}$-matrices from above. Then, for every $z \in \mathbb{C}$, there is a number $\gamma(z) \in [0, \infty)$, called the Lyapunov exponent (associated with the operator family) such that we have

$$\gamma(z) = \lim_{n\to\infty} \frac{1}{n} \ln \|M_{(\omega,k)}(n, z)\| \quad \text{for } \mathbb{P}\text{-a.e. } (\omega, k) \in \Omega. \tag{3.5}$$

The main purpose of this section is to show that γ is a fixed multiple of γ_0:

Proposition 3.1. *We have for every $z \in \mathbb{C}$,*

$$\gamma_0(z) = \langle L\rangle \gamma(z). \tag{3.6}$$

Proof. Let $\tilde{\Omega}_0$ be the full measure set of those $\omega \in \Omega_0$ such that (3.4) holds and also

$$\frac{1}{k} \sum_{i=1}^{k} |\omega_i| \to \langle L\rangle \quad \text{as } k \to \infty.$$

For $\omega \in \tilde{\Omega}_0$, it is easily seen that

$$\lim_{n\to\infty} \frac{1}{n} \ln \|M_{(\omega,1)}(n, z)\| = \frac{\gamma_0(z)}{\langle L\rangle}.$$

Since

$$\mathbb{P}\{(\omega, 1) : \omega \in \tilde{\Omega}_0\} = \frac{1}{\langle L\rangle} > 0,$$

we conclude from (3.5) that (3.6) holds. ☐

4. Floquet Solutions Associated With a Periodic Potential

In this and the following section we use various facts about algebraic functions, in particular that they have only finitely many roots. We refer to [14] for their general theory.

Let $V_{\mathrm{per}} : \mathbb{Z} \to \mathbb{R}$ be p-periodic, that is, $V_{\mathrm{per}}(n + p) = V_{\mathrm{per}}(n)$ for every $n \in \mathbb{Z}$. We start by collecting some facts from Floquet theory for the periodic operator $H_0 := \Delta + V_{\mathrm{per}}$, where $(\Delta u)(n) = u(n + 1) + u(n - 1)$. For any $z \in \mathbb{C}$, let $u_N(\cdot, z)$ and $u_D(\cdot, z)$ denote the solutions of

$$u(n + 1) + u(n - 1) + V_{\mathrm{per}}(n)u(n) = zu(n) \tag{4.7}$$

with $u_N(0) = u_D(1) = 1$ and $u_N(1) = u_D(0) = 0$. The transfer matrix of (4.7) from 1 to p is the matrix

$$g_0(z) = \begin{pmatrix} u_D(p+1,z) & u_N(p+1,z) \\ u_D(p,z) & u_N(p,z) \end{pmatrix} = \mathcal{T}(V_{\mathrm{per}}(p),z) \times \cdots \times \mathcal{T}(V_{\mathrm{per}}(1),z),$$

which is unimodular with polynomial entries in z. The eigenvalues of $g_0(z)$ are the roots of

$$\rho^2 - D(z)\rho + 1 = 0, \tag{4.8}$$

that is,

$$\rho_\pm(z) = \frac{D(z) \pm \sqrt{D(z)^2 - 4}}{2}, \tag{4.9}$$

where $D(z) = \mathrm{Tr}[g_0(z)]$. As roots of (4.8), the functions $\rho_\pm$ are algebraic with singularities at points with $D(z) = \pm 2$.

The spectrum of H_0, $\sigma(H_0)$, consists of a finite number of bands which are given by the set of real energies λ for which $|D(\lambda)| \le 2$. Let (a,b) be a stability interval of H_0, that is, a maximal interval such that $|D(\lambda)| < 2$ for every $\lambda \in (a,b)$. For real $\lambda \in (a,b)$, one has that

$$\rho_\pm(\lambda) = \tfrac{1}{2}\left(D(\lambda) \pm i\sqrt{4 - D(\lambda)^2}\right),$$

$|\rho_\pm(\lambda)| = 1$, and $\rho_-(\lambda) = \overline{\rho_+(\lambda)}$. Let

$$S := \{z \in \mathbb{C} : z = \lambda + i\eta \ \text{ where } \ a < \lambda < b \ \text{ and } \ \eta \in \mathbb{R}\}$$

be the vertical strip in the complex plane containing (a,b). For $z = \lambda + i\eta \in S$, one has that the following are equivalent:

$$\text{(i) } |\rho_\pm(z)| = 1, \ \ \text{(ii) } \eta = 0, \ \ \text{(iii) } D(z) \in (-2,2).$$

The implications (ii) ⇒ (iii) and (iii) ⇒ (i) are clear. To see that (i) ⇒ (ii), assume that (i) is true for some $z = \lambda + i\eta$, where $\eta \neq 0$. As both $\rho_\pm$ have modulus 1, all solutions of (4.7) are bounded. By Weyl's alternative, however, if $\eta \neq 0$, then there exists a solution in ℓ^2 near $+\infty$, the Weyl solution, while all other solutions are unbounded. This is a contradiction.

The arguments above imply that $\rho_\pm$ have analytic continuations to all of S. These continuations remain algebraic and the only possible singularities occur at a and b. In addition, as $\rho_+(z)\rho_-(z) = \det g_0(z) = 1$, we have by continuity of $|\cdot|$ that exactly one of ρ_+ and ρ_- satisfies $|\rho(z)| < 1$ in the upper half-plane. Without loss of generality, let us denote by ρ_+ the eigenvalue for which $|\rho_+(\lambda + i\eta)| < 1$ for all $\eta > 0$ and $\lambda \in (a,b)$. This corresponds to choosing a branch of the square root in (4.9). Then, ρ_- satisfies $|\rho_-(\lambda + i\eta)| > 1$ for all $\eta > 0$ and $\lambda \in (a,b)$. Since we also have $|\rho_\pm(\lambda)| = 1$ for $\lambda \in (a,b)$, it follows from the Schwarz reflection principle (apply a fractional linear transformation) that $|\rho_+(\lambda + i\eta)| > 1$ and $|\rho_-(\lambda + i\eta)| < 1$ for all $\eta < 0$ and $\lambda \in (a,b)$.

