Probab. Theory Relat. Fields (2008) 142:443–473
DOI 10.1007/s00440-007-0112-z

Shannon–McMillan theorems for discrete random fields along curves and lower bounds for surface-order large deviations

Julia Brettschneider

Received: 26 March 2006 / Revised: 10 October 2007 / Published online: 28 November 2007

Abstract The notion of a surface-order specific entropy $h_c(P)$ of a two-dimensional discrete random field P along a curve c is introduced as the limit of rescaled entropies along lattice approximations of the blowups of c. Existence is shown by proving a corresponding Shannon–McMillan theorem. We obtain a representation of $h_c(P)$ as a mixture of specific entropies along the tangent lines of c. As an application, the specific entropy along curves is used to refine Föllmer and Ort's lower bound for the large deviations of the empirical field of an attractive Gibbs measure from its ergodic behaviour in the phase-transition regime.

Mathematics Subject Classification (2000) Primary: 60F10; Secondary: 60G60 · 94A17 · 82B26 · 82B20

1 Introduction

The entropy of a stationary random field P indexed by a lattice is usually defined as a limit of entropies on an increasing sequence of boxes, rescaled by the volume of the boxes. Shannon–McMillan theorems describe this convergence on a deeper level, as $\mathcal{L}^1$-convergence of rescaled information quantities. In the context of large deviations for Gibbs measures the volume-order entropies may not provide enough information when a phase transition occurs. Instead, Föllmer and Ort [9] introduced the concept of surface-order entropy on boxes, proved a corresponding version of the

J. Brettschneider (✉)
Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK
e-mail: julia.brettschneider@warwick.ac.uk

J. Brettschneider
Department of Community Health and Epidemiology and Cancer Research Institute Division of Cancer Care and Epidemiology, Queen's University, Kingston, ON, K7L 3N6, Canada

Shannon–McMillan theorem and used it to estimate large-deviation probabilities. The construction of the Wulff shape by Dobrushin et al. [4] suggests that such estimates can be improved if boxes are replaced by more general shapes.

We investigate the problem of constructing entropies on general surfaces, proving appropriate versions of the Shannon–McMillan theorem, and using these constructions to refine the large deviation lower bound. We carry out this program in the two-dimensional case, where surfaces reduce to contour curves. The existence of a surface-order specific entropy does not simply follow from a subadditivity argument. Instead, we prove a corresponding Shannon–McMillan theorem, that is, $L^1(P)$-convergence of rescaled information quantities along lattice approximations of the successive blowups of the curve c. This is accomplished in three steps. The first main result is a Shannon–McMillan theorem along lines; cf. Theorem 3.5 for rational slopes and Theorem 3.6 for irrational slopes. The proof relies on *uniform* convergence in ergodic theorems for a suitable skew product transformation. The second step is to extend the results to polygons under the assumption of a strong 0-1 law on P. Finally, we obtain an explicit formula for the specific entropy $h_c(P)$ as a mixture of the conditional entropies of the random field restricted to the origin, given suitably defined *past*-σ-algebras along tangent lines of the curve (Theorem 4.3). Under certain conditions, this construction can be extended to relative entropies of one random field with respect to another. This will be the key to our proof of a refined lower bound for large deviations in the phase transition regime (Theorem 6.2).

The role of Shannon–McMillan theorems in the refined analysis of large deviations provided the original motivation for this work. It seems, however, that the study of entropies along surfaces may hold independent interest. This paper lays some of the groundwork for such a general theory of specific entropies along shapes. The key to this theory is a careful combination of probabilistic arguments and non-standard discrete geometrical constructions. We now explain our approach and our results in more detail.

Shannon–McMillan theorems and entropy. Consider a random sequence ω of letters from a finite alphabet Υ, modelled by a probability measure P on the space $\Omega := \Upsilon^{\{1,2,\dots\}}$. For any finite n, the *information* provided by the first n letters can be described by the function $-\log P[\,\omega_{\{1,\dots,n\}}]$, where $\omega_{\{1,\dots,n\}}$ denotes the restriction of ω to $\{1,\dots,n\}$, and $P\,[(y_1,\dots,y_n)]$ is the probability that the pattern $(y_1,\dots,y_n)$ appears in the first n trials. In the classical case, when the letters are independent and identically distributed according to a measure μ, the classical Shannon–McMillan theorem states that the rescaled information functions $-n^{-1}\log P[\,\omega_{\{1,\dots,n\}}]$, converge in $\mathcal{L}^1(P)$ to the *entropy* $H(\mu) := -\sum_{y\in\Upsilon}\mu(y)\log\mu(y)$ of the measure μ. The theorem can be extended to a general ergodic sequence. In the bilateral case, when P is an ergodic measure on $\Upsilon^{\mathbb{Z}}$, the limiting quantity takes the form $h(P) := E\,[H\,(P_0[\,\cdot\,|\,\mathcal{P}](\omega))]$, where $P_0[\,\cdot\,|\,\mathcal{P}]$ is the distribution of the random field in $(0,0)$ conditioned on the "past" σ-algebra $\mathcal{P}$ generated by the projection of ω to the set $\{-1,-2,\dots\}$.

These constructions can be extended to a spatial setting when the random field is given by a stationary probability measure P on a configuration space $\Upsilon^{\mathbb{Z}^d}$. Thouvenot [27] and Föllmer [6] proved spatial extensions of Shannon and McMillan's result. The *specific entropy* $h(P)$ is introduced as the limit of $|V_n|^{-1}H_{V_n}(P)$, where V_n is the set of all lattice sites in the box $[-n, n]^d$, and $H_{V_n}(P)$ is the entropy of the measure P restricted to V_n. The existence of the limit follows from the subadditivity of H_V with respect to V. What is more, the corresponding Shannon–McMillan theorem shows $\mathcal{L}^1(P)$-convergence of the rescaled information functions $-|V_n|^{-1} \log P[\omega_{V_n}]$. If P is ergodic we obtain the formula $h(P) = E\left[H\left(P_0[\,\cdot\,|\,\mathcal{P}^d](\omega)\right)\right]$, where $\mathcal{P}^d$ is a σ-algebra representing a spatial version of the "past". More precisely, $\mathcal{P}^d$ is generated by the projections of ω to the sites preceding the origin in the lexicographical ordering of $\mathbb{Z}^d$.

Surface entropy. Our goal is to derive refined versions of the Shannon–McMillan theorem, where the information functions are observed along surfaces. This was carried out in [9] for the surfaces of boxes parallel to the axes. In this paper, we focus on the two-dimensional case. We develop a construction of surface entropy where rectangles are replaced by general curves. More precisely, following a suggestion by Hans Föllmer, we introduce the specific entropy along sets generated by lattice approximations to lines, and then extend this to polygons and piecewise smooth curves.

Most of the work here goes into our first result, a Shannon–McMillan theorem for the *specific entropy* $h_\lambda(P)$ of a stationary random field P *along a line* (Theorem 3.5 if the slope is rational and Theorem 3.6 if the slope is irrational). We prove the $\mathcal{L}^1(P)$-convergence of the rescaled information functions along increasing segments of the line's *lattice approximation* $\{(z, [\lambda z + a]) \mid z \in \mathbb{Z}\}$, where $[x]$ denotes the integer part of x. If P fulfills a 0-1 law on the tail field, we obtain the formula

$$h_\lambda(P) = \int_0^1 E\left[H\left(P_0[\,\cdot\,|\,\mathcal{P}_{\lambda,t}](\omega)\right)\right] dt,$$

where $\mathcal{P}_{\lambda,t}$ is the σ-algebra generated by those approximating sites which precede 0 in the lexicographical ordering of $\mathbb{Z}^2$. If λ is rational, it suffices for P to be ergodic. Furthermore, the specific entropy along the line can be written as

$$h_\lambda(P) = \frac{1}{q} \sum_{\nu=0}^{q-1} E\left[H\left(P_0\left[\,\cdot\,|\mathcal{P}_{p/q,\,\nu p/q}\right](\omega)\right)\right],$$

where p/q is the unique representation of λ by integers $p \in \mathbb{N}$ and $q \in \mathbb{Z}$ having no common divisor. The past σ-algebras $\mathcal{P}_{p/q,\,\nu p/q}$ correspond to the q different possibilities to start the q-periodic pattern of the lattice approximation.

The idea to investigate such a specific entropy along lines has two precursors. The first is a volume-order *directional entropy*, which Milnor [16,17] introduced in the context of cellular automata. The second is the specific entropy along hyperplanes perpendicular to an axis, which was defined by [9] as a step toward their surface entropy along boxes. It may be noted that a distinction between rational and irrational slopes was also made by Sinai in his work [25] on Milnor's directional entropy for cellular automata. This construction was further developed by Park [19–21] and Sinai [26]. The original problem of continuity with respect to the direction was finally solved in Park [22].

Our next result is a Shannon–McMillan theorem along polygons (Theorem 4.2). In particular, we obtain a representation of the *specific entropy* of P *along a polygon* as a mixture of entropies along lines corresponding to its edges. The extension to polygons requires a strong form of the 0-1 law on the tail field, which was introduced in [9]. It says that, for any subset J of $\mathbb{Z}^2$, the σ-algebra generated by the sites in J does not increase if we add information about the tail behaviour outside of J; cf. Definition 2.2. Finally, we come to the main result in this section. Theorem 4.3 says that the *specific entropy along a curve* $c : [0, T] \longrightarrow \mathbb{R}^2$ is a mixture of entropies along the tangent lines:

$$h_c(P) = \int_0^T h_{c'(t)}(P)\, dt \tag{1}$$

Here, $h_{c'(t)}(P)$ denotes the specific entropy along a line having the same slope as the tangent of c in t; cf. (32) for the exact definition.

About the proofs of the Shannon–McMillan theorems. We will proceed in several steps, first for lines, then for polygons and finally for curves.

(i) *Lines:* The lattice approximation of a line with slope $\lambda \in [0, 1]$ and y-intercept a on an interval $[0, n]$ is defined by $L_{\lambda,a}(I) = \{(z, [\lambda z + a]) \mid z = 1, \ldots, n\}$. We want to prove the convergence of the rescaled information content $-(n+1)^{-1} \log P\left[\omega_{L_{\lambda,a}([0,n])}\right]$ along successively larger segments of the line. To make this problem accessible to ergodic theory, we have to find a transformation which captures the stair climbing pattern along the lattice approximation of the line. If the slope is rational, the steps become periodic, and we proceed by combining a finite number of different transformations. In the case of an irrational slope, no such simplification is possible. We need to keep track not only of the integer part but also of the fractional part $\{\lambda z + a\}$ in each step. We can realize this by introducing the skew product transformation

$$\begin{aligned} S_\lambda : \mathbb{T} \times \Omega &\longrightarrow \mathbb{T} \times \Omega \\ (t, \omega) &\longmapsto \left(\tau_\lambda(t), \vartheta_{(1,[t+\lambda])}\omega\right), \end{aligned}$$

where $\mathbb{T}$ is the one-dimensional torus, equipped with the Borel σ-algebra and the Haar measure, and τ_λ is the translation by λ. Using appropriate ergodic theorems on the product space with the skew product we obtain a Shannon–McMillan theorem along the lattice approximation of the line. In view of the extension of this result to polygons, we further need a variation of this result. Instead of lattice approximations of increasing parts of a line, we use lattice approximations of blow-ups of a line segment.

(ii) *Polygons:* Consider a polygon π, parametrized on $[0, T]$, and its blowups $B_n\pi(t) = n\pi\,(t/n)$, parametrized on $[0, nT]$. We study the sequence $-\log P\left[\omega_{L_n^\pi}\right]$ ($n \in \mathbb{N}$) of rescaled information of P restricted to its lattice approximations L_n^π. Conditioning site by site, the problem can be reduced to the Shannon–McMillan theorem along the edges, which is essentially covered by the Shannon–McMillan theorem along lines that we already established. One difficulty remains, which is getting around the corners. It can be overcome by the technique which [9] used in the case of boxes. We need the additional assumption of a strong form of a 0-1 law

(Definition 2.2). Under this condition, the entropy along a polygon is represented as a mixture of the entropies of its edges (Theorem 4.2).

(iii) *Curves:* Our last step is the entropy along a piecewise smooth curve. After having established the results for lines and polygons, this part has become easy. We obtain (1) by approximating the curve with polygons.

Relative entropy and large deviations. Shannon–McMillan theorems for the *specific relative entropy* $h(Q, P)$ of two probability measures Q and P on the sequence space $\Upsilon^{\{1,2,\dots\}}$ are based on the functions $-\log(dQ/dP)\,[\,\omega_{\{1,\dots,n\}}]\ (n \in \mathbb{N})$ describing the *relative information* gained from the first n trials of an experiment. They are a key tool in the search for estimates in the theory of large deviations. By a *large deviation* we mean a rare event, or an untypical behavior occurring in a random sequence. Consider the *empirical distributions* $\mu_n(\omega) := n^{-1}\sum_{i=1}^{n}\delta_{\omega_i}$ $(n \in \mathbb{N})$ of a stationary random sequence ω_i $(i \in \mathbb{N})$. If the measure P is ergodic then μ_n converges to the marginal distribution μ of P, P-almost surely and in $\mathcal{L}^1(P)$. Large deviations are events like $[\mu_n \in A]$, A being a set in the space of probability measures on Υ whose closure does not contain μ (cf. for instance [3]).

