# AN INTEGRAL REPRESENTATION FOR TOPOLOGICAL PRESSURE IN TERMS OF CONDITIONAL PROBABILITIES

BRIAN MARCUS AND RONNIE PAVLOV

Abstract. Given an equilibrium state $\mu$ for a continuous function $f$ on a shift of finite type $X$, the pressure of $f$ is the integral, with respect to $\mu$, of the sum of $f$ and the information function of $\mu$. We show that under certain assumptions on $f$, $X$ and an invariant measure $\nu$, the pressure of $f$ can also be represented as the integral with respect to $\nu$ of the same integrand. Under stronger hypotheses we show that this representation holds for all invariant measures $\nu$. We establish an algorithmic implication for approximation of pressure, and we relate our results to a result in thermodynamic formalism.

## 1. Introduction

Given a finite alphabet $\mathcal{A}$, the entropy of a shift-invariant measure $\mu$ on $\mathcal{A}^{\mathbb{Z}}$ is sometimes defined as the expected conditional entropy of the present given the past, or

$$h(\mu) = \int H(x_0 \mid x_{-1}, x_{-2}, \ldots)\; d\mu(x_{-1}, x_{-2}, \ldots),$$

where

$$H(x_0 \mid x_{-1}, x_{-2}, \ldots) = \sum_{a \in \mathcal{A}} -\mu(x_0 = a \mid x_{-1}, x_{-2}, \ldots) \log \mu(x_0 = a \mid x_{-1}, x_{-2}, \ldots).$$

Equivalently,

$$h(\mu) = \int -\log \mu(x_0 \mid x_{-1}, x_{-2}, \ldots)\; d\mu(x).$$

It is less well-known that there is an analogue of this formula for a shift-invariant measure $\mu$ on $\mathcal{A}^{\mathbb{Z}^d}$ which involves the notion of lexicographic past. Define $\mathcal{P} \subseteq \mathbb{Z}^d$ to be the set of sites lexicographically less than 0, or equivalently the set of $v \in \mathbb{Z}^d \setminus \{0\}$ whose last nonzero coordinate is negative. Then

$$h(\mu) = \int -\log \mu(x_0 \mid \{x_p\}_{p \in \mathcal{P}})\; d\mu \tag{1}$$

(see [6, Theorem 15.12] or [9, p. 283, Theorem 2.4]).

The integrand in (1), which we will denote by $I_\mu(x)$, is fundamental in ergodic theory and information theory and is known as the information function.

We will mostly be interested in implications of (1) on topological pressure. For any continuous function $f$ on a $\mathbb{Z}^d$ shift of finite type, it is a consequence of the variational principle ([16], [15]) that the measure-theoretic pressure function $P_\rho(f) = h(\rho) + \int f\; d\rho$ achieves its maximum on a nonempty set of measures; such

---

2000 *Mathematics Subject Classification.* Primary: 37D35, 37B50; Secondary: 37B10, 37B40.

*Key words and phrases.* multidimensional shifts of finite type; Markov random field; Gibbs measure; entropy; pressure; equilibrium state.

measures are called equilibrium states of $f$, and the maximum is called the topological pressure $P(f)$. When the function $f$ is locally finite, all such equilibrium states are examples of so-called finite-range Gibbs measures ([8], [14]). For more on equilibrium states and Gibbs measures, see Sections 2.2 and 2.6.

When applied to an equilibrium state $\mu$ for $f$, (1) clearly implies

$$P(f) = \int I_\mu(x) + f(x)\ d\mu. \tag{2}$$

For certain classes of equilibrium states and Gibbs measures, sometimes there are even simpler representations for the pressure. A recent example of this was given by Gamarnik and Katz in [5, Theorem 1], who showed that for any Gibbs measure $\mu$ which has a measure-theoretic mixing property, called strong spatial mixing, and whose support contains a so-called safe symbol 0 (a very strong topological mixing property, defined in Section 2.1),

$$P(f) = I_\mu(0^{\mathbb{Z}^d}) + f(0^{\mathbb{Z}^d}) \tag{3}$$

(here, $0^{\mathbb{Z}^d} \in \mathcal{A}^{\mathbb{Z}^d}$ is the configuration on $\mathbb{Z}^d$ which is 0 at every site of $\mathbb{Z}^d$). This result was the primary motivation for our paper.

They used this simple representation to give a polynomial time approximation algorithm for $P(f)$ in certain cases. Approximation schemes are very important because in most cases it is quite difficult (and sometimes impossible!) to obtain exact, closed form expressions for entropy and pressure ([7]), let alone the exact values of conditional $\mu$-measures of specific cylinder sets or exact values of the function $I_\mu$ that are needed to evaluate (3).

A consequence (Corollary 3.2; see also Corollary 3.3) of one of our main results is that under certain hypotheses on the equilibrium state $\mu$ and its support, the integrand in (2) yields $P(f)$ when integrated against *any* shift-invariant measure $\nu$ whose support is contained within the support of $\mu$:

$$P(f) = \int I_\mu(x) + f(x)\ d\nu. \tag{4}$$

For instance, (3) is the special case of (4) where $\nu$ is the point mass at $0^{\mathbb{Z}^d}$. This consequence is related to a result of Ruelle [14, 4.7b] which characterizes when two interactions have a common Gibbs measure; this connection is discussed in Section 5. The conclusion of Corollary 3.2 is well known when $d = 1$ and the support of $\mu$ is an irreducible shift of finite type.

We have two main results, Theorems 3.1 and 3.4, with different sets of hypotheses for representation of pressure of $f$ with respect to a given invariant measure $\nu$. For each theorem, our hypotheses are of three broad types: a weak topological mixing condition on the support of $\mu$ (see Section 2.5), a type of continuity assumption on $p_\mu$ over $\mathrm{supp}(\nu)$ (see Section 2.3), and a type of positivity assumption on $p_\mu$ over $\mathrm{supp}(\nu)$ (see Section 2.4).

All of these hypotheses, for both theorems, are weaker than those used in [5, Theorem 1], and so each implies (3) by taking $\nu$ to be the point mass at $0^{\mathbb{Z}^d}$.

From Theorem 3.1 we prove Proposition 4.1 which, for $d > 1$ and certain $f$, shows that $P(f)$ can be approximated to within tolerance $\epsilon$ in time $e^{O((\log \frac{1}{\epsilon})^{d-1})}$. This yields a polynomial time approximation scheme when $d = 2$ and generalizes some cases of the approximation algorithms given in [5]. This scheme is more efficient than the approximation scheme of [11], but requires the additional hypothesis that (4) holds for some “simple” $\nu$.

We summarize the remainder of this paper. In Section 2, we give relevant definitions, background and preliminary results. In Section 3, we prove our main results, Theorems 3.1 and 3.4. In Section 4, we describe a consequence of our pressure representation results that yields a method for approximating pressure, and we discuss connections with work of Gamarnik and Katz [5]. Finally, in Section 5, we describe a connection between our results and the thermodynamic formalism of Ruelle [14].

## 2. Definitions and Preliminary Results

### 2.1. **Subshifts.**

We view $\mathbb{Z}^d$ as a graph (the so-called **cubic lattice**), where vectors in $\mathbb{Z}^d$ are the vertices (also sometimes called **sites**), and two vertices $u, v \in \mathbb{Z}^d$ are said to be **adjacent**, and we write $u \sim v$, if $|u - v| = 1$, where $|\cdot|$ is the $L^1$ metric. For subsets $S, T$ of $\mathbb{Z}^d$, $d(S,T)$ denotes the distance between $S$ and $T$ using this metric.

We also will use the **lexicographic ordering** on $\mathbb{Z}^d$, where $v < v'$ if $v \neq v'$ and, for the largest $i$ for which $v_i \neq v_{i'}$, $v_i$ is strictly smaller than $v_{i'}$. The **lexicographic past** $\mathcal{P}$ is the set of all $v \in \mathbb{Z}^d$ smaller than the zero vector.

An **edge** is an unordered pair $(u, v)$ of adjacent sites in $\mathbb{Z}^d$. The **boundary** of a set $S \subset \mathbb{Z}^d$, denoted by $\partial S$, is the set of $v \in S^c$ which are adjacent to some element of $S$. In the case where $S$ is a singleton $\{v\}$, we call the boundary the **set of nearest neighbors** $N_v$. The **inner boundary** of $S$, denoted by $\underline{\partial} S$, is the set of sites in $S$ adjacent to some element of $S^c$, or $\partial(S^c)$.

For any integers $a < b$, we use $[a, b]$ to denote $\{a, a+1, \dots, b\}$.

An **alphabet** $\mathcal{A}$ is a finite set with at least two elements. A **configuration** $u$ on the alphabet $\mathcal{A}$ is any mapping from a non-empty subset $S$ of $\mathbb{Z}^d$ to $\mathcal{A}$, where $S$ is called the **shape** of $u$. For any configuration $u$ with shape $S$ and any $T \subseteq S$, denote by $u(T)$ the restriction of $u$ to $T$, i.e. the subconfiguration of $u$ occupying $T$. For $S, T \subset \mathbb{Z}^d$, $x \in \mathcal{A}^S$ and $y \in \mathcal{A}^T$, $xy$ denotes the configuration on $S \cup T$ defined by $(xy)(S) = x$ and $(xy)(T) = y$, which we call the **concatenation** of $x$ and $y$ (if $S \cap T \neq \emptyset$, this requires that $x(S \cap T) = y(S \cap T)$). For a symbol $a \in \mathcal{A}$ and a subset $S \subseteq \mathbb{Z}^d$, $a^S$ denotes the configuration on $S$ which takes value $a$ at all elements of $S$.

For any $d$, we use $\sigma$ to denote the natural **shift action** on $\mathcal{A}^{\mathbb{Z}^d}$ defined by $(\sigma_v(x))(u) = x(u+v)$.

For any alphabet $\mathcal{A}$, $\mathcal{A}^{\mathbb{Z}^d}$ is a topological space when endowed with the product topology (where $\mathcal{A}$ has the discrete topology), and any subsets will inherit the induced topology. A basis for the topology is the collection of **cylinder sets** which are sets of the form $[w] := \{x \in \mathcal{A}^{\mathbb{Z}^d} \ : \ x(S) = w\}$, where $w$ is a configuration with arbitrary finite shape $S \subseteq \mathbb{Z}^d$.

A **subshift** (or **shift space**) is a closed, translation invariant subset of $\mathcal{A}^{\mathbb{Z}^d}$. An equivalent definition is given as follows. Let $\mathcal{A}^*$ denote the set of all configurations on finite subsets of $\mathbb{Z}^d$, where we often identify two configurations if they differ by a translate. A subset $X$ of $\mathcal{A}^{\mathbb{Z}^d}$ is a subshift iff

$$X = \{x \in \mathcal{A}^{\mathbb{Z}^d} : \ x(S) \notin \mathcal{F} \ \text{ for all finite subsets } S \subset \mathbb{Z}^d\}$$

for some list $\mathcal{F} \subset \mathcal{A}^*$ of configurations on finite subsets. For a subshift $X$, when we wish to emphasize the dimension of the lattice we will refer to $X$ as a $\mathbb{Z}^d$**-subshift**.

In the case where $\mathcal{F}$ can be chosen to be finite, $X$ is called a **shift of finite type** (SFT). In the case where $\mathcal{F}$ consists of configurations only on edges, $X$ is called a **nearest-neighbor shift of finite type**.

The following are prominent examples of nearest neighbor SFT's.