For $z \in S$, let $v_\pm(z)$ be the eigenvectors of $g_0(z)$ corresponding to $\rho_\pm(z)$ with the second component normalized to be one, that is,

$$v_\pm(z) = \begin{pmatrix} 1 \\ c_\pm(z) \end{pmatrix}. \tag{4.10}$$

One may easily calculate that

$$c_\pm(z) = \frac{\rho_\pm(z) - u_D(p+1,z)}{u_N(p+1,z)}. \tag{4.11}$$

Here the denominator cannot vanish for $z \in S$: Suppose that $u_N(p+1,z) = 0$. Then z is an eigenvalue of the finite Jacobi matrix with diagonal $(V_{\mathrm{per}}(2), \dots, V_{\mathrm{per}}(p))$ and off-diagonal elements 1. Thus z and the solutions $u_D(\cdot,z)$ and $u_N(\cdot,z)$ are real. It follows from $1 = \det g_0(z) = u_D(p+1,z)u_N(p,z)$ that $\operatorname{Tr} g_0(z) = u_D(p+1,z) + u_D(p+1,z)^{-1} \geq 2$. Therefore z is either an endpoint of a stability interval or in a gap of H_0.

Thus, $v_\pm$ are analytic in S. In particular, as $u_N(p+1,\cdot)^{-1}$ is a rational function with at most a pole at a and b, we have that $c_\pm$, and therefore $v_\pm$ as well, are algebraic functions with, at worst, singularities at a and b. Let $\phi_\pm(\cdot,z)$ be the Floquet solutions of (4.7), that is, the solutions satisfying

$$\begin{pmatrix} \phi_\pm(1,z) \\ \phi_\pm(0,z) \end{pmatrix} = v_\pm(z). \tag{4.12}$$

We first note that $\phi_\pm(\cdot,\lambda+i\eta) \in \ell^2$ near $\pm\infty$ if $\eta > 0$, and $\phi_\pm(\cdot,\lambda+i\eta) \in \ell^2$ near $\mp\infty$ if $\eta < 0$. Thus in this setting, the Floquet solutions are the Weyl solutions. Secondly, for fixed n, $\phi_\pm(n,\cdot)$ are algebraic functions, which are analytic in S with, at most, singularities at a and b arising from the singularities in the initial conditions $v_\pm$. Lastly, $\{\phi_+(\cdot,z), \phi_-(\cdot,z)\}$ are a fundamental system of (4.7) for every $z \in S$, as $\rho_+(z) \neq \rho_-(z)$ on S.

5. Scattering with Respect to a Periodic Background

Let V_{per} be a p-periodic potential. We will insert a local perturbation. That is, given real numbers $W_1, \dots, W_m$, we define a potential V by

$$V(n) = \begin{cases} V_{\mathrm{per}}(n) & \text{if } n \leq 0 \\ W_n & \text{if } 1 \leq n \leq m \\ V_{\mathrm{per}}(n-m) & \text{if } n \geq m+1. \end{cases}$$

Consider the operator $H = \Delta + V$ in $\ell^2(\mathbb{Z})$. Take $z \in S$, where S is as above, and let u_+ be the solution of

$$u(n+1) + u(n-1) + V(n)u(n) = zu(n) \tag{5.13}$$

satisfying $u_+(n) = \phi_+(n)$ for negative n. Since $V(n)$ and $V_{\mathrm{per}}(n)$ coincide for $n \leq 0$, we get that $u_+(n) = \phi_+(n)$ for $n \leq 1$. Moreover, to the right of the perturbation, the solution u_+ can be written as a linear combination of $\phi_+(\cdot - m + p)$ and $\phi_-(\cdot - m + p)$, and this identity holds then for $n \geq m$ since $V(n) = V_{\mathrm{per}}(n-m+p)$ for $n \geq m+1$. In other words, there are numbers $a(z)$ and $b(z)$ such that

$$u_+(n,z) = \begin{cases} \phi_+(n,z) & \text{for } n \leq 1 \\ a(z)\phi_+(n-m+p,z) + b(z)\phi_-(n-m+p,z) & \text{for } n \geq m. \end{cases} \tag{5.14}$$

Since $\phi_\pm$ are linearly independent for $z \in S$, this defines $a(z)$ and $b(z)$ uniquely. The numbers $t(z) = 1/a(z)$ and $r(z) = b(z)/a(z)$ are discrete analogues of the classical transmission and reflection coefficients, at least if $V_{\mathrm{per}} = 0$. In particular, vanishing of b is equivalent to vanishing of the reflection coefficient. Thus b and u_+ take on the role of a (modified) reflection coefficient and Jost solution relative to the periodic background $V_{\rm per}$, respectively.

Since for $\lambda \in (a,b)$, we know $\phi_-(n,\lambda) = \overline{\phi_+(n,\lambda)}$ from (4.10), (4.11), and (4.12), by taking u_- to be the solution of (5.13) with $u_-(n,\lambda) = \phi_-(n,\lambda)$ for $n \leq 1$, we get that for $n \geq m$,

$$u_-(n,\lambda) = \overline{a(\lambda)}\phi_-(n-m+p,\lambda) + \overline{b(\lambda)}\phi_+(n-m+p,\lambda). \tag{5.15}$$

Using constancy of the non-zero Wronskian $\phi_+(n+1)\phi_-(n) - \phi_-(n+1)\phi_+(n)$, we arrive at the familiar relation

$$|a(\lambda)|^2 - |b(\lambda)|^2 = 1, \tag{5.16}$$

for $\lambda \in (a,b)$, corresponding to $|r|^2 + |t|^2 = 1$.

