Commun. Math. Phys. 220, 583 – 621 (2001)

Communications in
Mathematical Physics

Anderson Localization for Schrödinger Operators on $\mathbb{Z}$ with Potentials Given by the Skew–Shift

Jean Bourgain[1], **Michael Goldstein**[1,⋆], **Wilhelm Schlag**[2]

[1] Institute for Advanced Study, Olden Lane, Princeton, NJ 08540, USA.
E-mail: bourgain@math.ias.edu; mgold@math.ias.edu
[2] Department of Mathematics, Princeton University, Fine Hall, Princeton NJ 08544, USA.
E-mail: schlag@math.princeton.edu

Received: 19 September 2000 / Accepted: 15 February 2001

Dedicated to Yakov G. Sinai on the occasion of his 65th birthday

Abstract: In this paper we study one-dimensional Schrödinger operators on the lattice with a potential given by the skew shift. We show that Anderson localization takes place for most phases and frequencies and sufficiently large disorders.

1. Introduction

In this paper we study the positivity of the Lyapunov exponent, the regularity of the integrated density of states, and the nature of the spectrum for the Schrödinger operators,

$$H_{\omega,(x,y)}\psi_n = -\psi_{n+1} - \psi_{n-1} + v(T_\omega^n(x, y))\psi_n \ \text{ on } \ell^2(\mathbb{Z}), \tag{1.1}$$

where $T_\omega = (x + y, y + \omega) \pmod 1$ is the skew-shift on the two-dimensional torus $\mathbb{T}^2$. The number ω will be assumed to be Diophantine. The study of families of Schrödinger operators with potentials that are in some sense random has a long and rich history, starting with the famous work by P. Anderson [1]. It is not our intention to review this subject, as some of the history as well as many references can be found in [7]. Furthermore, the methods in this paper have little overlap with the work that has been done on the purely random case. Our approach is motivated by the recent works [3] and [7].

The main results in this paper are as follows. Fix a nonconstant real–analytic function v_0 on $\mathbb{T}^2$ and some small $\varepsilon > 0$. Then there exists a set $\Omega_\varepsilon \subset \mathbb{T}$ with $\text{mes}[\mathbb{T} \setminus \Omega_\varepsilon] < \varepsilon$, and a large constant $\lambda_0(\varepsilon, v_0)$ so that for any $\omega \in \Omega_\varepsilon$ and $\lambda \geq \lambda_0$, the equation (1.1) with $v = \lambda v_0$ has the following properties:

⋆ Permanent address: Dept. of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 1A1

- The Lyapunov exponents of (1.1) are positive for all energies, see Prop. 2.11.
- The integrated density of states is continuous with modulus of continuity
$$h(t) = \exp\Big(-c|\log t|^{\frac{1}{24}-}\Big),$$
see Prop. 2.13.
- The operators (1.1) display Anderson localization, i.e., there exists $\widetilde{\Omega_\varepsilon} \subset \mathbb{T}^2$ with $\mathrm{mes}[\mathbb{T}^2 \setminus \widetilde{\Omega_\varepsilon}] < \varepsilon$ so that for all $(x, y) \in \widetilde{\Omega_\varepsilon}$ the spectrum is pure point and the eigen functions decay exponentially, see Theorem 3.7.

2. A Large Deviation Theorem for the Monodromy Matrices and Positivity of the Lyapunov Exponents for Large Disorder

Consider the Schrödinger operator (1.1), where v is a trigonometric polynomial, say. An important example is $v(x, y) = \cos(2\pi x)$. Any solution of (1.1) is of the form
$$\begin{pmatrix} \psi_{n+1} \\ \psi_n \end{pmatrix} = M_n(x, y; E) \begin{pmatrix} \psi_1 \\ \psi_0 \end{pmatrix},$$
where $M_n(x, y; E) = \prod_{j=n}^{1} A_j(x, y; E)$ with ($T = T_\omega$ for simplicity)
$$A_j(x, y; E) = \begin{bmatrix} v(T^j(x, y)) - E & -1 \\ 1 & 0 \end{bmatrix}. \tag{2.1}$$
The matrix $M_n(x, y; E)$ is called the fundamental, or monodromy matrix of Eq. (1.1). As usual,
$$L_n(E) = \int_{\mathbb{T}^2} \frac{1}{n} \log \|M_n(x, y; E)\| \, dxdy$$
and $L(E) = \lim_{n\to\infty} L_n(E) = \inf_n L_n(E)$ denotes the Lyapunov exponent. Clearly, $L(E) \geq 0$ for all E. Kingman's subadditive ergodic theorem asserts that
$$\frac{1}{n} \log \|M_n(x, y; E)\| \to L(E) \ \text{ for a.e. } (x, y) \in \mathbb{T}^2 \ \text{ as } n \to \infty.$$
A more quantitative version of this convergence statement will be of particular importance in this paper. In fact, the goal of this section is to prove an estimate of the form
$$\sup_E \ \mathrm{mes}\Big[(x, y) \in \mathbb{T}^2 \ \Big| \ \Big|\frac{1}{n} \log \|M_n(x, y; E)\| - L_n(E)\Big| > n^{-\sigma}\Big] \leq C \exp\big(-n^{\sigma}\big) \tag{2.2}$$
for all positive integers n and some constant $\sigma > 0$, see Prop. 2.11 below for a more precise statement. These so-called "large deviation estimates" have been of central importance in some recent papers by the authors, see [3, 7], and [4]. They are a key ingredient in the proof of localization in [3] on the one hand, and are essential for proving regularity of the density of states as well as positivity of the Lyapunov exponent in [7]. The Schrödinger equations considered in [3] and [7] were of the form (1.1) with T given by the *shift rather than the skew-shift*, i.e., $T(x, y) = (x + \omega_1, y + \omega_2) \pmod{\mathbb{Z}^2}$ in the case of two dimensions. We want to emphasize that the methods from these papers do

not directly apply to the skew-shift and a completely new approach was required for the proof of Prop. 2.11 below. To understand the difficulty introduced by the skew-shift, let us briefly review some basic aspects of the techniques underlying the proof of the large deviation estimates in [3] for the case of the shift. Firstly, the map

$$u_n(z_1, z_2) = \frac{1}{n} \log \|M_n(z_1, z_2; E)\| \tag{2.3}$$

extends to a subharmonic function on a complex neighborhood of $\mathbb{T}^2$. Moreover, these subharmonic functions are bounded in that neighborhood uniformly in n. Using the standard Riesz–representation for subharmonic functions one obtains the decay of the Fourier coefficients

$$|\widehat{u_n}(\ell_1, \ell_2)| \leq \frac{C}{|\ell_1| + |\ell_2| + 1} \tag{2.4}$$

with some absolute constant C. The second important idea is to exploit the almost invariance of u_n under the transformation T. In fact, it follows immediately from the definition of M_n as a product that

$$\sup_{(x,y)\in\mathbb{T}^2} \Big| \frac{1}{K} \sum_{k=1}^{K} u_n(T^k(x, y)) - u_n(x, y) \Big| \leq C \frac{K}{n}. \tag{2.5}$$

Fourier expanding the sum in (2.5) leads to a series in which the main contributions are given by the resonances of the shift, i.e., those $\underline{k} \in \mathbb{Z}^2 \setminus \{0\}$ for which

$$\|\underline{k} \cdot \underline{\omega}\| \ll 1.$$

Since $\underline{\omega} = (\omega_1, \omega_2)$ is assumed to be Diophantine, such resonances only occur for a sparse set of frequencies $\underline{k}$ and the decay (2.4) then controls the size of these contributions (in [3] certain technical problems arise due to the non-ℓ^2 decay provided by (2.4), which however do not concern us here).

The difficulty one faces with this method in the case of the skew-shift derives from the failure of uniform boundedness of the subharmonic function (2.3). This is due to the fact that iteration of the skew-shift is given by

$$T^k(x, y) = (x + ky + k(k-1)\omega/2, y + k\omega) \mod \mathbb{Z}^2. \tag{2.6}$$

Complexifying in the variable y therefore produces an imaginary part of size about n in half of the factors of the product M_n, cf. (2.1). Therefore, most factors of M_n will be of size e^n rather than bounded as in the case of the shift. Instead of (2.4) one can only assert that

$$|\widehat{u_n}(\ell_1, \ell_2)| \leq \frac{Cn}{|\ell_1| + |\ell_2| + 1}. \tag{2.7}$$

However, since one typically has a resonance at the site $(0, n)$ the Fourier series argument based on the decay (2.7) does not even provide that $\|u_n - L_n\|_2 \to 0$.

Of course, the argument which we outlined above is rather crude as the structure of M_n only enters through the almost invariance (2.5). The tool that will allow us to exploit the structure of M_n more carefully is the “avalanche principle” from [7]. We now reproduce the statement of this principle from [7], but refer the reader to that paper for the proof.

Proposition 2.1. *Let $A_1, \dots, A_n$ be a sequence of arbitrary unimodular 2×2-matrices. Suppose that*

$$\min_{1 \le j \le n} \|A_j\| \ge \mu \ge n \quad \textit{and} \tag{2.8}$$

$$\max_{1 \le j < n} [\log \|A_{j+1}\| + \log \|A_j\| - \log \|A_{j+1} A_j\|] \le \frac{1}{2} \log \mu. \tag{2.9}$$

Then

$$\left| \log \|A_n \cdot \dots \cdot A_1\| + \sum_{j=2}^{n-1} \log \|A_j\| - \sum_{j=1}^{n-1} \log \|A_{j+1} A_j\| \right| < C \frac{n}{\mu}. \tag{2.10}$$

Proposition 2.1 will allow us to prove (2.2) inductively. More precisely, assume that (2.2) holds for some integers n and $2n$. Consider the monodromy matrix M_N with a choice of N which is basically subexponential in n. Let the matrices A_j be the matrices $M_n \circ T^{jn}$ so that

$$M_N(x, y; E) = \prod_{j=N/n}^{0} A_j(x, y; E).$$

By (2.2) conditions (2.8) and (2.9) will hold for all $(x, y) \in \mathbb{T}^2$ up to a set of measure not exceeding

$$\exp\bigl(-n^{\sigma}\bigr). \tag{2.11}$$

The advantage of passing to the much shorter monodromy matrices M_n instead of M_N lies with the fact that the size of their subharmonic extensions is only n rather than N. This allows one to prove that the averages appearing in (2.10) are close to their respective means up to a set which is subexponentially small in N, cf. Lemma 2.6 below. However, in order to apply the avalanche principle we had to remove a set of size given by (2.11), whereas the goal is to prove (2.2) for N. The key tool to circumvent this difficulty is the following BMO estimate for subharmonic functions, which have the additional property of being the sum of two functions, one of which is small in L^∞ and one that is small in L^1. This mechanism is really the new feature compared to the methods from [3].

2.1. Subharmonic functions with small BMO-norm.

Definition 2.2. *Throughout this paper $e(x) := e^{2\pi i x}$. For any $0 < \rho < 1$,*

$$\mathcal{A}_\rho := \{z \in \mathbb{C} \mid 1 - \rho < |z| < 1 + \rho\}.$$

For a function u defined on $\mathcal{A}_\rho$ we shall write $u(x)$ instead of $u(e(x))$. It will be clear from the context whether we mean $u(z)$ for complex z or $u(x) = u(e(x))$ for real x. For any positive integer d,

$$\mathbb{T}^d := \mathbb{R}^d / \mathbb{Z}^d$$

denotes the d-dimensional torus. $\mathrm{BMO}(\mathbb{T})$ *is the space of functions of bounded mean oscillation on* $\mathbb{T}$*, see [16]. Identifying functions that differ only by an additive constant, the norm on* $\mathrm{BMO}(\mathbb{T})$ *is given by*

$$\|f\|_{\mathrm{BMO}(\mathbb{T})} := \sup_{I\subset\mathbb{T}} \frac{1}{|I|}\int_I |f - \langle f\rangle_I|\,dx,$$

where $\langle f\rangle_I = \frac{1}{|I|}\int_I f(x)\,dx$*. The open unit disk will be denoted by* $\mathbb{D}$*.*

Lemma 2.3. *Suppose u is subharmonic on $\mathcal{A}_\rho$, with $\sup_{\mathcal{A}_\rho}|u| \le N$. Furthermore, assume that $u = u_0 + u_1$, where*

$$\|u_0 - \langle u_0\rangle\|_{L^\infty(\mathbb{T})} \le \varepsilon_0 \quad \textit{and} \quad \|u_1\|_{L^1(\mathbb{T})} \le \varepsilon_1. \tag{2.12}$$

Then for some constant C_ρ depending only on ρ,

$$\|u\|_{\mathrm{BMO}(\mathbb{T})} \le C_\rho\Big(\varepsilon_0 \log\big(N/\varepsilon_1\big) + \sqrt{N\varepsilon_1}\Big). \tag{2.13}$$

Proof. By Riesz's representation theorem, there is a positive measure μ with $\mathrm{supp}(\mu) \subset \mathcal{A}_{\rho/2}$ and a harmonic function h such that for any $z \in \mathcal{A}_{\rho/2}$,

$$u(z) = \int \log|z-\zeta|\,d\mu(\zeta) + h(z), \tag{2.14}$$

where

$$\mu(\mathcal{A}_{\rho/2}) + \|h\|_{L^\infty(\mathcal{A}_{\rho/4})} \le C_\rho\, N. \tag{2.15}$$

We first claim that one may assume

$$\mathrm{supp}(\mu) \subset \overline{\mathbb{D}} \cap \mathcal{A}_{\rho/2}. \tag{2.16}$$

Indeed, define μ^* by

$$\mu^*(E) = \mu(E\cap\overline{\mathbb{D}}) + \mu(E^*),$$

where

$$E^* = \Big\{\overline{z^{-1}} : z \in E\Big\}$$

for any measurable $E \subset \mathbb{C}$. For any $|z| = 1$,

$$\int \log|z-\zeta|\,d\mu(\zeta) - \int_{\overline{\mathbb{D}}} \log|z-\zeta|\,d\mu^*(\zeta) = \int_{\mathbb{C}\setminus\overline{\mathbb{D}}} \log|\zeta|\,d\mu(\zeta).$$

Since the term on the right-hand side is nonnegative and no larger than $C_\rho\, N$, subtracting this constant from u and u_0 changes N by at most a multiplicative constant, whereas both the hypothesis and the conclusion of the lemma remain unchanged. This implies claim (2.16). In particular, since

$$\int_{\mathbb{T}} \log|e(t)-\zeta|\,dt = 0 \text{ for all } |\zeta| \le 1$$

we can assume that

$$\langle u\rangle = \langle h\rangle = 0. \tag{2.17}$$

For $\zeta = r\cdot e(x)$ with $0\le r\le 1$ let

$$P_\zeta(y) = \frac{1-r^2}{1-2r\cos(2\pi(x-y))+r^2}$$

be the usual Poisson kernel. If $|\zeta|=1$, then $P_\zeta = \delta_\zeta$. For any $f\in L^1(\mathbb{T})$ with $\langle f\rangle = 0$, the anti-derivative $D^{-1}f$ is defined as

$$(D^{-1}f)(t) = \int_{t_0}^t f(x)\,dx \ \text{ where } t_0 \text{ is chosen so that } \langle D^{-1}f\rangle = 0 \tag{2.18}$$

for arbitrary $t\in\mathbb{T}$. The existence of t_0 is guaranteed by the mean value theorem. We shall also need (2.18) in case $f=\delta_{\theta_0}$, $\theta_0\in\mathbb{T}$. In that case let $t_0=\theta_0+\frac12 \pmod 1$. Observe that $D^{-1}f$ is unique whereas the choice of t_0 is not necessarily unique.