**Example 2.1.** *The **hard square shift** is the nearest-neighbor $\mathbb{Z}^d$-SFT with alphabet $\{0,1\}$ defined by forbidding 1's on any adjacent pair of sites.*

**Example 2.2.** *The $k$-**checkerboard** (or $k$-**coloring**) SFT $C_k^{(d)}$ is the nearest-neighbor $\mathbb{Z}^d$-SFT with alphabet $\{0,1,\ldots,k-1\}$ consisting of all configurations on $\mathbb{Z}^d$ such that letters at adjacent sites must be different.*

**Definition 2.3.** *For any $\mathbb{Z}^d$ subshift $X$, the **language** of $X$ is*

$$\mathcal{L}(X) = \bigcup_{\{S\subset\mathbb{Z}^d,\ |S|<\infty\}} \mathcal{L}_S(X)$$

*where*

$$\mathcal{L}_S(X) = \{x(S) : x \in X\}.$$

Given a forbidden list $\mathcal{F}$ that defines a subshift $X$ and $S \subseteq \mathbb{Z}^d$, every configuration on $S$ that does not contain any element of $\mathcal{F}$ is called **locally admissible**; every configuration on $S$ that extends to an element of $X$ is called **globally admissible**. Note that every locally admissible configuration is globally admissible, but not conversely. Clearly, a finite configuration is globally admissible if and only if it is in $\mathcal{L}(X)$.

**Definition 2.4.** *A nearest-neighbor SFT $X$ is **single-site fillable (SSF)** if for some forbidden list $\mathcal{F}$ of nearest neighbors that defines $X$ and every $\eta \in \mathcal{A}^{N_0}$, there exists $a \in \mathcal{A}^0$ such that $\eta a$ is locally admissible.*

It is easy to see that a nearest-neighbor SFT $X$ satisfies SSF if and only if for some forbidden list $\mathcal{F}$ of nearest neighbors that defines $X$, every locally admissible configuration is globally admissible.

In the definition of SSF above, the symbol $a$ may depend on the configuration $\eta$. This generalizes the concept of a **safe symbol**, which is a symbol $a \in \mathcal{A}^{\{0\}}$ such that $\eta a$ is locally admissible for every configuration $\eta \in \mathcal{A}^{N_0}$ (strictly speaking, the concept of safe symbol applies to a forbidden list on the alphabet of symbols that occur in a point of the subshift). The hard square shift has a safe symbol in every dimension. No checkerboard shift has a safe symbol, but for $k \geq 2d+1$, $C_k^{(d)}$ satisfies SSF.

### 2.2. Markov Random Fields and Gibbs Measures.

We will frequently speak of measures on $\mathcal{A}^{\mathbb{Z}^d}$, and all such measures in this paper will be Borel probability measures. This means that any $\mu$ is determined by its values on the cylinder sets. For notational convenience, rather than referring to a cylinder set $[w]$ within a measure or conditional measure, we just use the configuration $w$. For instance, $\mu(w \cap v \mid u)$ represents the conditional measure $\mu([w] \cap [v] \mid [u])$. By the **support**, $\mathrm{supp}(\mu)$, of $\mu$, we mean the topological support, i.e., the smallest closed set of full measure. Note that for a configuration $w$ on a finite set, $[w]$ intersects $\mathrm{supp}(\mu)$ iff $\mu(w) > 0$.

A measure $\mu$ on $\mathcal{A}^{\mathbb{Z}^d}$ is **shift-invariant** (or **stationary** or **translation-invariant**) if $\mu(A) = \mu(\sigma_v A)$ for all measurable sets $A$ and $v \in \mathbb{Z}^d$.

**Definition 2.5.** *A shift-invariant $\mathbb{Z}^d$-measure $\mu$ is a $\mathbb{Z}^d$ **Markov random field** (or MRF) if, for any finite $S \subset \mathbb{Z}^d$, any $\eta \in \mathcal{A}^S$, any finite $T \subset \mathbb{Z}^d$ s.t. $\partial S \subseteq T \subseteq \mathbb{Z}^d \setminus S$, and any $\delta \in \mathcal{A}^T$ with $\mu(\delta) > 0$,*

$$\mu(\eta \mid \delta(\partial S)) = \mu(\eta \mid \delta). \tag{5}$$

Informally, $\mu$ is an MRF if, for any finite $S \subset \mathbb{Z}^d$, configurations on the sites in $S$ and configurations on the sites in $\mathbb{Z}^d \setminus (S \cup \partial S)$ are $\mu$-conditionally independent given a configuration on the sites in $\partial S$.

**Definition 2.6.** *For any Markov random field $\mu$, any finite $S \subseteq \mathbb{Z}^d$, and any $\delta \in \mathcal{A}^{\partial S}$ with $\mu(\delta) > 0$, define the measure $\mu^\delta$ on $\mathcal{A}^S$ by*

$$\mu^\delta(w) = \mu(w \mid \delta)$$

*for every $w \in \mathcal{A}^S$.*

We will deal mostly with nearest-neighbor Gibbs measures, which are MRF's specified by nearest-neighbor interactions, defined below.

**Definition 2.7.** *A **nearest-neighbor interaction** is a shift-invariant function $\Phi$ from the set of configurations on edges in $\mathbb{Z}^d$ to $\mathbb{R} \cup \infty$. Here, shift-invariance means that $\Phi(\sigma_v w) = \Phi(w)$ for all configurations $w$ on edges and all $v \in \mathbb{Z}^d$.*

Clearly, a nearest-neighbor interaction is defined by only finitely many numbers, namely the values of the interaction on configurations on edges $\{0, e_i\}, i = 1, \ldots, d$.

For a nearest-neighbor interaction $\Phi$, we define its **underlying SFT** as follows:

$$X_\Phi = \{x \in \mathcal{A}^{\mathbb{Z}^d} \;:\; \Phi(x(\{v, v'\})) \neq \infty, \text{ for all } v \sim v' \text{ in } \mathbb{Z}^d\}.$$

Note that $X_\Phi$ is a nearest-neighbor SFT.

**Definition 2.8.** *For a nearest-neighbor interaction $\Phi$, any finite set $S \subset \mathbb{Z}^d$, and any $w \in \mathcal{A}^S$, the **energy function** of $w$ with respect to $\Phi$ is*

$$U^\Phi(w) := \sum_e \Phi(w(e)),$$

*where the sum ranges over all edges $e$ of $S$. The **partition function** of $S$ is*

$$Z^\Phi(S) := \sum_{w \in \mathcal{A}^S} e^{-U^\Phi(w)}.$$

*For $\delta \in \mathcal{A}^{\partial S}$ we define*

$$Z^{\Phi,\delta}(S) := \sum_{w \in \mathcal{A}^S} e^{-U^\Phi(w\delta)}.$$

*In all definitions, we adopt the convention that $\infty + x = \infty$ for all $x \in \mathbb{R}$.*

**Definition 2.9.** *For any nearest-neighbor interaction $\Phi$, an MRF $\mu$ is called a **Gibbs measure** for $\Phi$ if for any finite set $S \subset \mathbb{Z}^d$ and $\delta \in \mathcal{A}^{\partial S}$ for which $\mu(\delta) > 0$, we have $Z^{\Phi,\delta}(S) \neq 0$ and, for any $w \in \mathcal{A}^S$,*

$$\mu^\delta(w) = \frac{e^{-U^\Phi(w\delta)}}{Z^{\Phi,\delta}(S)}.$$

Note that for a Gibbs measure $\mu$, $\mathrm{supp}(\mu)$ is automatically contained in the underlying SFT $X_\Phi$; we have allowed the interaction to take on infinite values in order to allow our Gibbs measures to be supported on proper subsets of $\mathcal{A}^{\mathbb{Z}^d}$. In our main results, we will assume that $\Phi$ is a nearest-neighbor interaction and that $X_\Phi$ satisfies a topological mixing property (the $D$-condition or block $D$-condition described below) that guarantees $\mathrm{supp}(\mu) = X_\Phi$.

Given a nearest neighbor SFT $X$, and a forbidden list of nearest neighbor configurations, a **uniform Gibbs measure** on $X$ is a Gibbs measure corresponding to the nearest-neighbor interaction which is 0 on all nearest-neighbor configurations except the forbidden configurations (on which it is $\infty$).

Every nearest-neighbor interaction $\Phi$ has as least one Gibbs measure; this is a very special case of a general result of Ruelle [14]. Often there are multiple Gibbs measures for a single $\Phi$; this phenomenon is often called a phase transition. One type of condition which guarantees uniqueness of Gibbs measures is so-called spatial mixing, with two variants defined below.

### 2.3. **Spatial Mixing.**

Let $B_n = [-n, n]^d$, the d-dimensional cube of side length $2n + 1$ centered at the origin.

**Definition 2.10.** *For a function $f(n) : \mathbb{N} \to \mathbb{R}^+$, $\lim_{n\to\infty} f(n) = 0$, we say that an MRF $\mu$ satisfies **weak spatial mixing (WSM) with rate** $f(n)$ if for any finite set $S \subseteq B_n$ and any $w \in \mathcal{A}^S$, $\delta, \delta' \in \mathcal{A}^{\partial B_n}$ s.t. $\mu(\delta), \mu(\delta') > 0$,*

$$|\mu^\delta(w) - \mu^{\delta'}(w)| < |S| f(d(S, \partial B_n)).$$

**Definition 2.11.** *For a function $f(n) : \mathbb{N} \to \mathbb{R}^+$, $\lim_{n\to\infty} f(n) = 0$, we say that an MRF $\mu$ satisfies **strong spatial mixing (SSM) with rate** $f(n)$ if for any disjoint finite sets $S, T \subseteq B_n$ and any $v \in \mathcal{A}^T$, $w \in \mathcal{A}^S$, $\delta, \delta' \in \mathcal{A}^{\partial B_n}$ s.t. $\mu(\delta), \mu(\delta'), \mu^\delta(v), \mu^{\delta'}(v) > 0$,*

$$|\mu^\delta(w \mid v) - \mu^{\delta'}(w \mid v)| < |S| f(d(S, \partial B_n)).$$

Informally, weak spatial mixing means that conditioning on a boundary configuration does not have much effect on the measure of a configuration on a set $S$ far from the boundary, and strong spatial mixing means that this is still true even if one first conditions on a configuration on sites which may be close to $S$.

The factor of $|S|$ on the right-hand side is unavoidable in both definitions; without it weak spatial mixing would force a much more stringent condition on $\mu$ called $m$-dependence. (see [2])

It is well-known that for a nearest-neighbor interaction $\Phi$, weak spatial mixing at any rate implies that there is only one Gibbs measure defined by $\Phi$ [17, Proposition 2.2]. Well-known examples of Gibbs measures that satisfy SSM include the unique uniform Gibbs measures for the hard square $\mathbb{Z}^2$-SFT and the $k$-checkerboard $\mathbb{Z}^2$-SFT for $k \geq 12$ (see [5],[12]).

We will need only somewhat weaker spatial mixing conditions for our main results. These conditions can be formulated in terms of a function naturally associated to $\mu$ and described as follows.

Recall that $\mathcal{P}$ denotes the lexicographic past in $\mathbb{Z}^d$. For any shift-invariant measure $\mu$, define the function

$$p_\mu(x) := \mu(x(0) \mid x(\mathcal{P}))$$

which is defined $\mu$-a.e. on $\text{supp}(\mu)$. Note that $p_\mu(x)$ depends only on $x(\mathcal{P} \cup \{0\})$ for $\mu$-a.e. $x$. Recall that $I_\mu(x) := -\log p_\mu(x)$ is known as the information function of $\mu$.