The aim of large deviation theory is to find lower and upper bounds which describe the asymptotic decay of the probabilities of such large deviations. In the classical case of a sequence of independent and identically distributed random variables, the decay of large deviations of the empirical distribution is described by Sanov's theorem. Cramèr's theorem addresses similar questions for the empirical averages. As a third level for investigating large deviations, Donsker and Varadhan [5] initiated the investigation of large deviations of empirical processes.

We replace the random sequence by a random field, and the empirical processes by the *empirical fields* $R_n(\omega) := |V_n|^{-1}\sum_{i\in V_n}\delta_{\vartheta_i\omega}$ $(n \in \mathbb{N})$, where ϑ_i $(i \in \mathbb{Z}^d)$ denotes the group the shift transformations. Comets [2], Föllmer and Orey [8], and Olla [18] found the following large deviation principle for the empirical fields of a stationary Gibbs measure: For any open subset A of the space $\mathcal{M}_1(\Omega)$ of probability measures on $\Omega = \Upsilon^{\mathbb{Z}^d}$,

$$\liminf_{n\to\infty}\frac{1}{|V_n|}\log P[R_n \in A] \ge -\inf_{Q\in A}\ h(Q, P), \tag{2}$$

and for any closed set $C \in \mathcal{M}_1(\Omega)$,

$$\limsup_{n\to\infty}\frac{1}{|V_n|}\log P[R_n \in C] \le -\inf_{Q\in C}\ h(Q, P), \tag{3}$$

where the *rate function* is based on the specific relative entropy $h(Q, P) := \lim_{n\to\infty} \mid V_n \mid^{-1} H_{V_n}(Q, P)$.

Large deviations in the phase transition regime. In the case of phase transition, there exists more than one Gibbs measure with respect to the same potential. Due to the variational principle for Gibbs measures (cf. Lanford and Ruelle [15] and, in purely information theoretical terms, Föllmer [6]), the specific relative entropy of P to another stationary Gibbs measure Q with the same interaction potential vanishes. Thus, the relative entropy $h(Q, P)$ appearing in (2) and (3) may be zero even though Q is not contained in the closure of A. This suggests we need a refined analysis of large deviations in terms of surface-order rather than volume-order entropies. Assume in fact that the interaction satisfies the local Markov property. Then $H_V(Q, P) = H_{\partial V}(Q, P)$ for any finite subset V of $\mathbb{Z}^2$, where ∂V is the boundary of V, i.e., the set of all sites outside of V which have distance 1 to V (cf. the end of Sect. 2 in [9]). Consequently, this relative entropy is in fact a surface-order term, and so it should be rescaled not by the size of the volume $|V|$ but by the size of the surface $|\partial V|$. This observation was the main motivation for introducing the concept of surface entropy, and for proving the corresponding Shannon–McMillan theorem.

In the context of the two-dimensional Ising model, Schonmann [24] showed the existence of surface-order upper and lower bounds for the large deviations of the empirical means. For attractive models with a totally ordered state space, Föllmer and Ort [9] found a lower bound for the large deviations of the empirical field in terms of the relative surface entropies along boxes (recalled as Theorem 6.1 in this paper). In their detailed analysis of the two-dimensional Ising model, Dobrushin et al. [4] justified on the basis of local interactions, that the phase-separating curve has the form of a Wulff shape. They proved a large deviation principle with a rate function in terms of the surface tension along the Wulff shape. Using different methods, Ioffe [11,12] was able to extend their result up to the critical temperature. The appearance of such shapes suggests to revisit the approach of [9] but using the generalized surface entropies introduced in Sect. 4 instead of the surface entropies based on boxes.

This extension is carried out in the last section of this work, for the case of Gibbs measures with attractive interactions on a two-dimensional lattice. We study the large deviations for its empirical measure. The main result of this part of the paper is Theorem 6.2, which provides a lower bound in terms of the specific relative entropies along curves. One of the ingredients in the proof is the well known strategy of switching to a measure under which the large deviation becomes normal behavior, and then applying a Shannon–McMillan theorem. Making use of the global Markov property, we pass from densities restricted to the lattice points inside of a polygon to densities on the lattice approximations of its boundary. In this context, we prove an appropriate *relative* version of the Shannon–McMillan theorem, in analogy to the results in Sect. 4. Other major ingredients in the proof are geometrical observations concerning the interplay of the random field and the lattice approximations of curves. In particular, in Lemma 6.5 we compute the asymptotic ratio of the length of a line segment and its lattice approximation. These quantities merge into a factor in the lower bound in Theorem 6.2 involving the derivative of the curve. We will further touch on the case when the Markov property is only satisfied with respect to a boundary that is a *contour* in the sense of statistical mechanics.

Outline of the paper. The first section reviews some basic notions and properties of discrete random fields, information and entropy. In Sect. 2, we introduce a specific entropy of a random field along a line. In Sect. 4 we construct the specific entropy of a random field along a curve proceeding in three steps: line segments, polygons, curves. Section 5 recalls some basic notions about Gibbs measures and phase transitions. In Sect. 6 we prove the refined large deviations lower bound. The case of *contour* boundaries is briefly discussed in Sects. 4.4 and 6.

2 Random fields

Consider $\Omega := \Upsilon^{\mathbb{Z}^d}$, where Υ is a finite set. For any subset V of $\mathbb{Z}^d$ define $\Omega_V := \Upsilon^V$. Let ω_V be the projection of ω to V, P_V the distribution of ω_V with respect to P, and $\mathcal{F}_V := \sigma(\omega_V)$ the σ-algebra generated by this projection. A probability measure P on $(\Omega, \mathcal{F})$ is called a two-dimensional discrete *random field.* The transformations $(\theta_v)_{v\in\mathbb{Z}^d}$ defined by $\theta_v\omega(u) = \omega(u+v)$ $(u \in \mathbb{Z}^d)$ form a group of transformations on Ω, called *shift transformations*. We assume P is *stationary,* that is, invariant with respect to the shift transformations. There are different levels of Markov properties for random fields: when the subset of the lattice which generates the condition has to be *finite,* and when it can be any type of subset of the lattice. They both involve the *boundary* $\partial V := \{\, j \in \mathbb{Z}^d \backslash V \mid \mathrm{dist}_V(j) = 1\,\}$ of a subset V of the lattice $\mathbb{Z}^d$.

Definition 2.1 A random field P has the *local Markov property* if, for any finite $V \subset \mathbb{Z}^d$ and for any nonnegative $\mathcal{F}_V$-measurable ϕ, $E[\phi \mid \mathcal{F}_{\mathbb{Z}^d\backslash V}] = E[\phi \mid \mathcal{F}_{\partial V}]$. A random field P which fulfills the local Markov property is called a Markov field. If the local Markov property holds for all any $V \subset \mathbb{Z}^d$, then P has the *global Markov property.*

In Sect. 5, we will introduce the class of Gibbs measures in terms of interaction potentials. Any Gibbs measure belonging to a nearest-neighbor potential is a Markov field. Examples of random fields which have the local Markov property but not the global Markov property were given by Weizsäcker [28] and Israel [13]. P is called *tail-trivial* if it fulfills a 0-1 law on the *tail field*

$$\mathcal{T} := \bigcap_{V\subset\mathbb{Z}^d\ \text{finite}} \mathcal{F}_{\mathbb{Z}^d\backslash V} = \bigcap_{n\in\mathbb{N}} \mathcal{F}_{\mathbb{Z}^d\backslash V_n}, \tag{4}$$

where $V_n := ([-n, n] \cap \mathbb{Z})^d$. Due to the spatial structure of a random field, tail-triviality is equivalent (cf. Proposition 7.9 from [10]) to a mixing condition called *short-range correlations*:

$$\sup_{A\in\mathcal{F}_{Z^d\backslash V_n}} |\, P(A\cap B) - P(A)P(B)\,| \xrightarrow{n\to\infty} 0 \quad \text{for all } B \in \mathcal{F}.$$

The following strong version of a 0-1 law was introduced by Föllmer and Ort [9]. For $J = \emptyset$ it reduces to the classical 0-1 law on $\mathcal{F}$. Remark 3.2.3 from [9] shows that it implies the global Markov property provided P has the local Markov property.

Definition 2.2 P satisfies the strong 0-1 law if for any subset J of $\mathbb{Z}^d$ the σ-algebra $\mathcal{F}_J$ coincides modulo P with the σ-algebra

$$\mathcal{F}_J^* := \bigcap_{V\subset\mathbb{Z}^d\ \text{finite}} \mathcal{F}_{J\cup(\mathbb{Z}^d\backslash V)}.$$

Let V and W be subsets of $\mathbb{Z}^d$. The *information* in ω restricted to V, with respect to P, is given by the random variable $\mathcal{I}(P_V)(\omega) := -\log P[\omega_V]$, and the *information* *conditioned* on $\mathcal{F}_W$ is defined as $\mathcal{I}\,(P_V[\,\cdot\,|\mathcal{F}_W](\omega)) := -\log P[\omega_V\,|\,\omega_W]$. The *entropy* of P restricted to V is defined as

$$H_V(P) := E[\mathcal{I}(P_V)(\omega)] = -\sum_{\omega\in\Upsilon^{\mathbb{Z}^d}} P[\omega_V]\log P[\omega_V],$$

and the *conditional entropy* of P to $\mathcal{F}_W$ is $H_V(P[\,\cdot\,|\mathcal{F}_W]) := E[-\log P[\omega_V\,|\,\omega_W]] = H(P_V[\,\cdot\,|\mathcal{F}_W])$. The *specific entropy* $h(P)$ of P is defined as the limit of $|\,V_n|^{-1}H(P_{V_n})$ for n to infinity. Its existence can be proved by a subadditivity property (cf. for instance, Theorem 15.12 in [10]), but it also follows from a Shannon–McMillan theorem by Föllmer [6] and Thouvenot [27]. They showed that the specific entropy for ergodic P is $E\,[H\,(P_0[\,\cdot\,|\mathcal{P}](\omega))]$, where $P_0[\,\cdot\,|\mathcal{P}](\omega)$ is the conditional distribution of the random field in the origin with respect to the σ-algebra $\mathcal{P}$ generated by all sites which are smaller than the origin with respect to the lexicographical ordering on $\mathbb{Z}^d$. Moreover, they showed that for all stationary P

$$\frac{1}{|\,V_n|}\,\mathcal{I}(P_{V_n}) \xrightarrow{n\to\infty} E\left[H\,(P_0[\,\cdot\,|\mathcal{P}](\omega))\,\middle|\mathcal{J}\right] \quad \text{in } \mathcal{L}^1(P), \tag{5}$$

where $\mathcal{J}$ is the σ-algebra of all sets invariant with respect to the transformations θ_v $(v\in\mathbb{Z}^d)$.

A lemma from [9] will be used in some of the proofs in this paper. We recall it here for the reader's convenience.

Lemma 2.3 *Consider σ-algebras $\mathcal{B}_i \subseteq \mathcal{B}_i^*$ $(i\in\mathbb{N})$ increasing to $\mathcal{B}_\infty$, respectively decreasing to $\mathcal{B}_\infty^*$, and assume that $\mathcal{B}_\infty = \mathcal{B}_\infty^*$ mod P. Then for any $\phi\in\mathcal{L}^1(P)$,*

$$\lim_{i\to\infty}\ \sup_{\mathcal{B}_i\subseteq\mathcal{C}_i\subseteq\mathcal{B}_i^*} \left\|E[\phi\,|\,\mathcal{C}_i] - E[\phi\,|\,\mathcal{B}_\infty]\right\|_{\mathcal{L}^1(P)} = 0. \tag{6}$$

3 A Shannon–McMillan theorem along lines

From now on we will consider the two dimensional case. The line with slope λ and y-intercept a is described by the function

$$l_{\lambda,a}(x) := \lambda x + a \quad (x\in\mathbb{R}). \tag{7}$$

Using $[x]$ and $\{x\}$ for the integer and the fractional part of x, respectively, the two-sided sequences $\left([l_{\lambda,a}(z)]\right)_{z\in\mathbb{Z}}$ and $\left(\{l_{\lambda,a}(z)\}\right)_{z\in\mathbb{Z}}$ are the line's integer and fractional parts at the integer points $z\in\mathbb{Z}$. In the case when $0\le\lambda\le 1$, the *lattice approximation* of $l_{\lambda,a}$ on $U\subseteq\mathbb{Z}$ and on an interval $I\subseteq\mathbb{R}$ are given by

$$L_{\lambda,a}(U) := \left\{(z,[l_{\lambda,a}(z)])\,|\,z\in U\right\} \quad\text{and}\quad L_{\lambda,a}(I) := L_{\lambda,a}(I\cap\mathbb{Z}). \tag{8}$$

In the case when $-1\le\lambda<0$, we use the lattice approximation $L_{\lambda,a}(I) := -L_{-\lambda,a}(I)$. If $|\lambda|>1$, we represent the line as a function of the y-axis with the new slope $1/\lambda$ (or 0 in the case of a parallel to the y-axis) and proceed as before.

We want to show the $\mathcal{L}^1(P)$-convergence of the sequence of rescaled information along successively increased parts of the lattice approximation to the line

$$|L_{\lambda,a}([0,n])|^{-1}\,\mathcal{I}\left(P_{L_{\lambda,a}([0,n])}\right) \quad (n \in \mathbb{N}).$$

In order to make our problem accessible to ergodic theory, we need to create a transformation that follows the stair climbing pattern along the lattice approximation of the line. Lemma 3.2 will introduce a transformation that follows the desired path. To get there we need a mechanism to keep track not only of the integer part, but also of the fractional part. Let $\tau_\lambda(t) := \{t+\lambda\}$ ($t \in \mathbb{T}$) be the translation by λ on the torus $\mathbb{T} := [0,1]$ with its ends identified and equipped with the Borel σ-algebra $\mathcal{B}$ and the Lebesgue measure μ. Consider the product space $\mathbb{T} \times \Omega$, equipped with the product σ-algebra $\overline{\mathcal{F}}$, and the product measure $\overline{P} = \mu \otimes P$. We will now show a few technical lemmata that are needed to prove the main results in this section.