For any $\zeta = |\zeta|e(y)\in\mathbb{D}$ one has the elementary relation

$$\frac{d}{dx}\log|e(x)-\zeta| = \frac{2\pi|\zeta|\sin(2\pi(x-y))}{1-2|\zeta|\cos(2\pi(x-y))+|\zeta|^2} = Q_\zeta(x) = (\mathcal{H}P_\zeta)(x),$$

where $\mathcal{H}$ denotes the Hilbert transform and Q_ζ is the standard notation for the conjugate function of the Poisson kernel, cf. Katznelson [9]. In particular,

$$\log|e(x)-\zeta| = (D^{-1}\mathcal{H}P_\zeta)(x) = \Big(\mathcal{H}D^{-1}(P_\zeta-1)\Big)(x).$$

Hence (2.14), (2.16), and (2.17) imply that

$$u|_{\mathbb{T}} = \mathcal{H}\Big[D^{-1}\int (P_\zeta(\cdot)-1)\,d\mu(\zeta) + \mathcal{H}^{-1}h\Big]. \tag{2.19}$$

The anti-derivative appearing in (2.19) is a harmonic function on $\mathbb{D}$. In fact, if $z = r\cdot e(t)\in\mathbb{D}$, then

$$\begin{aligned}\Big(D^{-1}(P_r(\cdot)-1)\Big)(t) &= \int_{-\frac12}^t (P_r(x)-1)\,dx = \int_{-\frac12}^t 2\sum_{n=1}^\infty \cos(2\pi nx)r^n\,dx\\ &= \frac{-1}{\pi}\sum_{n=1}^\infty \frac1n \sin(2\pi nt)r^n\\ &= \frac{-1}{\pi}\Im\log(1-z) = -2\mathrm{Arg}(1-z),\end{aligned} \tag{2.20}$$

where Arg denotes the principal branch of the argument, i.e.,

$$\mathrm{Arg}(z) = x \ \text{ if and only if } \ z=|z|e(x) \text{ and } -\frac12\le x<\frac12.$$

In particular,

$$\Big(D^{-1}(P_1(\cdot)-1)\Big)(t) = \begin{cases} -t-\frac12 & \text{if } -\frac12\le t<0\\ \frac12 - t & \text{if } 0<t\le\frac12\end{cases} \tag{2.21}$$

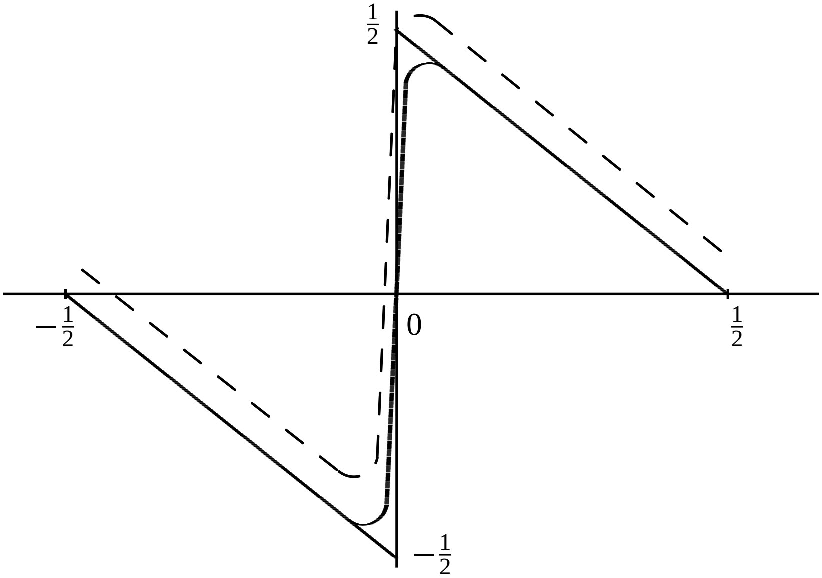

Fig. 1. The right-hand side of (2.24)

and similarly for P_ζ with arbitrary $|\zeta| = 1$. For any $|\zeta| \le 1$ denote

$$h_\zeta = D^{-1}(P_\zeta(\cdot) - 1).$$

The functions are harmonic in the sense of (2.20). Let $\chi \ge 0$ be a C^∞-function on the line with $\mathrm{supp}(\chi) \subset [-1, 1]$ and $\int \chi(x)\,dx = 1$. Let R be a large number to be determined below and set

$$\phi_R(x) = R\chi(Rx). \tag{2.22}$$

Clearly,

$$\sum_k |\widehat{\phi_R}(k)| \le CR. \tag{2.23}$$

We claim that for any $t \in \mathbb{T}$ and any $|\zeta| \le 1$,

$$(h_\zeta * \phi_R)(t - C_0R^{-1}) - C_1R^{-1} \le h_\zeta(t) \le (h_\zeta * \phi_R)(t + C_0R^{-1}) + C_1R^{-1}, \tag{2.24}$$

provided C_0, C_1 are suitable absolute constants. Since all the functions appearing in (2.24) are harmonic, it suffices by the maximum principle to prove the claim for $|\zeta| = 1$. By translation invariance, we may even set $\zeta = 1$. In that case, h_ζ is given by the saw–tooth function (2.21) for which (2.24) is evident, see Fig. 1 (the rounded–off curve lying inside the saw–tooth function represents the convolution of (2.21) with ϕ_R, whereas the dashed line is given by raising that smoothed out function and translating it to the left until it lies above the saw-tooth). Let h be the harmonic function given by (2.14). In view of (2.15) one has

$$\|(\mathcal{H}^{-1}h)'\|_{L^\infty(\mathbb{T})} \le C_\rho \|h\|_{L^\infty(\mathcal{A}_{\rho/4})} \le C_\rho\, N.$$

Therefore,

$$(h * \phi_R)(t - C_0 R^{-1}) - C_2 N R^{-1} \le (\mathcal{H}^{-1} h)(t) \le (h * \phi_R)(t + C_0 R^{-1}) + C_2 N R^{-1} \tag{2.25}$$

with the same constant C_0 as in (2.24), but a different choice of C_2 also depending on ρ. Let $F = [\dots]$ denote the expression in brackets on the right-hand side of (2.19). By construction, $\langle F \rangle = 0$. Integrating (2.24) over the positive measure $d\mu(\zeta)$ with mass controlled by (2.15) and adding (2.25) yields

$$(F * \phi_R)(t - C_0 R^{-1}) - C_\rho N R^{-1} \le F(t) \le (F * \phi_R)(t + C_0 R^{-1}) + C_\rho N R^{-1}, \tag{2.26}$$

for any $t \in \mathbb{T}$. Thus

$$\begin{aligned} \|F\|_\infty &\le \|(F - \mathcal{H}^{-1} u_0) * \phi_R\|_\infty + \|\mathcal{H}^{-1} u_0 * \phi_R\|_\infty + CNR^{-1} \\ &\le \sum_{k \ne 0} \Big| \big(F - \mathcal{H}^{-1} u_0\big)^\wedge(k) \Big| |\widehat{\phi_R}(k)| + \|u_0 * \mathcal{H}\phi_R\|_\infty + CNR^{-1}. \end{aligned} \tag{2.27}$$

Since $F = \mathcal{H}^{-1} u_0 + \mathcal{H}^{-1} u_1$ by (2.19), the sum in (2.27) can be estimated as follows:

$$\begin{aligned} \sum_{k \ne 0} \Big| \big(F - \mathcal{H}^{-1} u_0\big)^\wedge(k) \Big| |\widehat{\phi_R}(k)| &\le \sum_{k \ne 0} \Big| \widehat{\mathcal{H}^{-1} u_1}(k) \Big| |\widehat{\phi_R}(k)| \\ &\le \sum_{k \ne 0} \|u_1\|_1 |\widehat{\phi_R}(k)| \le C\varepsilon_1 R. \end{aligned}$$

Next we claim that $\|\mathcal{H}\phi_R\|_1 \le C \log R$. With $Q(y) = \pi \cot(\pi y)$ being the kernel of $\mathcal{H}$,

$$\begin{aligned} |(\mathcal{H}\phi_R)(y) - Q(y)| &= \Big| \int \big[Q(y - x) - Q(y) \big] \phi_R(x)\, dx \Big| \\ &\le \int \frac{C|x|}{|y|^2} \phi_R(x)\, dx \le \frac{C}{R|y|^2}, \end{aligned}$$

provided $R|y| \ge C$. Thus,

$$\int_{[R|y| > C]} \Big| (\mathcal{H}\phi_R)(y) \Big|\, dy \le C \log R. \tag{2.28}$$

On the other hand,

$$\int_{[R|y| \le C]} \Big| (\mathcal{H}\phi_R)(y) \Big|\, dy \le C \|\mathcal{H}\phi_R\|_2\, R^{-\frac{1}{2}} \le C \|\phi_R\|_2\, R^{-\frac{1}{2}} \le C. \tag{2.29}$$

Thus

$$\|u_0 * \mathcal{H}\phi_R\|_\infty \le \|u_0\|_\infty \|\mathcal{H}\phi_R\|_1 \le C\varepsilon_0 \log R.$$

In view of the preceding

$$\|F\|_\infty \le C\Big(\varepsilon_0 \log R + \varepsilon_1 R + NR^{-1} \Big). \tag{2.30}$$

The lemma follows from (2.19) and (2.30) by taking $R = \sqrt{N/\varepsilon_1}$. □

Remark 2.4. The main application of Lemma 2.13 in this paper will be to estimates on the measure of the set

$$\{x \in \mathbb{T} \mid |u(x) - \langle u\rangle| > \lambda\}.$$

In fact, by the well-known John–Nirenberg inequality [16], the measure of this set does not exceed

$$C \exp\Big(-\frac{c\lambda}{\|u\|_{\mathrm{BMO}}}\Big). \tag{2.31}$$

The exponential integrability of the Hilbert transform of a bounded function can be derived much more easily than by going through BMO and John–Nirenberg. Indeed, it is a classical, and rather simple fact that for any real-valued function f on $\mathbb{T}$ such that $|f| \le 1$, one has the bound

$$\int_{\mathbb{T}} \exp\Big(\alpha|(\mathcal{H}f)(t)|\Big)\, dt \le \frac{2}{\cos(\alpha\pi/2)}$$

for any $0 \le \alpha < 1$, see Theorem 1.9 in [9] (with $\alpha < 1$ being optimal). Using this bound in the previous proof instead of the deeper fact that $\mathcal{H} : L^\infty \to \mathrm{BMO}$ leads directly to the estimate (2.31) on the measure. Since the BMO-estimate (2.13) might be of interest in its own right, we have chosen to present Lemma 2.3 in this way.

Lemma 2.5. *Let $u : \mathbb{T}^2 \to \mathbb{R}$ satisfy $\|u\|_{L^\infty(\mathbb{T}^2)} \le 1$. Assume that u extends as a separately subharmonic function in each variable to a neighborhood of $\mathbb{T}^2$ such that for some $N \ge 1$ and $\rho > 0$,*

$$\sup_{z_1 \in \mathcal{A}_\rho} \sup_{z_2 \in \mathcal{A}_\rho} |u(z_1, z_2)| \le N.$$

Furthermore, suppose that $u = u_0 + u_1$ on $\mathbb{T}^2$ where

$$\|u_0 - \langle u\rangle\|_{L^\infty(\mathbb{T}^2)} \le \varepsilon_0 \;\; \textit{and} \;\; \|u_1\|_{L^1(\mathbb{T}^2)} \le \varepsilon_1$$

with $0 < \varepsilon_0, \varepsilon_1 < 1$. Here $\langle u\rangle := \int_{\mathbb{T}^2} u(x, y)\, dxdy$. Then for any $\delta > 0$,

$$\mathrm{mes}\Big[(x, y) \in \mathbb{T}^2 \;\Big|\; |u(x, y) - \langle u\rangle| > B^\delta \log(N/\varepsilon_1)\Big] \le CN^2\varepsilon_1^{-1} \exp\Big(-cB^{-\frac{1}{2}+\delta}\Big),$$

where $B = \varepsilon_0 \log(N/\varepsilon_1) + N^{\frac{3}{2}}\varepsilon_1^{\frac{1}{4}}$. The constants c, C only depend on ρ.

Proof. We may assume that $\langle u\rangle = 0$ without significantly changing the hypotheses. Let $M = \Big\lceil N^2\varepsilon_1^{-\frac{1}{2}}\Big\rceil$ and denote the Fejér-kernel on $\mathbb{T}$ with Fourier support $[-M+1, M-1]$ by F_M. Then

$$u *_1 F_M = u_0 *_1 F_M + u_1 *_1 F_M,$$

where $*_1$ denotes the convolution in x alone. It is clear that for fixed $x \in \mathbb{T}$,

$$\|u_0 *_1 F_M(x, \cdot)\|_{L^\infty_y} \le \varepsilon_0 \quad \text{and} \quad \|u_1 *_1 F_M(x, \cdot)\|_{L^1_y} \le M\varepsilon_1 \le 2N^2\sqrt{\varepsilon_1}.$$

Since $F_M \geq 0$, $(u *_1 F_M)(x, \cdot)$ extends to a subharmonic function in the second variable satisfying

$$\sup_{z\in\mathcal{A}_\rho} |u *_1 F_M(x, z)| \leq N.$$

Hence Lemma 2.3, in conjunction with the John–Nirenberg inequality, implies that for any $\lambda > 0$,

$$\sup_{x\in\mathbb{T}} \operatorname{mes}\Big[y \in \mathbb{T} \;\Big|\; \big|(u *_1 F_M)(x, y) - \langle (u *_1 F_M)(x, \cdot)\rangle\big| > \lambda\Big] \leq C \exp\Big(-\frac{c\lambda}{B}\Big), \tag{2.32}$$

$$\text{where} \quad B := C_\rho(\varepsilon_0 \log(N/\varepsilon_1) + N^{\frac{3}{2}} \varepsilon_1^{\frac{1}{4}}). \tag{2.33}$$

Observe that for any $x, x' \in \mathbb{T}$

$$\sup_{y\in\mathbb{T}} |(u *_1 F_M)(x, y) - (u *_1 F_M)(x', y)| \leq M \|u\|_{L^\infty(\mathbb{T}^2)} |x - x'| \leq M\, |x - x'|. \tag{2.34}$$

Let $\mathcal{N} \subset \mathbb{T}$ be a $M^{-1}\lambda/4$-net. In view of (2.32) and (2.34) one concludes that

$$\operatorname{mes}\Big[y \in \mathbb{T} \;\Big|\; \sup_{x\in\mathbb{T}} \big|(u *_1 F_M)(x, y) - \langle (u *_1 F_M)(x, \cdot)\rangle\big| > \frac{1}{2}\lambda\Big] \leq C\frac{M}{\lambda} \exp\Big(-\frac{c\lambda}{B}\Big). \tag{2.35}$$

Now let $\lambda = 2\sqrt{B}$ and denote the set on the left-hand side of (2.35) with this choice of λ by $\mathcal{B}_1$. Thus

$$\operatorname{mes}(\mathcal{B}_1) \leq C N^2 \varepsilon_1^{-\frac{1}{2}} B^{-\frac{1}{2}} \exp\Big(-cB^{-\frac{1}{2}}\Big) \leq C N^2 \varepsilon_1^{-1} \exp\Big(-cB^{-\frac{1}{2}}\Big). \tag{2.36}$$

Now fix some $y \in \mathbb{T} \setminus \mathcal{B}_1$ and consider the decomposition of $u(\cdot, y)$ as a function of the first variable given by

$$u(\cdot, y) = u(\cdot, y) - (u *_1 F_M)(\cdot, y) + (u *_1 F_M)(\cdot, y). \tag{2.37}$$