For any finite $S \subset \mathbb{Z}^d$, define the function

$$p_{\mu,S}(x) := \mu(x(0) \mid x(S)).$$

We will sometimes refer to the special case

$$p_{\mu,n}(x) = p_{\mu,\mathcal{P}_n}(x), \text{ where } \mathcal{P}_n = B_n \cap \mathcal{P}.$$

**Definition 2.12.** *We write* $\lim_{S\to\mathcal{P}} p_{\mu,S}(x) = L$ *to mean the limit in the "net" sense: for any* $\epsilon > 0$*, there exists* $n$ *such that for all finite* $S$ *satisfying* $\mathcal{P}_n \subset S \subset \mathcal{P}$*, we have* $|p_{\mu,S}(x) - L| < \epsilon$*.*

By martingale convergence, $\lim_{n\to\infty} p_{\mu,n}(x) = p_\mu(x)$ for $\mu$-a.e. $x \in \text{supp}(\mu)$. For this reason, for any $x \in \text{supp}(\mu)$, if $\lim_{S\to\mathcal{P}} p_{\mu,S}(x)$ exists, we will take $p_\mu(x)$ to be this limit.

Our main results, Theorem 3.1 and 3.4, will establish a representation for pressure in terms of a given shift-invariant measure $\nu$. For Theorem 3.1 we assume $\lim_{S\to\mathcal{P}} p_{\mu,S}(x) = p_\mu(x)$ uniformly on $\text{supp}(\nu)$, i.e. that $\forall\epsilon > 0\ \exists N > 0$ so that $\forall x \in \text{supp}(\nu)$, $\mathcal{P}_n \subset S \subset \mathcal{P} \implies |p_{\mu,S}(x) - p_\mu(x)| < \epsilon$.

For Theorem 3.4, we will need a stronger type of convergence.

**Definition 2.13.** *We write* $\lim_{S\to\mathcal{P},\ U\to+\infty} p_{\mu,S\cup U}(x) = L$ *to mean that for any* $\epsilon > 0$*, there exists* $n$ *such that for all finite* $S, U$ *satisfying* $\mathcal{P}_n \subset S \subset \mathcal{P}$ *and* $U \subset (B_n \cup \mathcal{P})^c$*, we have* $|p_{\mu,S\cup U}(x) - f(x)| < \epsilon$*.*

We note that this definition could also be written in terms of a single set; the only property required of $S \cup U$ is that $S \cup U$ contains $\mathcal{P}_n$ and is contained in $B_n^c \cup \mathcal{P}_n$. The definition is written with $S$ and $U$ decoupled only to make comparisons to $\lim_{S\to\mathcal{P}} p_{\mu,S}$ more clear.

Again, we will take $p_\mu(x)$ to be the value of this limit when it exists. For Theorem 3.4, we will assume $\lim_{S\to\mathcal{P},\ U\to+\infty} p_{\mu,S\cup U}(x) = p_\mu(x)$ uniformly on $\text{supp}(\nu)$. Clearly this implies $\lim_{S\to\mathcal{P}} p_{\mu,S}(x) = p_\mu(x)$ uniformly on $\text{supp}(\nu)$.

We have the following implication.

**Proposition 2.14.** *For an MRF* $\mu$*, if* $\mu$ *satisfies SSM at any rate, then* $\lim_{S\to\mathcal{P},U\to+\infty} p_{\mu,S\cup U}(x) = p_\mu(x)$ *uniformly on* $\text{supp}(\mu)$*.*

*Proof.* We find it convenient to define the vector-valued function

$$\hat{p}_\mu^n(x) := \mu(y(0) = \cdot \mid y(\partial S_n) = x(\partial S_n))$$

where $S_n = B_n \setminus \mathcal{P}_n$ (so, $\hat{p}_\mu^n(x)_a = \mu(y(0) = a \mid y(\partial S_n) = x(\partial S_n))$).

By SSM applied to $S = \{0\}$ and $T_n = \mathcal{P}_n = \mathcal{P} \cap B_n$. we see that given $\epsilon > 0$, for $n$ sufficiently large, if $x, x' \in \text{supp}(\mu)$, and $x(T_n) = x'(T_n)$, then $|\hat{p}_\mu^n(x) - \hat{p}_\mu^n(x')| < \epsilon$. For $m \geq n$, $\hat{p}_\mu^m(x)$ can be written as a weighted average of $\hat{p}_\mu^n(x')$ for finitely many $x'$ agreeing with $x$ on $T_n$. Thus, $|\hat{p}_\mu^n(x) - \hat{p}_\mu^m(x)| < \epsilon$. So, the sequence $\hat{p}_\mu^n$ is uniformly Cauchy and therefore uniformly convergent.

We can decompose $\partial S_n$ as a disjoint union

$$\partial S_n = U_n \cup C_n$$

where $U_n = (\partial S_n) \cap \mathcal{P}$ (the "upper layer" of $\mathcal{P}_n$) and $C_n = \partial S_n \setminus U_n$ (a"canopy" sitting over $U_n$).

Fix $n$ and let $S$ be a finite set satisfying $\mathcal{P}_n \subset S \subset \mathcal{P}$ and $U \subset (B_n \cup \mathcal{P})^c$. Then

$$(6) \quad p_{\mu,S\cup U}(x) = \sum_{\delta\in\mathcal{A}^{C_n}:\ \mu(x(S\cup U)\delta)>0} \mu(x(0) \mid x(S\cup U),\delta)\mu(\delta \mid x(S\cup U))$$
$$= \sum_{\delta\in\mathcal{A}^{C_n}:\ \mu(x(S\cup U)\delta)>0} \mu(x(0) \mid x(U_n),\delta)\mu(\delta \mid x(S\cup U))$$

$$= \sum_{\delta\in\mathcal{A}^{C_n}:\ \mu(x(S\cup U)\delta)>0} \mu(y(0) = x(0) \mid y(\partial S_n) = y_\delta(\partial S_n))\mu(\delta \mid x(S\cup U)),$$

where $y_\delta$ is any point in $\mathrm{supp}(\mu)$ such that $y_\delta(U_n) = x(U_n)$ and $y_\delta(C_n) = \delta$. (For instance, any $y \in [x(S\cup U)\delta]$.)

Let $g(x)$ denote the (vector-valued) uniform limit of $\hat{p}^n_\mu$ on $\mathrm{supp}(\mu)$. It follows from the above that given $\epsilon > 0$, for sufficiently large $n$,

$$|\mu(y(0) = x(0) \mid y(\partial S_n) = y_\delta(\partial S_n)) - g(x)_{x(0)}| < \epsilon.$$

Thus, $\lim_{S\to\mathcal{P},U\to+\infty} p_{\mu,S\cup U}(x) = g(x)_{x(0)}$ uniformly on $\mathrm{supp}(\mu)$ and $g(x)_{x(0)} = p_\mu(x)$ by our convention. $\square$

2.4. **Positivity of $p_\mu$.**

Positivity of $p_\mu$ and related functions will play an important role in our main results. We begin with an easy implication between two forms of positivity.

**Definition 2.15.** *For any shift-invariant $\mu$, make the notation*

$$c_\mu = \inf_{x\in\mathrm{supp}(\mu),\ S\subset\mathcal{P},\ |S|<\infty} p_{\mu,S}(x).$$

**Proposition 2.16.** *If $c_\mu > 0$, then $p_\mu$ is bounded away from zero $\mu$-a.e.*

*Proof.* If $c_\mu > 0$, then the functions $p_{\mu,n}$ are uniformly bounded away from zero on $\mathrm{supp}(\mu)$. Since $p_{\mu,n}$ converges to $p_\mu$ $\mu$-a.e., it follows that $p_\mu$ is bounded away from zero $\mu$-a.e. $\square$

Next, we show that SSF is sufficient for a stronger form of positivity.

**Proposition 2.17.** *If $\Phi$ is a nearest-neighbor interaction and $X_\Phi$ satisfies SSF, then for any Gibbs measure $\mu$ for $\Phi$, $c_\mu > 0$.*

*Proof.* We will show in fact that

$$\inf_{x\in\mathrm{supp}(\mu),\ S\subset\mathbb{Z}^d\setminus\{0\},\ |S|<\infty} p_{\mu,S}(x) > 0.$$

Recall that $N_0$ denotes the set of nearest neighbors of 0. Let $S \subset \mathbb{Z}^d \setminus \{0\}$ be a finite set. Let

- $U = N_0 \setminus S$
- $V = (\partial(\{0\} \cup U)) \setminus S$
- $S' = (\partial(\{0\} \cup U)) \cap S$

In particular, $\partial(\{0\} \cup U)$ is the disjoint union of $V$ and $S'$.

Let $x \in X$. There exists $v \in \mathcal{A}^V$ such that

$$\mu(v \mid x(S)) \geq |\mathcal{A}|^{-|V|}.$$

Let $L$ and $\ell$ be upper and lower bounds on finite values of $\Phi$. Since any locally admissible configuration is globally admissible, there exists $u \in \mathcal{A}^U$ such that $x(0)x(S)uv \in \mathcal{L}(X)$. Thus,

$$p_{\mu,S}(x) \geq \mu\left(y(0) = x(0), y(U) = u \mid y(S) = x(S)\right) \geq$$
$$\mu\left(y(0) = x(0), y(U) = u \mid y(S) = x(S), y(V) = v\right) \mu(y(V) = v \mid y(S) = x(S)) =$$
$$\mu\left(y(0) = x(0), y(U) = u \mid y(S') = x(S'), y(V) = v\right) \mu(y(V) = v \mid y(S) = x(S)) \geq$$
$$|\mathcal{A}|^{-|U|-1} e^{(4d^2)(\ell - L)} |\mathcal{A}|^{-|V|} \geq e^{(4d^2)(\ell - L)} |\mathcal{A}|^{-|N_0|-|N_1|-1}$$

where $N_1$ is the set of nearest neighbors of $N_0$, other than 0.

Since this lower bound is positive and independent of $x$ and $S$, we are done. $\square$

The preceding result applies to $X = C_k^{(2)}$ for any $k \geq 5$. In contrast we have:

**Proposition 2.18.** *Let $X = C_3^{(2)}$ and $\mu$ be a shift-invariant measure with* $\operatorname{supp}(\mu) = X$*. Then $c_\mu = 0$.*

*Proof.* Let $x^* \in X$ be defined by:

$$x^*(v) = \begin{cases} v_1 + v_2 \pmod 3 \text{ if } v \in \mathcal{P} \\ 1 + v_1 + v_2 \pmod 3 \text{ if } v \notin \mathcal{P} \end{cases}.$$

We claim that $x^* \in X$. To see this, we must show that for all $v$ and $i = 1, 2$, $x^*(v + e_i) \neq x^*(v)$. Clearly this holds when $v, v + e_i \in \mathcal{P}$ or $v, v + e_i \notin \mathcal{P}$ by the individual piecewise formulas. The last case is when $v \in \mathcal{P}$ and $v + e_i \notin \mathcal{P}$, and in this case $x^*(v + e_i) = 2 + x^*(v) \neq x^*(v) \mod 3$.

Fix any $n$, and let $W_{3n} = [1, 3n] \times [-3n, -1]$. For $1 \leq i \leq n - 1$, define $x^i = \sigma_{(-3i,0)} x^*$ and note that

$$x^i(W_{3n}) = x^*(W_{3n}), \; x^i(3i - 1, 0) = 2, \text{ and } x^i(3i, 0) = 1.$$

We now show that the sets $[x^i(W_{3n} \cup \{(3i - 1, 0), (3i, 0)\}]$, for $1 \leq i \leq n - 1$, are disjoint. Choose $1 \leq i < i' \leq n - 1$. Let $x \in [x^i(W_{3n} \cup \{(3i - 1, 0), (3i, 0))\}]$. We claim that for $3i \leq j \leq 3i'$, $x(j, 0) = j + 1 \pmod 3$. To see this, we argue by induction. For $j = 3i$, this is true by definition of $x^i$. Assume this is true for a given $j$. Then $x(j + 1, 0) \neq x(j, 0) = j + 1 \pmod 3$ and $x(j + 1, 0) \neq x(j + 1, -1) = x^*(j + 1, -1) = j \pmod 3$. Thus, $x(j + 1, 0) = j + 2 \pmod 3$, as desired, completing our proof by induction. But then $x(3i', 0) = 2$, and since $x^{i'}(3i', 0) = 1$, $x \notin [x^{i'}(W_{3n} \cup \{(3i' - 1, 0), (3i', 0))\}]$, and these cylinder sets are disjoint as claimed.