Proposition 5.1. *$a(\cdot)$ and $b(\cdot)$, defined on S as above, are algebraic functions with singularities only possible at the boundaries of stability intervals.*

Proof. Recall that u_+ is the solution of (5.13) with

$$\begin{pmatrix} u_+(1,z) \\ u_+(0,z) \end{pmatrix} = \begin{pmatrix} \phi_+(1,z) \\ \phi_+(0,z) \end{pmatrix} = v_+(z).$$

Thus $(u_+(m+1,z), u_+(m,z))^t$ is algebraic in S with singularities only possible at a and b, that is, the boundaries of the stability interval. As was determined before, the same is true for both $(\phi_\pm(p+1,z), \phi_\pm(p,z))^t$. By the definition of $a(z)$ and $b(z)$, we have

$$\begin{pmatrix} a(z) \\ b(z) \end{pmatrix} = \begin{pmatrix} \phi_+(p+1,z) & \phi_-(p+1,z) \\ \phi_+(p,z) & \phi_-(p,z) \end{pmatrix}^{-1} \begin{pmatrix} u_+(m+1,z) \\ u_+(m,z) \end{pmatrix}, \tag{5.17}$$

and so we are done. ☐

Lemma 5.2. *If $b(\lambda) = 0$ for all $\lambda \in (a,b)$, then $V = V_{\rm per}$.*

Proof. Suppose that $b(\lambda) = 0$ for all $\lambda \in (a,b)$, and hence all $z \in S$ by analyticity. We know then that for every $\lambda \in (a,b)$ and $\eta > 0$, the Jost solution $u_+(n, \lambda + i\eta)$ is $a(\lambda + i\eta)\phi_+(n-m+p, \lambda+i\eta)$ (for all $n \geq m$), that is, u_+ is exponentially decaying in this region of the upper half-plane. Thus u_+ is the Weyl solution for the perturbed equation (5.13). We may therefore calculate the Weyl-Titchmarsh m-function, m_V, for (5.13) on the half-line $[1,\infty)$,

$$\begin{aligned} m_V(\lambda + i\eta) &= \frac{u_+(1,\lambda+i\eta)}{u_+(0,\lambda+i\eta)} \\ &= \frac{\phi_+(1,\lambda+i\eta)}{\phi_+(0,\lambda+i\eta)} \\ &= m_{V_{\rm per}}(\lambda+i\eta), \end{aligned}$$

where the latter is the m-function of (4.7) on $[1,\infty)$ (which is the same as the m-function of (5.13) on $[m+1,\infty)$). As the m-functions are analytic in the entire upper half-plane, we conclude that $m_V(z) = m_{V_{\rm per}}(z)$ for all $z \in \mathbb{C}^+$. Thus, by standard results from inverse spectral theory (see, e.g., [17]), we conclude that $V = V_{\rm per}$ on $[1, \infty)$. Since they coincide on $(-\infty, 0]$ by definition, we get $V = V_{\rm per}$. $\square$

Thus, if we restrict our attention to the case $V \neq V_{\rm per}$, we know that $\{\lambda \in (a, b) : b(\lambda) = 0\}$ is finite.

Now we consider a gap. Take α such that $-\infty \leq \alpha < a < b$ and (α, a) is a maximal, non-trivial gap in the spectrum of H_0 (if $a = \inf \sigma(H_0)$, then $\alpha = -\infty$). We note that there is also a gap (b, ∞) where $b = \sup(\sigma(H_0))$. The analysis of this gap is identical to the case $\alpha = -\infty$ below, excepting that we analytically continue to the right rather than the left. Consider the following split strip:

$$S' := \{z = \lambda + i\eta : \alpha < \lambda < b, \eta \in \mathbb{R}\} \setminus [\mathrm{a}, \mathrm{b}).$$

For $i = 1, 2$, let $\rho_i(z)$ be the branches of (4.8) with $|\rho_1(z)| < 1$ and $|\rho_2(z)| > 1$ for all $z \in S'$, which are well defined since $[a, b)$ is excluded. We first note that $\rho_1 = \rho_+$ on the upper half of S, but $\rho_1 = \rho_-$ on the lower half of S. Secondly, as before, it can be seen that ρ_i are algebraic in S'. They have, at most, singularities at α, a, and b, and therefore they may be continued analytically across (a, b). In particular, ρ_i is the analytic continuation of ρ_j, where $i, j \in \{1, 2\}$ with $i \neq j$.

For $z \in S'$, choose eigenvectors $v_i(z) = (c_i(z), 1)^t$ of $g_0(z)$ corresponding to $\rho_i(z)$, analogously to (4.10). Taking ϕ_i to be the solutions of (4.7) with $(\phi_i(1, z), \phi_i(0, z))^t = v_i(z)$ for $z \in S'$, we see again that $\phi_1(\cdot, z)$ (resp., $\phi_2(\cdot, z)$) is in ℓ^2 near $+\infty$ (resp., $-\infty$), that is, they are the Weyl solutions. Set u_i to be the Jost solutions of (5.13) satisfying

$$u_i(n, z) = \begin{cases} \phi_i(n, z) & n \leq 1 \\ a_i(z)\phi_i(n - m + p, z) + b_i(z)\phi_j(n - m + p, z) & n \geq m \end{cases} \tag{5.18}$$

for $z \in S'$ and the same i, j convention used above. As in (5.17) above, one sees that $a_1(z)$ and $b_1(z)$ are algebraic in S'. In fact, in the upper half S_+ of S, they coincide with $a(z)$ and $b(z)$, since for $z \in S_+$, we have that $v_+(z)$ and $v_1(z)$ coincide. Thus $a_1(z)$ and $b_1(z)$ are analytic continuations of the restrictions of $a(z)$ and $b(z)$ to S_+. Similarly, it is seen that $a_2(z)$ and $b_2(z)$ are analytic continuations of the restrictions of $a(z)$ and $b(z)$ to the lower half S_- of S. Thus, $a_1(z)$, $b_1(z)$, $a_2(z)$, and $b_2(z)$ are algebraic and have, at most, singularities at α, a, and b. In particular, when $V \neq V_{\rm per}$, they cannot vanish identically, and hence the set

$$\{\lambda \in (\alpha, b) : a_1(\lambda)b_1(\lambda)a_2(\lambda)b_2(\lambda) = 0\} \tag{5.19}$$

is finite, even in the case $\alpha = -\infty$.