From the Riesz representation

$$u(z, y) = \int \log|z - \zeta|\, d\mu(\zeta) + h(z) \quad \text{with} \quad \mu(\mathcal{A}_{\rho/2}) + \|h\|_{L^\infty(\mathcal{A}_{\rho/4})} \leq C_\rho\, N,$$

it is standard to deduce that the Fourier coefficients

$$\hat{u}(\ell, y) := \int_{\mathbb{T}} u(x, y)\, e(-\ell x)\, dx$$

decay as follows:

$$|\hat{u}(\ell, y)| \leq \frac{C_\rho N}{|\ell|}.$$

In particular, by definition of F_M and because of our choice of y, see (2.35),

$$\begin{aligned}&\|u(\cdot,y)-(u*_1 F_M)(\cdot,y)\|_2 \le C_\rho\, N M^{-\frac{1}{2}} \quad \text{and}\\ &\sup_{x\in\mathbb{T}} \big|(u*_1 F_M)(x,y)-\langle (u*_1 F_M)(x,\cdot)\rangle\big| \le \sqrt{B}.\end{aligned} \tag{2.38}$$

The mean appearing in the second term is uniformly small. In fact, for all $x\in\mathbb{T}$,

$$\begin{aligned}\big|\langle (u*_1 F_M)(x,\cdot)\rangle\big| &\le \int_{\mathbb{T}} |(u_0*_1 F_M)(x,y)|\,dy + \int_{\mathbb{T}} |(u_1*_1 F_M)(x,y)|\,dy\\ &\le \|u_0\|_{L^\infty(\mathbb{T}^2)} + M\|u_1\|_{L^1(\mathbb{T}^2)} \le \varepsilon_0 + 2N^2\sqrt{\varepsilon_1}.\end{aligned} \tag{2.39}$$

Assuming as we may that $B\le 1$, one checks from (2.33) that the bound in (2.39) is no larger than $C\sqrt{B}$. Hence (2.38) implies that for any $y\in\mathbb{T}\setminus\mathcal{B}_1$ (recall that $M=\Big\lceil N^2\varepsilon_1^{-\frac{1}{2}}\Big\rceil$)

$$\|u(\cdot,y)-(u*_1 F_M)(\cdot,y)\|_1 \le C_\rho\, \varepsilon_1^{\frac{1}{4}} \;\text{ and }\; \sup_{x\in\mathbb{T}} \big|(u*_1 F_M)(x,y)\big| \le C\sqrt{B}.$$

Applying Lemma 2.3 to the function $u(\cdot,y)$ with the decomposition given by (2.37) therefore yields

$$\sup_{y\in\mathbb{T}\setminus\mathcal{B}_1} \|u(\cdot,y)\|_{\mathrm{BMO}} \le C_\rho(\sqrt{B}\log(N/\varepsilon_1) + N^{\frac{1}{2}}\varepsilon_1^{\frac{1}{8}}) \le C_\rho\sqrt{B}\log(N/\varepsilon_1). \tag{2.40}$$

It remains to be shown that

$$v(y) := \langle u(\cdot,y)\rangle = \int_{\mathbb{T}} u(x,y)\,dx$$

is close to zero for most y. Clearly, v extends to a subharmonic function on $\mathcal{A}_\rho$ such that

$$\sup_{z\in\mathcal{A}_\rho} |v(z)| \le N \quad \text{and} \quad \langle v\rangle = \langle u\rangle = 0.$$

With $v_0(y) := \langle u_0(\cdot,y)\rangle$ and $v_1(y) := \langle u_1(\cdot,y)\rangle$ one has

$$\|v_0\|_{L^\infty(\mathbb{T})} \le \varepsilon_0 \quad \text{and} \quad \|v_1\|_{L^1(\mathbb{T})} \le \varepsilon_1.$$

Therefore, Lemma 2.3 implies that

$$\|v\|_{\mathrm{BMO}} \le C\Big(\varepsilon_0\log(N/\varepsilon_1) + \sqrt{N\varepsilon_1}\Big) \le CB.$$

Thus

$$\operatorname{mes}\Big[y\in\mathbb{T}\;\Big|\; |v(y)| > \sqrt{B}\Big] \le C\exp\Big(-cB^{-\frac{1}{2}}\Big). \tag{2.41}$$

Denoting the set on the left-hand side by $\mathcal{B}_2$, let $\mathcal{B} := \mathcal{B}_1\cup\mathcal{B}_2$. One concludes from (2.36), (2.41), and (2.40) by means of the John–Nirenberg inequality that

$$\operatorname{mes}\Big[(x,y)\in\mathbb{T}^2\;\Big|\; |u(x,y)| > B^\delta\log(N/\varepsilon_1)\Big] \le \operatorname{mes}(\mathcal{B}) + C\,\exp\Big(-cB^{-\frac{1}{2}+\delta}\Big),$$

and the lemma follows. □

2.2. Averages of subharmonic functions over orbits of the skew-shift. In what follows we assume that $\omega \in (0, 1)$ is Diophantine in the sense that

$$\|n\omega\| \geq \varepsilon\, n^{-1}(1 + \log n)^{-2} \text{ for any } n \in \mathbb{Z}^+, \tag{2.42}$$

where $\varepsilon > 0$ is some arbitrary but fixed small number. Let Ω_ε be the set of those ω that satisfy (2.42). It is clear that

$$\mathrm{mes}[\mathbb{T} \setminus \Omega_\varepsilon] < C\varepsilon$$

with an absolute constant C. The choice of logarithm in (2.42) is mainly for convenience. A very small power loss is also acceptable. Throughout this section we will use Ω_ε in this sense. Let $T_\omega : \mathbb{T}^2 \to \mathbb{T}^2$, $T_\omega(x, y) = (x + y, y + \omega) \pmod{\mathbb{Z}^2}$ be the skew-shift. Observe that the iterates of T_ω are given by

$$T_\omega^k(x, y) = (x + ky + k(k-1)\omega/2,\, y + k\omega) \quad \bmod \mathbb{Z}^2 \tag{2.43}$$

for any $k \in \mathbb{Z}$.

Lemma 2.6. *Let $u : \mathbb{T}^2 \to \mathbb{R}$ extend to some neighborhood of $\mathbb{T}^2$ as a separately subharmonic function in each variable so that for some $\rho > 0$,*

$$\sup_{z_1 \in \mathcal{A}_\rho} \sup_{z_2 \in \mathcal{A}_\rho} |u(z_1, z_2)| \leq 1. \tag{2.44}$$

Fix a small $\varepsilon > 0$ and let $\omega \in \Omega_\varepsilon$, see (2.42)*. For any $\delta > 0$ there exist constants c, C such that*

$$\mathrm{mes}\Big[(x, y) \in \mathbb{T}^2 \;\Big|\; \Big|\frac{1}{K}\sum_{k=1}^{K} u \circ T_\omega^k(x, y) - \langle u\rangle\Big| > K^{-\frac{1}{12}+2\delta}\Big] \leq C\exp\Big(-cK^\delta\Big), \tag{2.45}$$

for any positive integer K. Here $\langle u\rangle = \int_{\mathbb{T}^2} u(x, y)\, dxdy$ and the constants depend only on $\rho, \delta, \varepsilon$.

Proof. Let $\hat{u}(\ell, y) = \int_0^1 u(x, y)e(-\ell x)\, dx$ denote the Fourier coefficient with respect to the first variable. As above one deduces by means of the Riesz representation of the subharmonic function $z \mapsto u(z, y)$ and from (2.44) that

$$\sup_{y \in \mathbb{T}} |\hat{u}(\ell, y)| \leq C_\rho\, |\ell|^{-1}. \tag{2.46}$$

With some positive integer p_1 to be determined, let

$$\begin{aligned} u(x, y) &= \sum_{|\ell_1| \leq p_1} \hat{u}(\ell_1, y)e(\ell_1 x) + \sum_{|\ell_1| > p_1} \hat{u}(\ell_1, y)e(\ell_1 x) \\ &=: u_1(x, y) + \widetilde{u_1}(x, y), \end{aligned} \tag{2.47}$$

where u_1 and $\widetilde{u_1}$ are the respective sums on the right-hand side of (2.47). By (2.46),

$$\sup_{y \in \mathbb{T}} \|\widetilde{u_1}(\cdot, y)\|_{L^2_x} \leq C p_1^{-\frac{1}{2}}. \tag{2.48}$$

With some positive integer p_2 to be determined below, let

$$\begin{aligned} u_1(x,y) &= \sum_{\substack{|\ell_1|\le p_1\\ |\ell_2|>p_2}} \hat{u}(\ell_1,\ell_2)e(\ell_1 x+\ell_2 y) + \sum_{\substack{|\ell_1|\le p_1\\ |\ell_2|\le p_2}} \hat{u}(\ell_1,\ell_2)e(\ell_1 x+\ell_2 y) \\ &=: u_2(x,y)+u_3(x,y). \end{aligned} \tag{2.49}$$

Using the Riesz representation in the second variable one derives from (2.44) that

$$|\hat{u}(\ell_1,\ell_2)| \le \int_{\mathbb{T}}\Big|\int_{\mathbb{T}} e(-\ell_2 y)u(x,y)\,dy\Big|\,dx \le \frac{C_\rho}{1+|\ell_2|}. \tag{2.50}$$

Therefore,

$$\|u_2\|_{L^2(\mathbb{T}^2)} \le \sum_{|\ell_1|\le p_1}\Big\|\sum_{|\ell_2|>p_2}\hat{u}(\ell_1,\ell_2)e(\ell_2 y)\Big\|_{L^2_y} \le C\,p_1 p_2^{-\frac{1}{2}}.$$

In particular,

$$\begin{aligned} &\mathrm{mes}\Big[y\in\mathbb{T}\;\Big|\;\int_{\mathbb{T}}\frac{1}{K}\Big|\sum_{k=1}^{K}u_2\circ T^k(x,y)\Big|\,dx > K^{-1}\Big] \\ &\qquad\le K\int_{\mathbb{T}^2}\frac{1}{K}\Big|\sum_{k=1}^{K}u_2\circ T^k(x,y)\Big|\,dxdy \\ &\qquad\le K\|u_2\|_{L^1(\mathbb{T}^2)} \le C\,Kp_1p_2^{-\frac{1}{2}}. \end{aligned} \tag{2.51}$$

Let $\mathcal{B}$ be the set on the left-hand side of (2.51). In view of (2.43),

$$\begin{aligned} &\sup_{x,y\in\mathbb{T}^2}\Big|\frac{1}{K}\sum_{k=1}^{K}u_3\circ T^k(x,y)-\langle u\rangle\Big| \\ &\quad\le \frac{1}{K}\sum_{\substack{|\ell_1|\le p_1,|\ell_2|\le p_2\\ |\ell_1|+|\ell_2|\ne 0}}\frac{C}{1+|\ell_2|}\Big|\sum_{k=1}^{K}e\big[\ell_1(ky+\omega k(k-1)/2)+\ell_2 k\omega\big]\Big| \\ &\quad\le \frac{1}{K}\sum_{\ell_2=1}^{p_2}\frac{C}{\ell_2}\Big|\sum_{k=1}^{K}e(\ell_2 k\omega)\Big| \\ &\qquad + \frac{1}{K}\sum_{\ell_1=1}^{p_1}\sum_{\ell_2=0}^{p_2}\frac{C}{1+\ell_2}\left(\sum_{m=1}^{K-1}\min\Big(K,\|m\ell_1\omega\|^{-1}\Big)\right)^{\frac{1}{2}} \end{aligned} \tag{2.52}$$

$$\begin{aligned} &\quad\le \frac{C}{K}\sum_{\ell_2=1}^{p_2}\frac{1}{\ell_2}\min(K,\|\ell_2\omega\|^{-1}) \\ &\qquad + \frac{C}{K}\sqrt{p_1}\log p_2\left(\sum_{\ell_1=1}^{p_1}\sum_{m=1}^{K-1}\min\Big(K,\|m\ell_1\omega\|^{-1}\Big)\right)^{\frac{1}{2}} \\ &\quad=: S_1+S_2. \end{aligned} \tag{2.53}$$

To obtain the second term in line (2.52), one uses the well-known method of Weyl–differencing, cf. Montgomery [11, Chap. 3]. In fact,

$$\Big|\sum_{k=1}^{K} e\big[\ell_1(ky+\omega k(k-1)/2)+\ell_2 k\omega\big]\Big|^2 \le K + 2\sum_{m=1}^{K-1}\min\Big(K, \frac{2}{|1-e(\ell_1\omega m)|}\Big)$$
$$\le C\sum_{m=1}^{K-1}\min\big(K, \|\ell_1\omega m\|^{-1}\big),$$

which leads to (2.52). In view of (2.42) (with $a \sim b$ denoting $b \le a \le 2b$), for any positive integer R,

$$\sum_{\ell=1}^{R}\frac{1}{\ell}\min(1, K^{-1}\|\ell\omega\|^{-1}) \le \sum_{1\le 2^j\le K}\sum_{\ell=1}^{R}\chi_{[\|\ell\omega\|\sim 2^{-j}]}\frac{1}{\ell}\min(1, K^{-1}2^j)$$
$$+\sum_{\ell=1}^{R}\chi_{[\|\ell\omega\|\le K^{-1}]}\frac{1}{\ell}$$
$$\le C\sum_{1\le 2^j\le K}\frac{2^j}{K}\, j^2\, 2^{-j}\log R + C\sum_{\ell=1}^{R}\frac{(\log K)^2}{\ell K}$$
$$\le C\frac{(\log K)^2}{K}\,\log R.$$

Here the constants depend on ε. Thus,

$$S_1 \le C_\varepsilon\,\frac{(\log K)^2}{K}\,\log p_2. \tag{2.54}$$

By Dirichlet's principle there is an integer $1 \le q \le K$ and an integer p so that $\gcd(p, q) = 1$ and $|\omega - \frac{p}{q}| \le \frac{1}{qK}$. In view of (2.42), one also has $q \ge c_\varepsilon \frac{K}{(\log K)^2}$. By means of the standard bound on the divisor function and the usual estimates for reciprocal sums, cf. [11, Chap. 3],

$$\sum_{\ell_1=1}^{p_1}\sum_{m=1}^{K-1}\min(K, \|m\ell_1\omega\|^{-1}) \le C_{\varepsilon_2}(p_1K)^{\varepsilon_2}\sum_{k=1}^{p_1K}\min(K, \|k\omega\|^{-1})$$
$$\le C_{\varepsilon_2}(p_1K)^{\varepsilon_2}\Big(\frac{p_1K^2}{q} + p_1K\log q + K + q\log q\Big)$$
$$\le C_{\varepsilon_2}(p_1K)^{1+2\varepsilon_2}, \tag{2.55}$$

where $\varepsilon_2 > 0$ is an arbitrarily small parameter. One obtains from (2.53), (2.54), and (2.55) that

$$\sup_{x,y\in\mathbb{T}^2}\Big|\frac{1}{K}\sum_{k=1}^{K} u_3\circ T^k(x,y) - \langle u\rangle\Big| \le S_1 + S_2 \le Cp_1^{1+\varepsilon_2}K^{-\frac{1}{2}+\varepsilon_2}\log p_2 \tag{2.56}$$

with a constant that depends both on ε and ε_2. Fix some small $\delta > 0$ and choose $p_1 = K^{\frac{1}{3}}$ and $p_2 = \exp\left(4K^{\delta}\right)$. The conclusion from the preceding is as follows, cf. (2.48), (2.51), and (2.56): There exists a subset $\mathcal{B} \subset \mathbb{T}$ of measure

$$\text{mes}(\mathcal{B}) \leq CK^{\frac{4}{3}} \exp\left(-2K^{\delta}\right) \leq C \exp\left(-K^{\delta}\right), \tag{2.57}$$

such that (choosing $2\varepsilon_2 < \delta$)

$$\begin{aligned} &\sup_{y\in\mathbb{T}\setminus\mathcal{B}} \left\| \frac{1}{K} \sum_{k=1}^{K} u \circ T^k(\cdot, y) - \langle u \rangle \right\|_{L^1_x} \\ &\quad \leq \sup_{y\in\mathbb{T}} \left\| \frac{1}{K} \sum_{k=1}^{K} \widetilde{u_1} \circ T^k(x, y) \right\|_{L^1_x} + \sup_{y\in\mathbb{T}\setminus\mathcal{B}} \int_{\mathbb{T}} \frac{1}{K} \left| \sum_{k=1}^{K} u_2 \circ T^k(x, y) \right| dx \\ &\qquad + \sup_{(x,y)\in\mathbb{T}^2} \left| \frac{1}{K} \sum_{k=1}^{K} u_3 \circ T^k(x, y) - \langle u \rangle \right| \\ &\quad \leq CK^{-\frac{1}{6}} + K^{-1} + C\,K^{\frac{1}{3}+\varepsilon_2} K^{\delta-\frac{1}{2}+\varepsilon_2} \leq C\,K^{-\frac{1}{6}+2\delta} \end{aligned} \tag{2.58}$$

with constants that depend on both δ and ε. To obtain (2.45), one uses Lemma (2.3) to convert the L^1-bound (2.58) into an L^∞-bound at the cost of removing an exponentially small set. For any fixed $y \in \mathbb{T} \setminus \mathcal{B}$, consider the bounded subharmonic function

$$v_y(z) := \frac{1}{K} \sum_{k=1}^{K} u \circ T^k(z, y) \quad \text{with} \quad z \in \mathcal{A}_\rho.$$