We now decompose $\mu(x^i(\{(3i - 1, 0), (3i, 0)\} \mid x^*(W_{3n})))$ as

$$\mu(x^i(3i - 1, 0) \mid x^*(W_{3n})))\mu(x^i(3i, 0) \mid x^*(W_{3n}), x^i(3i - 1, 0)).$$

Since $W_{3n} - (3i - 1, 0) \subseteq \mathcal{P}_{3n}$ and $x^*(W_{3n}) = x^i(W_{3n})$, the first factor can be rewritten as $p_{\mu, W_{3n} - (3i-1,0)}(\sigma_{(3i-1,0)} x^i)$. Similarly, the second factor can be expressed as $p_{\mu, \{(-1,0)\} \cup (W_{3n} - (3i,0))}(\sigma_{(3i,0)} x^i)$. Both are greater than or equal to $c_\mu$ by definition, so

$$\mu(x^i(\{(3i - 1, 0), (3i, 0)\} \mid x^*(W_{3n}))) \geq c_\mu^2.$$

By disjointness of $\{[x^i(W_{3n} \cup \{(3i - 1, 0), (3i, 0))\}]\}$, we obtain $c_\mu^2(n - 1) \leq 1$. Since this is true for all $n$, $c_\mu = 0$. $\square$

We remark that there is a fully-supported nearest-neighbor uniform Gibbs measure on $C_3^{(2)}$ ([3]), and so the proposition is not vacuously true. It is clear from the proof that this result actually holds for any shift-invariant measure whose support contains the point $x^*$ (and therefore all $x^i$ as well.)

2.5. **D-condition.**
We will frequently make use of a topological mixing condition, defined by Ruelle [14, Section 4.1] and given in the definition below. For this, we use the following notation. Let $\Lambda_n$ be a sequence of finite sets. We write $\Lambda_n \nearrow \infty$ if $\cup_n \Lambda_n = \mathbb{Z}^d$ and for each $v \in \mathbb{Z}^d$,

$$\lim_{n\to\infty} \frac{|\Lambda_n \ \Delta \ (\Lambda_n + v)|}{|\Lambda_n|} = 0$$

where $\Delta$ denotes symmetric difference.

**Definition 2.19.** *An SFT $X$ satisfies the **D-condition** if there exist sequences of finite subsets $(\Lambda_n)$, $(M_n)$ of $\mathbb{Z}^d$ such that $\Lambda_n \nearrow \infty$, $\Lambda_n \subseteq M_n$, $\frac{|M_n|}{|\Lambda_n|} \to 1$, and, for any $v \in \mathcal{L}_{\Lambda_n}(X)$ and finite $S \subset M_n^c$ and $w \in \mathcal{L}_S(X)$, $[v] \cap [w] \neq \varnothing$.*

Roughly speaking, this means that there exists an exhaustive sequence of shapes, $\Lambda_n$, for which relatively few elements of $\Lambda_n$ are near $\partial\Lambda_n$, and "collars," $(M_n \setminus \Lambda_n)$, around the shapes of comparatively small size such that it is possible to "fill in" between any legal configurations inside and outside the collar.

We will make use of the following property of a sequence $\Lambda_n \nearrow \infty$.

**Lemma 2.20.** *Let $\Lambda_n \nearrow \infty$. Given $n, N \in \mathbb{N}$, let*

$$\Lambda_n^N = \{v \in \Lambda_n : \ v + \mathcal{P}_N \subset \Lambda_n\}.$$

*Then for fixed $N$,*

$$\lim_{n\to\infty} \frac{|\Lambda_n^N|}{|\Lambda_n|} = 1.$$

*Proof.* If $v \in \Lambda_n$ and $v + \mathcal{P}_N \not\subset \Lambda_n$, then for some $w \in \mathcal{P}_N$, $v \in \Lambda_n \setminus (\Lambda_n - w)$. Thus,

$$\frac{|\Lambda_n \setminus \Lambda_n^N|}{|\Lambda_n|} \leq \frac{\sum_{w\in\mathcal{P}_N} \ |\Lambda_n \setminus (\Lambda_n - w)|}{|\Lambda_n|}.$$

But the right hand side tends to 0 as $n \to \infty$. □

For most known examples, one can choose $\Lambda_n$ to be rectangular prisms, or cubes. This motivates the following variation:

**Definition 2.21.** *An SFT $X$ satisfies the **block D-condition** if there exists a sequence of integers $(R_n)$ such that $\frac{R_n}{n} \to 0$ and for any rectangular prism $B = \prod_{i=1}^d [-n_i, n_i]$, any integers $m_i \geq R_{n_i}$, and any $w \in \mathcal{L}_{\partial\left(\prod_{i=1}^d [-(n_i+m_i), n_i+m_i]\right)}(X)$ and $v \in \mathcal{L}_B(X)$, $[v] \cap [w] \neq \varnothing$.*

For a nearest-neighbor SFT, the block D-condition can be expressed in the following equivalent form.

**Lemma 2.22.** *A nearest-neighbor SFT $X$ satisfies the block D-condition if and only if there exists a sequence of integers $(R_n)$ such that $\frac{R_n}{n} \to 0$ and for any rectangular prism $B = \prod_{i=1}^d [-n_i, n_i]$, any integers $m_i \geq R_{n_i}$, any finite set $S \subset \left(\prod_{i=1}^d [-(n_i + m_i), n_i + m_i]\right)^c$, if $w \in \mathcal{L}_S(X)$ and $v \in \mathcal{L}_B(X)$, then $[v] \cap [w] \neq \varnothing$.*

*Proof.* Let $T = \prod_{i=1}^{d}[-(n_i+m_i), n_i+m_i]$. If $w \in \mathcal{L}_S(X)$ for such a set $S$, then we can extend $w$ to a globally admissible configuration $w'$ on $S \cup \partial T$. Let $w'' = w'(\partial T)$. By the block D-condition $[v] \cap [w''] \neq \varnothing$. Since $X$ is a nearest-neighbor SFT, it follows that $[v] \cap [w'] \neq \varnothing$ and thus $[v] \cap [w] \neq \varnothing$. □

It follows, by choosing $\Lambda_n = \prod_{i=1}^{d}[-n, n]$ and $M_n = \prod_{i=1}^{d}[-(n+R_n), n+R_n]$, that for a nearest-neighbor SFT, the block D-condition does indeed imply the D-condition, and this implication can be easily generalized to any SFT. To distinguish between these definitions we sometimes refer to the D-condition as the classical D-condition.

For Theorem 3.1, we will assume the classical D-condition. For Theorem 3.4, we will need the block D-condition. However, we are not aware of any example which satisfies the classical D-condition and not the block D-condition; in fact in some works (e.g., [13]) the D-condition is stated with the assumption that the sets $\Lambda_n$ and $M_n$ are cubes or rectangular blocks.

We will make use of the following result that is well known under a much weaker hypothesis ([14, Remark 1.14]). We give a proof for completeness.

**Proposition 2.23.** *If $\Phi$ is a nearest-neighbor interaction and $X_\Phi$ satisfies the D-condition, then for any Gibbs measure $\mu$ for $\Phi$,* $\mathrm{supp}(\mu) = X_\Phi$.

*Proof.* As mentioned earlier, $\mathrm{supp}(\mu) \subseteq X_\Phi$. Let $X = X_\Phi$.

Let $S$ be a finite subset of $\mathbb{Z}^d$ and $w \in \mathcal{L}_S(X)$. By the $D$-condition there exists a finite set $T$ containing $S$ such that for any $\delta \in \mathcal{L}_{\partial T}(X)$, we have $w\delta \in \mathcal{L}(X)$. Since there exists some $\delta \in \mathcal{A}^{\partial T}$ such that $\mu(\delta) > 0$ and since $\mathrm{supp}(\mu) \subseteq X$, we have $\delta \in \mathcal{L}_{\partial T}(X)$, Thus, $w\delta \in \mathcal{L}(X)$ and so $\mu(w|\delta) > 0$. Since

$$\mu(w) = \sum_{\eta \in \mathcal{A}^{\partial T}:\ \mu(\eta)>0} \mu(w|\eta)\mu(\eta)$$

we have $\mu(w) > 0$. □

**Proposition 2.24.** *Any $\mathbb{Z}^d$ SFT that satisfies SSF must satisfy the block D-condition.*

*Proof.* It is clear that the block D-condition is satisfied with each $m_i = 1$, since for any dimensions $n_i$, $1 \leq i \leq d$, and configurations $w \in \mathcal{L}_{\partial \prod_{i=1}^{d}[-(n_i+1), n_i+1]}(X)$ and $v \in \mathcal{L}_{\prod_{i=1}^{d}[-n_i, n_i]}(X)$, the concatenation $vw$ is locally admissible, therefore globally admissible by SSF, and so $[v] \cap [w] \neq \varnothing$. □

Recall that $d = 2$ and $k \geq 5$, the $k$-checkerboard SFT satisfies SSF and therefore satisfies the block-D condition. On the other hand, the 3-checkerboard SFT does not even satisfy the classical D-condition.

**Proposition 2.25.** *The* 3*-checkerboard SFT does not satisfy the classical D-condition for any $d \geq 2$.*

*Proof.* For any $d$, let $x_d$ be defined by $x_d(v) = (\sum_i v_i) \pmod 3$ for all $v \in \mathbb{Z}^d$. Then $x_d \in C_3^{(d)}$.

We will show that there are no points $x \in C_3^{(d)}$, $x \neq x_d$, which agree with $x_d$ on all but finitely many sites. (Such points are sometimes called **frozen**.)

We argue for $d = 2$, which implies the same for all $d$ since the restriction of $x_d$ to any translate of $\{v \in \mathbb{Z}^d : v_i = 0 \text{ for all } i > 2\}$ agrees with a shift of $x_2$

Suppose that $y \in C_3^{(2)}$ agrees with $x_2$ on the complement of a finite set of sites. Let $S$ be the set of sites at which $y$ and $x_2$ disagree. Consider the leftmost site $v$ in the top row of $S$. Its neighbors $v - e_1$ and $v + e_2$ are in $S^c$, and by definition of $x_2$, $y(v - e_1) = x_2(v - e_1) \neq x_2(v + e_2) = y(v + e_2)$. Therefore, there is only one legal choice for $y(v)$, which is $x_2(v)$, contrary to the fact that $v \in S$.

In particular, this implies that for any $d$ and finite $S \subseteq \mathbb{Z}^d$, any boundary configuration $x_d(\partial S)$ has only one valid completion to all of $S \cup \partial S$, which precludes the classical D-condition (for instance, one cannot "fill in" between the configuration on any cube which consists of alternating 0's and 1's and the restriction of $z_d$ to the boundary of any larger cube). $\square$

2.6. **Entropy, Pressure and Equilibrium States.**

For a shift-invariant measure $\mu$ on $\mathcal{A}^{\mathbb{Z}^d}$, we define its entropy as follows.

**Definition 2.26.** *The **measure-theoretic entropy** of a shift-invariant measure $\mu$ on $\mathcal{A}^{\mathbb{Z}^d}$ is defined by*

$$h(\mu) = \lim_{j_1, j_2, \ldots, j_d \to \infty} \frac{-1}{j_1 j_2 \cdots j_d} \sum_{w \in \mathcal{A}^{\prod_{i=1}^d [1, j_i]}} \mu(w) \log(\mu(w)),$$

*where terms with $\mu(w) = 0$ are omitted.*

We define topological pressure for both interactions and functions on a shift space $X$. In order to discuss connections between these viewpoints, we need a mechanism for turning an interaction (which is a function on finite configurations) into a continuous function on the infinite configurations in $X$. Following Ruelle, we do this as follows for the special case of nearest-neighbor interactions $\Phi$. Define for $x \in X_\Phi$

$$A_\Phi(x) := -\sum_{i=1}^d \Phi(\{x_0, x_{e_i}\}).$$

We now give the two definitions of topological pressure.