Lemma 3.1 *For $\lambda \in \mathbb{R}$, $z, \widetilde{z} \in \mathbb{Z}$, $a \in \mathbb{T}$ and $I \subset \mathbb{Z}$ we have the following*:

(i) $\tau_\lambda^z(a) = \{l_{\lambda,a}(z)\}$.
(ii) *The function τ_λ^z has a unique zero. More explicitly: If z and λ are both positive or both negative, the zero is at $a = 1 - \{z\lambda\}$. If one is negative and the other is positive, the zero is at $a = -\{z\lambda\}$. If one of them is zero, then the unique zero is at $a = 0$.*
(iii) $l_{\lambda,a}(z+\widetilde{z}) = l_{\lambda,a}(z) + \lambda\widetilde{z}$ and $l_{\lambda,a+\widetilde{z}}(z) = l_{\lambda,a}(z) + \widetilde{z}$.
(iv) $[l_{\lambda,a}(z+\widetilde{z})] = [l_{\lambda,a}(z)] + [l_{\lambda,\tau_\lambda^z(a)}(\widetilde{z})]$ and $[l_{\lambda,a+\widetilde{z}}(z)] = [l_{\lambda,a}(z)] + \widetilde{z}$.
(v) $L_{\lambda,a}(I+z) = L_{\lambda,\tau_\lambda^z(a)}(I) + L_{\lambda,a}(z)$.
(vi) $L_{\lambda,a+z}(I) = L_{\lambda,a}(I) + (0,z)$.

Proof (i) By induction over z. (ii) The case $z=0$ and the case $\lambda=0$ are trivial. Let $z \in \mathbb{Z}\backslash\{0\}$. By (i), τ_λ^z has a zero at a if and only if $\{a+z\lambda\} = 0$. The latter is equivalent to $a + \{z\lambda\} \in \mathbb{Z}$ $(*)$. If z and λ are both positive then $0 < a + \{z\lambda\} < 2$. Therefore, condition $(*)$ is equivalent to $a = 1 - \{z\lambda\}$. If z and λ are both negative, $\{z\lambda\}$ is again positive and the argument works as well. If one is negative and the other positive, then $-1 < a + \{z\lambda\} < 1$, and $(*)$ is equivalent to $a + \{z\lambda\} = 0$. (iii) $l_{\lambda,a}(z+\widetilde{z}) = \lambda(z+\widetilde{z}) + a = l_{\lambda,a}(z) + \lambda\widetilde{z}$ and $l_{\lambda,a+\widetilde{z}}(z) = \lambda z + a + \widetilde{z} = l_{\lambda,a}(z) + \widetilde{z}$. (iv) $[l_{\lambda,a}(z+\widetilde{z})] = \big[[l_{\lambda,a}(z)] + \{l_{\lambda,a}(z)\} + \lambda\widetilde{z}\big] = [l_{\lambda,a}(z)] + [\tau_\lambda^z(a) + \lambda\widetilde{z}] = [l_{\lambda,a}(z)] + [l_{\lambda,\tau_\lambda^z(a)}(\widetilde{z})]$, using (iii), (i) and (7). The second statement is an immediate consequence of (iii). (v) Using (8) and (iv), $L_{\lambda,a}(I+z) = \big\{(\widetilde{z}, [l_{\lambda,a}(\widetilde{z})])\big|\,\widetilde{z} \in I+z\big\} = \big\{(\widetilde{z}+z, [l_{\lambda,a}(\widetilde{z}+z)])\big|\,\widetilde{z} \in I\big\} = \Big\{(\widetilde{z}, [l_{\lambda,\tau_\lambda^z(a)}(\widetilde{z})]) + (z, [l_{\lambda,a}(z)])\big|\,\widetilde{z} \in I\Big\} = L_{\lambda,\tau_\lambda^z(a)}(I) + L_{\lambda,a}(z)$. (vi) $L_{\lambda,a+z}(I) = \big\{(\widetilde{z}, [l_{\lambda,a+z}(\widetilde{z}))\,|\,\widetilde{z} \in I\big\} = \big\{(\widetilde{z}, [l_{\lambda,a+z}(\widetilde{z})) + (0,z)\,|\,\widetilde{z} \in I\big\} = L_{\lambda,a}(I) + (0,z)$, using (8) and (iv). □

Lemma 3.2 *The iterates of the transformation $S_\lambda(a,\omega) := (\tau_\lambda(a), \theta_{(1,[a+\lambda])}\omega)$ ($a \in \mathbb{T}, \omega \in \Omega$) are given by $S_\lambda^n(a,\omega) = \left(\tau_\lambda^n(a), \theta_{L_{\lambda,a}(n)}\omega\right)$ for all $n \in \mathbb{N}_0$.*

Proof With $\kappa(a) := (1, [a+\lambda])$, we get $S_\lambda(a,\omega) = (\tau_\lambda(a), \theta_{\kappa(a)}\omega)$ and $S_\lambda^n(a,\omega) = (\tau_\lambda^n(a), \theta_{\kappa_n(a)}\omega)$ where $\kappa_n; = \sum_{i=0}^{n-1} \kappa \circ \tau_\lambda^i$ ($n \in \mathbb{N}$). It remains to show that $\kappa_n(a) = (n, [l_{\lambda,a}(n)])$ for all $a \in \mathbb{T}$. For the first component this is obvious. For the second component it follows by induction: Trivial for $n = 0$. Then, because of Lemma 3.1(iv), $\kappa^{(2)}_{n+1}(a) = \kappa^{(2)}_n(a) + \kappa^{(2)}(\tau^n_\lambda(a)) = [l_{\lambda,a}(n)] + [\tau^n_\lambda(a) + \lambda] = [l_{\lambda,a}(n)] + [l_{\lambda,\tau^n_\lambda(a)}] = [l_{\lambda,a}(n+1)]$. □

We use the short forms $\mathcal{P}_{\lambda,a} := \mathcal{F}_{L_{\lambda,a}((-\infty,-1])}$ and $\mathcal{P}^{(i)}_{\lambda,a} := \mathcal{F}_{L_{\lambda,a}((-i,-1])}$ $(i \in \mathbb{N})$ for the various *past* σ-algebren indexed by the lattice approximation along of line with slope λ and intercept a. They play a central role in the representation of the limits in the following two Shannon–McMillan type theorems. We further define the functions

$$\begin{aligned} F(a,\omega) &:= \mathcal{I}\left(P_0[\,\cdot\,|\mathcal{P}_{\lambda,a}]\right)(\omega),\\ F_i(a,\omega) &:= \mathcal{I}\left(P_0[\,\cdot\,|\mathcal{P}^{(i)}_{\lambda,a}]\right)(\omega)\ \ (i \in \mathbb{N}) \qquad \text{with } a \in \mathbb{T}, \omega \in \Omega. \end{aligned} \tag{9}$$

Lemma 3.3 *For all* $a \in \mathbb{T}, \omega \in \Omega$, *and* $n \in \mathbb{N}$, $\mathcal{I}(P_{L_{\lambda,a}([0,n])})(\omega) = \sum_{i=0}^n F_i \circ S^i_\lambda(a,\omega)$.

Proof By conditioning and shifting

$$\begin{aligned} P[\omega_{L_{\lambda,a}([0,n])}] &= \prod_{i=0}^n P[\omega_{L_{\lambda,a}(i)}|\omega_{L_{\lambda,a}([0,i-1])}]\\ &= \prod_{i=0}^n P[\omega_{(0,0)}|\omega_{L_{\lambda,a}([0,i-1])-L_{\lambda,a}(i)}] \circ \theta_{L_{\lambda,a}(i)}. \end{aligned}$$

By Lemma 3.1(v), $L_{\lambda,a}([0,i-1]) - L_{\lambda,a}(i) = L_{\lambda,\tau^i_\lambda(a)}([-i,-1])$, so

$$\mathcal{I}(P_{L_{\lambda,a}([0,n])})(\omega) = \sum_{i=0}^n \mathcal{I}\left(P_0[\,\cdot\,|\mathcal{P}^{(i)}_{\lambda,\tau^i_\lambda(a)}]\right)(\theta_{L_{\lambda,a}(i)}\omega),$$

and Lemma 3.2 concludes the proof. □

We will also need the following result about the asymptotic behavior of the functions F_i as i goes to infinity and about their properties as functions of the first parameter. Recall that P_0 is the marginal distribution of P restricted to the origin.

Lemma 3.4 *Assume that*

$$\forall\, A \in \mathcal{F}_{\mathbb{Z}^2\setminus\{(0,0)\}}: \quad P_0[\,\cdot\,|A] > 0. \tag{10}$$

Then, $(F_i)_{i\in\mathbb{N}}$ *converge to* F, P*-almost surely and in* $\mathcal{L}^1(P)$. *For any fixed* $\omega \in \Omega$, *the functions* $F_i(\cdot,\omega)$ $(i \in \mathbb{N})$ *are piecewise constant on* $\mathbb{T}$. *If* λ *is rational then* $F(\cdot,\omega)$ *is piecewise constant as well, and the convergence is uniform in* t. *If* λ *is irrational,* *if P fulfills the strong* 0-1 *law and if*

$$\exists\, c > 0 \;\; \forall\, A \in \mathcal{F}_{\mathbb{Z}^2\setminus\{(0,0)\}} : \quad P_0[\,\cdot\,|A] > c, \tag{11}$$

then $F(\cdot\,,\omega)$ is Riemann-integrable on $\mathbb{T}$, and the convergence is uniform in t.

More precisely, the set of points where the function $F_i(\cdot,\omega)$ may be discontinuous is given by $\left\{t_{\lambda,\nu}\,\middle|\, \nu = -1,\ldots,-i\right\}$, where $t_{\lambda,\nu}$ is the unique zero of $\tau_{\lambda,\nu}$ on $\mathbb{T}$ (cf. Lemma 3.1(ii)). For rational λ, the set of potential points of discontinuities of $F_i(\cdot,\omega)$ and $F(\cdot,\omega)$ is $\left\{t_{\lambda,\nu}\,\middle|\, \nu = -1,\ldots,-(q\wedge i)\right\}$, where p/q is the unique representation of λ with integers $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ having no common divisor.

Proof of the Lemma Fix any $t \in \mathbb{T}$. Since the σ-algebras $\left(\mathcal{F}_{L_{\lambda,t}([-1,-i])}\right)_{i\in\mathbb{N}}$ form an increasing family, $\left(P\left[\omega_{(0,0)}\,\middle|\,\omega_{L_{\lambda,t}([-1,-i])}\right]\right)_{i\in\mathbb{N}}$ is a martingale, so that we obtain by the convergence theorem for martingales that

$$P\left[\omega_{(0,0)}\,\middle|\,\omega_{L_{\lambda,t}([-1,-i])}\right] \xrightarrow{i\to\infty} P\left[\omega_{(0,0)}\,\middle|\,\omega_{L_{\lambda,t}([-1,-\infty))}\right]$$

P-almost surely and in $\mathcal{L}^1(P)$. By (10) this remains true when we take logarithms on both sides, which proves the first assertion of the lemma.

Now fix $\omega \in \Omega$ and $t,\widetilde{t} \in \mathbb{T}$, and find a (sufficient) condition under which $F_i(t,\omega) = F_i(\widetilde{t},\omega)$. The only influence that the variable t actually has on F_i is its effect on the set $L_{\lambda,t}([-1,-i])$ of sites we condition on. By (8), $L_{\lambda,t}\,([-1,-i]) = L_{\lambda,\widetilde{t}}\,([-1,-i])$ if and only if

$$[l_{\lambda,t}(\nu)] = [l_{\lambda,\widetilde{t}}(\nu)] \quad \text{for all } \nu \in \{-1,\ldots,-i\}. \tag{12}$$

By Lemma 3.1(i), $[l_{\lambda,t}(\nu)] = -\lambda\nu - \tau_\lambda^\nu(t)$, so the equality in (12) is equivalent to $t - \widetilde{t} = \tau_\lambda^\nu(t) - \tau_\lambda^\nu(\widetilde{t})$. This is fulfilled if and only if t and $\widetilde{t}$ are both either smaller than $t_{\lambda,\nu}$ or both larger than $t_{\lambda,\nu}$. Applying this argument to all $\nu \in \{-1,\ldots,-i\}$, we see that the function $F_i(\,\cdot\,,\omega)$ is piecewise constant, and the set of possible jumps is given by $D_i = \left\{t_{\lambda,\nu}\,\middle|\, \nu = -1,\ldots,-i\right\}$.

If λ is rational, these sets actually become independent of i, for i large enough. With the unique representation $\lambda = p/q$ used above, and the periodicity of the sequence $(t_{\lambda,\nu})_{\nu\in\mathbb{N}}$, we obtain $D_i = \left\{t_{\lambda,\nu}\,\middle|\, \nu = -1,\ldots,-(q\wedge i)\right\}$. In particular, the convergence is uniform and the limit F is piecewise constant in the first variable.