It is important to notice that y is real. Otherwise $T^k(z, y) \notin \mathcal{A}_\rho \times \mathcal{A}_\rho$ for large k, see (2.43). One has the decomposition

$$\frac{1}{K} \sum_{k=1}^{K} u \circ T^k(\cdot, y) = \langle u \rangle + \frac{1}{K} \sum_{k=1}^{K} u \circ T^k(\cdot, y) - \langle u \rangle.$$

In view of (2.58) one obtains from Lemma 2.3 (with $N = 1$, $\varepsilon_0 = 0$, and $\varepsilon_1 = K^{-\frac{1}{6}+2\delta}$) that

$$\left\| \frac{1}{K} \sum_{k=1}^{K} u \circ T^k(\cdot, y) \right\|_{\text{BMO}_x} \leq C_\delta\, K^{-\frac{1}{12}+\delta}.$$

By the John–Nirenberg inequality thus

$$\sup_{y\in\mathbb{T}\setminus\mathcal{B}} \text{mes}\left[x \in \mathbb{T} \;\middle|\; |v_y(x) - \langle v_y \rangle| > C_\delta\, K^{-\frac{1}{12}+2\delta}\right] \leq C \exp\left(-K^{\delta}\right). \tag{2.59}$$

Since (2.58) implies that $|\langle v_y \rangle - \langle u \rangle| \leq C_\delta\, K^{-\frac{1}{6}+2\delta}$, the lemma follows from (2.59) and (2.57). $\square$

Remark 2.7. It will be important in the proof of localization below that the previous lemma requires only finitely many conditions on ω. More precisely, the arithmetic nature of ω only enters into the estimate of S_1 and S_2. Furthermore, what is required for the bound on S_1 is the following: If for some $K^{-1} \le \kappa < 1$ and some positive distinct integers ℓ, ℓ',

$$\|\ell\,\omega\| < \kappa \ \text{ and } \ \|\ell'\,\omega\| < \kappa,$$

then $|\ell - \ell'| \ge c_\varepsilon \kappa^{-1}(\log\kappa)^{-2}$. This clearly requires the Diophantine condition (2.42) only for $1 \le k \le K$. As far as S_2 is concerned, it is evident from the estimate of S_2 that (2.42) is used only in the range $1 \le k \le p_1 K \le K^2$.

2.3. The main inductive step in the proof of the large deviation theorem. Consider equations of the form

$$-\psi_{n+1} - \psi_{n-1} + \lambda v(T_\omega^n(x,y))\psi_n = E\psi_n, \tag{2.60}$$

where $T_\omega : \mathbb{T}^2 \to \mathbb{T}^2$, $T_\omega(x,y) = (x+y, y+\omega) \pmod{\mathbb{Z}^2}$ is the skew-shift, and v is a nonconstant real–analytic function on $\mathbb{T}^2$ satisfying some further conditions that will specified below. Let

$$A_j(x,y;\lambda,E) = \begin{bmatrix} \lambda v(T_\omega^j(x,y)) - E & -1 \\ 1 & 0 \end{bmatrix}.$$

The matrix $M_n(x,y;\lambda,E) = \prod_{j=n}^{1} A_j(x,y;\lambda,E)$ denotes the monodromy matrix of Eq. (2.60). As usual,

$$L_n(\lambda,E) = \frac{1}{n}\int_{\mathbb{T}^2} \log\|M_n(x,y;\lambda,E)\|\,dxdy$$

and $L(\lambda,E) = \lim_{n\to\infty} L_n(\lambda,E)$ denotes the Lyapunov exponent. Introduce a scaling factor

$$S(\lambda,E) = \log(C_v + |\lambda| + |E|) \ge 1, \tag{2.61}$$

where C_v is a constant depending only on the potential v so that for all n

$$\sup_{z\in\mathcal{A}_\rho}\sup_{y\in\mathbb{T}} \frac{1}{n}\log\|M_n(z,y;\lambda,E)\| + \sup_{z_1\in\mathcal{A}_\rho}\sup_{z_2\in\mathcal{A}_\rho} \frac{1}{n^2}\log\|M_n(z_1,z_2;\lambda,E)\| \le S(\lambda,E). \tag{2.62}$$

Here $\rho = \rho_v$ is determined by v. Observe that (2.62) basically requires the function v to extend in the first variable to an analytic function on $\mathbb{C}\setminus\{0\}$ such that

$$\sup_{y\in\mathbb{T}} |v(z,y)| \le C(|z|^d + |z|^{-d})$$

with some constant d, see (2.43). For example, any trigonometric polynomial

$$v(x,y) = \sum_{|k|+|\ell|\le d} a_{k,\ell}\, e(kx+\ell y)$$

satisfies this requirement. Another possibility, which is slightly more technical to state, but applies to any analytic function on a neighborhood of $\mathbb{T}^2$, is as follows: For all n,

$$\sup_{z_1\in\mathcal{A}_\rho}\ \sup_{z_2\in\mathcal{A}_{\rho/n}} \frac{1}{n}\log\|M_n(z_1,z_2;\lambda,E)\| \le S(\lambda,E). \tag{2.63}$$

The difference from (2.62) here is that in the second term the z_2-variable only needs to be taken in an annulus of thickness $\frac{\rho}{n}$. Observe that (2.63) can be stated for any potential v that extends analytically to a neighborhood of $\mathbb{T}^2$ of size ρ. This is essential for real-analytic v. The reason (2.63) is sufficient for our purposes is the following simple fact. Suppose u is a subharmonic function on $\mathcal{A}_{\rho/n}$ bounded by one. Then there is the Riesz representation

$$u(z) = \int \log|z-\zeta|\,d\mu(\zeta) + h(z),$$

where

$$\mu(\mathcal{A}_{\rho/(2n)}) + \|h\|_{L^\infty(\mathcal{A}_{\rho/(4n)})} \le C_\rho\, n. \tag{2.64}$$

In particular, one has the decay of the Fourier coefficients

$$|\hat{u}(\ell)| \le \frac{Cn}{\ell}. \tag{2.65}$$

The reader will easily verify that (2.64), (2.65) are all that is required in the proof of the following lemma.

The following lemma provides the inductive step in the proof of the large deviation theorem. It is based on the avalanche principle and all our previous lemmas.

Lemma 2.8. *Fix $\varepsilon > 0$ small and let $\omega \in \Omega_\varepsilon$, see* (2.42). *Suppose n and $N > n$ are positive integers such that*

$$\mathrm{mes}\Big[(x,y)\in\mathbb{T}^2 \;\Big|\; \Big|\frac{1}{n}\log\|M_n(x,y;\lambda,E)\| - L_n(\lambda,E)\Big| > S(\lambda,E)\frac{\gamma}{10}\Big] \le N^{-10}, \tag{2.66}$$

$$\mathrm{mes}\Big[(x,y)\in\mathbb{T}^2 \;\Big|\; \Big|\frac{1}{2n}\log\|M_{2n}(x,y;\lambda,E)\| - L_{2n}(\lambda,E)\Big| > S(\lambda,E)\frac{\gamma}{10}\Big] \le N^{-10}. \tag{2.67}$$

Assume that

$$\min(L_n(\lambda,E), L_{2n}(\lambda,E)) \ge \gamma S(\lambda,E), \tag{2.68}$$

$$L_n(\lambda,E) - L_{2n}(\lambda,E) \le \frac{\gamma}{40} S(\lambda,E), \tag{2.69}$$

$$9\gamma n S \ge 10\log(2N) \quad \text{and} \quad n^2 \le N. \tag{2.70}$$

Then there is some absolute constant C_0 with the property that (with $L_N = L_N(\lambda,E)$ etc.)

$$\begin{aligned} &L_N \ge \gamma S(\lambda,E) - 2(L_n - L_{2n}) - C_0\, S(\lambda,E) n N^{-1} \\ &\text{and} \;\; L_N - L_{2N} \le C_0\, S(\lambda,E) n N^{-1}. \end{aligned} \tag{2.71}$$

Furthermore, for any $\sigma < \frac{1}{24}$ *there is* $\tau = \tau(\sigma) > 0$ *so that*

$$\operatorname{mes}\Big[(x,y) \in \mathbb{T}^2 \;\Big|\; \Big|\frac{1}{N}\log\|M_N(x,y;\lambda,E)\| - L_N(\lambda,E)\Big| > S(\lambda,E)N^{-\tau}\Big] \le C\,\exp\big(-N^{\sigma}\big) \tag{2.72}$$

with some constant $C = C(\sigma,\varepsilon)$.

Proof. We shall fix ω, λ, and E for the purposes of this proof and suppress these variables in the notation. In particular, $S = S(\lambda,E)$. Denote the set on the left-hand side of (2.66) by $\mathcal{B}_n$ and the set on the left-hand side of (2.67) by $\mathcal{B}_{2n}$. For any $(x,y) \in \mathbb{T}^2 \setminus \mathcal{B}_n$,

$$\|M_n(x,y)\| \ge \exp(n\gamma S - \frac{\gamma}{10}Sn) = \exp\Big(\frac{9\gamma}{10}Sn\Big) =: \mu.$$

By (2.70), $\mu \ge 2N$. Furthermore, for any $(x,y) \notin \mathcal{B}_n \cup T^{-n}\mathcal{B}_n \cup \mathcal{B}_{2n}$, (2.66)–(2.69) imply

$$\log\|M_n \circ T^n(x,y)\| + \log\|M_n(x,y)\| - \log\|M_{2n}(x,y)\| \le 2n(L_n - L_{2n}) + \frac{4\gamma}{10}Sn \le \frac{9\gamma}{20}Sn = \frac{1}{2}\log\mu.$$

Applying Prop. 2.1 N times yields a set $\mathcal{B}_1 \subset \mathbb{T}^2$ with measure

$$\operatorname{mes}(\mathcal{B}_1) \le 4N \cdot N^{-10} = 4N^{-9} \tag{2.73}$$

so that for any $(x,y) \in \mathbb{T}^2 \setminus \mathcal{B}_1$,

$$\Bigg|\frac{1}{N}\log\|M_N(x,y)\| + \frac{1}{N}\sum_{j=1}^{N}\frac{1}{n}\log\|M_n \circ T^j(x,y)\| - \frac{2}{N}\sum_{j=1}^{N}\frac{1}{2n}\log\|M_{2n}\circ T^j(x,y)\|\Bigg| \le C\Big(\frac{Sn}{N} + \frac{1}{\mu}\Big) \le CSnN^{-1}. \tag{2.74}$$

Integrating (2.74) over $\mathbb{T}^2$ yields

$$|L_N + L_n - 2L_{2n}| \le C\,SnN^{-1} + 16SN^{-9}, \tag{2.75}$$

which implies the first inequality in (2.71). To obtain the second inequality in (2.71), observe that by virtue of (2.70) all arguments so far apply equally well to M_{2N} instead of M_N. Subtracting (2.75) from the analogous inequality involving L_{2N} yields the desired bound. Denote

$$u_N(x,y) = \frac{1}{N}\log\|M_N(x,y)\|,$$

and similarly with n and $2n$. In view of (2.63), both u_n and u_{2n} extend to separately subharmonic functions in both variables such that

$$\sup_{z_1\in\mathcal{A}_\rho}\;\sup_{z_2\in\mathcal{A}_{(\rho/n)}}\Big[|u_n(z_1,z_2)| + |u_{2n}(z_1,z_2)|\Big] \le CS.$$

Applying Lemma 2.6 to u_n/S and $u_{2n}/(2S)$ (cf. the comments following (2.63), in particular (2.64) and (2.65)) therefore implies that there is a set $\mathcal{B}_2 \subset \mathbb{T}^2$ with measure ($\delta > 0$ is a fixed small number)

$$\text{mes}(\mathcal{B}_2) \le C \exp\Big(-N^\delta\Big), \tag{2.76}$$

such that for any $(x, y) \in \mathcal{G} := \mathbb{T}^2 \setminus (\mathcal{B}_1 \cup \mathcal{B}_2)$,

$$|u_N(x, y) + L_n - 2L_{2n}| \le CSnN^{-1} + C_\delta\, SN^{-\frac{1}{12}+2\delta}, \tag{2.77}$$

see (2.74). For small δ the second term is the larger one since $N \ge n^2$. Fix such an integer N. Consider the following decomposition of $u := u_N$ as a function on $\mathbb{T}^2$:

$$u = u\chi_{\mathcal{G}} + L_N\chi_{\mathcal{G}^c} + u\chi_{\mathcal{G}^c} - L_N\chi_{\mathcal{G}^c} =: u_0 + u_1.$$

Here u_0 is the sum of the first two terms (and $\mathcal{G}^c := \mathbb{T}^2 \setminus \mathcal{G}$). In view of (2.77) and (2.75),

$$\begin{aligned} \|u_0 - \langle u\rangle\|_\infty = \|u_0 - L_N\|_\infty &= \|u - L_N\|_{L^\infty(\mathcal{G})} \\ &\le \|u_N + L_n - L_{2n}\|_{L^\infty(\mathcal{G})} + |L_N + L_n - L_{2n}| \\ &\le C_\delta\, SN^{-\frac{1}{12}+2\delta}. \end{aligned} \tag{2.78}$$

On the other hand, (2.73) and (2.76) imply that

$$\|u_1\|_1 \le 2S\,\text{mes}(\mathcal{G}^c) \le CS\big[N^{-9} + \exp\big(-N^\delta\big)\big] \le C_\delta\, SN^{-9}. \tag{2.79}$$

Applying Lemma 2.5 to the function u/S with ε_0 and ε_1 given by (2.78) and (2.79), respectively, proves (2.72). Indeed, in this case the quantity B from Lemma (2.5) satisfies

$$B \le C_\delta\, N^{-\frac{1}{12}+2\delta} \log(N^{10}) + CN^{\frac{3}{2}} N^{-\frac{9}{4}},$$

which gives the value of σ stated above. □

Remark 2.9. In view of Remark 2.7 it is clear that Prop. 2.11 only requires the Diophantine condition (2.42) in the range $1 \le k \le N^2$. This will be relevant in the proof of localization below.