**Definition 2.27.** *For a nearest-neighbor interaction $\Phi$ the **(topological) pressure** of $\Phi$ is defined as*

$$P(\Phi) = \lim_{n_1, \ldots, n_d \to \infty} \frac{1}{\prod n_i} \log Z^\Phi \left( \prod [1, n_i] \right).$$

It is well-known [14, Corollary 3.13] that for any sequence $\Lambda_n \nearrow \infty$,

$$P(\Phi) = \lim_{n \to \infty} \frac{1}{|\Lambda_n|} \log Z^\Phi (\Lambda_n).$$

**Definition 2.28.** *For any continuous real-valued function $f$ on a $\mathbb{Z}^d$ SFT $X$, the **(topological) pressure** of $f$ is defined as*

$$P(f) = \sup_\mu \; h(\mu) + \int f \, d\mu,$$

*where the supremum ranges over all shift-invariant measures $\mu$ supported on $X$. Any $\mu$ achieving this supremum is called an **equilibrium state** for $f$.*

*When $f = 0$, $P(f)$ is called the **topological entropy** $h(X)$ of $X$, and any equilibrium state for $f$ is called a **measure of maximal entropy** for $X$.*

It is well-known that for an irreducible nearest-neighbor $\mathbb{Z}$-SFT $X$, there is a unique uniform Gibbs measure; this measure is the unique measure of maximal entropy on $X$ [10, Section 13.3] (which is an irreducible (first-order) Markov chain).

The celebrated Variational Principle [16] [15] implies that the definitions we have given are equivalent in the sense that $P(\Phi) = P(A_\Phi)$. It is well-known that any continuous $f$ has at least one equilibrium state [15]. And in the case that $X_\Phi$ satisfies the $D$-condition, a measure on $X_\Phi$ is a Gibbs state for $\Phi$ iff it is an equilibrium state for $A_\Phi$ [4], [8], [14, Theorem 4.2]. We will discuss connections, in a more general context, between Gibbs states and equilibrium states in Section 5.

## 3. Main results

**Theorem 3.1.** *If $\Phi$ is a nearest-neighbor interaction with underlying SFT $X = X_\Phi$, $\mu$ is a Gibbs measure for $\Phi$, $\nu$ is a shift-invariant measure with* $\operatorname{supp}(\nu) \subseteq X$*,*

(A1) *$X$ satisfies the classical $D$-condition,*
(A2) $\lim_{S\to\mathcal{P}}\ p_{\mu,S}(x) = p_\mu(x)$ *uniformly over* $x \in \operatorname{supp}(\nu)$*, and*
(A3) $c_\mu > 0$*,*

*then*

$$P(\Phi) = \int I_\mu(x) + A_\Phi(x)\ d\nu = \int I_\mu(x) - \sum_{i=1}^{d} \Phi(x(\{0, e_i\}))\ d\nu.$$

Theorem 1 of [5], which motivated our paper, shows that if $\mu$ is a Gibbs measure for a nearest-neighbor interaction $\Phi$ and $X_\Phi$ has a safe symbol $a$ and satisfies SSM, then the pressure representation above holds for the point mass $\nu$ on $a^{\mathbb{Z}^d}$. We remark that Theorem 3.1 generalizes this result, with weaker hypotheses and a stronger conclusion. To see this, first recall that the existence of a safe symbol is even stronger than SSF, which implies (A1) and (A3) by Propositions 2.24 and 2.17; second, recall from Proposition 2.14 that SSM implies (A2).

*Proof.* Recall from Proposition 2.23 that $\operatorname{supp}(\mu) = X$. So, $p_\mu$ is defined $\mu$-a.e. on $X$ and for any finite set $S$, $w \in \mathcal{L}_S(X)$ iff $\mu(w) > 0$.

Choose $\ell < 0$ and $L > 0$ to be lower and upper bounds respectively on finite values of $\Phi$. Let $\Lambda_n, M_n$ be as in the definition of the D-condition.

We begin by proving that

$$\frac{1}{|\Lambda_n|}\left(\log Z^\Phi(\Lambda_n) + \log \mu(x(\Lambda_n)) + U^\Phi(x(\Lambda_n))\right) \to 0 \tag{7}$$

uniformly in $x \in X$ (though we will need this only for $x \in \operatorname{supp}(\nu)$). For this, we will only use the D-condition (A1).

Fix $n$ and let $R_n = |M_n| - |\Lambda_n|$. Note that for any $w \in \mathcal{L}_{\Lambda_n}(X)$,

$$\mu(w) = \sum_{\delta \in \mathcal{L}_{\partial M_n}(X)} \mu(w \mid \delta)\mu(\delta). \tag{8}$$

For any such $w$ and $\delta$, by the D-condition there exists $y_{w,\delta} \in \mathcal{L}_{M_n \setminus \Lambda_n}(X)$ such that $w y_{w,\delta} \delta \in \mathcal{L}(X)$. Then there is a constant $C_d > 0$ such that

$$\mu(w \mid \delta) \geq \mu(wy_{w,\delta} \mid \delta) = \frac{e^{-U^\Phi(wy_{w,\delta}\delta)}}{\sum_{u \in \mathcal{L}_{M_n}(X)} e^{-U^\Phi(u\delta)}}$$
$$\geq \frac{e^{-U^\Phi(w)-C_dR_nL}}{\sum_{v \in \mathcal{L}_{\Lambda_n}(X)} e^{-U^\Phi(v)}|\mathcal{A}|^{C_dR_n}e^{-C_dR_n\ell}} = \frac{e^{-U^\Phi(w)}}{Z^\Phi(\Lambda_n)} e^{R_n(C_d\ell - C_dL - C_d \log|\mathcal{A}|)}.$$

Let $y_{\max}$ achieve $\max \mu(wy \mid \delta)$ over all $y \in \mathcal{L}_{M_n \setminus \Lambda_n}(X)$. Then,

$$\mu(w \mid \delta) = \sum_{y \in \mathcal{L}_{M_n \setminus \Lambda_n}(X)} \mu(wy \mid \delta) \leq |\mathcal{A}|^{R_n} \mu(wy_{\max} \mid \delta)$$
$$= |\mathcal{A}|^{R_n} \frac{e^{-U^\Phi(wy_{\max}\delta)}}{\sum_{u \in \mathcal{L}_{M_n}(X)} e^{-U^\Phi(u\delta)}} \leq |\mathcal{A}|^{R_n} \frac{e^{-U^\Phi(w)-C_dR_n\ell}}{\sum_{v \in \mathcal{L}_{\Lambda_n}(X)} e^{-U^\Phi(v)-C_dR_nL}}$$
$$= \frac{e^{-U^\Phi(w)}}{Z^\Phi(\Lambda_n)} e^{R_n(C_dL - C_d\ell + \log|\mathcal{A}|)}.$$

Since $\sum_\delta \mu(\delta) = 1$, we can combine the three formulas above to see that

$$\gamma^{-R_n} \leq \mu(w) Z^\Phi(\Lambda_n) e^{U^\Phi(w)} \leq \gamma^{R_n},$$

where $\gamma := e^{C_d \log|\mathcal{A}| + C_dL - C_d\ell} > 0$. Since $\frac{R_n}{|\Lambda_n|} \to 0$, this implies (7).

We use (7) to represent pressure:

$$P(\Phi) = \lim_{n\to\infty} \frac{\log Z^\Phi(\Lambda_n)}{|\Lambda_n|} = \lim_{n\to\infty} \int \frac{\log Z^\Phi(\Lambda_n)}{|\Lambda_n|}\, d\nu = \lim_{n\to\infty} \int \frac{-U^\Phi(x(\Lambda_n)) - \log\mu(x(\Lambda_n))}{|\Lambda_n|}\, d\nu.$$

(Here the second equality comes from the fact that $\frac{\log Z^\Phi(\Lambda_n)}{|\Lambda_n|}$ is independent of $x$, and the third from (7).) Since $\nu$ is shift-invariant and $\Lambda_n \nearrow \infty$,

$$\lim_{n\to\infty} \int \frac{-U^\Phi(x(\Lambda_n))}{|\Lambda_n|}\, d\nu = \int A_\Phi(x)\, d\nu,$$

and so we can write

$$P(\Phi) = \int A_\Phi(x)\, d\nu - \lim_{n\to\infty} \int \frac{\log\mu(x(\Lambda_n))}{|\Lambda_n|}\, d\nu.$$

It remains to show that $\lim_{n\to\infty} \int \frac{-\log\mu(x(\Lambda_n))}{|\Lambda_n|}\, d\nu = \int I_\mu(x)\, d\nu$. We do this by decomposing $\mu(x(\Lambda_n))$ as a product of conditional probabilities. Denote by $(s_i^{(n)})_{i=1}^{|\Lambda_n|}$ the sites of $\Lambda_n$, ordered lexicographically. For any $1 \leq i \leq |\Lambda_n|$, denote by $S_i^{(n)}$ the set $\{s_j^{(n)} :\ 1 \leq j \leq i-1\}$. (This means that $S_1^{(n)} = \varnothing$). Then for any $x \in \mathrm{supp}(\nu)$, we can write

(9)
$$-\log\mu(x(\Lambda_n)) = \sum_{i=1}^{|\Lambda_n|} -\log\mu\left(x(s_i^{(n)}) \mid x(S_i^{(n)})\right) = \sum_{i=1}^{|\Lambda_n|} -\log p_{\mu, S_i^{(n)} - s_i^{(n)}}(\sigma_{s_i^{(n)}} x).$$

Clearly, each term $-\log p_{\mu, S_i^{(n)} - s_i^{(n)}}(\sigma_{s_i^{(n)}} x)$ is lower bounded by 0. To get an upper bound, we use shift-invariance of $\mu$, the fact that $S_i^{(n)} - s_i^{(n)} \subset \mathcal{P}$ and Assumption (A3) to conclude that for any $x \in \mathrm{supp}(\nu)$ and all $n, i$,

$$p_{\mu, S_i^{(n)} - s_i^{(n)}}(\sigma_{s_i^{(n)}} x) \geq c_\mu.$$

Fix $\epsilon > 0$ and $N \in \mathbb{N}$.

By Lemma 2.20, for sufficiently large $n$,

$$\frac{|\Lambda_n^N|}{|\Lambda_n|} > 1 - \epsilon.$$

(recall that $\Lambda_n^N = \{v \in \Lambda_n : \ v + \mathcal{P}_N \subset \Lambda_n\}$.)