6. Furstenberg's Theorem and Positivity of the Lyapunov Exponent

To prove positivity of $\gamma_0(\lambda)$ for $\lambda \in \mathbb{R}$, away from a finite set, we will investigate properties of the transfer matrices. We assumed in (NC) that there are two words w_0, w_1 in $\mathrm{supp}(\nu)$ which do not commute, and we will therefore work with the following pair of transfer matrices: the "free" matrix

$$g_0(\lambda) = M(w_0, \lambda),$$

corresponding to the periodic problem (4.7) and the matrix

$$g_1(\lambda) = M(w_1, \lambda)$$

describing the local perturbation, corresponding to the perturbed difference equation (5.13). In correspondence with the previous sections, we take H_0 to be the operator with p-periodic potential $V_{\rm per}$, where $p = |w_0|$, such that $V_{\rm per}(n) = w_0(n)$, $1 \le n \le p$, and H to be the operator generated by the perturbed potential V results from $V_{\rm per}$ by inserting the m numbers $W_1, \dots, W_m$, where $m = |w_1|$ and $W_n = w_1(n)$, $1 \le n \le m$. Of course, this convention is arbitrary and the roles of the two transfer matrices could be interchanged.

Lemma 6.1. *Assume that* (NC) *holds with non-commuting words* w_0, w_1. *Then, we have for the potentials* $V_{\rm per}, V$ *defined above,*

$$V_{\rm per} \neq V. \tag{6.20}$$

Proof. Assume that (6.20) fails. Then, we get that w_0 is a power, that is, there is some v, which itself is not a power, and some $s \ge 2$ in $\mathbb{N}$ such that $w_0 = v^s = vv\dots v$. We have then that

$$w_1 vvv\dots = vvvv\dots \tag{6.21}$$

by considering the restrictions of $V_{\rm per}$ and V to $\mathbb{N}$ as one-sided infinte words. If the length of w_1 is not an integer multiple of the length of v, we can again argue that v must be a power, and hence get a contradiction. Thus, the length of w_1 is an integer multiple of the length of v. By inspection of (6.21), this implies that w_1, too, is a power of v. Thus $w_1 = v^{\hat{s}}$ and hence we get a contradiction to (NC) since $w_0 w_1 = v^{s+\hat{s}} = w_1 w_0$. □

Set $G(\lambda)$ to be the closed subgroup of $\mathrm{SL}(2,\mathbb{R})$ generated by $\{M(w,\lambda) : w \in \mathrm{supp}(\nu)\}$. Let $P(\mathbb{R}^2)$ be the projective space, that is, the set of the directions in $\mathbb{R}^2$ and $\overline{v}$ be the direction of $v \in \mathbb{R}^2 \setminus \{0\}$. Note that $\mathrm{SL}(2,\mathbb{R})$ acts on $P(\mathbb{R}^2)$ by $g\overline{v} = \overline{gv}$. We say that $G \subset \mathrm{SL}(2,\mathbb{R})$ is strongly irreducible if and only if there is no finite G-invariant set in $P(\mathbb{R}^2)$.

It follows from Furstenberg's theorem [2] that $\gamma_0(\lambda) > 0$ if $G(\lambda)$ is non-compact and strongly irreducible.

The main result of this section is the following:

Theorem 6.2. *Assume that* (NC) *holds. Then there exists a finite set* $M \subset \mathbb{R}$, *such that* $G(\lambda)$ *is non-compact and strongly irreducible for all* $\lambda \in \mathbb{R} \setminus M$. *In particular,* $\gamma_0(\lambda) > 0$ *for all* $\lambda \in \mathbb{R} \setminus M$.

Proof. The proof of Theorem 6.2 is analogous to the proof of Theorem 2.3 in [6] so we only sketch the argument briefly.

When non-compact, it is known that the group G is strongly irreducible if and only if for each $\overline{v} \in P(\mathbb{R}^2)$,

$$\#\{g\overline{v} : g \in G\} \ge 3; \tag{6.22}$$

see [2]. Note that both non-compactness of$G(\lambda)$ and (6.22) are properties which are preserved if the set of underlying words is enlarged. Thus one can assume without loss of generality that $G(\lambda)$ is generated by $g_0(\lambda)$ and $g_1(\lambda)$.

By an analysis which is almost identical to the one in [6], non-compactness and (6.22) can be shown away from the roots of b (finitely many in each of the finitely many stability intervals), the roots of D (one per stability interval), the endpoints of stability intervals, and the set in (5.19); and hence away from a finite set.

For λ in a gap, non-compactness follows since $g_0(\lambda)$ has an eigenvalue of modulus larger than 1. The set (5.19) is avoided to verify (6.22).

If λ is in a stability interval, then the crucial link to scattering coefficients is given by the fact that $G(\lambda)$ is conjugate to the group $\tilde{G}(\lambda)$ generated by the two matrices

$$\tilde{g}_0(\lambda) = \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix}, \; s(\lambda) = \begin{pmatrix} \mathrm{Re}[a(\lambda)+b(\lambda)] & \mathrm{Im}[a(\lambda)+b(\lambda)] \\ -\mathrm{Im}[a(\lambda)-b(\lambda)] & \mathrm{Re}[a(\lambda)-b(\lambda)] \end{pmatrix},$$

where $\omega = \omega(\lambda) \in (0,\pi)$ such that $\rho_+(\lambda) = e^{i\omega}$; see [6, Lemma 2.4]. This allows for a very explicit analysis of the group $\tilde{G}(\lambda)$ in terms of $a(\lambda)$, $b(\lambda)$ and ω. From $b(\lambda) \neq 0$ one can conclude non-compactness, and $D(\lambda) \neq 0$ means $\omega \neq \pi/2$, implying (6.22). We refer to [6] for details. □

7. Proof of Localization via Multiscale Analysis

The proofs of Theorems 2.2 and 2.3 involve the technical machinery of multiscale analysis, including the verification of the necessary ingredients which are known to make this whole apparatus work. Still, given the results which were established in the previous sections, we can conclude with a brief sketch of the remaining arguments. This is due to the fact that, from here on, the strategy which was developed by Carmona et al. [3] for the special case of the Anderson model can also be followed for the more general random word models studied here. The main point is that one has to stay away from the finite set of exceptional energies.