2.4. The initial condition via large disorder. Let $V_j = v \circ T^j(x, y)$ and define

$$f_n(x, y; \lambda, E) = \det \begin{bmatrix} \lambda V_1 - E & -1 & 0 & 0 & . & . & . & . & 0 \\ -1 & \lambda V_2 - E & -1 & 0 & 0 & . & . & . & 0 \\ 0 & -1 & \lambda V_3 - E & -1 & 0 & 0 & . & . & 0 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & -1 \\ 0 & 0 & . & . & . & . & 0 & -1 & \lambda V_n - E \end{bmatrix}. \tag{2.80}$$

Recall the simple property

$$M_n(x, y; \lambda, E) = \begin{bmatrix} f_n(x, y; \lambda, E) & -f_{n-1}(T(x, y); \lambda, E) \\ f_{n-1}(x, y; \lambda, E) & -f_{n-2}(T(x, y); \lambda, E) \end{bmatrix}. \tag{2.81}$$

Finally, let

$$D_n(x, y; \lambda, E) = \text{diag}(\lambda V_1 - E, \ldots, \lambda V_n - E). \tag{2.82}$$

Lemma 2.10. *There exist constants λ_0 and B depending only on v such that for any positive integer n,*

$$\sup_E \operatorname{mes}\Big[(x, y) \in \mathbb{T}^2 \;\Big|\; \Big|\frac{1}{n}\log\|M_n(x, y; \lambda, E)\| - L_n(\lambda, E)\Big| \geq \frac{1}{20}S(\lambda, E)\Big] \leq n^{-50}, \tag{2.83}$$

provided $\lambda \geq \lambda_0 \vee n^B$. Furthermore, for those λ and all E,

$$L_n(\lambda, E) \geq \frac{1}{2}S(\lambda, E) \;\text{ and }\; L_n(\lambda, E) - L_{2n}(\lambda, E) \leq \frac{1}{80}S(\lambda, E).$$

Proof. The matrix on the right-hand side of (2.80) can be written in the form $D_n + B_n$, where D_n is given by (2.82). Clearly, $\|B_n\| = 2$ and

$$\frac{1}{n}\log|\det D_n(x, y; \lambda, E)| = \log\lambda + \frac{1}{n}\sum_{j=1}^{n}\log|v(T^j(x, y)) - E/\lambda|. \tag{2.84}$$

It is a well-known property of nonconstant real-analytic functions v that there exist constants $b > 0$ and C depending on v such that

$$\operatorname{mes}\big[(x, y) \in \mathbb{T}^2 \;\big|\; |v(x, y) - h| < t\big] \leq Ct^b \tag{2.85}$$

for all $-2\|v\|_\infty \leq h \leq 2\|v\|_\infty$ and $t > 0$, see for example Lemma 11.4 in [7]. Therefore, for any $|E| \leq 2\lambda\|v\|_\infty$,

$$\operatorname{mes}\Big[(x, y) \in \mathbb{T}^2 \;\Big|\; \frac{1}{n}\sum_{j=1}^{n}\log|v \circ T^j(x, y) - E/\lambda| < -\rho\Big] < n\,Ce^{-b\rho}. \tag{2.86}$$

One also has the upper bound

$$\sup_{(x,y)\in\mathbb{T}^2}\frac{1}{n}\sum_{j=1}^{n}\log|v(x, y) - E/\lambda| \leq \log(3\|v\|_\infty). \tag{2.87}$$

Since

$$\|D_n(x, y; \lambda, E)^{-1}\| \leq \lambda^{-1}\max_{1\leq j\leq n}|v \circ T^j(x, y) - E/\lambda|^{-1},$$

(2.85) implies that

$$\begin{aligned}\operatorname{mes}\Big[(x, y) \in \mathbb{T}^2 \;\Big|\; \|D_n(x, y; \lambda, E)^{-1}\| > \frac{1}{4}\Big] &\\ \leq n\operatorname{mes}\big[(x, y) \in \mathbb{T}^2 \;\big|\; |v(x, y) - E/\lambda| < 4\lambda^{-1}\big] &\\ \leq Cn\lambda^{-b}. &\end{aligned} \tag{2.88}$$

Hence

$$\operatorname{mes}\Big[(x, y) \in \mathbb{T}^2 \;\Big|\; \|D_n(x, y; \lambda, E)^{-1}B_n\| > \frac{1}{2}\Big] \leq Cn\lambda^{-b}. \tag{2.89}$$

In view of (2.80), (2.84), (2.86), (2.87), and (2.88),

$$\begin{aligned}
&\Big|\frac{1}{n}\log|f_n(x,y;\lambda,E)| - \log\lambda\Big| \\
&\le \Big|\frac{1}{n}\sum_{j=1}^{n}\log|v(T^j(x,y)) - E/\lambda|\Big| + \Big|\frac{1}{n}\log|\det(I + D_n(x,y;\lambda,E)^{-1}B_n)|\Big| \\
&\le \rho + \log(3\|v\|_\infty) + \log 2
\end{aligned} \tag{2.90}$$

up to a set of measure not exceeding

$$Cne^{-b\rho} + Cn\lambda^{-b}. \tag{2.91}$$

Now let $\rho = \frac{1}{400}\log\lambda$ and assume $(6\|v\|_\infty)^{400} \le \lambda$. Then the right-hand side of (2.90) is no larger than $\frac{1}{200}\log\lambda$. Under these assumptions the measure given by (2.91) is on the order of $Cn\lambda^{-\frac{b}{400}}$. Choosing

$$\lambda \ge n^B$$

for some B depending only on v implies

$$\sup_{|E|\le 2\lambda\|v\|_\infty} \operatorname{mes}\Big[(x,y)\in\mathbb{T}^2 \;\Big|\; \Big|\frac{1}{n}\log|f_n(x,y;\lambda,E)| - \log\lambda\Big| \ge \frac{1}{200}\log\lambda\Big] \le n^{-100}.$$

In view of (2.81) one therefore obtains

$$\sup_{|E|\le 2\lambda\|v\|_\infty} \operatorname{mes}\Big[(x,y)\in\mathbb{T}^2 \;\Big|\; \Big|\frac{1}{n}\log\|M_n(x,y;\lambda,E)\| - \log\lambda\Big| \ge \frac{1}{199}\log\lambda\Big] \le 4n^{-100}. \tag{2.92}$$

In particular,

$$|L_n(\lambda,E) - \log\lambda| \le \frac{1}{199}\log\lambda + 4S(\lambda,E)n^{-100} \le \frac{1}{198}S(\lambda,E), \tag{2.93}$$

provided $n \ge 2$. Since

$$\log\lambda \ge \frac{99}{100}\sup_{|E|\le 2\lambda\|v\|_\infty} S(\lambda,E)$$

for large λ_0, (2.93) implies the second statement of the lemma in this range of E. Replacing $\log\lambda$ with L_n in (2.92) yields

$$\sup_{|E|\le 2\lambda\|v\|_\infty} \operatorname{mes}\Big[(x,y)\in\mathbb{T}^2 \;\Big|\; \Big|\frac{1}{n}\log\|M_n(x,y;\lambda,E)\| - L_n(\lambda,E)\Big| \ge \frac{1}{90}S(\lambda,E)\Big] \le 4n^{-100}. \tag{2.94}$$

If $|E| > 2\lambda\|v\|_\infty$ and λ_0 is sufficiently large, then the set in (2.83) is empty. In fact, for such E,

$$\Big|\frac{1}{n}\log|\det D_n(x,y;\lambda,E)| - \log|E|\Big| \le 2,$$

and thus

$$\Big|\frac{1}{n}\log|f_n(x,y;\lambda,E)| - \log|E|\Big| \le 4$$

which implies that for large λ,

$$\Big|\frac{1}{n}\log\|M_n(x,y;\lambda,E)\| - \log|E|\Big| \le 8 \le \frac{1}{200}S(\lambda,E).$$

Hence

$$|L_n(\lambda,E) - \log|E|| \le \frac{1}{200}S(\lambda,E),$$

and the lemma follows. □

2.5. The proof of the large deviation estimate and positivity of the Lyapunov exponent.

Proposition 2.11. *Fix $\varepsilon > 0$ small and let $\omega \in \Omega_\varepsilon$, see* (2.42). *Assume v is a nonconstant real–analytic function on $\mathbb{T}^2$. Then for all $\sigma < \frac{1}{24}$ there exist $\tau = \tau(\sigma) > 0$ and constants λ_1 and n_0 depending only on ε, v and σ such that*

$$\sup_E \operatorname{mes}\Big[(x,y)\in\mathbb{T}^2 \;\Big|\; \Big|\frac{1}{n}\log\|M_n(x,y;\lambda,E)\| - L_n(\lambda,E)\Big| > S(\lambda,E)n^{-\tau}\Big] \le C\exp\big(-n^\sigma\big) \quad (2.95)$$

for all $\lambda \ge \lambda_1$ and $n \ge n_0$. Furthermore, for those ω, λ and all E,

$$L(\lambda,E) = \inf_n L_n(\lambda,E) \ge \frac{1}{4}\log\lambda.$$

Proof. Fix $\sigma < \frac{1}{24}$ throughout the proof and let $\tau = \tau(\sigma) > 0$ be as in (2.72). Moreover, let $\lambda \ge \lambda_0 \vee n_0^B =: \lambda_1$ be as in Lemma 2.10. In this proof we shall require n_0 to be sufficiently large at various places, but of course n_0 will be assumed fixed. In view of Lemma 2.10 the hypotheses of Lemma 2.8 are satisfied with $\gamma = \gamma_0 = \frac{1}{2}$,

$$n_0^2 \le N \le n_0^5, \quad (2.96)$$

provided

$$9n_0 \ge 20\log(2n_0^{10}), \quad (2.97)$$

cf. (2.70) (recall that $S(\lambda,E) \ge 1$). It is clear that (2.97) holds if n_0 is large. Applying Lemma 2.8 one obtains (suppressing λ, E for simplicity)

$$\begin{aligned} L_N &\ge (\frac{1}{2} - \frac{1}{40})S - C_0\, SN^{-1}n_0 \ge \gamma_1 S \\ \text{and}\;\; L_N - L_{2N} &\le C_0\, SN^{-1}n_0 \le \frac{\gamma_1}{40}S \end{aligned} \quad (2.98)$$

with $\gamma_1 = \frac{1}{3}$. Moreover, with some constant $C_1 \geq 1$ depending on ε,

$$\operatorname{mes}\Big[(x, y) \in \mathbb{T}^2 \;\Big|\; \Big|\frac{1}{N} \log \|M_N(x, y; \lambda, E)\| - L_N(\lambda, E)\Big| > S(\lambda, E) N^{-\tau}\Big] \leq C_1 \exp\big(-N^\sigma\big) \quad (2.99)$$

for all N in the range given by (2.96). In particular, (2.99) implies that

$$\operatorname{mes}\Big[(x, y) \in \mathbb{T}^2 \;\Big|\; \Big|\frac{1}{N} \log \|M_N(x, y; \lambda, E)\| - L_N(\lambda, E)\Big| > S(\lambda, E) \frac{\gamma_1}{10}\Big] \leq C_1 \exp\big(-N^\sigma\big) \leq \bar{N}^{-10},$$

provided n_0 is large and

$$N^2 \leq \bar{N} \leq C_1^{\frac{1}{10}} \exp\Big(\frac{1}{10} N^\sigma\Big).$$

The first inequality was added to satisfy (2.70). In view of (2.96), one thus has the range

$$n_0^4 \leq \bar{N} \leq \exp\Big(\frac{1}{10} n_0^{5\sigma}\Big) \quad (2.100)$$

of admissible $\bar{N}$. Moreover,

$$\begin{aligned} L_{\bar{N}} &\geq \gamma_1 S - 2C_0\, S N^{-1} n_0 - C_0\, S \bar{N}^{-1} N \ \text{ and} \\ L_{\bar{N}} - L_{2\bar{N}} &\leq C_0\, S \bar{N}^{-1} N. \end{aligned} \quad (2.101)$$

At the next stage of this procedure, observe that the left end-point of the range of admissible indices starts at n_0^8, which is less than the right end-point of the range (2.100) (for n_0 large). Therefore, from this point on the ranges will overlap and cover all large integers. To ensure that the process does not terminate, simply note the rapid convergence of the series given by (2.101). □

Remark 2.12. Herman's method [8] for proving positivity of the Lyapunov exponent for potentials given by trigonometric polynomials also applies to the skew-shift. However, it is well-known that his bound only involves the coefficient of the highest frequency of the trigonometric polynomial. In particular, it does not generalize to analytic functions covered by Prop. 2.11. On the other hand, for the important example $v(x, y) = \cos(2\pi x)$, it gives the superior lower bound

$$\inf_E \; L(\lambda, E) \geq \log(\lambda/2).$$

Finally, in [2] the first author has recently shown that for this choice of v and all *sufficiently small* $\lambda > 0$ there is $\omega_0(\lambda) > 0$ and a subset $\mathcal{E}_\lambda \subset [-2, 2]$ with the property that $\operatorname{mes}([-2, 2] \setminus \mathcal{E}_\lambda) \to 0$ as $\lambda \to 0$ and such that

$$\inf_{E \in \mathcal{E}_\lambda} L(\omega, E) > 0 \ \text{ provided } \ 0 < \omega < \omega_0.$$

Here $L(\omega, E)$ denotes the Lyapunov exponent for the skew-shift $T_\omega(x, y) = (x + y, y + \omega)$. Observe that this behavior is the exact opposite of the one displayed by the well-known almost Mathieu equation as $\lambda \to 0$. The approach in [2] is based on Kotani's theorem [10, 14], Aubry-duality, and a perturbative argument for the almost Mathieu equation.

2.6. Regularity of the integrated density of states.. Let $E_{\Lambda,j}(\lambda, x, y)$, $j = 1, \ldots, b - a + 1 = |\Lambda|$ be the eigenvalues of the restriction of (2.60) to the interval $\Lambda = [a, b]$ with zero boundary conditions, $\psi(a-1) = \psi(b+1) = 0$. Consider

$$N_\Lambda(\lambda, E, x, y) = \frac{1}{|\Lambda|} \sum_j \chi_{(-\infty,E)}(E_{\Lambda,j}).$$

It is well-known that the weak limit (in the sense of measures)

$$\lim_{a\to-\infty, b\to+\infty} dN_\Lambda(\lambda, \cdot, x, y) = dN(\lambda, \cdot)$$

exists and does not depend on $(x, y) \in \mathbb{T}^2$ (up to a set of measure zero). The distribution function $N(\lambda, \cdot)$ is called the integrated density of states. It is connected with the Lyapunov exponent via the Thouless formula

$$L(\lambda, E) = \int \log|E - E'|\, dN(\lambda, E'). \tag{2.102}$$

In this subsection we show that for large λ both L and N have a modulus of continuity which is at least as good as

$$h(t) = \exp\Big(-c|\log t|^{\frac{1}{24}-}\Big). \tag{2.103}$$

This improves on various well-known continuity properties of L and N that hold for very general classes of transformations T. So far nothing better was known for the skew-shift than log-continuity, which corresponds to replacing the power of $\log t$ in (2.103) with $\log\log t$, see Figotin, Pastur [5] and the references therein.

For the proof of (2.103) we follow the approach from [4], which only requires a large deviation estimate and the avalanche principle. The latter does not depend on the transformation, and the former is given by Prop. 2.11. In particular, our assumption of large disorder is made necessary by that proposition. Since it is rather straightforward to apply the technique from [4] here, we shall be somewhat brief.

Proposition 2.13. *Let ω, v, and λ_1 be as in Prop. 2.11. For $\lambda > \lambda_1$ both $N(\lambda, E)$ and $L(\lambda, E)$ are continuous in E with modulus of continuity given by* (2.103).