Now assume that λ is irrational. To show that F is Riemann-integrable it suffices to show that the set of continuity points has full measure. We will prove that $F(\cdot,\omega)$ is continuous on $\mathbb{T}\setminus D_\infty$, where $D_\infty := \left\{t_{\lambda,\nu}\,\middle|\, \nu = -1,-2,\ldots\right\}$. Fix $t_0 \in \mathbb{T}\setminus D_\infty$ and let be $\varepsilon > 0$. We apply Lemma 2.3 with $\mathcal{B}_k = L_{\lambda,t}([-k,-1])$, and $\mathcal{B}_k^* = L_{\lambda,t}([-k,-1]) \cup (\mathbb{Z}^2\setminus V_k)$, where $V_k = [-k,k]^2$. This gives us a $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ and $t \in \mathbb{T}$ with $L_{\lambda,t_0}([-k,-1]) = L_{\lambda,t}([-k,-1])$ $(**)$ we obtain

$$\left|\, P\left[\omega_{(0,0)}\,\middle|\,\omega_{L_{\lambda,t_0}([-k,-1])}\right] - P\left[\omega_{(0,0)}\,\middle|\,\omega_{L_{\lambda,t}([-k,-1])}\right] \,\right| < \varepsilon.$$

By definition of t_0, $\delta := \min\left\{|\,t_0 - t_{\lambda,\nu}|\,\middle|\, \nu = -1, \ldots, -k\right\}$ is larger than 0, and by Lemma 3.1(i), $(**)$ is true for all $t \in \mathbb{T}$ for which $|t - t_0| < \delta$. So, $P\left[\omega_{(0,0)}\middle|\,\omega_{L_{\lambda,t}([-k,-1])}\right]$ is continuous in $t = t_0$. Note that, for all $i \in \mathbb{N}$, $D_i \subset D_{i+1}$, and that the maximal height of new jumps added in step i, that is, the steps in $D_{i+1}\backslash D_i$, decreases with i. This implies the uniformity of the convergence. Finally, by taking logarithms and using (11), the above statements are also true for the sequence $(F_i(\cdot\,,\omega))_{i\in\mathbb{N}}$. □

Theorem 3.5 *Let λ be rational and p/q its unique representation with $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ having no common divisor. Assume that P fulfills condition* (10). *Then for all $a \in \mathbb{R}$,*

$$\frac{1}{n+1}\mathcal{I}(P_{L_{\lambda,a}([0,n])}) \xrightarrow{n\to\infty} \frac{1}{q}\sum_{\nu=0}^{q-1} E\left[H\left(P_0\left[\,\cdot\,|\mathcal{P}_{p/q,\,\nu p/q}\right](\omega)\right)\middle|\mathcal{J}_{q,p}\right] \tag{13}$$

P-almost surely and in $\mathcal{L}^1(P)$, where $\mathcal{J}_{q,p}$ is the σ-algebra of all sets which are invariant with respect to $\theta_{(q,p)}$. In particular, if P is ergodic with respect to $\theta_{(q,p)}$ then the limit equals

$$\frac{1}{q}\sum_{\nu=0}^{q-1} E\left[H\left(P_0\left[\,\cdot\,|\mathcal{P}_{p/q,\,\nu p/q}\right](\omega)\right)\right].$$

Proof By Lemma 3.3, the left-hand side of (13) equals $1/(n+1)\sum_{i=0}^{n} F_i \circ S_\lambda^i(a,\omega)$. Because of Lemma 3.4 and Maker's version of the ergodic theorem (cf. Theorem 7.4 in Chap. 1 of [14]), it suffices to show that

$$\frac{1}{n+1}\sum_{i=0}^{n} F \circ S_\lambda^i(a,\omega) \tag{14}$$

converges to the limit in (13). For each $n \in \mathbb{N}$ there are unique $m \in \mathbb{N}$ and $\eta \in \{0, 1, \ldots, q-1\}$ such that $n = mq + \eta$. The last term can be rewritten as

$$\frac{m}{n+1}\sum_{\nu=0}^{q-1}\frac{1}{m}\sum_{j=0}^{m-1} F \circ S_\lambda^{jq+\nu} + \frac{1}{n+1}\sum_{\nu=0}^{\eta} F \circ S_\lambda^{mq+\nu}.$$

The second addend converges to 0 as n (and therefore also m) goes to infinity. The first factor converges to $1/q$, so it remains to study the asymptotic behavior of

$$A_m^{(\nu)}F(a,\omega) := \frac{1}{m}\sum_{j=0}^{m-1} F \circ S_\lambda^{jq+\eta}(a,\omega).$$

Use κ_n $(n \in \mathbb{N})$ defined as in the proof of Lemma 3.2. Since $\tau_\lambda^q = \mathrm{Id}$ we obtain $\kappa_{jq+\nu} = j\kappa_q + \kappa_\nu$, and $S_\lambda^{jq+\nu} = \left(\tau_\lambda^\nu, \theta_{\kappa_\nu} \circ (\theta_{\kappa_q})^j\right)$. This yields

$$A_m^{(\nu)} F(a, \omega) = \frac{1}{m} \sum_{j=0}^{m-1} F\left(\tau^\nu(a), \theta^{\kappa_\nu(a)} \circ (\theta^{\kappa_q(a)})^j \omega\right).$$

For a fixed, applying Birkhoff's ergodic theorem to the function $f_a^{(\nu)}(\omega) := F\,(\tau^\nu(a), \theta_{\kappa_\nu(a)}\omega)$ yields

$$\begin{aligned}\lim_{m\to\infty} A_m^{(\nu)} F(a, \cdot) &= E[f_t^{(\nu)} | \mathcal{J}_{(q,p)}] \\ &= E\left[H\left(P_0\left[\,\cdot\,\middle|\mathcal{P}_{p/q,\,\nu p/q}\right](\omega)\right)\middle|\mathcal{J}_{(q,p)}\right] \quad P\text{-a.s. and in } \mathcal{L}^1(P).\end{aligned}$$

In the ergodic case $\mathcal{J}_{(q,p)}$ is trivial. Using the invariance of P under θ, the last expression reduces to $E\,[F\,(\tau^\nu(a), \,\cdot\,)]$. □

For the case of an irrational slope λ we can show the corresponding result provided P fulfills some additional conditions, in particular the strong 0-1 law (cf. Definition 2.2).

Theorem 3.6 *Let λ be irrational. Assume that P fulfills condition* (11) *and the strong* 0-1 *law. Then for all $a \in \mathbb{R}$,*

$$\frac{1}{n+1} \mathcal{I}(P_{L_{\lambda,a}([0,n])}) \xrightarrow{n\to\infty} \int_0^1 E\left[H\left(P_0[\,\cdot\,|\mathcal{P}_{\lambda,t}](\omega)\right)\right] dt \quad \textit{in } \mathcal{L}^1(P). \tag{15}$$

Proof As in the proof of Theorem 3.5 the left-hand side of (13) can be rewritten as an ergodic average involving the skew product S_λ. Using Lemma 3.4 and Maker's version of the ergodic theorem (cf. Theorem 7.4 in Chap. 1 of [14]), it suffices to show that for all $i \in \mathbb{N}$,

$$1/(n+1) \sum_{i=0}^{n} F_k \circ S_\lambda^i(a, \omega) \xrightarrow{n\to\infty} \int_0^1 E\left[H\left(P_0[\,\cdot\,|\mathcal{P}_{\lambda,t}^{(i)}](\omega)\right)\right] dt \quad \text{in } \mathcal{L}^1(P) \tag{16}$$

For $i \in \mathbb{N}$ fixed, F_i is piecewise constant in t, so there is an $R \in \mathbb{N}$, intervals $U_1, \ldots, U_R$ and measurable functions $f_1, \ldots, f_R$ on $(\Omega, \mathcal{F})$ such that $F_i(t, \omega) = \sum_{r=1}^R 1_{U_r}\, f_r(\omega)$. By approximation, this can further be reduced to the case of indicator functions at the place of the f_r's. It remains to show the convergence for functions of the form $1_U(t)\, 1_A(\omega)$ for an interval $U \subset \mathbb{T}$ and a set $A \in \mathcal{F}$.

Let $F_i(t, \omega) = 1_U(t)\, 1_A(\omega)$. We will show the convergence in $\mathcal{L}^2(P)$. Using

$$\int_\Omega 1_A(\theta_{\kappa_i(t)}(\omega) 1_A(\theta_{\kappa_j(t)}\omega)\, P(d\omega) = P\left(\theta_{\kappa_i(t)-\kappa_j(t)}^{-1} A \cap A\right),$$

we see that

$$\left\| \frac{1}{n+1} \sum_{i=0}^{n} F\left(\tau^i(t), \theta_{\kappa_i(t)} \cdot\right) - \int_0^1 E[F]\, dt \right\|^2_{\mathcal{L}^2(P)}$$

$$= \frac{1}{(n+1)^2} \sum_{i=0}^{n} 1_U(\tau^i(t)) 1_U(\tau^j(t)) P\left(\theta^{-1}_{\kappa_i(t)-\kappa_j(t)} A \cap A\right) + \mu(U)^2 P(A)^2 \tag{17}$$

$$-\frac{2}{(n+1)^2} \sum_{i=0}^{n} \left[1_U(\tau^i(t)) \int_\Omega 1_A(\theta_{\kappa_i(t)}\omega)\, P(d\omega) \right] \mu(U) P(A) \tag{18}$$

The next step is to show that the first addend in line (17) may be replaced by

$$\left(\frac{1}{(n+1)} \sum_{i=0}^{n} 1_U(\tau^i(t)) \right)^2 P(A)^2 \tag{19}$$

without affecting the asymptotic behavior of (17). We may bound

$$\left| \frac{1}{(n+1)^2} \sum_{i,j=0}^{n} 1_U(\tau^i(t)) 1_U(\tau^j(t)) \left(P\left(\theta^{-1}_{\kappa_i(t)-\kappa_j(t)} A \cap A\right) - P(A)^2 \right) \right|$$

$$\leq \frac{1}{(n+1)^2} \sum_{i,j=0}^{n} \left| P\left(\theta^{-1}_{\kappa_i(t)-\kappa_j(t)} A \cap A\right) - P(A)^2 \right|.$$

Fix $\varepsilon > 0$. Let $\| \cdot \|$ denote the maximum norm on $\mathbb{Z}^2$. As a special case of the strong 0-1 law, P fulfills a 0-1 law on the tail field. By (2) there is an $m \in \mathbb{N}$ such that

$$\left| P\left(\theta_k^{-1} A \cap A\right) - P(A)^2 \right| < \frac{\varepsilon}{2} \qquad \text{for all } k \in \mathbb{Z}^2 \text{ with } \|k\| > m.$$

Note that $\|\kappa_i(t) - \kappa_j(t)\| \geq |i-j|$ for all $t \in \mathbb{T}$. Since $\lim_{n\to\infty} n^{-2} \left| \{1 \leq i, j \leq n \mid |i-j| \leq m\} \right| = 0$ for all $m \in \mathbb{N}$, we can find an $n_0 \in \mathbb{N}$ such that

$$\frac{1}{(n+1)^2} \sup_{t\in\mathbb{T}} \left| \left\{ 1 \leq i, j \leq n \,\middle|\, \|\kappa_i(t) - \kappa_j(t)\| \leq m \right\} \right| < \frac{\varepsilon}{2} \quad \text{for all } n \geq n_0.$$

This demonstrates that hat the difference created by the by replacing the first addend in (17) by (19) converges to 0 uniformly with respect to t.

Since U is an interval, the function 1_U is integrable in the sense of Riemann. Applying a variant of Weyl's theorem for Riemann integrable functions (cf. for instance, Theorem 2.6 in Chap. 1 of [14]) implies that, for n to infinity, $1/(n+1) \sum_{i=0}^{n} 1_U(\tau^i(t))$ converges to $\mu(U)$ uniformly in t. So, asymptotically, the sum of the two addends in line (17) equals $2\,\mu(U)^2 P(A)^2$.

The expression in (18) can be simplified to $-2/(n+1)\sum_{i=0}^{n} 1_U(\tau^i(t))\mu(U)P(A)^2$. Applying a variant of Weyl's theorem for Riemann integrable functions again, this expression converges to $-2\,\mu(U)^2P(A)^2$, which concludes the proof. $\square$

4 A Shannon–McMillan theorem along general shapes

Let P be a stationary random field that satisfies the strong 0-1 law and the condition (11). The goal of this section is a Shannon–McMillan theorem for a stochastic field along the lattice approximations of the blowups of a curve c, and an explicit formula for the limit $h_c(P)$, the *specific entropy* of P *along* c. This will be done in three steps: for linear segments, for polygons and for curves. In Sect. 4.4, we will further introduce lattice approximations that are *contours* in the sense of statistical mechanics and sketch corresponding results for them.

Let $c = \left(c^{(1)}, c^{(2)}\right)$ be a piecewise differentiable planar curve parametrized by $t \in [0, T]$. Assume that the trace of c does not contain the origin and that it hits the y-axis in $t = 0$. Let c' denote the right derivative of c. The blowups of the curve c are given by

$$B_\eta c : [0, \eta T) \longrightarrow \mathbb{R}^2, \quad B_\eta c(t) = \eta\, c(t/\eta) \qquad (\eta > 0). \tag{20}$$

It will follow from the construction in Sect. 4.2 that is enough to consider curves that are described by the graph of a function ϕ on a segment of one of the axes. Suppose that ϕ is a function on the interval $[x, \widetilde{x}]$ of the x-axis. (The case of the y-axis can be treated analogously.) More precisely, $x = c^{(1)}(0)$ and $\widetilde{x} = c^{(1)}(T)$. The interval $[x, \widetilde{x}]$ contains a finite number u of integers $z, \dots, z + u$, where $u = [x - \widetilde{x}]$ or $u = [x - \widetilde{x}] - 1$. In the same way, the blowups $B_n c$ can be represented as graphs of functions ϕ_n of intervals $[x_n, \widetilde{x}_n]$. We obtain by (20), $x_n = B_n c^{(1)}(0) = nx = nc^{(1)}(0)$ and $\widetilde{x}_n = B_n c^{(1)}(nT) = n\widetilde{x} = nc^{(1)}(T)$. Again, the interval $[x_n, \widetilde{x}_n]$ contains a finite number u_n of integers $z_n, z_n + 1, \dots, z_n + u_n$, where

$$u_n = [n(\widetilde{x} - x)] \quad \text{or} \quad u_n = [n(\widetilde{x} - x)] - 1. \tag{21}$$

In particular, the sequence $(u_n)_{n\in\mathbb{N}}$ goes to infinity.