First of all, for energy intervals in which positivity of γ can be established through Furstenberg's theorem, that is, where the group $G(\lambda)$ is non-compact and strongly irreducible, an analysis of the λ-dependence can be performed which yields Hölder continuity of γ_0, and thus γ. Details on this, which other than smooth dependence of the transfer matrices on λ uses only general facts from the theory of products of independent $SL(2,\mathbb{R})$-matrices, can be found in [3] and [4]; see also [6], where it is shown that this analysis also applies to continuum models.

Through the well known connection of γ and the integrated density of states (IDS) of ergodic discrete Schrödinger operators provided by the Thouless formula, one then obtains Hölder continuity of the IDS.

Combined with the positivity of γ, the latter is the basis for proving a *Wegner estimate* as well as an *initial length scale estimate*, both away from exceptional energies, suitable for starting the multiscale analysis. The details of this are straightforward extensions of the arguments provided in [3] for the Anderson model.

That multiscale analysis, given Wegner and initial length scale estimates, proves exponential localization as claimed in Theorem 2.2, has been known since the 1980's and is stated in [3] together with all the fundamental references. Improved versions, which show that strong dynamical localization as in Theorem 2.3 is obtained through multiscale analysis, were recently provided in [8] and [10]. While these papers are written for continuum operators, they both note that their results apply to lattice models as well.

APPENDIX

The aim of this appendix is to prove Proposition 2.1. We will start by providing a semi-algebra which generates the $\mathbb{P}$-measurable sets and justifying that $\mathbb{P}$ is uniquely determined by (2.1). We will then check that T, given by (2.2), is measure-preserving and show the asserted ergodicity properties by working on the semi-algebra; see Theorems 1.1 and 1.17 of [18]. The semi-algebra we construct consists of suitable cylinder sets. With these explicit sets we are able to demonstrate that the main idea of the classical proof of ergodicity of shifts, namely that two given cylinder sets become independent if their non-trivial components are shifted into disjoint regions, still holds. Due to the varying word length, the details of this argument become more cumbersome, which is our reason for including them here.

Let $\mathcal{B}_j$ denote the Borel sets in $\mathcal{W}_j$, $j = 1, \ldots, m$, and $\mathcal{B}$ the Borel sets in $\mathcal{W}$ (i.e., the ν-measurable sets). The set of all measurable cylinders of the form

$$\prod_{i=-\infty}^{-(n+1)} \mathcal{W} \times \prod_{i=-n}^{-1} A_i \times \widehat{A}_0 \times \prod_{i=1}^{n} A_i \times \prod_{i=n+1}^{\infty} \mathcal{W}, \tag{A.1}$$

where $n \in \mathbb{N}$, $A_0 \in \mathcal{B}_j$, and $A_j \in \mathcal{B}$, for $1 \leq |i| \leq n$, is a semi-algebra in $\{\omega \in \Omega_0 : |\omega_0| = j\}$; denote it by $\mathcal{S}_j$. Here the $\widehat{}$ symbol denotes the zero position in $\mathcal{W}^{\mathbb{Z}}$. It follows that

$$\mathcal{S} := \bigcup_{j=1}^{m} \{\mathcal{S}_j \times \{1\}, \ldots, \mathcal{S}_j \times \{j\}\} \tag{A.2}$$

is a semi-algebra in $\Omega = \bigcup_{j=1}^{m} \Omega_j$, where $\Omega_j = \{\omega \in \Omega_0 : |\omega_0| = j\} \times \{1, \ldots, j\}$.

We define a function $\mathbb{P} : \mathcal{S} \to \mathbb{R}^+$ by $\mathbb{P}(A \times \{k\}) = \mathbb{P}_0(A)/\langle L \rangle$ whenever $A \in \mathcal{S}_j$, and $1 \leq k \leq j$. $\mathbb{P}$ is countably additive since $\mathbb{P}_0$ is countably additive and

$$\sum_{j=1}^{m} \sum_{k=1}^{j} \mathbb{P}\left(\left(\prod_{i=-\infty}^{-1} \mathcal{W} \times \widehat{\mathcal{W}_j} \times \prod_{i=1}^{\infty} \mathcal{W}\right) \times \{k\}\right) = 1. \tag{A.3}$$

Thus $\mathbb{P}$ can be uniquely extended to a probability measure $\mathbb{P}$ on the σ-algebra $\mathcal{F}$ in Ω generated by $\mathcal{S}$. By construction, we have that $\mathbb{P}(A \times \{k\}) = \mathbb{P}_0(A)/\langle L \rangle$ for every $\mathbb{P}_0$-measurable $A \subset \{\omega \in \Omega_0 : |\omega_0| = j\}$ and $1 \leq k \leq j$. Define $T : \Omega \to \Omega$ by (2.2).

Lemma A.1. *T is a measure-preserving bijection.*

Proof. T is a bijection with

$$T^{-1}(\omega, k) = \begin{cases} (\omega, k-1) & \text{if } k > 1 \\ (T_0^{-1}\omega, |\omega_{-1}|) & \text{if } k = 1. \end{cases} \tag{A.4}$$

To prove that T is measure-preserving it suffices to show that $\mathbb{P}(T^{-1}M) = \mathbb{P}(M)$ for all $M \in \mathcal{S}$. Let $M = A \times \{k\}$, $A \in \mathcal{S}_j$, $1 \leq k \leq j$. Thus $\mathbb{P}(M) = \mathbb{P}_0(A)/\langle L \rangle$.

If $k > 1$, then $T^{-1}M = A \times \{k-1\}$ and $\mathbb{P}(T^{-1}M) = \mathbb{P}(M)$.