Proof. We shall prove this for L. It is standard to deduce the statement about N from that on L by means of (2.102), see [7, Sect. 10]. For the sake of simplicity we shall suppress λ in the notation. Fix any positive $\sigma < \frac{1}{24}$. Let N be a large integer and set $n = \lfloor C_0(\log N)^{\frac{1}{\sigma}} \rfloor$ with some large constant C_0. One deduces from the avalanche principle and (2.95) that

$$\begin{aligned} |L_N(E) - 2L_{2n}(E) + L_n(E)| &\le \frac{Cn}{N}, \\ |L_{2N}(E) - 2L_{2n}(E) + L_n(E)| &\le \frac{Cn}{N}. \end{aligned} \tag{2.104}$$

The point is that (2.95) insures that the hypotheses (2.8) and (2.9) in Prop. 2.1 are satisfied up to a set of measure less than $CN\exp(-n^\sigma)$. This measure can therefore be made less than N^{-1} by taking C_0 large enough. Taking the difference of the two inequalities in (2.104) yields

$$|L_N(E) - L_{2N}(E)| \le \frac{Cn}{N},$$

which after summing over dyadic N gives

$$|L_N(E) - L(E)| \le \frac{Cn}{N}. \tag{2.105}$$

Inserting (2.105) into (2.104) leads to

$$|L(E) - 2L_{2n}(E) + L_n(E)| \le \frac{Cn}{N}. \tag{2.106}$$

It is clear that the derivatives of $L_{2n}(E)$ and $L_n(E)$ in E are at most of size e^{Cn}. In view of this fact (2.106) implies that for any nearby E, E',

$$|L(E) - L(E')| \le \frac{Cn}{N} + e^{Cn}|E - E'| \le C \exp\Big(-c\Big|\log|E - E'|\Big|^{\sigma}\Big), \tag{2.107}$$

if one sets $|E - E'| = \exp(-2Cn)$.

3. Localization

The purpose of this section is to show that the operator (2.60) has pure point spectrum with exponentially decaying eigen functions for most $\omega, x, y \in \mathbb{T}$ (i.e., up to a set of small measure) provided λ is sufficiently large, see Theorem 3.7 below. We will follow the scheme from [3]. The basic idea behind the proof is to start with a generalized eigen function with energy E, whose existence is guaranteed by the Shnol–Simon theorem, and then to show that it in fact decays exponentially. It is well-known that for this to hold one needs the Green's functions $G_I(x, y; E)$ on most intervals

$$I \subset \mathbb{Z} \ \text{ with } \ \mathrm{dist}(I, 0) \sim |I| \tag{3.1}$$

to possess exponential off–diagonal decay. This in turn is the case provided the monodromy matrices corresponding to those intervals I have norms which are on the order of $e^{L(E)|I|}$, $L(E)$ being the Lyapunov exponent. By the large deviation estimate (2.95), the bad set of $(x, y) \in \mathbb{T}$, where any given one of these monodromy matrices has too small norm is exponentially small in $|I|$. The difficulty that arises here is of course that the sets of bad parameters depend on E. In principle, one would therefore need to remove the union over E of all these bad sets which might amount to the entire parameter set.

The approach in [3] is to consider the set of parameters where there is some energy E with the property that, on the one hand, for some interval $J \subset \mathbb{Z}$ centered at 0 the Green's function $G_J(x, y; E)$ has very large norm and, on the other hand, the Green's function $G_I(x, y; E)$ fails to have the necessary off–diagonal decay. Here I is an arbitrary interval as in (3.1), whose length and position is related to the length of J, see the proof of Theorem 3.7 below for details. Using the large deviation theorem it is possible to show that this set of parameters has small measure, see Lemma 3.6 below. It was observed in [3] that estimating the measure of the set of parameters that produce these "double resonances" can be accomplished provided one has some control on its complexity. This can be made precise in terms of semi-algebraic sets, which we also use here. The main technical statement in this context is Lemma 3.3 below. That lemma is in turn based on a general fact about the number of lattice points that can fall into a semi-algebraic set of not too large degree and small measure, see Lemma 3.2 for the exact statement. However, the proof of Lemma 3.3 also heavily exploits the structure of the skew-shift. It remains to be seen to what extent this method applies to other transformations.

The arguments in this section do not directly invoke the lemmas from the previous section. We do, of course, use Proposition 2.11 in an essential way.

3.1. An estimate on the number of lattice points falling into a small set of bounded complexity. We begin by introducing some notation that will be used repeatedly in this section.

Definition 3.1. *For any $a, b > 0$ let $a \lesssim b$ denote $C\,a \le b$ for some absolute constant C. The case where C is very large will be written as $a \ll b$. Finally, $a \sim b$ means that both $a \lesssim b$ and $a \gtrsim b$.*

The following lemma will be important in the process of elimination of the energy. It is basically contained in Sect. 13 of [3].

Lemma 3.2. *Let $S \subset [0,1] \times [0,1]$ be an open set with the following three properties:*

$$\mathrm{mes}(S) < e^{-B^\sigma} \quad \textit{for some } \sigma > 0, \tag{3.2}$$

$$\partial S \textit{ is contained in the union of at most } B \textit{ algebraic curves } \Gamma = [P = 0] \tag{3.3}$$
$$\textit{of degree } \deg P < B,$$

$$\textit{for any line } \mathcal{L},\ S \cap \mathcal{L} \textit{ has at most } B \textit{ connected components}. \tag{3.4}$$

Suppose M and B are related by the inequalities

$$\log\log M \ll \log B \ll \log M. \tag{3.5}$$

Then

$$\#\left\{(m_1, m_2) \in \mathbb{Z}^2 \;\middle|\; |m_i| < M \textit{ and } \left(\frac{m_1}{M}, \frac{m_2}{M}\right) \in S\right\} < B^C M. \tag{3.6}$$

Furthermore, assume that

$$\#\left\{(m_1, m_2) \in \mathbb{Z}^2 \;\middle|\; |m_i| < M \textit{ and } \left(\frac{m_1}{M}, \frac{m_2}{M}\right) \in S\right\} > M^{1-10^{-7}}. \tag{3.7}$$

Then S contains a line segment $\mathcal{L}$ of length

$$|\mathcal{L}| > M^{-1+10^{-2}}$$

which is parallel to some integer vector with coordinates bounded by $M^{10^{-6}}$ and which contains a point of the form $\left(\frac{m_1}{M}, \frac{m_2}{M}\right)$.

Proof. See [3]. □

3.2. On the number of times a generic orbit of the skew-shift visits a small set of bounded complexity.

Lemma 3.3. *Denote by* $T_\omega : \mathbb{T}^2 \longrightarrow \mathbb{T}^2$ *the* ω*-skew-shift on* $\mathbb{T}^2$*. Let* $S \subset \mathbb{T}^4 \times \mathbb{R}$ *be a semi-algebraic set of degree at most* B *such that*

$$\operatorname{mes}(\operatorname{Proj}_{\mathbb{T}^4} S) < e^{-B^\sigma} \quad \text{for some} \quad \sigma > 0. \tag{3.8}$$

Under the assumption (3.5) *on* M *and* B*,*

$$\operatorname{mes}\Big[(y_0, \omega) \in \mathbb{T}^2 \;\Big|\; \big(y_0, \omega, T_\omega^j(0, y_0)\big) \in \operatorname{Proj}_{\mathbb{T}^4} S \;\text{ for some } j \sim M\Big] < M^{-10^{-8}}. \tag{3.9}$$

Proof. Let $\omega \in (0, 1)$ be fixed and choose some $y_0 \in [0, 1)$. Then there are $(x, y) \in [0, 1)^2$ such that mod $\mathbb{Z}^2$ (with $\equiv$ denoting congruence mod $\mathbb{Z}^2$)

$$\begin{aligned} (x, y) \equiv T_\omega^j(0, y_0) &\equiv \Big(jy_0 + \frac{j(j-1)}{2}\omega, y_0 + j\omega\Big) \\ &\equiv \Big(jy_0 + \frac{j-1}{2}(y - y_0 + \nu'), y_0 + j\omega\Big) \\ &\equiv \Big(\frac{j-1}{2}y + \frac{j+1}{2}y_0 + \frac{\nu}{2}, y_0 + j\omega\Big), \end{aligned} \tag{3.10}$$

where $\nu, \nu' \in \{0, 1\}$. Assume $j \sim M$. Rewriting the congruences (3.10) as equalities in $\mathbb{R}$ yields

$$\begin{cases} x = \frac{\nu}{2} + \frac{j-1}{2}y + \frac{j+1}{2}y_0 + m_1 \\ y = y_0 + j\omega + m_2 \end{cases} \tag{3.11}$$

with $|m_i| \lesssim M$. Solving (3.11) for y_0, ω one obtains

$$\begin{cases} y_0 = \frac{2}{j+1}\big(-\frac{\nu}{2} + x - \frac{j-1}{2}y - m_1\big) = \frac{2x-\nu}{j+1} - \frac{j-1}{j+1}y - \frac{2}{j+1}m_1 \\ \omega = \frac{1}{j}(y - y_0 - m_2) = \frac{\nu-2x}{j(j+1)} + \frac{2}{j+1}y + \frac{2}{j(j+1)}m_1 - \frac{m_2}{j}. \end{cases} \tag{3.12}$$

Denoting

$$\pi(S) = \operatorname{Proj}_{\mathbb{T}^4}(S)$$

we shall estimate

$$\int_{\mathbb{T}^2}\Big[\sum_{j\sim M}\chi_{\pi(S)}\big(y_0, \omega, T_\omega^j(0, y_0)\big)\Big]dy_0 d\omega. \tag{3.13}$$

Using the change of variables given by (3.12) one obtains that the integral (3.13) is no larger than

$$\begin{aligned} &\sum_{\substack{j\sim M \\ |m_i|\lesssim M}} \int \begin{vmatrix} \frac{2}{j+1} & -\frac{j-1}{j+1} \\ -\frac{2}{j(j+1)} & \frac{2}{j+1} \end{vmatrix} \chi_{\pi(S)}(y_0(x, y), \omega(x, y), x, y)\, dxdy \\ &\sim M^{-2}\int_{\mathbb{T}^2}\sum_{\substack{j\sim M \\ |m_i|\lesssim M}} \chi_{\pi(S)_{x,y}} \\ &\Big(\tfrac{2x-\nu}{j+1} - \tfrac{j-1}{j+1}y - \tfrac{2}{j+1}m_1, \tfrac{\nu-2x}{j(j+1)} + \tfrac{2}{j+1}y + \tfrac{2}{j(j+1)}m_1 - \tfrac{m_2}{j}\Big)\, dxdy. \end{aligned} \tag{3.14}$$

Here $\pi(S)_{x,y}$ denotes the slice of $\pi(S)$ for fixed (x, y). Restrict $(x, y) \in \mathbb{T}^2$ to the set where

$$\operatorname{mes}(\pi(S)_{x,y}) < e^{-\frac{1}{2}B^\sigma}. \tag{3.15}$$

By (3.8), the complementary set contributes to the integral (3.14) an amount not exceeding

$$e^{-\frac{1}{2}B^\sigma} M < e^{-\frac{1}{3}B^\sigma}. \tag{3.16}$$

For fixed x, y, the set $S_{x,y} \subset \mathbb{T}^2 \times \mathbb{R}$ is still semi-algebraic of degree at most B. Therefore, condition (3.3) of Lemma 3.2 holds for $\pi(S_{x,y}) = \pi(S)_{x,y}$, with B^C instead of B. Moreover, for any line $\mathcal{L}$ in $[0, 1]^2$

$$\pi(S)_{x,y} \cap \mathcal{L} = \pi\big(S_{x,y} \cap (\mathcal{L} \times \mathbb{R})\big)$$

has at most B^C connected components, each of which is an interval. Thus condition (3.4) holds with B replaced by B^C. Fix a point $(x, y) \in [0, 1]^2$ satisfying (3.15) and assume

$$\sum_{j\sim M;|m_i|\lesssim M} \chi_{\pi S_{x,y}}\left(\tfrac{2x-\nu}{j+1} - \tfrac{j-1}{j+1}y - \tfrac{2}{j+1}m_1, \tfrac{\nu-2x}{j(1+1)} + \tfrac{2}{j+1}y + \tfrac{2}{j(j+1)}m_1 - \tfrac{m_2}{j}\right) > \kappa M^2, \tag{3.17}$$

where

$$\kappa = M^{-10^{-7}}.$$

Fix $j \sim M$ and consider the affine transformation of $\mathbb{R}^2$

$$A(z_1, z_2) := \left(\frac{2x-\nu}{j+1} - \frac{j-1}{j+1}y - \frac{2j}{j+1}z_1, \frac{\nu-2x}{j(j+1)} + \frac{2}{j+1}y + \frac{2}{j+1}z_1 - z_2\right) \tag{3.18}$$

for which

$$|\det(DA)| = \begin{vmatrix} -\frac{2j}{j+1} & 0 \\ \frac{2}{j+1} & -1 \end{vmatrix} \sim 1.$$

Thus the set $A^{-1}\pi(S)_{x,y}$ still satisfies conditions (3.2)–(3.4) of Lemma 3.2. Therefore, in view of (3.6),

$$\sum_{|m_1|\lesssim M, |m_2|\lesssim M} \chi_{\pi(S)_{x,y}}(A(m_1/j, m_2/j)) < B^C M.$$

In conjunction with (3.17) this implies that there exists a subset $\mathcal{J} \subset \{j \sim M\}$ such that

$$\#\mathcal{J} > B^{-C}\kappa M \quad \text{and} \tag{3.19}$$

$$\sum_{|m_i|\lesssim M} \chi_{\pi(S)_{x,y}}(A(m_1/j, m_2/j)) > \frac{\kappa}{2}M = \frac{1}{2}M^{1-10^{-7}} \tag{3.20}$$

for any choice of $j \in \mathcal{J}$. In view of (3.20), condition (3.7) of Lemma 3.2 holds for the set $A^{-1}\pi(S)_{x,y}$. Hence, for any $j \in \mathcal{J}$ there exists a vector $\underline{v} \in \mathbb{Z}^2\backslash\{0\}$ such that

$$|v_1| + |v_2| < M^{10^{-6}} \tag{3.21}$$

and a lattice point $\underline{m} \in \mathbb{Z}^2$, $|\underline{m}| \lesssim M$ such that

$$P + t\underline{v} := \underline{m}/j + t\underline{v} \in A^{-1}\pi(S)_{x,y}$$

for all

$$0 < t < M^{-1+\frac{1}{200}}. \tag{3.22}$$

Applying the affine transformation A given by (3.18) yields

$$\begin{aligned}\Big(\frac{2x-\nu}{j+1} - \frac{j-1}{j+1}y - \frac{2j}{j+1}(tv_1 + P_1),\ \frac{\nu - 2x}{j(j+1)}\\ + \frac{2}{j+1}y + \frac{2}{j+1}(tv_1+P_1) - (tv_2+P_2)\Big) \in \pi(S)_{x,y}\end{aligned} \tag{3.23}$$

for all t as in (3.22). Here $\underline{v} = (v_1, v_2)$ and $P = (P_1, P_2) = \underline{m}/j$ depend on j.