For all $i$ such that $s_i^{(n)} \in \Lambda_n^N$, we have $\mathcal{P}_N \subset S_i^{(n)} - s_i^{(n)} \subset \mathcal{P}$. By assumption (A2), we then have that for sufficiently large $N$, all $i$ such that $s_i^{(n)} \in \Lambda_n^N$ and all $x \in \operatorname{supp}(\nu)$,

$$\left| p_{\mu, S_i^{(n)} - s_i^{(n)}}(\sigma_{s_i^{(n)}} x) - p_\mu(\sigma_{s_i^{(n)}} x) \right| < \epsilon.$$

Since $p_{\mu, S_i^{(n)} - s_i^{(n)}}$ is bounded from below by $c_\mu$ and $p_{\mu,n}$ converges to $p_\mu$ on $\operatorname{supp}(\nu)$, it follows that $p_\mu$ is also bounded below by $c_\mu$ on $\operatorname{supp}(\nu)$. Thus,

$$\left| - \log p_{\mu, S_i^{(n)} - s_i^{(n)}}(\sigma_{s_i^{(n)}} x) - I_\mu(\sigma_{s_i^{(n)}} x) \right| < \epsilon / c_\mu.$$

Therefore,

$$\left| \int - \log p_{\mu, S_i^{(n)} - s_i^{(n)}}(\sigma_{s_i^{(n)}} x) \ d\nu - \int I_\mu(\sigma_{s_i^{(n)}} x) \ d\nu \right| < \epsilon / c_\mu.$$

Since $\nu$ is shift-invariant, we have

$$\left| \int - \log p_{\mu, S_i^{(n)} - s_i^{(n)}}(\sigma_{s_i^{(n)}} x) \ d\nu - \int I_\mu(x) \ d\nu \right| < \epsilon / c_\mu.$$

Now, we are prepared to give bounds on (9). By the preceding,

$$\left| \sum_{s_i^{(n)} \in \Lambda_n^N} \int - \log p_{\mu, S_i^{(n)} - s_i^{(n)}}(\sigma_{s_i^{(n)}} x) \ d\nu - |\Lambda_n^N| \int I_\mu(x) \ d\nu \right| \leq |\Lambda_n| (\epsilon / c_\mu).$$

Also, since $0 \leq \int - \log p_{\mu, S_i^{(n)} - s_i^{(n)}}(\sigma_{s_i^{(n)}} x) \ d\nu \leq - \log c_\mu$,

$$0 \leq \sum_{s_i^{(n)} \notin \Lambda_n^N} \int - \log p_{\mu, S_i^{(n)} - s_i^{(n)}}(\sigma_{s_i^{(n)}} x) \ d\nu \leq |\Lambda_n \setminus \Lambda_n^N| (- \log c_\mu) \leq \epsilon |\Lambda_n| (- \log c_\mu).$$

Therefore, by (9),

$$-|\Lambda_n| \epsilon / c_\mu \leq \int - \log \mu(x(\Lambda_n)) \ d\nu - |\Lambda_n^N| \int I_\mu(x) \ d\nu \leq |\Lambda_n| (\epsilon / c_\mu - \epsilon \log c_\mu).$$

By dividing by $|\Lambda_n|$ and letting $n \to \infty$, we see that

$$- \epsilon / c_\mu + \int I_\mu(x) \ d\nu \leq \liminf_{n \to \infty} \frac{\int - \log \mu(x(\Lambda_n))}{|\Lambda_n|} \ d\nu \text{ and}$$
$$\limsup_{n \to \infty} \frac{\int_U - \log \mu(x(\Lambda_n))}{|\Lambda_n} | \ d\nu \leq (\epsilon / c_\mu - \epsilon \log c_\mu) + \int I_\mu(x) \ d\nu.$$

Since $\epsilon > 0$ was arbitrary,

$$\lim_{n \to \infty} \frac{\int - \log \mu(x(\Lambda_n))}{|\Lambda_n|} \ d\nu = \int I_\mu(x) \ d\nu,$$

completing the proof.

$\square$

We now apply Theorem 3.1 to obtain a result which gives a integral representation of $P(\Phi)$ for *every* invariant measure $\nu$.

**Corollary 3.2.** *If $\Phi$ is a nearest-neighbor interaction with underlying SFT $X$, $\mu$ is a Gibbs measure for $\Phi$,*

(B1) *$X$ satisfies the classical D-condition,*
(B2) $\lim_{S\to\mathcal{P}}\ p_{\mu,S}(x) = p_\mu(x)$*, uniformly over* $x \in \mathrm{supp}(\mu)$*, and*
(B3) $c_\mu > 0$*,*

*then*

$$P(\Phi) = \int I_\mu(x) + A_\Phi(x)\ d\nu = \int I_\mu(x) - \sum_{i=1}^{d} \Phi(x(\{0, e_i\}))\ d\nu$$

*for every shift-invariant measure $\nu$ with* $\mathrm{supp}(\nu) \subseteq X$*.*

*Proof.* This follows immediately from Theorem 3.1. □

**Corollary 3.3.** *If $\Phi$ is a nearest-neighbor interaction with underlying SFT $X = X_\Phi$, $\mu$ is a Gibbs measure for $\Phi$,*

(1) *$X$ satisfies SSF, and*
(2) *$\mu$ satisfies SSM,*

*then*

$$P(\Phi) = \int I_\mu(x) + A_\Phi(x)\ d\nu = \int I_\mu(x) - \sum_{i=1}^{d} \Phi(x(\{0, e_i\}))\ d\nu$$

*for every shift-invariant measure $\nu$ with* $\mathrm{supp}(\nu) \subseteq X$*.*

*Proof.* This follows from Corollary 3.2, Proposition 2.24, Proposition 2.14 and Proposition 2.17. □

Next, we state and prove a more difficult version of Theorem 3.1.

**Theorem 3.4.** *If $\Phi$ is a nearest-neighbor interaction with underlying SFT $X$, $\mu$ is a Gibbs measure for $\Phi$, $\nu$ is a shift-invariant measure with* $\mathrm{supp}(\nu) \subseteq X$*,*

(C1) *$X$ satisfies the block D-condition,*
(C2) $\lim_{S\to\mathcal{P}, U\to+\infty} p_{\mu,S\cup U}(x) = p_\mu(x)$ *uniformly on* $\mathrm{supp}(\nu)$*, and*
(C3) *$p_\mu$ is positive over* $\mathrm{supp}(\nu)$*,*

*then* $P(\Phi) = \int I_\mu(x) + A_\Phi(x)\ d\nu = \int I_\mu(x) - \sum_{i=1}^{d} \Phi(x(\{0, e_i\}))\ d\nu$*.*

In assumption (C3), $p_\mu$ is taken, by our convention, to mean the limit of $p_{\mu,n}$, which exists on $\mathrm{supp}(\nu)$ by assumption (C2).

In comparing Theorems 3.1 and 3.4, note the tradeoffs in the assumptions.

(1) (C1) of 3.4 clearly implies (A1) of 3.1.
(2) (C2) of 3.4 clearly implies (A2) of 3.1.
(3) (A2), (A3) of 3.1 together imply (C3) of 3.4: by (A3), $c_\mu > 0$ is a lower bound for $\{p_{\mu,n}(x) :\ x \in \mathrm{supp}(\nu)\}$, and by (A2), for $x \in \mathrm{supp}(\nu)$, $p_{\mu,n}(x)$ approaches $p_\mu(x)$, which must also be bounded from below by $c_\mu$.

*Proof.* As in the proof of Theorem 3.1, since the D-condition holds on $X$, it suffices to show that $\lim_{n\to\infty} \int \frac{-\log \mu(x(B_n))}{(2n+1)^d}\, d\nu = \int I_\mu\, d\nu$. For our purposes, it will be more convenient to show that $\lim_{n\to\infty} \int \frac{-\log \mu(x(K_n))}{n^d}\, d\nu = \int I_\mu\, d\nu$, where $K_n = [0, n+1]^d$. Clearly, since $\frac{|K_n|}{n^d} \to 1$, this still suffices.

Analogous to the proof of Theorem 3.1, we will decompose $\mu(x(K_n))$ as a product of conditional probabilities. However, we no longer have a positive lower bound on $p_{\mu,S}(x)$ over $x \in \mathrm{supp}(\nu)$ and finite subsets $S$ of $\mathcal{P}$. By using a more complicated decomposition of $K_n$ and the block D-condition, we obtain a much weaker (but still strong enough for our purposes) lower bound on the conditional probabilities involving sites near the boundary of $K_n$. The following lemma gives this bound.

**Lemma 3.5.** *Let $\Phi$ be a nearest-neighbor interaction with underlying SFT $X = X_\Phi$. Suppose that $S, T, U$ are finite subsets of $\mathbb{Z}^d$ such that $S \subseteq T \subseteq U$, $U$ is connected, $S \cap \underline{\partial} U = \varnothing$, and $[v] \cap [w] \neq \varnothing$ for all $v \in \mathcal{L}_S(X)$ and $w \in \mathcal{L}_{U\setminus T}(X)$. Then, for any $x \in \mathcal{L}_U(X)$,*

$$\mu(x(U \setminus T) \mid x(\underline{\partial} U)) \geq \gamma^{-(|U|-|S|)},$$

*where $\gamma = |\mathcal{A}|e^{2dL-\ell}$ for $\ell$ and $L$ the minimum and maximum, respectively, of finite values of $\Phi$.*

*Proof.* Consider any such $\Phi$, $S$, $T$, and $U$. Then for every $v \in \mathcal{L}_S(X)$ and $w \in \mathcal{L}_{U\setminus T}(X)$, there exists $u_{v,w} \in \mathcal{L}_{T\setminus S}(X)$ so that $vu_{v,w}w \in \mathcal{L}(X)$. Fix $x \in \mathcal{L}_U(X)$ and denote $y = x(U \setminus (T \cup \underline{\partial} U))$, $\delta = x(\underline{\partial} U)$ and $\bar{u} = u_{v,y\delta}$. Since $\mu$ is a Gibbs measure,

$$\mu(y \mid \delta) = \frac{\sum_{z\in\mathcal{L}_{U\setminus\underline{\partial}U}(X) \text{ s.t. } z\delta\in\mathcal{L}(X) \text{ and } z(U\setminus(T\cup\underline{\partial}U))=y} e^{-U^\Phi(z\delta)}}{\sum_{z\in\mathcal{L}_{U\setminus\underline{\partial}U}(X) \text{ s.t. } z\delta\in\mathcal{L}(X)} e^{-U^\Phi(z\delta)}}$$

$$\geq \frac{\sum_{v\in\mathcal{L}_S(X)} e^{-U^\Phi(v\bar{u}y\delta)}}{\sum_{v\in\mathcal{L}_S(X)} |\mathcal{A}|^{|U|-|S|} \max_{u\in\mathcal{L}_{U\setminus(S\cup\underline{\partial}U)}(X) \text{ s.t. } uv\delta\in\mathcal{L}(X)} e^{-U^\Phi(uv\delta)}}$$

$$\geq \frac{\sum_{v\in\mathcal{L}_S(X)} e^{-U^\Phi(v)-2dL(|U|-|S|)}}{\sum_{v\in\mathcal{L}_S(X)} |\mathcal{A}|^{|U|-|S|} e^{-U^\Phi(v)-\ell(|U|-|S|)}} = (|\mathcal{A}|e^{2dL-\ell})^{-(|U|-|S|)} = \gamma^{-(|U|-|S|)}.$$

□

We decompose $K_n$ into geometric shapes, and for those near the boundary will use Lemma 3.5 to show that the conditional probability of filling the shape in a certain way cannot be too small.

By (C2), for any $\epsilon > 0$, there exists $k := k_\epsilon$ so that for any finite sets $S, U$ with $\mathcal{P}_n \subseteq S \subset \mathcal{P}$ and $U \subseteq (B_k \cup \mathcal{P})^c$, $|p_{\mu,S\cup U}(x) - p_\mu(x)| < \epsilon$ for all $x \in \mathrm{supp}(\nu)$. (Equivalently, we could just require that $S \cup U$ contain $\mathcal{P}_k$ and be contained in $B_k^c \cup \mathcal{P}$.)

Since $X$ satisfies the block D-condition, there exists a sequence $R_n$ of integers s.t. $\frac{R_n}{n} \to 0$ and, for any rectangular prism $P = \prod_i [1, n_i]$, any $r \geq R_{\max n_i}$, and any configurations $v \in \mathcal{L}_P(X)$ and $w \in \mathcal{L}_S(X)$ for any finite $S \subset (\prod_i [1-r, n_i + r])^c$, we have $[v] \cap [w] \neq \varnothing$ (see Lemma 2.22). We may assume without loss of generality that $R_n$ is non-decreasing by redefining each $R_n$ as $\max_{i\leq n} R_i$.