If $k = 1$, represent A in the form (A.1) and decompose

$$A_{-1} = \bigcup_{j=1}^{m} C_j, \quad C_j \in \mathcal{B}_j.$$

Then, $T^{-1}(A \times \{1\}) = \bigcup_{j=1}^{m}(\Lambda_j \times \{j\})$, where

$$\Lambda_j = \prod_{i=-\infty}^{-n} \mathcal{W} \times \prod_{i=-n+1}^{-1} A_{i-1} \times \widehat{C}_j \times \prod_{i=1}^{n+1} A_{i-1} \times \prod_{i=n+2}^{\infty} \mathcal{W}.$$

Since $\mathbb{P}_0$ is additive and T_0 is measure-preserving, we get

$$\mathbb{P}(T^{-1}(A \times \{1\})) = \sum_{j=1}^{m} \mathbb{P}(\Lambda_j \times \{j\}) = \sum_{j=1}^{m} \frac{\mathbb{P}_0(\Lambda_j)}{\langle L \rangle} = \frac{\mathbb{P}_0(A)}{\langle L \rangle} = \mathbb{P}(A \times \{1\}),$$

concluding the proof. □

Proof of Proposition 2.1. (a) We assume that J is relatively prime. In order to show that T is strongly mixing, we need to prove that

$$\lim_{\ell \to \infty} \mathbb{P}(T^{-\ell}(A, k_A) \cap (B, k_B)) = \mathbb{P}(A, k_A) \cdot \mathbb{P}(B, k_B) = \frac{\mathbb{P}_0(A) \cdot \mathbb{P}_0(B)}{\langle L \rangle^2}. \tag{A.5}$$

Here $(A, k_A) \subset \Omega$ and $(B, k_B) \subset \Omega$ are arbitrary sets of the form

$$\begin{aligned} A &= \cdots \times \mathcal{W} \times C_{-n} \times \cdots \times \widehat{C}_0 \times \cdots \times C_n \times \mathcal{W} \times \cdots, \\ B &= \cdots \times \mathcal{W} \times D_{-n} \times \cdots \times \widehat{D}_0 \times \cdots \times D_n \times \mathcal{W} \times \cdots, \end{aligned} \tag{A.6}$$

where $n \in \mathbb{N}$, $C_j \in \mathcal{B}_{a_j}$, $D_j \in \mathcal{B}_{b_j}$, $j = -n, \ldots, n$, $1 \le k_A \le a_0$, and $1 \le k_B \le b_0$. The sets (A, k_A) with A as in (A.6) are also a generating semi-algebra $\tilde{\mathcal{S}}$ for $\mathcal{F}$ (take finite disjoint unions to get $\mathcal{S}$). Thus strong mixing follows from (A.5) and [18, Theorem 1.17 (iii)].

In order to calculate the left-hand side of (A.5), we will have to write $T^{-\ell}(A, k_A)$ as a disjoint union of sets (C, k) with C of type (A.1). This is simple for $\ell = \ell_0 + a_{-1} + \cdots + a_{-n} - 1$, where

$$T^{-\ell_0}(A, k_A) = (A_0, 1), \tag{A.7}$$

and

$$A_0 = \cdots \times \mathcal{W} \times \widehat{C}_{-n} \times \cdots \times C_n \times \mathcal{W} \times \cdots. \tag{A.8}$$

For $\ell > \ell_0$, we have $T^{-\ell}(A, k_A) = T^{-(\ell-\ell_0)}(A_0, 1)$. Splitting sufficiently many of the $\mathcal{W}$-factors in (A.8) into their disjoint components $\mathcal{W}_j$, we see that $T^{-(\ell-\ell_0)}(A_0, 1)$ is a disjoint union of sets of the form $(A(j_1, \ldots, j_r), k)$ where

$$A(j_1, \ldots, j_r) := \cdots \times \mathcal{W} \times \widehat{\mathcal{W}_{j_1}} \times \cdots \times \mathcal{W}_{j_r} \times C_{-n} \times \cdots \times C_n \times \mathcal{W} \times \cdots. \tag{A.9}$$

We only need to determine the $(A(j_1, \ldots, j_r), k)$ in $T^{-(\ell-\ell_0)}(A_0, 1)$ with $k = k_B$, since otherwise the intersection with (B, k_B) is empty. Comparing (A.8) and (A.9) and counting the length of the words shows that $(A(j_1, \ldots, j_r), k_B) \subset T^{-(\ell-\ell_0)}(A_0, 1)$ if and only if r and $j_1, \ldots, j_r$ are such that

$$k_B \le j_1 \le m, \quad 1 \le j_2, \ldots, j_r \le m, \quad j_1 + \ldots + j_r = k_B + (\ell - \ell_0) - 1. \tag{A.10}$$

This shows that

$$T^{-\ell}(A, k_A) \cap (B, k_B) = \bigcup (A(j_1, \ldots, j_r) \cap B, k_B), \tag{A.11}$$

where the disjoint union is taken over all $r, j_1, \dots, j_r$ as in (A.10). If ℓ is sufficiently large, then we have for all the sets on the right-hand side of (A.11), that

$$A(j_1,\dots,j_r)\cap B = \cdots \times \mathcal{W} \times D_{-n} \times \cdots \times D_{-1} \times (\widehat{D_0 \cap \mathcal{W}_{j_1}}) \times \cdots$$
$$\cdots \times (D_n \cap \mathcal{W}_{j_{n+1}}) \times \mathcal{W}_{j_{n+2}} \times \cdots \times \mathcal{W}_{j_r} \times C_{-n} \times \cdots \times C_n \times \mathcal{W} \times \cdots .$$

Hence,

$$\mathbb{P}\left(T^{-\ell}(A,k_A)\cap(B,k_B)\right) = \frac{1}{\langle L\rangle^2}\mathbb{P}_0\left(\bigcup (A(j_1,\dots,j_r)\cap B)\right) =$$
$$= \frac{1}{\langle L\rangle^2}\sum \mathbb{P}_0\left(B\cap(\cdots\times\mathcal{W}\times\widehat{\mathcal{W}}_{j_1}\times\cdots\times\mathcal{W}_{j_r}\times\mathcal{W}\times\cdots)\right)\cdot\mathbb{P}_0(A).$$