4.1 Line segments

The main result of this section is the convergence of the sequence of renormalized information functions

$$\frac{1}{|L_n(a)|}\,\mathcal{I}\left(P_{L_n(a)}\right)(\omega) \tag{22}$$

along the lattice approximation converges to the entropy $h_\lambda(P)$ of P along a line with slope λ. If c is a line segment, the functions ϕ_n are of the form $\phi_n(x) = \lambda x + a_n$ $(x \in [nc^{(1)}(0), nc^{(1)}(T)])$ where $\lambda = (c^{(2)}(T) - c^{(2)}(0))/(c^{(1)}(T) - c^{(1)}(0))$ and $a_n = n(c^{(2)}(0) - \lambda c^{(1)}(0))$. Assume $0 \le \lambda \le 1$. As explained at the beginning of Sect. 3, the other cases can be reduced to this case. At first sight it seems we could just apply the results for the specific entropy along a line from the last section. But the blowups of the line segments move in space, which has the following consequences:

(i) There is a sequence $(a_n)_{n\in\mathbb{N}}$ instead of a constant a.
(ii) The sequence is real-valued, as opposed to the constant having values in $\mathbb{T}$.
(iii) The positions of the lattice points in each step are more difficult to describe. In the case of the line we simply looked at approximating points with x-values between 0 and n. Now, x-values of the lattice points lie in the interval between z_n and $z_n + u_n$.

The last problem forces us to apply, at each step n, an additional shift to ω which brings the line segment close to the origin. These shifts do not affect the $\mathcal{L}^1(P)$-convergence since the limit is shift invariant. The number of points in the nth step is given by $(u_n)_{n\in\mathbb{N}}$, instead of simply $n+1$ as in the last section, but this is irrelevant as long as the sequence goes to infinity. The second problem requires another shift in each step n. The first point is the most delicate. It is here that we need the convergence of the ergodic averages in *all* t, rather than just *almost* all t.

Theorem 4.1 *In $\mathcal{L}^1(P)$ and uniformly in $a \in \mathbb{R}$,*

$$\frac{1}{|L_n(a)|}\mathcal{I}(P_{L_n(a)}) \xrightarrow{n\to\infty} h_\lambda(P).$$

Proof We start with some technical preparations. Set $a := a_1$. Using the notation for $[x, \widetilde{x}]$ described above (21), define for $n \in \mathbb{N}$, $L_n(a) := L_{\lambda,a_n}(z_n, \ldots, z_n + u_n)$. The total number of sites in $L_n(t)$ is $u_n + 1$. To transform (22) into some sort of ergodic average we first condition on successively smaller parts of $L_n(a)$. A new step begins at i if and only if $\tau_\lambda^{z_n+i-1}(\{a\}) \geq 1-\lambda$. For all $z, i \in \mathbb{Z}$, and $a \in \mathbb{R}$,

$$L_{\lambda,a}(z+i) - L_{\lambda,a}(z) = L_{\lambda,\tau_\lambda^z(\{a\})}(i). \tag{23}$$

To prove this, first apply the second equation in Lemma 3.1(v) with $a = \{a\}$ and $z = [a]$, and then apply the first equation with $\widetilde{z} = i$,

$$\begin{aligned}
L_{\lambda,a}(z+i) &= (z+i, [l_{\lambda,\{a\}}(z+i)]) + (0, [a]) \\
&= (z+i, [l_{\lambda,\{a\}}(z)] + [l_{\lambda,\tau_\lambda^z(\{a\})}(i)]) + (0, [a]) \\
&= \big(z, [l_{\lambda,a}(z)]\big) + \Big(i, [l_{\lambda,\tau_\lambda^z(\{a\})}(i)]\Big) = L_{\lambda,a}(z) + L_{\lambda,\tau_\lambda^z(\{a\})}(i).
\end{aligned}$$

We calculate the information in (22) by conditioning site by site along $L_n(a)$. We use $\omega(i)$ instead of ω_i for easier reading. Shifting ω to the origin, applying (23) and using the functions defined in (9) yields

$$\begin{aligned}
\mathcal{I}(P_{L_n(a)})(\omega) &= -\sum_{i=0}^{u_n} \log P\left[\omega\left(L_{\lambda,a_n}(z_n+i)\right) \,\middle|\, \omega\Big(L_{\lambda,a_n}(z_n+i-1,\ldots,z_n)\Big)\right] \\
&= -\sum_{i=0}^{u_n} \log P\left[\omega(0,0) \,\middle|\, \omega\left(L_{\lambda,a_n}(z_n+i-1,\ldots,z_n) - L_{\lambda,a_n}(z_n+i)\right)\right] \\
&\qquad \circ\ \theta_{L_{\lambda,a_n}(z_n+i)} \qquad\qquad (24)\\
&= -\sum_{i=0}^{u_n} \log P\left[\omega(0,0) \,\middle|\, \omega\left(L_{\lambda,\tau_\lambda^{z_n+i}(\{a_n\})}(-1,\ldots,-i)\right)\right] \\
&\qquad \circ\ \theta_{L_{\lambda,\tau_\lambda^{z_n}(\{a_n\})}(i)} \circ \theta_{L_{\lambda,a_n}(z_n)} \\
&= -\sum_{i=0}^{u_n} F_i\left(\tau_\lambda^i(\tau_\lambda^{z_n}(\{a_n\})), \theta_{L_{\lambda,\tau_\lambda^{z_n}(\{a_n\})}(i)} \circ \theta_{L_{\lambda,a_n}(z_n)}\,\omega\right).
\end{aligned}$$

Applying Lemma 3.2, putting all together back in (24) and renormalizing yields

$$\frac{1}{u_n+1}\mathcal{I}(P_{L_n(a)})(\omega) = \frac{1}{u_n+1}\sum_{i=0}^{u_n} F_i \circ S_\lambda^i\left(\tau_\lambda^{z_n}(\{a_n\}), \theta_{L_{\lambda,a_n}(z_n)}\,\omega\right). \tag{25}$$

To prove the convergence, we have to distinguish the case when λ is rational from the case when it is irrational, because this determines whether τ_λ is periodic or uniquely ergodic. We proceed as in the proof of Theorem 3.5 and Theorem 3.6, respectively. □

4.2 Polygons

The next step is to define the entropy along a polygon, that is a piecewise linear curve $\pi : [0, T] \to \mathbb{R}^2$. Assume further that π fulfills the other assumptions on c stated at the beginning of this section. Let R be the number of edges of π. We can find slopes $\lambda^{(r)} \in (-1, 1]$, constants $t^{(r)} \in \mathbb{R}$, and intervals $I^{(r)}$ of the x- or the y-axis such that

$$\pi\,([0,T]) = \bigcup_{r=1}^{R} l_{\lambda^{(r)},t^{(r)}}(I^{(r)}), \tag{26}$$

with $l_{\lambda,t}$ as defined in (7) as a function of the x- or of the y-axis. Proceeding the same way for the blowups $B_n\pi$ $(n \in \mathbb{N})$ as defined in (20), we choose $t_n^{(r)} \in \mathbb{R}$ and $I_n^{(r)} \subset \mathbb{R}$, such that

$$B_n\pi\,([0,T]) = \bigcup_{r=1}^{R} l_{\lambda^{(r)},t_n^{(r)}}(I_n^{(r)}),$$

The lattice approximations of the edges combine to a lattice approximation of $B_n\pi$:

$$L_n^\pi := \bigcup_{r=1}^{R} L_{\lambda^{(r)},t_n^{(r)}} \tag{27}$$

Theorem 4.2 *The lattice approximations converge. More precisely,*

$$\frac{1}{\text{length}\, L_n^\pi}\,\mathcal{I}(P_{L_n^\pi}) \xrightarrow{n\to\infty} \frac{1}{\text{length}\,\pi}\sum_{r=1}^{R} \text{length}\,\pi^{(r)}\, h_{\lambda^{(r)}}(P) \quad \textit{in } \mathcal{L}^1(P). \tag{28}$$

In some contexts it is more convenient to express the limit as an integral with respect to t rather than as a sum. Let $\pi'(t)$ denote the right derivative of π. Then the limit can be written as

$$\frac{1}{\text{length}\,\pi}\int_0^T h_{\pi'(t)}(P)\,dt. \tag{29}$$

Proof of the theorem Use $\omega(i)$ for ω_i and define the sets $E_n^{(r)} := L_{\lambda^{(r)},t_n^{(r)}}(I_n^{(r)})$ ($r \in \{1,\dots,R\}$). By conditioning,

$$\mathcal{I}\left(P_{L_n^\pi}\right)(\omega) = \sum_{r=1}^{R} \log P\left[\omega(E_n^{(r)})\,\Big|\,\omega(E_n^{(r-1)},\dots,E_n^{(1)})\right]. \tag{30}$$

Fix $r \in \{1,\dots,R\}$. Omit the index r when there is no risk of confusion (for example $\lambda := \lambda^{(r)}$, $t_n := t_n^{(r)}$, $E_n := E_n^{(r)}$) and use the short form $\breve{E}_n := E_n^{(r-1)} \cup \cdots \cup E_n^{(1)}$ for the lattice approximations of the edges of the polygon which come prior to $E_n^{(r)}$. We will condition successively on the elements of $E_n^{(r)}$. Denoting the integers in $I_n^{(r)}$ by $z_n, z_n+1,\dots,z_n+u_n$ as in (21),we obtain for the rth addend in (30)

$$\begin{aligned}&\log P\left[\omega(E_n)\,\Big|\,\omega\left(\breve{E}_n\right)\right]\\ &\quad= \sum_{i=0}^{u_n} \log P\left[\omega\left(L_{\lambda,t_n}(z_n+i)\right)\,\Big|\,\omega\left(L_{\lambda,t_n}(z_n+i-1,\dots,z_n),\,\breve{E}_n\right)\right].\end{aligned}$$

Shifting by $v_{ni}(t) := L_{\lambda,t_n}(z_n+i)$ yields

$$\sum_{i=0}^{u_n} \log P\left[\omega(0,0)\Big|\omega\left(L_{\lambda,t_n}(z_n+i-1,\dots,z_n) - v_{ni}(t),\,\breve{E}_n - v_{ni}(t)\right)\right]\circ\theta_{v_{ni}(t)}.$$

By (23), this equals

$$\sum_{i=0}^{u_n} \log P\left[\omega(0,0) \,\Big|\, \omega\left(L_{\lambda,\tau_\lambda^{z_n+i}(\{t_n\})}(-1,\ldots,-i),\, \breve{E}_n - v_{ni}(t)\right)\right] \circ \theta_{L_{\lambda,\tau_\lambda^{z_n}(\{t_n\})}(i)} \circ \theta_{L_{\lambda,t_n}(z_n)}. \tag{31}$$

This expression is similar to (24) except for the additional conditionings on the sites $\breve{E}_n - v_{ni}(t)$. We will show that these conditions disappear asymptotically. The argument will be given in detail for the first summand; it is similar for the remaining ones. Let α be the minimum angle between any neighboring edges of the polygon π and let d_n be the minimum distance between an edge of the nth blowup $B_n\pi$ of the polygon and any of its nonneighboring edges. Also, let H_n be the hexagon defined as follows: H_n is symmetric around $E_n^{(r)}$, two sides are parallel to $E_n^{(r)}$ at a distance $d_n/2$. The other sides reach from the endpoint of the first two to the endpoints of $E_n^{(r)}$, and they intersect at an angle α. Observe that $\breve{E}_n \subset \mathbb{Z}^2 \backslash H_n$, and therefore $\breve{E}_n - v_{ni}(t) \subset \mathbb{Z}^2 \backslash (H_n - v_{ni}(t))$. Define the σ-algebras $\mathcal{B}_i(t) := \mathcal{F}(L_{\lambda,\tau_\lambda^i(\{t\})}(-1,\ldots,-i))$, $\mathcal{B}_\infty(t) := \mathcal{F}(L_{\lambda,\tau_\lambda^i(\{t\})}(-1,-2,\ldots))$ and $\mathcal{B}_i^*(t) := \mathcal{F}(L_{\lambda,\tau_\lambda^i(\{t\})}(-1,\ldots,-i) \cup \mathbb{Z}^2 \backslash (H_n - v_{ni}(t)))$. The sequence $(\mathcal{B}_i(t))_{i\in\mathbb{N}}$ is increasing to $\mathcal{B}_\infty(t)$, and the sequence $(\mathcal{B}_i^*(t))_{i\in\mathbb{N}}$ is decreasing to $\mathcal{B}_\infty^*(t) := \bigcap_{i\in\mathbb{N}} \mathcal{B}_i^*(t)$. By the strong 0-1 law, $\mathcal{B}_\infty^*(t) = \mathcal{B}_\infty(t)$ mod P. By Lemma 2.3,

$$\lim_{i\to\infty} \Big\| \log P\left[\omega(0,0) \,\Big|\, \omega\left(L_{\lambda,\tau_\lambda^i(t)}(-1,\ldots,-i),\, \breve{E}_n - v_{ni}(t)\right)\right] - \log P\left[\omega(0,0) \,\Big|\, \mathcal{F}\left(L_{\lambda,\tau_\lambda^i(\{t\})}(-1,\ldots,-i)\right)\right](\omega) \Big\|_{\mathcal{L}^1(P)} = 0.$$