Due to the technicality of the decomposition, we will begin with $d = 2$, and then give an outline of how to extend to larger $d$ via induction. We are therefore decomposing $K_n = [0, n+1]^2$. Our construction requires a parameter $m := m_n$, which we choose to be equal to $\lfloor\sqrt{nR_n}\rfloor$, and so $\frac{R_n}{n} \to 0$, $\frac{m_n}{n} \to 0$, and $\frac{R_n}{m_n} \to 0$.. We only consider $n$ large enough so that $m > 2k$ and $\frac{R_n}{n} < 1$. Define sets

$$C_0 = (\underline{\partial} K_n),$$

$$\begin{aligned} C_1 &= \{(i,j) \ : \ 1 \leq j \leq m, 1 \leq i \leq k(m+1-j)\}, \\ C_2 &= \{(i,j) \ : \ 1 \leq j \leq m, k(m+1-j) < i < n-kj\}, \\ C_3 &= \{(i,j) \ : \ 1 \leq j \leq m, n-kj \leq i \leq n\}, \end{aligned}$$

$$\begin{aligned} C_{3t+1} &= C_1 + (0,tm) \text{ for all } 1 \leq t \leq \lfloor \tfrac{n}{m} \rfloor - 2, \\ C_{3t+2} &= C_2 + (0,tm) \text{ for all } 1 \leq t \leq \lfloor \tfrac{n}{m} \rfloor - 2, \\ C_{3t+3} &= C_3 + (0,tm) \text{ for all } 1 \leq t \leq \lfloor \tfrac{n}{m} \rfloor - 2, \text{ and} \end{aligned}$$

$$C_{3\lfloor \frac{n}{m} \rfloor - 2} = [1,n] \times [m\lfloor \tfrac{n}{m} \rfloor - m + 1, n].$$

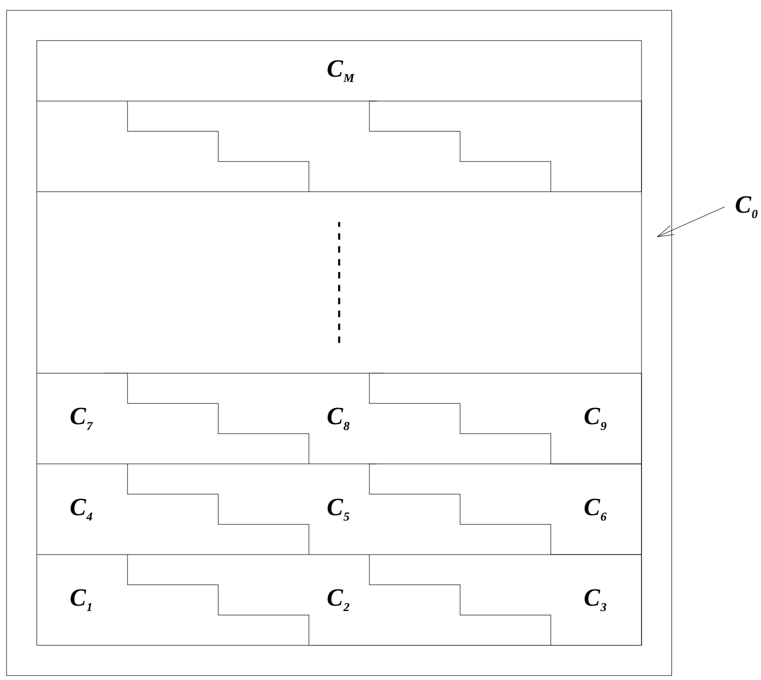


Figure 1. Decomposing $K_n$ (here, $M = 3\lfloor \frac{n}{m} \rfloor - 2$)

As illustrated in Figure 1, $C_0$ is the inner boundary of $K_n$. For each $t$, $C_{3t+1}$, $C_{3t+2}$, and $C_{3t+3}$ form a partition of the strip $[1,n] \times [tm+1,(t+1)m]$ of height $m$, made of two (discrete) trapezoids and a (discrete) parallelogram. And $C_{3\lceil \frac{n}{m} \rceil - 2}$ is simply a single "leftover" strip at the top of $[1,n]^2 = K_n \setminus \underline{\partial} K_n$. For every $t \in [1, 3\lceil \frac{n}{m} \rceil - 2]$, we define $D_t = \bigcup_{s=0}^{t-1} C_s$, and define $D_0 = \varnothing$. Note that by the choice of $m$, for large $n$, the bulk of $K_n$ is comprised of $\{C_i; \ i = 2 \mod 3\}$.

We decompose $\mu(x(K_n))$ as

$$\prod_{i=0}^{3\lceil \frac{n}{m} \rceil - 2} \mu(x(C_i) \mid x(D_i)). \tag{10}$$

We begin by giving a lower bound for $\mu(x(C_0) \mid x(D_0)) = \mu(x(C_0))$. For any $n$, letting $R := R_n$, then by definition $S = [R+1, n-R]^2$, $T = K_n \setminus C_0$ and $U = [-R, n+R]^2$ satisfy the hypotheses of Lemma 3.5. There clearly exists $\delta_n \in \mathcal{L}_{\underline{\partial} U}(X)$ s.t. $\mu(\delta_n) \geq |\mathcal{A}|^{-|\underline{\partial} U|}$. For any $x \in X$, by definition of $R$, there exists $y \in \mathcal{L}_{U \setminus (K_n \cup \underline{\partial} U)}$ s.t. $x(C_0) y \delta_n \in \mathcal{L}(X)$. So, by Lemma 3.5,

$$\mu\left(x(C_0) y \mid \delta_n\right) \geq \gamma^{-(|U|-|S|)} \geq \gamma^{-4(2R+1)(n+2R+1)}.$$

Therefore,

$$\begin{aligned} \mu(x(C_0)) &\geq \mu(x(C_0) y \delta_n) = \mu(\delta_n) \mu(x(C_0) y \mid \delta_n) \\ &\geq |\mathcal{A}|^{-4(n+2R+1)} \gamma^{-4(2R+1)(n+2R+1)} \geq \gamma^{-4(2R+2)(n+2R+1)}. \end{aligned} \tag{11}$$

(here the last inequality follows from $\gamma > |\mathcal{A}|$.)

The final factor of (10) is easy to bound from below. Note that the sets $S = T = \varnothing$ and $U = C_{3\lfloor \frac{n}{m} \rfloor - 2} \cup \partial C_{3\lfloor \frac{n}{m} \rfloor - 2}$ satisfy the hypotheses of Lemma 3.5. Then,

$$\begin{aligned} &\mu(x(C_{3\lfloor \frac{n}{m} \rfloor - 2}) \mid x(D_{3\lfloor \frac{n}{m} \rfloor - 2})) \\ &= \mu(x(C_{3\lfloor \frac{n}{m} \rfloor - 2}) \mid x(\partial C_{3\lfloor \frac{n}{m} \rfloor - 2})) = \mu(x(U \setminus T) \mid x(\underline{\partial} U)) \geq \gamma^{-(|U|-|S|)} \geq \gamma^{-(n+2)(2m+2)}. \end{aligned} \tag{12}$$

(here the first equality follows from the fact that $\mu$ is an MRF.)

We next deal with the terms in (10) of the form $\mu(x(C_{3t+1}) \mid x(D_{3t+1}))$. We wish to apply Lemma 3.5 for $U = [0, n+1] \times [tm, n+1]$, $T = U \setminus (\underline{\partial} U \cup C_{3t+1})$, and $S = ([km+R+1, n-R] \times [tm+R+1, (t+1)m]) \cup ([R+1, n-R] \times [(t+1)m+R+1, n-R])$. (See Figure 2.)

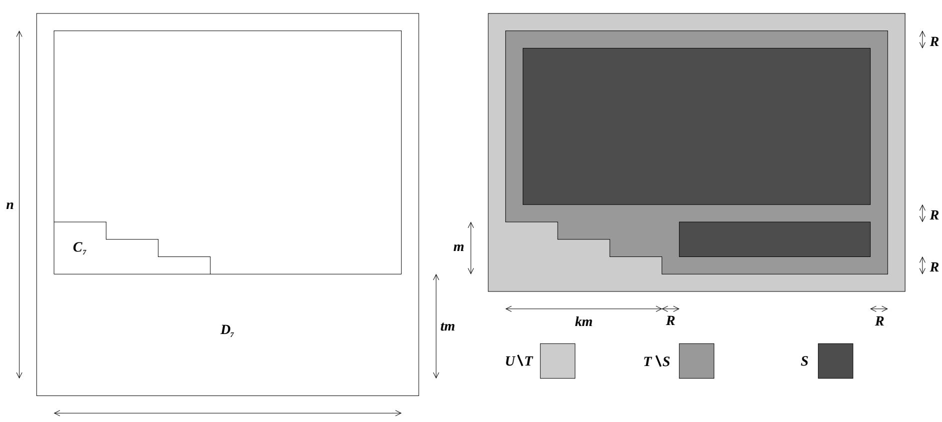


FIGURE 2. $S$, $T$, and $U$ for $\mu(x(C_{3t+1}) \mid x(D_{3t+1}))$

To check that Lemma 3.5 can be used, we must show that for any configurations $v \in \mathcal{L}_S(X)$ and $w \in \mathcal{L}_{\underline{\partial} U \cup C_{3t+1}}(X)$, $[v] \cap [w] \neq \varnothing$. This requires two applications of the block D-condition. Write $v$ as a concatenation $pq$ for $p \in \mathcal{L}_{[km+R+1,n-R]\times[tm+R+1,(t+1)m]}(X)$ and $q \in \mathcal{L}_{[R+1,n-R]\times[(t+1)m+R+1,n-R]}(X)$. By the block D-condition, there exists a configuration $w'$ which extends $w$ and $p$. A second application of the block D-condition gives a configuration $w''$ that extends $q$ and $w'$ and hence $p, q$ and $w$. Thus, $[p] \cap [q] \cap [w] = [v] \cap [w] \neq \varnothing$.

Since $\mu$ is an MRF, Lemma 3.5 implies that

(13)
$$\mu(x(C_{3t+1}) \mid x(D_{3t+1})) = \mu(x(C_{3t+1}) \mid x(\underline{\partial}([0,n+1]\times[tm,n+1]))) \geq \gamma^{-(|U|-|S|)} \geq \gamma^{-(km^2+5(R+1)(n+2))}.$$

We next deal with the terms in (10) of the form $\mu(x(C_{3t+3}) \mid x(D_{3t+3}))$. We wish to apply Lemma 3.5 for $U = (([0,n+1]\times[tm,n+1]) \cup C_{3t+3}) \cup \partial C_{3t+3}$, $T = U \setminus (\underline{\partial}U \cup C_{3t+3})$, and $S = [R+1,n-R]\times[(t+1)m+R+1,n-R]$. (See Figure 3.)

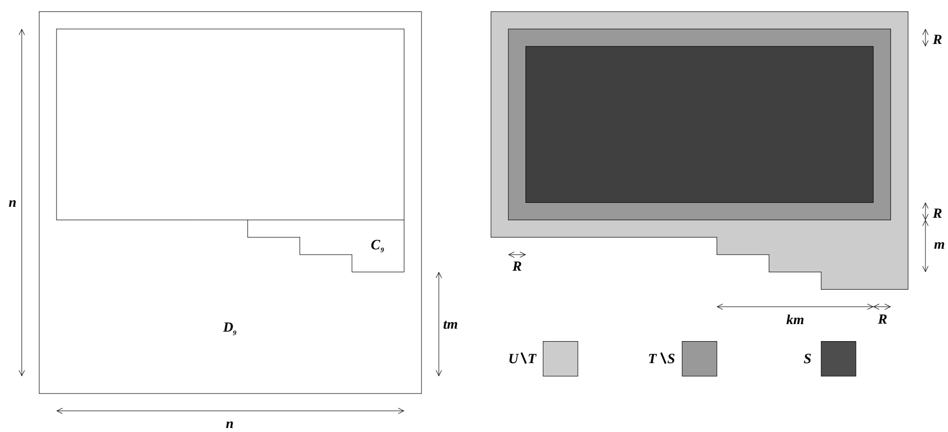


Figure 3. $S$, $T$, and $U$ for $\mu(x(C_{3t+3}) \mid x(D_{3t+3}))$

To check that Lemma 3.5 can be used, we must show that for any configurations $v \in \mathcal{L}_S(X)$ and $w \in \mathcal{L}_{\underline{\partial}U\cup C_{3t+3}}(X)$, $[v] \cap [w] \neq \varnothing$. This is a straightforward application of the block D-condition.