Taking into account further that

$$B\cap(\cdots\times\mathcal{W}\times\widehat{\mathcal{W}}_{j_1}\times\cdots\times\mathcal{W}_{j_r}\times\mathcal{W}\times\dots)\neq\emptyset$$

only if $(b_0,\dots,b_n)=(j_1,\dots,j_{n+1})$, in which case

$$\mathbb{P}_0\left(B\cap(\cdots\times\mathcal{W}\times\widehat{\mathcal{W}}_{j_1}\times\cdots\times\mathcal{W}_{j_{n+1}}\times\mathcal{W}\times\cdots)\right)=\mathbb{P}_0(B),$$

we conclude

$$\mathbb{P}\left(T^{-\ell}(A,k_A)\cap(B,k_B)\right)=\frac{1}{\langle L\rangle}\sum \nu_{j_{n+2}}\cdot\ldots\cdot\nu_{j_r}\cdot\mathbb{P}_0(A)\mathbb{P}_0(B), \tag{A.12}$$

where the sum runs over $1\le j_{n+2},\dots,j_r\le m, j_{n+2}+\ldots+j_r=\ell-\ell_{A,B}$. Here we have set $\ell_{A,B}:=\ell_0+b_0+\ldots+b_n+1-k_B$ and $\nu_j:=\nu(\mathcal{W}_j)$. Thus, after an index shift, (A.5) becomes equivalent to

$$\lim_{\ell\to\infty}\sum_{\substack{1\le j_1,\dots,j_s\le m,\\ j_1+\ldots+j_s=\ell}}\nu_{j_1}\cdot\ldots\cdot\nu_{j_s}=\frac{1}{\langle L\rangle}=\frac{1}{\sum_{j=1}^m j\cdot\nu_j}. \tag{A.13}$$

Since $\sum_{j=1}^m \nu_j = 1$ and J is relatively prime, this is exactly the result proven in Lemma A.2 (a) below. This finishes the proof of part (a).

(b) If J is not relatively prime, then by Lemma A.2 (b) below, the left-hand side of (A.13) converges to $\langle L\rangle^{-1}$ in Cesàro mean. Our previous arguments then yield that

$$\lim_{d\to\infty}\frac{1}{d}\sum_{\ell=1}^{d}\mathbb{P}\left(T^{-\ell}(A,k_A)\cap(B,k_B)\right)=\mathbb{P}(A,k_A)\cdot\mathbb{P}(B,k_B),$$

for all (A,k_A) and (B,k_B) in $\tilde{\mathcal{S}}$. By [18, Theorem 1.17 (i)] this implies ergodicity of T.

To complete the proof of part (b), it remains to check that in the latter case, T is not weakly mixing. To this end, note that by the non-triviality condition (NC), there exist $j\in\{1,\dots,m\}$ and $C\in\mathcal{B}_j$ such that $0<\nu_j(C)<1$. Choose $A:=\cdots\times\mathcal{W}\times\widehat{C}\times\mathcal{W}\cdots$, that is, $A\times\{1\}\in\tilde{\mathcal{S}}$. If ℓ is not a multiple of the greatest common divisor $D>1$ of J, then by considerations as above one sees that

$$\mathbb{P}\left(T^{-\ell}(A,1)\cap(A,1)\right)=0.$$

Since $\mathbb{P}(A,1)=\nu_j(C)/\langle L\rangle$, this implies that

$$\liminf_{d\to\infty}\frac{1}{d}\sum_{\ell=1}^{d}\left|\mathbb{P}\left(T^{-\ell}(A,1)\cap(A,1)\right)-\mathbb{P}(A,1)^2\right|\ge\frac{1}{2}\left(\frac{\nu_j(C)}{\langle L\rangle}\right)^2>0.$$

Thus T is not weakly mixing; see [18, Theorem 1.17 (ii)]. $\square$

In the above we have used a combinatorial lemma, which we state and prove below. Let $m \in \mathbb{N}$ be fixed. For any $\ell \in \mathbb{N}$, let the set of unordered partitions, or compositions, of ℓ by natural numbers less than or equal to m be denoted by

$$P(\ell, m) := \left\{ (j_1, \ldots, j_s) \in \{1, \ldots, m\}^s : \ s \in \mathbb{N} \text{ and } \sum_{r=1}^{s} j_r = \ell \right\},$$

and set $\mathbf{j_s} := (j_1, \ldots, j_s) \in P(\ell, m)$.

Lemma A.2. *For $j = 1, \ldots, m$, suppose $0 \le \nu_j \le 1$ are given with $\sum_{j=1}^m \nu_j = 1$.*
(a) *If $J := \{j : \nu_j > 0\}$ is relatively prime, then*

$$\lim_{\ell \to \infty} \sum_{\mathbf{j_s} \in P(\ell,m)} \nu_{j_1} \cdot \nu_{j_2} \cdot \ldots \cdot \nu_{j_s} = \frac{1}{\sum_{j=1}^m j \cdot \nu_j}. \tag{A.14}$$

(b) *If J is not relatively prime, then*

$$\lim_{d \to \infty} \frac{1}{d} \sum_{\ell=1}^{d} \sum_{\mathbf{j_s} \in P(\ell,m)} \nu_{j_1} \cdot \nu_{j_2} \cdot \ldots \cdot \nu_{j_s} = \frac{1}{\sum_{j=1}^m j \cdot \nu_j}. \tag{A.15}$$

Proof. For each $\ell \in \mathbb{N}$, define

$$A_\ell := \sum_{\mathbf{j_s} \in P(\ell,m)} \nu_{j_1} \cdot \nu_{j_2} \cdot \ldots \cdot \nu_{j_s}, \tag{A.16}$$

and for any $z \in \mathbb{C}$ with $|z| < 1$, set

$$\tilde{\phi}(z) := \sum_{\ell=1}^{\infty} A_\ell z^\ell. \tag{A.17}$$

Clearly,

$$\begin{aligned}
\tilde{\phi}(z) &= \sum_{\ell=1}^{\infty} \sum_{\mathbf{j_s} \in P(\ell,m)} \nu_{j_1} \cdot \nu_{j_2} \cdot \ldots \cdot \nu_{j_s} z^\ell \\
&= \sum_{s=1}^{\infty} \sum_{j_1=1}^{m} \cdots \sum_{j_s=1}^{m} \nu_{j_1} \cdot \nu_{j_2} \cdot \ldots \cdot \nu_{j_s} z^{\sum_{r=1}^{s} j_r} \\
&= \sum_{s=1}^{\infty} \left(\sum_{j=1}^{m} \nu_j z^j \right)^s \\
&= \frac{\sum_{j=1}^m \nu_j z^j}{1 - \sum_{j=1}^m \nu_j z^j}.
\end{aligned}$$

Take $A_0 := 0$ and consider

$$\phi(z) := (1 - z)\tilde{\phi}(z).$$

It is clear that both $\tilde{\phi}$ and ϕ are analytic in $|z| < 1$.