Because of (3.19) and (3.21), there is a subset $\mathcal{J}' \subset \mathcal{J}$, so that

$$\#\mathcal{J}' > M^{-2.10^{-6}}\#\mathcal{J} > M^{-3.10^{-6}}M \tag{3.24}$$

and for which all choices of $j \in \mathcal{J}'$ have the same vector $\underline{v}$. We first consider the case where $\underline{v}$ lies on the line

$$v_2 = 2v_1. \tag{3.25}$$

Denoting by $\mathcal{L}^{(j)}$ the line segment given by (3.23), assume that for some choice of $j \neq j'$ in $\mathcal{J}'$,

$$\operatorname{dist}(\mathcal{L}^{(j)}, \mathcal{L}^{(j')}) < \tau.$$

Thus there exist t, t' as in (3.22) so that

$$\begin{aligned}\Big|\Big(\frac{2x-\nu}{j+1} - \frac{j-1}{j+1}y - \frac{2}{j+1}m_1 - \frac{2j}{j+1}tv_1\Big)\\ - \Big(\frac{2x-\nu}{j'+1} - \frac{j'-1}{j'+1}y - \frac{2}{j'+1}m_1' - \frac{2j'}{j'+1}t'v_1\Big)\Big| < \tau\end{aligned} \tag{3.26}$$

and

$$\begin{aligned}\Big|\Big(\frac{\nu-2x}{j(j+1)} + \frac{2}{j+1}y + \frac{2m_1}{j(j+1)} - \frac{m_2}{j} + t\Big(\frac{2v_1}{j+1} - v_2\Big)\Big)\\ - \Big(\frac{\nu-2x}{j'(j'+1)} + \frac{2}{j'+1}y + \frac{2m_1'}{j'(j'+1)} - \frac{m_2'}{j'} + t'\Big(\frac{2v_1}{j'+1} - v_2\Big)\Big)\Big| < \tau.\end{aligned} \tag{3.27}$$

Since by (3.25)

$$-\frac{2j}{j+1}v_1 = \frac{2v_1}{j+1} - v_2, \tag{3.28}$$

subtracting (3.26) from (3.27) and multiplying the resulting expression by $j(j+1)j'(j'+1)$ yields

$$\begin{aligned}&\|2jj'(j'+1)x-(j-1)jj'(j'+1)y-2jj'(j+1)x+jj'(j+1)(j'-1)y\\&\quad+2j'(j'+1)x-2jj'(j'+1)y-2j(j+1)x+2j(j+1)j'y\|\lesssim\tau M^4,\end{aligned}$$

which is the same as

$$\|2(j'-j)(1+j)(1+j')x\|\lesssim M^4\tau. \tag{3.29}$$

Here $\|\cdot\|$ denotes the distance to the nearest integer. The points $x\in\mathbb{T}$ for which (3.29) holds for an arbitrary choice of distinct $1\le j,j'\lesssim M$ form a set of measure $\lesssim M^6\tau$. Taking

$$\tau=M^{-100},$$

one concludes that the contribution of those points x to the integral (3.14) is at most

$$M^{-90}. \tag{3.30}$$

Excluding those points, one can therefore assume that for any choice of $j\neq j'$ in $\mathcal{J}'$,

$$\operatorname{dist}(\mathcal{L}^{(j)},\mathcal{L}^{(j')})>\tau, \tag{3.31}$$

where the line segments $\mathcal{L}^{(j)}\subset\pi(S)_{x,y}$. We will show that this leads to a contradiction.

For an arbitrary set $\Omega\subset\mathbb{R}^2$ denote by $\mathcal{N}(\Omega,\tau)$ the number of τ-balls needed to cover the set Ω. $\mathcal{N}$ is also referred to as “entropy”. In view of (3.31), (3.24), and the property that $|\mathcal{L}^{(j)}|>M^{-1+\frac{1}{100}}$,

$$\begin{aligned}\mathcal{N}\Big(\pi(S)_{x,y},\frac{\tau}{10}\Big)>\mathcal{N}\Big(\bigcup_{j\in\mathcal{J}'}\mathcal{L}^{(j)},\frac{\tau}{10}\Big)&\gtrsim\#\mathcal{J}'\tau^{-1}M^{-(1-\frac{1}{100})}\\&\gtrsim M^{1-3\cdot10^{-6}}\tau^{-1}M^{\frac{1}{100}-1}\\&\gtrsim M^{\frac{1}{200}}\tau^{-1}.\end{aligned} \tag{3.32}$$

On the other hand, $\pi(S)_{x,y}$ lies within a $e^{-\frac{1}{4}B^{\sigma}}$-neighborhood of at most B^C many algebraic curves Γ of degree not exceeding B^C. By our assumption (3.5), $\tau\gg e^{-\frac{1}{4}B^{\sigma}}$. Therefore,

$$\begin{aligned}\mathcal{N}(\Gamma,\tau)&\lesssim\tau^{-1}\ell(\Gamma)<B^C\tau^{-1},\\\mathcal{N}(\pi(S)_{x,y},\tau)&\lesssim B^C\tau^{-1}.\end{aligned} \tag{3.33}$$

Because of $\log M\gg\log B$ this contradicts (3.32).

It remains to consider the case where the vector $\underline{v}\in\mathbb{Z}^2\backslash\{0\}$ satisfies (3.21) but

$$v_2\neq 2v_1.$$

It follows from (3.23) that the segment $\mathcal{L}^{(j)}$ is oriented in the direction

$$\frac{\frac{2}{j+1}-\frac{v_2}{v_1}}{-\frac{2j}{j+1}}=\frac{s(j+1)-1}{j}\quad\text{where } s:=\frac{v_2}{2v_1}\neq1,\quad\text{in fact, } |s-1|\ge\frac{1}{2|v_1|}\ge\frac{1}{M}.$$

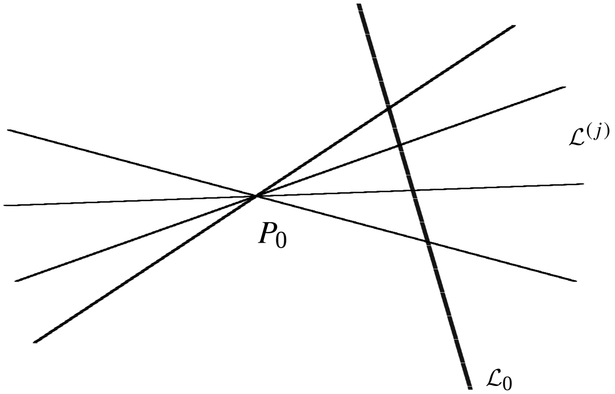

Fig. 2. The bush $\mathcal{L}^{(j)}$

Thus for any choice of $j \neq j'$,

$$\left|\frac{s(j+1)-1}{j} - \frac{s(j'+1)-1}{j'}\right| \gtrsim \frac{|1-s|}{M^2} \geq M^{-3}. \tag{3.34}$$

One now again considers the system of lines $\{\mathcal{L}^{(j)} | j \in \mathcal{J}'\}$. Let $\mathcal{L}_\tau^{(j)}$ denote a τ-neighborhood of $\mathcal{L}^{(j)}$. Then, on the one hand,

$$\int_{\mathbb{T}^2} \sum_{j\in\mathcal{J}'} \chi_{\mathcal{L}_\tau^{(j)}}\, dxdy \gtrsim \#\mathcal{J}'\, M^{-1+\frac{1}{100}}\tau \gtrsim M^{\frac{1}{200}}\tau. \tag{3.35}$$

On the other hand, since each $\mathcal{L}_\tau^{(j)}$ is contained in a τ-neighborhood of $\pi(S)_{x,y}$, (3.33) implies that

$$\int_{\mathbb{T}^2} \sum_{j\in\mathcal{J}'} \chi_{\mathcal{L}_\tau^{(j)}}\, dxdy \lesssim \Big\| \sum_{j\in\mathcal{J}'} \chi_{\mathcal{L}_\tau^{(j)}} \Big\|_\infty \tau^2 \mathcal{N}(\pi(S)_{x,y}, \tau) \lesssim \tau B^c \Big\| \sum_{j\in\mathcal{J}'} \chi_{\mathcal{L}_\tau^{(j)}} \Big\|_\infty. \tag{3.36}$$

One concludes from (3.35) and (3.36) that

$$\Big\| \sum_{j\in\mathcal{J}'} \chi_{\mathcal{L}_\tau^{(j)}} \Big\|_\infty \gtrsim M^{\frac{1}{200}} B^{-C} \gtrsim M^{\frac{1}{300}}. \tag{3.37}$$

Hence there is a subsystem $\{\mathcal{L}^{(j)} | j \in \mathcal{J}''\}$ of cardinality

$$\#\mathcal{J}'' \gtrsim M^{\frac{1}{300}}$$

such that the tubes $\{\mathcal{L}_\tau^{(j)} | j \in \mathcal{J}''\}$ have a common point P_0. It follows from (3.34) that

$$\angle(\mathcal{L}^{(j)}, \mathcal{L}^{(j')}) \gtrsim M^{-3} \text{ for any choice of } j \neq j'. \tag{3.38}$$

Choose a line $\mathcal{L}_0$ that crosses the majority of lines in the bush $\{\mathcal{L}^{(j)} | j \in \mathcal{J}''\}$ transversely. Recalling that $\pi(S)_{x,y} \cap \mathcal{L}_0$ has at most $B^C \ll M^{\frac{1}{300}}$ many components, one obtains two distinct $j, j' \in \mathcal{J}''$ for which the points

$$\mathcal{L}_0 \cap \mathcal{L}^{(j)}, \quad \mathcal{L}_0 \cap \mathcal{L}^{(j')} \in \pi(S)_{x,y}$$

belong to the same component of $\pi(S)_{x,y} \cap \mathcal{L}_0$. In view of (3.38) and (3.22) this implies that

$$\mathrm{mes}_{\mathcal{L}_0}(\pi(S)_{x,y} \cap \mathcal{L}_0) \gtrsim M^{-4}.$$

Since one can translate $\mathcal{L}_0$ by an amount M^{-1}, one finally obtains

$$\mathrm{mes}(\pi(S)_{x,y}) \gtrsim M^{-5},$$

which again contradicts (3.15). We have reached the conclusion that our assumption (3.17) fails. Recalling estimates (3.16), (3.30) on the exceptional (x, y)-sets, this implies that

$$(3.13),\ (3.14) \lesssim \frac{1}{M^2}\big(e^{-\frac{1}{3}B^\sigma} + M^{-99} + \kappa M^2\big) < 2M^{-10^{-7}},$$

which proves (3.9). □

3.3. Averaging the monodromy matrix over long orbits. For the remainder of this paper we shall assume that there is a large deviation estimate as in Prop. 2.11, without specifying λ in our notation. More precisely, we shall write the large deviation estimate in the form

$$\sup_E \ \mathrm{mes}\Big[(x, y) \in \mathbb{T}^2 \ \Big| \ \Big|\frac{1}{n} \log \|M_n(x, y; E)\| - L_n(E)\Big| > n^{-\sigma}\Big] \le C \exp\big(-n^\sigma\big). \tag{3.39}$$

By Prop. 2.11 this holds provided $\sigma > 0$ is sufficiently small and for all $n \ge n_1(\lambda, v, \varepsilon)$, where $\omega \in \Omega_\varepsilon$. Moreover, for the sake of simplicity v will be assumed to be a trigonometric polynomial. The extension to real–analytic potentials is straightforward.

Lemma 3.4. *Let T_ω be the ω-skew-shift, ω satisfying*

$$\|k\omega\| \ge c_\varepsilon\, |k|^{-1-\varepsilon} \ \textit{for all } k \in \mathbb{Z}, 0 < |k| < N. \tag{3.40}$$

Then, denoting

$$u_{N_0}(x, y) := \frac{1}{N_0} \log \Bigg\| \prod_{j=N_0}^{1} \begin{pmatrix} v(T_\omega^j(x, y)) - E & -1 \\ 1 & 0 \end{pmatrix} \Bigg\|$$

there exist constants $\sigma > 0$, $C > 1$ so that for $N > N_0^C$ one has the uniform bound

$$\Bigg\| \frac{1}{N} \sum_{j=1}^{N} \Big| u_{N_0} \circ T_\omega^j - \int_{\mathbb{T}^2} u_{N_0}(x, y)\, dx dy \Big| \ \Bigg\|_{L^\infty(\mathbb{T}^2)} < N_0^{-\sigma}. \tag{3.41}$$

Proof. By the large deviation theorem, the set

$$\Omega := \Big[(x,y) \in \mathbb{T}^2 \,\Big|\, |u_{N_0}(x,y) - \int u_{N_0}| > N_0^{-\sigma} \Big]$$

satisfies

$$\operatorname{mes}(\Omega) < e^{-N_0^\sigma}. \tag{3.42}$$

Since v is a trigonometric polynomial, Ω is clearly semi-algebraic expressed by polynomials in (x, y) of degree not exceeding N_0^C. Hence $\partial\Omega$ is contained in the union of no more than N_0^C many algebraic curves Γ of degree bounded by N_0^C. Therefore, one has the entropy bounds

$$\mathcal{N}(\Gamma, \tau) \lesssim N_0^C \tau^{-1},$$

and since, by (3.42)

$$\sup_{(x,y)\in\Omega} \operatorname{dist}((x,y), \partial\Omega) \lesssim e^{-\frac{1}{2}N_0^\sigma},$$

one also has

$$\mathcal{N}(\Omega, \tau) \lesssim N_0^C \tau^{-1}, \tag{3.43}$$

provided $\tau > e^{-\frac{1}{3}N_0^\sigma}$. It clearly suffices to prove (3.41) for $N < e^{\frac{1}{10}N_0^\sigma}$. Consider the expression

$$\begin{aligned} &\frac{1}{N^2} \sum_{j\neq j'=1}^{N} \|T^j(x,y) - T^{j'}(x,y)\|^{-2} \\ &\quad\sim \frac{1}{N^2} \sum_{j\neq j'=1}^{N} \Big[\|(j-j')y + \big(j(j-1) - j'(j'-1)\big)\omega/2\| + \|(j-j')\omega\| \Big]^{-2}, \end{aligned} \tag{3.44}$$

where $\|\cdot\|$ denotes both the natural distance on $\mathbb{T}^2$ and $\mathbb{T}$. Setting $k = j - j'$ and $\ell = j + j' - 1$, (3.44) can be rewritten in the form

$$\begin{aligned} &\frac{1}{N^2} \sum_{\substack{0<|k|\le N \\ |\ell|\le 2N}} [\|k(y + \tfrac{1}{2}\ell\omega)\| + \|k\omega\|]^{-2} \\ &\qquad \le N^{-2} \sum_{0<|k|\le N} \Big\{ \sup_z \sum_{|\ell'|\le N} [\|z + \ell' k\omega\| + \|k\omega\|]^{-2} \Big\}. \end{aligned} \tag{3.45}$$

Let

$$\|\theta\| = \|k\omega\| = \delta > N^{-1+\varepsilon}.$$

Then the inner sum in (3.45) is at most

$$\delta N^{1+\varepsilon}\sum_{s=0}^{1/\delta}(\delta^2+s^2\delta^2)^{-1}\sim\frac{N^{1+\varepsilon}}{\delta}=\frac{N^{1+\varepsilon}}{\|k\omega\|}. \tag{3.46}$$

Summing (3.46) over $0<|k|\le N$ implies that

$$(3.44), (3.45)\lesssim N^{\varepsilon}, \tag{3.47}$$

again invoking (3.40). Fixing $(x,y)\in\mathbb{T}^2$, we shall estimate $\#\mathcal{J}$, where

$$\mathcal{J}=\{j=1,\dots,N\mid T^j(x,y)\in\Omega\}.$$

Let $\tau>\frac{1}{N}$ and choose a collection of disks $\{D(P_s,\tau)|s=1,\dots,r\}$ covering Ω, where by (3.43)

$$r\lesssim N_0^C\tau^{-1}. \tag{3.48}$$

Since by (3.47)

$$N^{-2}\sum_{j\ne j'=1}^{N}\|T^j(x,y)-T^{j'}(x,y)\|^{-2}\lesssim N^{\varepsilon}, \tag{3.49}$$

we obtain in particular that

$$\tau^{-2}N^{-2}\sum_{s=1}^{r}\#\Big[j\ne j'\ \Big|\ T^j(x,y)\in D(P_s,\tau),\,T^{j'}(x,y)\in D(P_s,\tau)\Big]\lesssim N^{\varepsilon}. \tag{3.50}$$

Define for $s=1,\dots,r$,

$$\mathcal{J}_s=\{j=1,\dots,N\mid T^j(x,y)\in D(P_s,\tau)\}$$

so that $\mathcal{J}\subset\bigcup_s\mathcal{J}_s$. Clearly, (3.48) implies that

$$\#\mathcal{J}\lesssim N_0^C\tau^{-1}+\sum_{\#\mathcal{J}_s>1}\#\mathcal{J}_s. \tag{3.51}$$

Furthermore, by (3.50),

$$\sum_{\#\mathcal{J}_s>1}(\#\mathcal{J}_s)^2\lesssim\tau^2N^{2+\varepsilon}. \tag{3.52}$$

It follows from (3.51), and (3.52) that

$$\#\mathcal{J}\lesssim N_0^C\tau^{-1}+\sqrt{r}\,\tau N^{1+\varepsilon}\lesssim N_0^C\tau^{-1}+N_0^C\tau^{1/2}N^{1+\varepsilon}.$$

Optimizing in τ yields

$$\#\mathcal{J}\lesssim N_0^CN^{\frac{2}{3}+\varepsilon}. \tag{3.53}$$

Since u_{N_0} is bounded, (3.53) implies that

$$\frac{1}{N}\sum_{j=1}^{N}\Big|u_{N_0}(T^j(x,y))-\int_{\mathbb{T}^2}u_{N_0}\Big|\lesssim N_0^{-\sigma}+CN^{-1}\#\mathcal{J}\lesssim N_0^{-\sigma}+N_0^CN^{-\frac{1}{3}+\varepsilon}.$$

Inequality (3.41) follows provided $N>N_0^{C_1}$ with some large C_1. $\square$

The somewhat technical assumption (3.40), which requires only finitely many conditions on ω in terms of k, was made in order to insure that Lemma 3.3 can be applied. This will be important in the proof of localization, see Theorem 3.7 below. The previous lemma turns out to have several applications, one of which is the following uniform upper bound on the norm of the monodromy matrices.