Proceeding with (31) as with (24) and using that, for all $r \in \{1,\ldots,R\}$, length $E_n^{(r)}$/ length L_n^π asymptotically equals length $\pi^{(r)}$/length π concludes the proof. □

4.3 Curves

To use the results from the previous section, we need to relate the derivatives of the curve with the slopes of lines. Let $v \in S^1 = \{w \in \mathbb{R}^2 \mid |w| = 1\}$, and α the angle from the positive x-axis to the vector v. If $|\alpha| \le \pi/4$ or $|\alpha| \ge 3\pi/4$ then describe the line in the direction of v by a function of the x-axis; otherwise describe it as a function of the y-axis. We assign any $v \in S^1$ a specific entropy

$$h_v(P) := h_{\lambda(v)}(P), \quad \text{where } \lambda(v) := \min(|tg\,\alpha|, |ct\,\alpha|). \tag{32}$$

Theorem 4.3 *Let $c : [0,T] \longrightarrow \mathbb{R}^2$ be a piecewise continuously differentiable curve. Assume the trace does not contain the origin. Let $\pi_n : [0,nT] \longrightarrow \mathbb{R}^2$ ($n \in \mathbb{N}$) be a* *sequence of polygons such that*

$$\frac{1}{\operatorname{length}\pi_n}\sup_{t\in[0,nT]}\left|(B_n^{-1}\pi_n)'(t)-c'(t)\right| \xrightarrow{n\to\infty} 0. \tag{33}$$

Then, in $\mathcal{L}^1(P)$,

$$\frac{1}{\operatorname{length}\pi_n}\mathcal{I}(P_{L_n^\pi}) \xrightarrow{n\to\infty} \frac{1}{\operatorname{length} c}\int_0^T h_{c'(t)}(P)\,dt.$$

Proof We have to show that

$$\lim_{n\to\infty}\left\|\frac{1}{\operatorname{length}\pi_n}\mathcal{I}(P_{L_n^\pi})-\frac{1}{\operatorname{length} c}\int_0^T h_{c'(t)}(P)\,dt\right\|_{\mathcal{L}^1(P)}=0. \tag{34}$$

Without loss of generality we can assume that c has no self-intersections and that c is parametrized by arc length. As can be seen by the construction of the entropy for polygons,

$$\left\|\frac{1}{\operatorname{length}\pi_n}\mathcal{I}(P_{L_n^\pi})-h_{\pi_n}(P)\right\|_{\mathcal{L}^1(P)}$$

converges to 0. By the representation formula in Remark 29 and since $\pi_n'(t) = ((B_n\pi_n)^{-1})'(t/n)$, for all $t\in[0,nT]$, we obtain

$$h_{\pi_n}(P)=\frac{1}{\operatorname{length}\pi_n}\int_0^{nT} h_{\pi_n'(r)}(P)\,dr=\frac{n}{\operatorname{length}\pi_n}\int_0^T h_{((B_n\pi_n)^{-1})'(t)}(P)\,dt,$$

and by (33) and Lemma 3.4, the integral converges to $\int_0^T h_{c'(t)}\,dt$. Use $\|\cdot\|$ for the euclidian norm in the plane. Using

$$\frac{1}{n}\operatorname{length}\pi_n=\int_0^T\|\pi_n'(nt)\|\,dt=\int_0^T\|(B_n^{-1})'(t)\|\,dt,$$

we obtain by (33)

$$\lim_{n\to\infty}\frac{1}{n}\operatorname{length}\pi_n=\int_0^T\|c'(t)\|\,dt=\operatorname{length} c.$$

This implies that

$$\left\| h_{\pi_n}(P) - \frac{1}{\text{length}\, c} \int\limits_0^T h_{c'(t)}(P)\, dt \right\|_{\mathcal{L}^1(P)}$$

converges to 0 as well, and (34) follows by the triangle inequality. □

Note that the limits do not depend on the sequence of polygons we used to approximate the curve and that any approximation of the curve by lattice points can be described by a lattice approximation of a suitable polygon. This justifies the

Definition 4.4 Let P and be as in Theorem 4.3. Then

$$h_c(P) := \frac{1}{\text{length}\, c} \int\limits_0^T h_{c'(t)}(P)\, dt$$

is called *specific entropy of* P *along* c.

Note that the condition that the trace of c does not contain the origin is no real restriction for the definition as the expression on the left-hand side only depends on the derivative of c. This is intuitive, because we assumed that P is stationary. Note the following property for the entropies of the blowups of a curve defined in (20). The proof is a simple scaling argument.

Corollary 4.5 Let $c : [0, T] \longrightarrow \mathbb{R}^2$ be a piecewise continuously differentiable, and let $B_\eta c : [0, \eta T] \longrightarrow \mathbb{R}^2$ with $B_\eta c(t) = \eta\, c(\frac{t}{\eta})$ $(\eta > 0)$ be the family of its blowups. Then $h_{B_\eta c}(P) = h_c(P)$ for all $\eta > 0$.

4.4 Contour approximation

In statistical mechanics, a *contour* is a set of sites corresponding unambiguously to a chain of bonds. Note that the lattice approximation is not a contour in this sense. The last site before a new step is catercornered from the first site of the step, so not connected by a bond, and closing the gap is not uniquely defined. We define the *contour approximation* by adding, at each new step, the site which is one unit below it. A new step begins in $i + 1$ if and only if $\tau_\lambda^i(\{a\}) \geq 1 - \lambda$, and the site we add in this case is $L_{\lambda,a}(I) - (0, 1)$. For $z \in \mathbb{Z}$ and $u \in \mathbb{N}$,

$$\begin{aligned} &\widehat{L}_{\lambda,a}(z, \ldots, z + u) \\ &\quad := L_{\lambda,a}(z, \ldots, z + u) \cup \left\{ L_{\lambda,a}(i) - (0, 1) \,\middle|\, 0 \leq i \leq u - 1 \wedge \tau_\lambda^{z+i}(\{a\}) \geq 1 - \lambda \right\} \end{aligned}$$

be the contour approximation of the line segment $l_{\lambda,a}(I)$. Lemmas 3.1 and 3.4 translate immediately to $\widehat{L}$. With little modifications we can prove contour versions of the limit theorems derived earlier in this section. We will sketch the results here and refer to [1] for details and proofs.

For $n \in \mathbb{N}$, define $\widehat{L}_n(a) := \widehat{L}_{\lambda,a_n}(z_n, \ldots, z_n + u_n)$, and for $a \in \mathbb{R}$ and $z, \widetilde{z} \in \mathbb{Z}$ with $\widetilde{z} \geq z$ define

$$\widehat{L}^{\sharp}_{\lambda,a}(\widetilde{z}, \ldots, z) := \begin{cases} \widehat{L}_{\lambda,a}(\widetilde{z}-1, \ldots, z) \cup \left(L_{\lambda,a}(\widetilde{z}) - (0,1)\right) & \text{if } \tau_{\lambda}^{\widetilde{z}-1}(\{a\}) \geq 1-\lambda, \\ \widehat{L}_{\lambda,a}(\widetilde{z}-1, \ldots, z) & \text{otherwise.} \end{cases} \tag{35}$$

The specific contour entropy along a line $\widehat{h}_\lambda(P)$ with slope λ is defined as

$$\frac{1}{1+\lambda}\left(\int_0^1 E\left[H\left(P_0[\cdot\,|\,\mathcal{F}\left(\widehat{L}^{\sharp}_{\lambda,t}(-\mathbb{N})\right)](\omega)\right)\right] dt \right.$$
$$\left. + \int_{1-\lambda}^{1} E\left[H\left(P_0[\cdot\,|\,\mathcal{F}\left(\widehat{L}_{\lambda,t}(-\mathbb{N}) \cup \{(0,1)\}\right)](\omega)\right)\right] dt\right).$$

We can show a contour version of Theorem 4.1:

$$\frac{1}{|\widehat{L}_n(a)|}\mathcal{I}(P_{\widehat{L}_n(a)}) \xrightarrow{n\to\infty} \widehat{h}_\lambda(P) \quad \text{in } \mathcal{L}^1(P) \text{ and uniformly in } a \in \mathbb{R}.$$

The formulations of contour versions of Theorems 4.2 and 4.3 are now obvious.

5 Gibbs measures and specific entropies

A collection $(U_V)_{V\subset\mathbb{Z}^d \text{ finite}}$ of functions on Ω is called *stationary summable interaction potential* if the following three conditions are fulfilled: (i) U_V is measurable with respect to $\mathcal{F}_V$ for all $V \subset \mathbb{Z}^d$. (ii) For all $i \in \mathbb{N}$ and all finite $V \subset \mathbb{Z}^d$, $U_{V+i} = U_V \circ \theta_i$. (iii) $\sum_{V\subset\mathbb{Z}^d \text{ finite}: 0\in V} \| U_V \|_\infty < \infty$. Let $\xi, \eta \in \Omega$ be two configurations. The *conditional energy* of ξ on V given the environment η on $\mathbb{Z}^d \backslash V$ is defined as

$$E_V(\xi|\eta) = \sum_{A\subset\mathbb{Z}^d \text{ finite}: A\cap V\neq\emptyset} U_A\left((\xi,\eta)_V\right),$$

where $(\xi,\eta)_V$ is the element of Ω given by $(\xi,\eta)_V(i) := \xi(i)$, for $i \in V$, and $(\xi,\eta)_V(i) := \eta(i)$ for $i \in \mathbb{Z}^d \backslash V$. P is called *Gibbs measure* with respect to U if for any finite subset V of $\mathbb{Z}^d$ the conditional distribution of ω_V under P with respect to $\mathcal{F}_{\mathbb{Z}^d\backslash V}$ is given by

$$P[\omega_V = \xi_V\,|\,\mathcal{F}_{\mathbb{Z}^d\backslash V}](\eta) = \frac{1}{Z_V(\eta)} e^{-E_V(\xi|\eta)}, \text{ where } Z_V(\eta) := \int_\Omega e^{-E_V(\xi|\eta)}\, P(d\xi)$$

is called *partition function*. We say that there is a *phase transition* if there is more than one Gibbs measure with respect to the same interaction potential.

Assume that Υ is furnished with a total order $\leq$, and denote by $-$ the minimal and by $+$ the maximal element in Υ. Suppose that U is *attractive* with respect to the order on Υ, in the sense of (9.7) in [23]. Let P^- and P^+ denote the minimal and the maximal Gibbs measure with respect to U, and let $P^\alpha = \alpha P^- + (1-\alpha)P^+$ $(0 < \alpha < 1)$ be their mixtures. Both P^- and P^+ are ergodic and, as follows from [7], they fulfill the strong 0-1 law and the global Markov property. Föllmer and Ort [9] define the *specific relative entropy* based on hyperspaces by

$$s(P^-, P^+) = \frac{1}{d}\sum_{l=1}^{d}\int_\Omega H\left(P_0^-\left[\cdot\middle|\mathcal{F}^{(l)}\right](\omega), P_0^+\left[\cdot\middle|\mathcal{F}^{(l)}\right](\omega)\right) P^-(d\omega), \tag{36}$$

where $\mathcal{F}^{(l)}$ is the σ-algebra generated by those coordinates in $\{(i^{(1)},\ldots,i^{(d)}) \in \mathbb{Z}^d \mid i^{(l)} = 0\}$ which precede 0 in the lexicographical order on $\mathbb{Z}^d$.

In the two-dimensional case, the conditions in (36) are simply along the coordinate axes. Based on the work in Sect. 4, we can now extend this definition to a surface-order entropy along any direction $v \in S^1$. Furthermore, we can introduce an entropy along curves.

Definition 5.1 Let $v \in S^1$. Let $c : [0, T] \longmapsto \mathbb{R}^2$ be a piecewise differentiable curve parametrized by arc length with right derivative c'.

$$h_v(P^-, P^+) := \int_0^1\int_\Omega H\left(P_0^-\left[\cdot\middle|\mathcal{P}_{\lambda(v),t}\right](\omega), P_0^+\left[\cdot\middle|\mathcal{P}_{\lambda(v),t}\right](\omega)\right) P^-(d\omega)dt$$

is called specific relative entropy of P^- with respect to P^+ *in direction* v.

$$h_c(P^-, P^+) := \frac{1}{\text{length}\, c}\int_0^T h_{c'(t)}(P^-, P^+)\, dt$$

is called specific relative entropy of P^- with respect to P^+ along c.

The order on Υ induces an order on the set $\mathcal{M}_1(\Upsilon)$ of probability measures on Υ : We say that μ is *larger* then ν if the density $\frac{d\mu}{d\nu}$ is an increasing function with respect to the order on Υ, and in this case we write $\mu \geq \nu$. In particular, ν is absolutely continuous with respect to μ. The following inverse triangle inequality for relative entropies was shown in the proof of Theorem 4.2 in [9]. For the reader's convenience we state it in the following form.