Observe that the set $J := \{j : \nu_j > 0\}$ is relatively prime if and only if the equation $\sum_{j=1}^m \nu_j z^j = 1$ has exactly one solution ($z = 1$) on the unit circle.

We may conclude that if J is relatively prime, then ϕ is analytic in a neighborhood of $\{z \in \mathbb{C} : |z| \leq 1\}$, and

$$\lim_{\ell \to \infty} A_\ell = \phi(1) = \frac{1}{\sum_{j=1}^m j \cdot \nu_j}.$$

This completes the proof of (a).

If the set J is not relatively prime, then let D be the greatest common divisor of J, that is, suppose each $j \in J$ can be written as $j = Dk_j$, and let $K := \{k_j : j = Dk_j \in J\}$.

Rescale the ν_j as follows: for each $1 \leq j \leq m$, define

$$\tilde{\nu}_j := \begin{cases} 0 & \text{if } j \notin K \\ \nu_{Dj} & \text{if } j \in K. \end{cases}$$

Note that the set $\tilde{J} := \{j : \tilde{\nu}_j > 0\}$ is relatively prime by construction. Taking

$$\tilde{A}_\ell := \sum_{\mathbf{j_s} \in P(\ell,m)} \tilde{\nu}_{j_1} \cdot \tilde{\nu}_{j_2} \cdot \ldots \cdot \tilde{\nu}_{j_s},$$

one sees that

$$A_\ell = \begin{cases} 0 & \text{if } \frac{\ell}{D} \notin \mathbb{N} \\ \tilde{A}_j & \text{if } \frac{\ell}{D} = j \in \mathbb{N}. \end{cases}$$

Thus

$$\lim_{d \to \infty} \frac{1}{d} \sum_{\ell=1}^d A_\ell = \lim_{d \to \infty} \frac{1}{d} \sum_{j \in \mathbb{N} : Dj \leq d} A_{Dj}.$$

Letting $m = d/D$, we have that

$$\lim_{d \to \infty} \frac{1}{d} \sum_{\ell=1}^d A_\ell = \frac{1}{D} \lim_{m \to \infty} \frac{1}{m} \sum_{j=1}^m \tilde{A}_j = \frac{1}{\sum_{j=1}^m j \cdot \nu_j},$$

which completes our proof. □

REFERENCES

[1] S. de Bièvre and F. Germinet, Dynamical localization for the random dimer Schrödinger operator, *J. Stat. Phys.* **98** (2000), 1135–1148

[2] P. Bougerol and J. Lacroix, *Products of Random Matrices with Applications to Schrödinger Operators*, Birkhäuser, Boston–Stuttgart (1985)

[3] R. Carmona, A. Klein, and F. Martinelli, Anderson localization for Bernoulli and other singular potentials, *Commun. Math. Phys.* **108** (1987), 41–66

[4] R. Carmona and J. Lacroix, *Spectral Theory of Random Schrödinger Operators*, Birkhäuser, Basel–Berlin (1990)

[5] D. Damanik, Gordon-type arguments in the spectral theory of one-dimensional quasicrystals, in *Directions in Mathematical Quasicrystals*, Eds. M. Baake, R. V. Moody, CRM Monograph Series **13**, AMS, Providence, RI (2000), 277–305

[6] D. Damanik, R. Sims, and G. Stolz, Localization for one-dimensional, continuum, Bernoulli-Anderson models, *Duke Math. J.* **114** (2002), 59–100

[7] D. Damanik and B. Solomyak, Some high-complexity Hamiltonians with purely singular continuous spectrum, *Ann. Henri Poincaré* **3** (2002), 99–105

[8] D. Damanik and P. Stollmann, Multi-scale analysis implies strong dynamical localization, *Geom. Funct. Anal.* **11** (2001), 11–29

[9] H. von Dreifus and A. Klein, A new proof of localization in the Anderson tight binding model, *Commun. Math. Phys.* **124** (1989), 285–299

[10] F. Germinet and A. Klein, Bootstrap multiscale analysis and localization in random media, *Commun. Math. Phys.* **222** (2001), 415–448

[11] S. Jitomirskaya, H. Schulz-Baldes, and G. Stolz, Delocalization in random polymer models, preprint (2002), available at mp-arc 02-267

[12] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin–Heidelberg (1966)

[13] W. Kirsch, On a class of random Schrödinger operators, *Adv. in Appl. Math.* **6** (1985), 177–187

[14] K. Knopp, *Theory of Functions*, Vol. II, Dover, New York (1947)

[15] K. Petersen, *Ergodic theory*, Cambridge University Press, Cambridge (1983)

[16] C. Shubin, R. Vakilian, and T. Wolff, Some harmonic analysis questions suggested by Anderson-Bernoulli models, *Geom. Funct. Anal.* **8** (1998), 932–964

[17] G. Teschl, *Jacobi Operators and Completely Integrable Nonlinear Lattices*, Mathematical Surveys and Monographs **72**, AMS, Providence, RI (2000)

[18] P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, New York–Berlin (1982)

Department of Mathematics 253-37, California Institute of Technology, Pasadena, CA 91125, USA; Email: damanik@its.caltech.edu

Department of Mathematics, Princeton University, Princeton, NJ 08544, USA; Email: rsims@princeton.edu

Department of Mathematics, University of Alabama at Birmingham, Birmingham, AL 35394-1170, USA; Email: stolz@math.uab.edu