Corollary 3.5. *Assume ω satisfies the Diophantine condition* (3.40). *For any $N > N_0^C$, there is a uniform estimate for all $E \in \mathbb{R}$,*

$$\sup_{(x,y)\in\mathbb{T}^2} \frac{1}{N} \log \|M_N(x, y; E)\| < L_{N_0}(E) + N_0^{-\sigma}. \tag{3.54}$$

3.4. Double resonances occur with small probability. Fix $\varepsilon > 0$ small and let $\omega \in \Omega_\varepsilon$, see (2.42). Since we are assuming that the disorder λ is large, Prop. 2.11 guarantees that

$$\inf_E L(E) > c_0 > 0. \tag{3.55}$$

The purpose of this subsection is to prove the following lemma, which asserts in effect that double resonances occur with small probability. An analogous statement for the shift can be found in [3]. The importance of double resonances is of course a standard fact in the theory of localization, cf. Sinai [15] and Fröhlich, Spencer, Wittwer [6]. In what follows, $H_{[-N_1,N_1]}(\omega, x, y)$ denotes the operator given by the left-hand side of (2.60) (with $T = T_\omega$) restricted to the interval $[-N_1, N_1]$ with Dirichlet boundary conditions. We shall also write $L_N(\omega, E)$ instead of $L_N(E)$ to indicate the dependence on ω.

Lemma 3.6. *Fix a small $\varepsilon > 0$. Let N be an arbitrary positive integer and let $C_2 \geq 1$ be some constant. Define $S = S_N \subset \mathbb{T}^4 \times \mathbb{R}$ to be the set of those (ω, y_0, x, y, E) for which there exists some $N_1 < N^{C_2}$ so that*

$$\|k\omega\| \geq \varepsilon\, |k|^{-1}(1 + \log k)^{-2} \ \text{ for all } \ 0 < k < N, \tag{3.56}$$

$$\left\| \big(H_{[-N_1,N_1]}(\omega, 0, y_0) - E\big)^{-1} \right\| > e^{C_3 N}, \tag{3.57}$$

$$\tfrac{1}{N} \log \|M_N(\omega, x, y, E)\| < L_N(\omega, E) - c_0/10. \tag{3.58}$$

Here c_0 is the constant from (3.55) *and C_3 will be a sufficiently large constant depending on v. Then*

$$\mathrm{mes}(\mathrm{Proj}_{\mathbb{T}^4} S) \lesssim \exp\Big(-\frac{1}{2} N^\sigma\Big). \tag{3.59}$$

Moreover, S is contained in a set S' satisfying the measure estimate (3.59) *and which is semi-algebraic of degree at most N^C for some constant C depending on v, ε, C_2 and C_3.*

Proof. Fix some sufficiently large N. Firstly, recall that the large deviation estimate (3.39) for $n = N$ holds under the condition (3.56) on ω, see Remark 2.9. Now fix some ω as in (3.56) and let $y_0 \in \mathbb{T}$ be arbitrary. If E satisfies (3.57), then by self-adjointness of H,

$$|E - E'| < e^{-C_3 N} \tag{3.60}$$

for some $E' \in \mathrm{Spec}\Big(H_{[-N_1,N_1]}(\omega, 0, y_0)\Big)$. Observe that these eigenvalues E' do not depend on (x, y). It follows from (3.60) with sufficiently large C_3 and (3.58) that

$$\tfrac{1}{N} \log \|M_N(\omega, x, y, E')\| < L_N(\omega, E') - c_0/20. \tag{3.61}$$

This can be seen by differentiating the functions on the left-hand side of (3.61) in the energy. In view of (3.39), the measure of the set of $(x, y) \in \mathbb{T}^2$ for which (3.61) holds with fixed E' does not exceed e^{-N^σ}. This proves that

$$\mathrm{mes}\big(\mathrm{Proj}_{\mathbb{T}^4} S\big) \lesssim N_1^2\, e^{-N^\sigma} \lesssim e^{-\frac{1}{2}N^\sigma}, \tag{3.62}$$

as claimed.

It remains to be shown that conditions (3.57), and (3.58) can be replaced by inequalities involving only polynomials of degree at most N^C for some C, without increasing the measure estimate (3.59) by more than a factor of two, say. We will not provide all details, since they can be readily found in [3]. Using Hilbert–Schmidt norms in (3.57) and expressing the inverse in terms of Cramer's rule shows that condition (3.57) is semi-algebraic of degree at most CN_1^3. Using Lemma 3.4, we may express the Lyapunov exponent

$$L_N(\omega, E) = \frac{1}{N} \int_{[0,1]^2} \log \|M_N(\omega, x, y, E)\|\, dxdy$$

appearing in (3.58) as a discrete average

$$L_N(\omega, E) = R^{-1} \sum_{j=1}^{R} \frac{1}{N} \log \|M_N(\omega, T_\omega^j(0,0), E)\| + o(1)$$

with $R < N^C$. Therefore, one obtains a semi-algebraic condition in ω, x, y, E of degree at most N^C by rewriting (3.58) in the form

$$\|M_N(\omega, x, y; E)\|^{2R} \le e^{-NRc_0/10} \prod_{j=1}^{R} \|M_N(\omega, T_\omega^j(0,0); E)\|^2.$$

Finally, the measure of the set S does not change by more than a factor in this process.
□

3.5. The proof of localization for the skew-shift with large disorder. The following theorem is the main result of this section.

Theorem 3.7. *Fix $\varepsilon > 0$ small. Let $v = v(x, y)$ be a nonconstant trigonometric polynomial on $\mathbb{T}^2$ and let $\lambda_1 = \lambda_1(v, \varepsilon)$ be as in Prop. 2.11. Let $T_\omega(x, y) = (x + y, y + \omega) \pmod{\mathbb{Z}^2}$ denote the ω-skew-shift on $\mathbb{T}^2$. Then for every $\lambda > \lambda_1$ and all $(\omega, x, y) \in \mathbb{T}^3$ up to a set of measure ε, the operator*

$$\big(H_{\omega,(x,y)}\psi\big)_n := -\psi_{n-1} - \psi_{n+1} + \lambda v(T_\omega^n(x, y))\psi_n \ \text{ on } \ell^2(\mathbb{Z})$$

displays Anderson localization for all energies.

Proof. Let $\omega \in \Omega_\varepsilon$, see (2.42). For large N, let S_N be as in Lemma 3.6. Then Lemma 3.3 applies to S_N and setting $\bar{N} = e^{(\log N)^2}$ it follows that

$$\operatorname{mes}\Big[(y_0, \omega) \in \mathbb{T}^2 \;\Big|\; (y_0, \omega, T_\omega^j(0, y_0)) \in \operatorname{Proj}_{\mathbb{T}^4}(S_N) \;\text{ for some }\; j \sim \bar{N}\Big] < \bar{N}^{-10^{-8}}. \tag{3.63}$$

Let $\mathcal{B}_N$ denote the set on the left-hand side of (3.63) and define

$$\mathcal{B}^{(0)} := \limsup_{N\to\infty} \mathcal{B}_N.$$

Thus $\operatorname{mes}(\mathcal{B}^{(0)}) = 0$. Since $T^\ell(x, y) = x + T^\ell(0, y) \pmod 1$, this construction applied to the potential $v(x + \cdot, \cdot)$ instead of v produces a set $\mathcal{B}^{(x)}$ of measure zero. Finally, set

$$\mathcal{B} := \Big\{(\omega, x, y) \;\Big|\; (y, \omega) \in \mathcal{B}^{(x)}\Big\},$$

which is again of measure zero. It is for all $(\omega, x, y) \in \Omega_\varepsilon \times \mathbb{T}^2 \setminus \mathcal{B}$ that we shall prove localization.

Fix such a choice of (ω, x, y) and any $E \in \operatorname{Spec}\big(H_{\omega,(x,y)}\big)$. By the Shnol–Simon theorem [12, 13] there exists a generalized eigenfunction ξ, i.e.,

$$(H_{\omega,(x,y)} - E)\xi = 0 \;\text{ and }\; |\xi_n| \lesssim 1 + |n| \;\text{ for all } n \in \mathbb{Z}. \tag{3.64}$$

Furthermore, we normalize $|\xi_0| + |\xi_1| = 1$. Fix some large integer N and assume that (3.57) holds. By our choice of (ω, x, y),

$$\frac{1}{N'} \log \|M_{N'}\big(T_\omega^j(x, y); E\big)\| > L(E) - c_0/10$$

for all $N' \sim N$ and $j \sim \bar{N} = e^{(\log N)^2}$, cf. (3.58). It follows from the avalanche principle that then also

$$\frac{1}{N_2} \log \|M_{N_2}\big(T_\omega^j(x, y); E\big)\| > L(E) - c_0/10 \qquad \text{if} \qquad \frac{\bar{N}}{2} < |j| < \bar{N} \;\text{ and }\; N^2 < N_2 < \frac{\bar{N}}{10}. \tag{3.65}$$

As usual, let

$$G_\Lambda(\omega, x, y; E) := \Big(H_\Lambda(\omega, x, y) - E\Big)^{-1}$$

be the Green's function. As before, H_Λ denotes the restriction of H to the interval Λ with Dirichlet boundary conditions. Consider intervals

$$\Lambda = \Big[j, j + \frac{\bar{N}}{10}\Big], \;\text{ where }\; \frac{\bar{N}}{2} < |j| < \bar{N}.$$

By definition of G_Λ and because of (3.64), it will suffice to prove that

$$\max_{\ell\in\partial\Lambda}\Big|G_\Lambda(\omega, x, y; E)(k, \ell)\Big| \lesssim \exp(-c_1\bar{N}) \;\text{ for all } k \in \Lambda \;\text{ with }\; \operatorname{dist}(k, \partial\Lambda) > \frac{1}{4}|\Lambda|. \tag{3.66}$$

Here $c_1 > 0$ is some fixed constant. The proof of (3.66) follows from (3.65) by a standard argument. In fact, it is a simple consequence of Cramer's rule and the representation of the Hamiltonian as the matrix appearing on the right-hand side of (2.80) that for any n and $1 \leq k, \ell \leq n$,

$$G_{[1,n]}(x, y; E)(k, \ell) = \frac{f_{k-1}(x, y; E) f_{n-\ell-1}(T^{\ell}(x, y); E)}{f_n(x, y; E)}.$$

In conjunction with (2.81), Corollary 3.5, and (3.65), this implies (3.66) as desired. Recall, however, that we made the assumption that (3.57) holds. To establish this condition it suffices to show that

$$|\xi_{N_1+1}| + |\xi_{-N_1-1}| \lesssim e^{-2C_3 N}$$

for some $N_1 \sim N^{C_2}$. In view of (3.64) this estimate holds provided both Green's functions

$$G_{[j-4C_3N,\, j+4C_3N]}(\omega, x, y; E) = G_{[-4C_3N, 4C_3N]}\big(\omega, T^j(x, y); E\big) \text{ with } j = N_1, -N_1$$

satisfy an exponential decay estimate as in (3.66). In view of the preceding argument involving (3.66) it remains to show that for some $j \sim N^{C_2}$ one has the property

$$\frac{1}{4C_3N} \log \|M_{4C_3N}\big(T_{\omega}^{j}(x, y), E\big)\| > L(E) - c_0/10$$

and similarly for $-j$. That, however, is an immediate consequence of Lemma 3.4. □

Acknowledgement. The authors thank Thomas Spencer and Yakov Sinai for helpful discussions. The third author wishes to thank Thomas Wolff for the suggestion that the Hilbert transform might lead to better bounds in the paper [7], and for pointing out Theorem 1.9 in [9], and Charles Fefferman for sketching to him the proof of an important BMO estimate for logarithms of polynomials. The second author was supported by a grant of the NEC Research Institute, Inc. during his stay at the Institute for Advanced Study, Princeton. The third author was supported in part by the NSF, grant number DMS-9706889.

References

1. Anderson, P.: Absence of diffusion in certain random lattices. Phys. Rev. **109**, 1492–1501 (1958)
2. Bourgain, J.: Positive Lyapunov exponents for most energies. *Geometric Aspects of Functional Analysis*. Lecture Note sin Math. **1745**. Berlin–Heidelberg–New York: Springer, 2000, pp. 37–66
3. Bourgain, J., Goldstein, M.: On nonperturbative localization with quasiperiodic potential. Annals of Math. **152**, (3), 835–879 (2000)
4. Bourgain, J., Schlag, W.: Anderson localization for Schrödinger operators on $\mathbb{Z}$ with strongly mixing potentials. Commun. Math. Phys. **215**, 143–175 (2000)
5. Figotin, A., Pastur, L.: *Spectra of random and almost–periodic operators*. Grundlehren der mathematischen Wissenschaften **297**, Berlin–Heidelberg–New York: Springer, 1992
6. Fröhlich, J., Spencer, T., Wittwer, P.: Localization for a class of one dimensional quasi-periodic Schrödinger operators. Commun. Math. Phys. **132**, 5–25 (1990)
7. Goldstein, M., Schlag, W.: Hölder continuity of the integrated density of states for quasiperiodic Schrödinger equations and averages of shifts of subharmonic functions. To appear in Annals of Math.
8. Herman, M.: Une méthode pour minorer les exposants de Lyapounov et quelques exemples montrant le charactère local d'un theoreme d'Arnold et de Moser sur le tore de dimension 2. Comment. Math. Helv. **58** no. 3, 453–502 (1983)
9. Katznelson, Y.: *An introduction to harmonic analysis*. New York: Dover, 1976
10. Kotani, S.: Ljapunov indices determine absolutely continuous spectra of stationary random one-dimensional Schrödinger operators. In: *Stochastic analysis (Kotata/Kyoto, 1982)*, North-Holland Math. Library, **32**, Amsterdam–New York: North-Holland, 1984, pp. 225–247

11. Montgomery, H.: *Ten lectures on the interface between harmonic analysis and analytic number theory*. CBMS Regional conference series in mathematics **84**, Providence, RI: AMS, 1994
12. Shnol, I.: On the behavior of the Schroedinger equation. Mat. Sb. 273–286 (1957) Russian
13. Simon, B.: Spectrum and continuum eigenfunctions of Schroedinger Operators. J. Funct. Anal. **42**, 66–83 (1981)
14. Simon, B.: Kotani theory for one-dimensional stochastic Jacobi matrices. Commun. Math. Phys. **89**, 227–234 (1983)
15. Sinai, Y.G.: Anderson localization for one-dimensional difference Schrödinger operator with quasi-periodic potential. J. Stat. Phys. **46**, 861–909 (1987)
16. Stein, E.: *Harmonic analysis*. Princeton Mathematical Series **43**, Princeton, NJ: Princeton University Press, 1993

Communicated by P. Sarnak