Lemma 5.2 *Let $\lambda \geq \mu \geq \nu$, and assume that μ is bounded below by a positive constant. Then $H(\nu, \lambda) \geq H(\nu, \mu) + H(\mu, \lambda)$.*

6 Lower bound

Let P^- and P^+ be the minimal and maximal Gibbs measure defined in Sect. 5. The following lower bound for the large deviations of the *empirical field* $R_n(\omega) := \sum_{i\in V_n} \delta_{\theta_i\omega}$ of P^+ was proved by Föllmer and Ort [9]. Recall that V_n is the set of all lattice sites in $[-n, n]^d$, and that the boundary of a subset V of $\mathbb{Z}^d$ is defined as $\partial V = \left\{i \in \mathbb{Z}^d \backslash V \,\middle|\, \text{dist}(i, V) = 1\right\}$.

Theorem 6.1 *For any open* $A \in \mathcal{M}_1(\Omega)$,

$$\liminf_{n\to\infty} \frac{1}{|\partial V_n|} \log P^+[R_n \in A] \geq -\inf_{\alpha: P_\alpha \in A} \sqrt{\alpha}\, s(P^-, P^+).$$

The aim of this section is to improve the lower bound by replacing the boxes by more general shapes in the two-dimensional case. The corresponding Shannon–McMillan theorems developed in Sect. 4 will be the key to the proof. For a closed curve c let $\text{int}\, c$ be the subset of $\mathbb{R}^2$ surrounded by c. Define the sets

$$C_\alpha := \Big\{ c \,\Big|\, c : [0, T] \longrightarrow \mathbb{R}^2 \text{ closed piecewise } C^1\text{-curve parametrized by arc length, without self-intersections, and with area int } c = \alpha \Big\}. \tag{37}$$

Theorem 6.2 *For any open* $A \in \mathcal{M}_1(\Omega)$,

$$\liminf_{n\to\infty} \frac{1}{|\,\partial V_n\,|} \log P^+[R_n \in A] \geq -\inf_{\alpha: P_\alpha \in A} \inf_{c \in \mathcal{C}_\alpha} \frac{1}{4} \int_0^T \frac{dt}{\sqrt{1 + \lambda(c'(t))^2}}\, h_c(P^-, P^+).$$

Remark 6.3 Replacing the class C_α by squares with area α this bound coincides with the bound in Theorem 6.1: Let π be a square parametrized by arc length and with area $\text{int}\, \pi = \alpha$. Then the length of every edge is $\sqrt{\alpha}$. For the two horizontal edges of the square the slope λ (cf. (32)) is 0 with respect to the x-axis, and for the vertical edges it is 0 with respect to the y-axis. Therefore, the integral equals $4\sqrt{\alpha}$. The entropy $h_\pi(P^-, P^+)$ equals $s(P^-, P^+)$, since the σ-algebras $\mathcal{P}_{0,t}$ coincide with $\mathcal{F}^{(2)}$ for the horizontal edges and with $\mathcal{F}^{(1)}$ for the vertical edges.

Remark 6.4 In the case where the Markov property holds only with respect to the contour boundary we can state a bound similar to the one in Theorem 6.2 by replacing the lattice approximation by the contour approximation. The proof is essentially the same as for Theorem 6.2 (cf. [1] for details).

The rest of this section is devoted to the proof of Theorem 6.2. We will need some properties of the geometry of lattice approximations of polygons and their interplay with the random field. To begin with, we restate explicitly the global Markov property for random fields in the case when the conditioning is concentrated on a set of sites surrounded by a closed polygon π without self-intersections. We use the notation $\Gamma(c) := \operatorname{int} c \cap \mathbb{Z}^2$ to indicate the set of lattice points surrounded by a closed curve c. By the definition (27), $\partial\left(\mathbb{Z}^2\backslash\Gamma(\pi)\right) = L^\pi$, and the global Markov property (cf. Definition (2.1)) with $V = \mathbb{Z}^2\backslash\Gamma(\pi)$, we have that for any $\mathcal{F}(\mathbb{Z}^2\backslash\Gamma(\pi))$-measurable nonnegative function Φ,

$$E\left[\phi \,\middle|\, \mathcal{F}_{\Gamma(\pi)}\right] = E\left[\phi \,\middle|\, \mathcal{F}_{L^\pi}\right]. \tag{38}$$

We will further need two lemmata that compute the asymptotic fractions of the lengths of the blowups of a line segment, or a polygon, and the sizes of their lattice approximations.

Lemma 6.5 *Let I be a real interval, $l(x) = \lambda x + a$ be a linear function with slope λ, and B_k $(k \in \mathbb{N})$ be the sequence of its blowups restricted to I. If L_k is the lattice approximation of B_k then*

$$\lim_{k\to\infty} \frac{|\,L_k\,|}{\text{length } B_k} = \frac{1}{\sqrt{1+\lambda^2}}.$$

Proof We consider only the case when $0 \le \lambda \le 1$, that is, when the lattice approximation is given by $L(z) = (z, [l(z)])$ $(z \in I \cap \mathbb{Z})$. Other cases only differ in terms of notation. For any $k \in \mathbb{N}$, $|L_k|$ is either $[\text{length } b_k]$ or $[\text{length } b_k] + 1$, where b_k is the projection of B_k to the x-axis. We can ignore the second case, since the additional point does not matter for the limit. Observe that $(\text{length } B_k)^2 = (\text{length } b_k)^2 + (\lambda \text{ length } b_k)^2$. Consequently, $\text{length } b_k = \text{length } B_k \,/\, (\sqrt{1+\lambda^2})$, which proves the convergence. □

Lemma 6.6 *Let π be a polygon with edges π_r $(r = 1, \dots, R)$ and λ_r $(r = 1, \dots, R)$ their slopes as defined in (32). Let $B_k\pi$ $(k \in \mathbb{N})$ be the blowups of π and $L_k\pi$ $(k \in \mathbb{N})$ their lattice approximations. Then*

$$\lim_{k\to\infty} \frac{|\,L_k\pi\,|}{\text{length } B_k\pi} = \sum_{r=1}^{R} \frac{1}{\sqrt{1+\lambda_r^2}} \frac{\text{length } \pi_r}{\text{length } \pi}.$$

Proof Using $\text{length } B_k\pi_r = k \text{ length } \pi_r$, we obtain

$$\frac{|\,L_k\pi\,|}{\text{length } B_k\pi} = \sum_{r=1}^{R} \frac{|\,L_k\pi\,|}{\text{length } B_k\pi_r} \frac{\text{length } B_k\pi_r}{\text{length } B_k\pi} = \sum_{r=1}^{R} \frac{|\,L_k\pi\,|}{\text{length } B_k\pi_r} \frac{\text{length } \pi_r}{\text{length } \pi}.$$

By the previous lemma applied to the individual sides, the first factors converge to $1\big/\sqrt{1+\lambda_r^2}$, which prove the statement of the lemma. □

Proof of Theorem 6.2 Let be $0 < \alpha \le 1$, such that $P_\alpha \in A$. Since A is open, we can choose open neigborhoods A^- and A^+ of P^- respectively P^+ in $\mathcal{M}_1(\Omega)$ such that $\alpha A^- + (1-\alpha)A^+ \subseteq A$. Without loss of generality we may assume that A^- and A^+ are in $\mathcal{F}_{V_p}$ for some $p \in \mathbb{N}$. Define the set $\Pi_\alpha := \{\pi \,|\, \pi$ closed polygon without self-intersections, area int $\pi = \alpha\}$. Let $\pi \in \Pi_{4\alpha}$ with $0 \in \operatorname{int}\pi$, and let $(B_n\pi)_{n\in\mathbb{N}}$ be the sequence of blowups of π. For $\alpha = 1$ take $C_n := V_n$. Otherwise, define

$$C_n := \Gamma(B_{k(n)}\pi) \quad \text{and} \quad D_n := V_n \backslash \Gamma(B_{l(n)}\pi),$$

where $k(n)$ and $l(n)$ are chosen such that $k(n) \le l(n)$, $l(n) - k(n) \xrightarrow{n\to\infty} \infty$,

$$\lim_{n\to\infty} \frac{|C_n|}{|V_n|} = \alpha \quad \text{and} \quad \lim_{n\to\infty} \frac{|D_n|}{|V_n|} = 1 - \alpha. \tag{39}$$

To see that such sequences exist we show that $k(n) = \left[\sqrt{\alpha|V_n|/\text{area int}\,\pi}\,\right]$ and $l(n) = \left[k(n) + \sqrt{n}\,\right]$ fulfill the conditions. Obviously, both the sequences and their difference tend to infinity as n goes to infinity. Using area int $(B_k\pi) = k^2$area int π,

$$\frac{|\Gamma(B_k\pi)|}{\text{area int } B_k\pi} \xrightarrow{k\to\infty} 1 \quad \text{and} \quad \frac{k(n)^2}{|V_n|} \xrightarrow{n\to\infty} \frac{\alpha}{\text{area int}\,\pi}. \tag{40}$$

We obtain for the first expression in (39)

$$\lim_{n\to\infty} \frac{|C_n|}{|V_n|} = \lim_{n\to\infty} \frac{\text{area int }(B_{k(n)}\pi)}{|V_n|} = \lim_{n\to\infty} \frac{k(n)^2 \text{ area int}\,\pi}{|V_n|} = \alpha.$$

Similarly, we see for the second expression in (39)

$$\lim_{n\to\infty} \frac{|D_n|}{|V_n|} = 1 - \lim_{n\to\infty} \frac{|\Gamma(B_{l(n)})|}{|V_n|} = 1 - \lim_{n\to\infty} \frac{l(n)^2 \text{ area int}\,\pi}{|V_n|}.$$

By definition of $l(n)$, $l(n)^2 = k(n)^2 + [2k(n)\sqrt{n}] + n$. The last two addends are of order n and will go to 0 when divided by $|V_n|$. It remains to study $k(n)^2$ area int $\pi/|V_n|$, but we already know from the second statement in (40) that this converges to α.

Define

$$R_n^- = \frac{1}{|C_{n,p}|} \sum_{i\in C_{n,p}} \delta_{\theta_i}\omega \quad \text{and} \quad R_n^+ = \frac{1}{|D_{n,p}|} \sum_{i\in D_{n,p}} \delta_{\theta_i}\omega,$$

where $C_{n,p} := \Gamma(B_{k(n)-p}\pi)$ and $D_{n,p} := V_{n-p} \backslash \Gamma(B_{l(n)+p}\pi)$. Then $\{R_n^- \in A^-\} \in \mathcal{F}_{C_n}$, $\{R_n^+ \in A^+\} \in \mathcal{F}_{D_n}$, for large enough n, $\{R_n \in A\} \supseteq \{R_n^- \in A^-\} \cap \{R_n^+ \in A^+\} := \Lambda_n$. Define the measures

$$Q_n = P_{C_n}^- \otimes P_{\mathbb{Z}^2\backslash C_n}^+ \quad (n \in \mathbb{N}).$$

Q_n coincides with P^- on $\mathcal{F}_{C_n}$ and with P^+ on $\mathcal{F}_{D_n}$, and makes these σ-fields independent. Thus we obtain $Q_n[\Lambda_n] = P^-[R_n^- \in A^-]\, P^+[R_n^+ \in A^+]$, and by the ergodic behaviour of P^- and P^+, the sequence $Q_n[\Lambda_n]$ $(n \in \mathbb{N})$ converges to 1.

Let ϕ_n denote the density of Q_n with respect to P^+ on $\mathcal{F}_{C_n \cup D_n}$. Then for $\gamma > 0$, $\varepsilon > 0$, and for large enough n,

$$\begin{aligned} P^+[R_n \in A] &\geq P^+[\Lambda_n] \\ &\geq \int 1_{\Lambda_n \cap \{|\partial V_n|^{-1} \log \phi_n \leq \gamma + \varepsilon\}} \phi_n^{-1} \, dQ_n \\ &\geq \exp(-(\gamma+\varepsilon)\,|\partial V_n|)\, Q_n \left[\Lambda_n \cap \left\{ |\partial V_n|^{-1} \log \phi_n \leq \gamma + \varepsilon \right\} \right]. \end{aligned}$$

By the convergence of the $Q_n[\Lambda_n]$, the lower bound $\liminf_{n\to\infty} |\partial V_n|^{-1} \log P^+[R_n \in A] \geq -\gamma$ follows if γ is chosen such that, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} Q_n \left[|\partial V_n|^{-1} \log \phi_n \leq \gamma + \varepsilon \right] = 1. \tag{41}$$

We will show that (41) holds with

$$\gamma = \sum_{r=1}^{R} \frac{1}{\sqrt{1+\lambda_r^2}} \frac{\text{length}\, \pi_r}{8} \; h_\pi(P^-, P^+).$$

Since $Q_n = P^+$ on D_n, and the fact that both P^- and P^+ are Gibbs measures with respect to the same potential we obtain

$$\phi_n(\omega) = \frac{P^-[\omega_{C_n}]P^+[\omega_{D_n}]}{P^+[\omega_{C_n \cup D_n}]} = \frac{P^-[\omega_{C_n}]}{P^-[\omega_{C_n} \mid \omega_{D_n}]}.$$

Let L_n be the lattice approximation of $B_n\pi$. By (38),

$$\begin{aligned} P^-[\omega_{C_n} \mid \omega_{D_n}] &= P^-[\omega_{D_n} \mid \omega_{C_n}] \frac{P^-[\omega_{D_n}]}{P^-[\omega_{C_n}]} \\ &= P^-[\omega_{D_n} \mid \omega_{L_{k(n)}}] \frac{P^-[\omega_{D_n}]}{P^-[\omega_{C_n}]} = P^-[\omega_{L_{k(n)}} \mid \omega_{D_n}] \frac{P^-[\omega_{L_{k(n)}}]}{P^-[\omega_{C_n}]}, \end{aligned}$$

and thus

$$\phi_n(\omega) = \frac{P^-[\omega_{L_{k(n)}}]}{P^-[\omega_{L_{k(n)}} \mid \omega_{D_n}]} = \frac{P^-[\omega_{L_{k(n)}}]}{P^+[\omega_{L_{k(n)}} \mid \omega_{D_n}]}. \tag{42}$$

Going around the R sides of $L_{k(n)}$, and conditioning site by site as in the proof of Theorem 4.2, we obtain

$$\frac{1}{|V_n|} \log \phi_n(\omega) = \frac{1}{|V_n|} \sum_{r=1}^{R} \Psi^{(r)},$$

where the $\Psi^{(r)}$ corresponds to the rth side of the polygon. Similar to the calculation between (30) to (31) we obtain

$$\Psi^{(r)} = \sum_{i=0}^{u_n} Z_{n,i,t} \circ \theta_{L_{\lambda,\tau_\lambda^{z_n}(\{t_n\})}(i)} \circ \theta_{L_{\lambda,t_n}(z_n)},$$

where λ is the slope of the rth side of the polygon, t_n and r_n are as in Sect. 4.2, and $Z_{n,i,t} = X_{n,i,t} - Y_{n,i,t}$, with

$$X_{n,i,t} = \log P_0^- \left[\omega(0,0) \,\Big|\, \omega \left[\widehat{L}^{\sharp}_{\lambda,\tau_\lambda^{z_n+i}(\{t_n\})}(-1,\ldots,-i) \cup A_{n,i,t} \right] \right],$$

$$\text{and } Y_{n,i,t} = \log P_0^+ \left[\omega(0,0) \,\Big|\, \omega \left[\widehat{L}^{\sharp}_{\lambda,\tau_\lambda^{z_n+i}(\{t_n\})}(-1,\ldots,-i) \cup B_{n,i,t} \right] \right].$$

To simplify notation we have omitted the index r. For the sets in the conditional expectations we have $A_{n,i,t} \subseteq B_{n,i,t} \subseteq \mathbb{Z}^2 \setminus \left(H_n - L_{\lambda,t_n}(z_n+i) \right)$. $A_{n,i,t}$ is obtained by shifting a subset of $L_n \subseteq C_n$. H_n is constructed as in the paragraph above, but using the minimum of the diameter d_n and the distance $l(n) - k(n)$ in place of d_n.

To prove convergence, we study the X and Y-parts separately. Because of the way the sets $A_{n,i,t}$ are constructed, the behavior of $X_{n,i,t}$ under Q_n is the same as under P^-. But the proof of Theorem 4.2 shows that

$$\sum_{i=0}^{u_n} X_{n,i,t} \circ \theta_{L_{\lambda,\tau_\lambda^{z_n}(\{t_n\})}(i)} \circ \theta_{L_{\lambda,t_n}(z_n)}$$

converges to $-h_\pi(P^-)$ in $\mathcal{L}^1(P^-)$. The convergence remains true when we replace $X_{n,i,t}$ by

$$X^-_{n,i,t} := \log P_0^- \left[\omega(0,0) \,\Big|\, \omega \left[L_{\lambda,\tau_\lambda^{z_n+i}(\{t_n\})}(-1,\ldots,-i) \right]^- \right],$$

where, for a subset L of $\mathbb{Z}^2$, the element $\omega(L)^-$ equals ω on L and assumes the minimal state in Υ outside of $H_n - L_{\lambda,t_n}(z_n+i)$. To control the behavior of $Y_{n,i,t}$ under Q_n define $Z^-_{n,i,t} = X^-_{n,i,t} - Y_{n,i,t}$. Use the law of large numbers for martingales with bounded increments in its $\mathcal{L}^2$-form to replace

$$\frac{1}{|L_{k(n)}|} \sum_{i=0}^{u_n} Z^-_{n,i,t} \circ \theta_{L_{\lambda,\tau_\lambda^{z_n}(\{t_n\})}(i)} \circ \theta_{L_{\lambda,t_n}(z_n)}$$

by

$$\frac{1}{|L_{k(n)}|} \sum_{i=0}^{u_n} E\left[Z^-_{n,i,t} \circ \theta_{L_{\lambda,\tau_\lambda^{z_n}(\{t_n\})}(i)} \circ \theta_{L_{\lambda,t_n}(z_n)} \,\Big|\, \mathcal{A}_{n,i,t} \right],$$

where $\mathcal{A}_{n,i,t}$ is the σ-field generated by the sites in D_n and those sites of $L_{k(n)}$ which precede i in the canonical ordering of $L_{k(n)}$. These conditional expectations can be written as the relative entropy $H(\nu, \mu)$, with the random measures

$$\mu(\omega) := P_0^- \left[\cdot \,|\, \omega \left(L_{\lambda, t_\lambda^{z_n+i}(\{t_n\})}(-1, \ldots, -i) \right)^- \right]$$
$$\text{and} \quad \nu(\omega) := P_0^+ \left[\cdot \,|\, \omega \left(L_{\lambda, \tau_\lambda^{z_n+i}(\{t_n\})}(-1, \ldots, -i) \cup B_{n,i,t} \right) \right].$$

Now we want to replace μ by a measure η for which

$$\left(|\, L_{\lambda, \tau_\lambda^{z_n+i}(\{t_n\})}(-1, \ldots, -i) \,| \right)^{-1} \sum_{i=0}^{u_n} H(\nu, \eta) \circ \theta_{L_{\lambda, \tau_\lambda^{z_n}(\{t_n\})}(i)} \circ \theta_{L_{\lambda, t_n}(z_n)} \tag{43}$$

converges to $h_\lambda(P^-, P^+)$, in $\mathcal{L}^1(P^-)$, as n goes to infinity. Define $\omega(L)^+$ in analogy to $\omega(L)^-$. Since for *all* ω,

$$P_0^- \left[\cdot \,\middle|\, \omega \left(L_{\lambda,t}(-1, \ldots, -i) \right)^+ \right] \xrightarrow{i \to \infty} P_0^- \left[\cdot \,\middle|\, \mathcal{P}_{\lambda,t} \right],$$

we obtain (43) by taking

$$\eta(\omega) := P_0^- \left[\cdot \,\middle|\, \omega \left(L_{\lambda, \tau_\lambda^{z_n+i}(\{t_n\})}(-1, \ldots, -i) \right)^+ \right].$$

By Lemma 5.2, $H(\nu(\omega), \mu(\omega)) \leq H(\nu(\omega), \eta(\omega))$. Summing over $r = 1, \ldots, R$, and passing from convergence in $\mathcal{L}^1(P^-)$ to stochastic convergence with respect to Q_n yields

$$\lim_{n \to \infty} Q_n \left[|\, L_{k(n)} \,|^{-1} \phi_n > h_\pi(P^-, P^+) + \varepsilon \right] = 0$$

for any $\varepsilon > 0$. To derive (41) it remains to show that

$$\lim_{n \to \infty} \frac{|\, L_{k(n)} \,|}{|\, \partial V_n \,|} = \sum_{r=1}^{R} \frac{1}{\sqrt{1 + \lambda_r^2}} \frac{\text{length}\, \pi_r}{8}. \tag{44}$$

The fraction on the left-hand side can be written as a product:

$$\frac{|\, L_{k(n)} \,|}{|\, \partial V_n \,|} = \frac{|\, L_{k(n)} \,|}{\text{length}\, B_{k(n)} \pi} \cdot \frac{k(n)\, \text{length}\, \pi}{|\, \partial V_n \,|}.$$

We first study the asymptotics of the second factor: As we are only interested in the limit behaviour, we can drop the brackets for the integer part in the definition of $k(n)$ and just use $\sqrt{\alpha |\, V_n \,| / \text{area int}\, \pi}$. The denominator equals $4 \,|\, V_n \,|$. Reducing the fraction yields $1/8$. The limit of the first factor in (6) was computed in Lemma 6.6. So, the second factor in (6) converges to length $\pi/8$. Combining all yields (44).

Now, we replace the polygon π by the polygon $\widetilde{\pi} = B_{1/2}\pi$. Since length $\widetilde{\pi}_r =$ length $\pi_r/2$, and since, by Corollary 4.5, $h_{\widetilde{\pi}}(P^-, P^+) = h_\pi(P^-, P^+)$,

$$\gamma = \sum_{r=1}^{R} \frac{1}{\sqrt{1+\lambda_r^2}} \frac{\text{length}\,\widetilde{\pi}_r}{4} h_{\widetilde{\pi}}(P^-, P^+).$$

Finally, by Lemma 3.4, the infimum of that function over all polygons $\widetilde{\pi} \in \Pi_\alpha$ equals the infimum over all curves $c \in C_\alpha$. □

Acknowledgements This work was a part of my PhD thesis, and I thank my supervisor Hans Föllmer for suggesting to extend the concept of surface entropies to general shapes via a stepwise approach, and to use them to refine the large deviation lower bounds in the phase-transition regime. I am grateful to Hans-Otto Georgii for noticing and correcting a mistake in the constant in the lower bound, and I thank Dima Ioffe for interesting discussions. I also want to thank the referee for insightful and supportive advice. Finally, I am very grateful to Kristan Aronson for asking a very good question about this work.

References

1. Brettschneider, J.: Shannon–McMillan theorems for random fields along curves and lower bounds for surface-order large deviations. Eurandom. Technical report 2001-018
2. Comets, F.: Grandes deviations pour des champs de gibbs sur F^d. C. R. Acad. Sci. Paris Ser. I **303**, 511–513 (1986)
3. Deuschel, J.-D., Stroock, D.W.: Large Deviations. Academic, Boston (1989)
4. Dobrushin, R., Kotecky, R., Shlosman, S.: Wulff construction. A global shape from local interaction. In: AMS Translations of Mathematical Monographs, vol. 104. Providence (1992)
5. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time. II. Commun. Pure Appl. Math. **28**, 279–301 (1975)
6. Föllmer, H.: On entropy and information gain in random fields. Z. Wahrscheinlichkeitstheor. Verw. Geb. **26**, 207–217 (1973)
7. Föllmer, H.: On the global Markov property. In: Quantum Fields—Algebras, Processes, Proc. Symp., Bielefeld 1978, pp. 293–302. Springer, Heidelberg (1980)
8. Föllmer, H., Orey, S.: Large deviations for the empirical field of a Gibbs measure. Ann. Probab. **16**(3), 961–977 (1988)
9. Föllmer, H., Ort, M.: Large deviations and surface entropy for Markov fields. Astérisque **288**, 173–190 (1988)
10. Georgii, H.-O.: Gibbs Measures and Phase Transitions. W. de Gruyter, Berlin (1988)
11. Ioffe, D.: Large deviations for the 2D Ising model: a lower bound without cluster expansion. J. Stat. Phys. **74**, 411–432 (1994)
12. Ioffe, D.: Exact large deviation bounds up to T_c for the Ising model in two dimensions. Probab. Theory Relat. Fields **102**(3), 313–330 (1995)
13. Israel, R.B.: Some examples concerning the global Markov property. Commun. Math. Phys. **105**, 669–673 (1986)
14. Krengel, U.: Ergodic Theorems. W. de Gruyter, Berlin (1985)
15. Lanford, O.E., Ruelle, D.: Observables at infinity and states with short range correlations in statistical mechanics. Commun. Math. Phys. **13**, 194–215 (1969)
16. Milnor, J.: Directional entropies of cellular automaton-maps. In: Disorderd Systems and Biological Organization (Les Houches, 1985), vol. 20, pp. 113–115. NATO Adv. Sci. Inst. Ser. F: Comut. Systems Sci., New York (1986)
17. Milnor, J.: On the entropy geometry of cellular automata. Complex Syst. **2**(3), 357–385 (1988)
18. Olla, S.: Large deviations for Gibbs random fields. Probab. Theory Relat. Fields **77**(3), 343–357 (1988)
19. Park, K.K.: Continuity of directional entropy. Osaka J. Math. **31**(3), 613–628 (1994)
20. Park, K.K.: Continuity of directional entropy for a class of $\mathbb{Z}^2$-actions. J. Korean Math. Soc. **32**(3), 573–582 (1995)

21. Park, K.K.: Entropy of a skew product with a $\mathbb{Z}^2$-action. Pac. J. Math. **172**(1), 227–241 (1996)
22. Park, K.K.: On directional entropy functions. Isr. J. Math. **113**, 243–267 (1999)
23. Preston, C.: Random Fields. In: LNM, vol. 534. Springer, New York (1976)
24. Schonmann, R.: Second order large deviation estimates for ferromagnetic systems in the phase coexistence region. Commun. Math. Phys. **112**(3), 409–422 (1987)
25. Sinai, Ya.G.: An answer to a question by J. Milnor. Comment. Math. Helv. **60**, 173–178 (1985)
26. Sinai, Ya.G.: Topics in Ergodic Theory. Princeton University Press, Princeton (1994)
27. Thouvenot, J.-P.: Convergence en moyenne de l'information pour l'action de $\mathbb{Z}^2$. Z. Wahrscheinlichkeitstheor. Verw. Geb. **24**, 135–137 (1972)
28. Weizsäcker, H.V.: Exchanging the order of taking suprema and countable intersections of σ-algebras. Ann. Inst. Henri Poincaré B **19**(1), 91–100 (1983)