Since $\mu$ is an MRF, we can use Lemma 3.5 to show that

$$(14)\quad \mu(x(C_{3t+3}) \mid x(D_{3t+3})) = \mu(x(C_{3t+3}) \mid x(\underline{\partial}U)) \geq \gamma^{-(|U|-|S|)} \geq \gamma^{-(km^2+4(R+1)(n+2))}.$$

It remains to deal with factors of the form $\mu(x(C_{3t+2}) \mid x(D_{3t+2}))$. For each $t$, denote the sites of $C_{3t+2}$ in lexicographic order as $s_i^{(3t+2)}$, $1 \leq i \leq |C_{3t+2}|$, for each such $i > 1$, define $S_i^{(3t+2)} = \bigcup_{j=1}^{i-1}\{s_j^{(3t+2)}\}$, and define $S_1^{(3t+2)} = \varnothing$. We first decompose $\mu(x(C_{3t+2} \mid D_{3t+2}))$ as

$$(15)\qquad \prod_{i=1}^{|C_{3t+2}|} \mu\left(x(s_i^{(3t+2)}) \mid x(S_i^{(3t+2)} \cup D_{3t+2})\right).$$

Let $t \geq 1$. For every $i$,

$$\mu\left(x(s_i^{(3t+2)}) \mid x(S_i^{(3t+2)} \cup D_{3t+2})\right) \quad = \quad p_{\mu,(S_i^{(3t+2)} \cup D_{3t+2}) - s_i^{(3t+2)}}(\sigma_{s_i^{(3t+2)}} x).$$

The discrete parallelogram structure of $C^{(3t+2)}$ guarantees that each set $S_i^{(3t+2)} \cup D_{3t+2} - s_i^{(3t+2)}$ contains $\mathcal{P}_k$ and is contained in $\mathcal{P} \cup B_k^c$. By definition of $k$, we then have

$$\left| p_{\mu,(S_i^{(3t+2)} \cup D_{3t+2}) - s_i^{(3t+2)}}(\sigma_{s_i^{(3t+2)}} x) - p_\mu(\sigma_{s_i^{(3t+2)}} x) \right| < \epsilon. \tag{16}$$

We claim that (16) also holds for $t = 0$ (and every $i$). In this case, $S_i^{(2)} \cup D_2 - s_i^{(2)}$ need not contain $\mathcal{P}_k$. However, $S_i^{(2)} \cup D_2 \cup [0, n+1] \times [-k, -1] - s_i^{(2)}$ does contain $\mathcal{P}_k$ (and is contained in $\mathcal{P} \cup B_k^c$). Moreover, since $\mu$ is an MRF and $D_2 \supseteq C_0 = \underline{\partial} K_n$, we have

$$p_{\mu,S_i^{(2)} \cup D_2 - s_i^{(2)}}(\sigma_{s_i^{(2)}} x) = p_{\mu,S_i^{(2)} \cup D_2 \cup [0,n+1] \times [-k,-1] - s_i^{(2)}}(\sigma_{s_i^{(2)}} x)$$

which is within $\epsilon$ of $p_\mu(\sigma_{s_i^{(2)}} x)$, proving the claim.

It follows from (C2) that $p_\mu$ is the uniform limit of continuous functions on $\mathrm{supp}(\nu)$; since, by (C3), it is also positive on $\mathrm{supp}(\nu)$, it has a lower bound $c > 0$ there. We can therefore integrate with respect to $\nu$ to see that

$$\left| \int -\log p_{\mu,(S_i^{(3t+2)} \cup D_{3t+2}) - s_i^{(3t+2)}}(\sigma_{s_i^{(3t+2)}} x) \, d\nu - \int I_\mu(x) \, d\nu \right| < \epsilon c^{-1}. \tag{17}$$

We now combine (10), (11), (12), (13), (14), (15), and (17) to see that

$$\begin{aligned}
&\left| \int -\log \mu(x(K_n)) \, d\nu - \int I_\mu(x) \, d\nu \left( \sum_{t=0}^{\lfloor \frac{n}{m} \rfloor - 2} |C_{3t+2}| \right) \right| \\
&\leq n^2 \epsilon c^{-1} + \frac{n}{m}(9(R+1)(n+2) + 2km^2) \log \gamma + (4(2R+2) + 2m + 2)(n + 2R + 1) \log \gamma \\
&\quad \leq n^2 \epsilon c^{-1} + n(n + 2R + 1) \log \gamma \left( \frac{9(R+1)}{m} + \frac{2km}{n+2} + \frac{4(2R+2) + 2m + 2}{n} \right).
\end{aligned}$$

Note that $\sum_t |C_{3t+2}| \geq n^2 - 2nm - 2\frac{n}{m} m(km+1) = n^2 - 2n((k+1)m+1)$. Therefore,

$$\begin{aligned}
\left| \int \frac{-\log \mu(x(K_n))}{n^2} \, d\nu - \int I_\mu(x) \, d\nu \right| &\leq \frac{2(k+1)(m+1)}{n} \left| \int I_\mu(x) \, d\nu \right| + \epsilon c^{-1} \\
&+ \log \gamma \left( 1 + \frac{2R}{n} + \frac{1}{n} \right) \left( \frac{9(R+1)}{m} + \frac{2km}{n+2} + \frac{4(2R+2) + 2m + 2}{n} \right).
\end{aligned}$$

Recalling that $m = \lfloor \sqrt{nR} \rfloor$, $\frac{R}{n} \to 0$, and $k$ is a constant, the right-hand side of this inequality approaches $\epsilon c^{-1}$ as $n \to \infty$. Therefore,

$$-\epsilon c^{-1} + \int I_\mu(x) \, d\nu \leq \liminf_{n \to \infty} \frac{\int -\log \mu(x(K_n))}{n^2} \, d\nu \text{ and}$$

$$\limsup_{n \to \infty} \frac{\int -\log \mu(x(K_n))}{n^2} \, d\nu \leq \epsilon c^{-1} + \int I_\mu(x) \, d\nu.$$

By letting $\epsilon \to 0$, we see that

$$\lim_{n\to\infty} \frac{\int -\log \mu(x(K_n))}{n^2}\ d\nu = \int I_\mu(x)\ d\nu,$$

completing the proof for $d = 2$.

We now outline an inductive proof for $d > 2$. We first note that we can broadly describe the above construction as follows: for every $n$, $K_n = [0, n+1]^d$ was partitioned as $\bigcup_{i=0}^M C_i$, which was broken into two classes $B$ (for "big") and $S$ (for "small"); there, $B$ consisted of $\{C_i : i = 2 \mod 3\}$. In fact, to set up our induction, we will need to assume that the partition could be done for $[0, n+1]^{d-1} \times [0, n-\ell+1]$ for an arbitrary parameter $\ell$ (dependent on $n$) such that $\frac{\ell}{n} \to 0$ as $n \to \infty$. We will always assume $C_0$ to be $\underline{\partial}([0, n+1]^{d-1} \times [0, n-\ell+1])$ and assign $C_0 \in S$. For technical reasons, we will also need a condition whose significance was not obvious in the $d = 2$ case:

**(i)** $\frac{M}{n} \to 0$ as $n \to \infty$.

We then dealt with the conditional probabilities $\mu(x(C_i) \mid x(\bigcup_{j=0}^{i-1} C_j))$ via two methods. For $C_i \in B$, we broke $C_i$ into its sites, written in lexicographic order, and used the uniform convergence of $p_{\mu,U\cup S}$ by fixing a $k$ so that for any set $S \cup U$ which contains $\mathcal{P}_k$ and is contained in $B_k^c \cup \mathcal{P}$, $|p_{\mu,S\cup U}(x) - p_\mu(x)| < \epsilon$. We then required the following two properties:

**(ii)** For any $C_i \in B$ and any $v \in C_i$, $(\mathcal{P}_k + v) \setminus C_i \subseteq ([0, n+1]^{d-1} \times [0, n-\ell+1])^c \cup \bigcup_{q=0}^{i-1} C_q$. (This ensures that for $v \in C_i$, when we compute

$$\mu\left(x(v) \mid x\left((\bigcup_{q=0}^{i-1} C_q) \cup (C_i \cap (\mathcal{P} + v))\right)\right)$$

$$= \mu\left(x(v) \mid x\left(([0, n+1]^{d-1} \times [0, n-\ell+1])^c \cup (\bigcup_{q=0}^{i-1} C_q) \cup (C_i \cap (\mathcal{P} + v))\right)\right),$$

we've already conditioned on the sites in $\mathcal{P}_k + v$.)

**(iii)** For any $C_i \in B$ and any $v \in C_i$, $(\mathcal{F}_k + v) \cap \bigcup_{q=0}^{i-1} C_q = \varnothing$, where $\mathcal{F}_k = [-k, k]^d \setminus (\mathcal{P}_k \cup \{0\})$. (This ensures that for each site $v \in C_i$, when we compute $\mu\left(x(v) \mid x((\bigcup_{q=0}^{i-1} C_q) \cup (C_i \cap (\mathcal{P} + v)))\right)$, we have not yet conditioned on any site in $\mathcal{F}_k + v$.)

For $C_i \in S$, we conditioned on $x(C_i)$ "all at once" by using Lemma 3.5. For this, we needed shapes $S_i \subseteq \bigcup_{q=i+1}^M C_q$ with the following property:

**(iv)** Arbitrary configurations on $S_i$ and $\bigcup_{q=0}^i C_q$ can be extended to a point of $x$: For any $v \in \mathcal{L}_{S_i}(X)$ and $w \in \mathcal{L}_{\bigcup_{q=0}^i C_q}(X)$, there exists $x \in X$ s.t. $x(S_i) = v$ and $x\left(\bigcup_{q=0}^i C_i\right) = w$.

This property suffices for the use of Lemma 3.5 because we can take $T_i = \bigcup_{q=i+1}^M C_q$ and $U_i = C_i \cup T_i \cup \partial T_i$. Then $\underline{\partial} U_i \subseteq C_i \cup \partial T_i$, which is disjoint from $S_i$. Also, $U_i \setminus T_i \subseteq C_i \cup \partial T_i \subseteq \bigcup_{q=0}^i C_q$, so (iv) implies the hypothesis of Lemma 3.5.

Finally, the contributions from uses of Lemma 3.5 need to be negligible as $n \to \infty$. For our choices made above, $U_i \setminus S_i = C_i \cup \partial \left( \bigcup_{q=i+1}^M C_q \right) \cup \left( \left( \bigcup_{q=i+1}^M C_q \right) \setminus S_i \right)$, so the following condition is sufficient.

**(v)** As $n \to \infty$,

$$n^{-d} \sum_{C_i \in S} |C_i| \to 0, n^{-d} \sum_{C_i \in S} \left| \partial \left( \bigcup_{q=i+1}^M C_q \right) \right| \to 0, n^{-d} \sum_{C_i \in S} \left| \left( \bigcup_{q=i+1}^M C_q \right) \setminus S_i \right| \to 0.$$

(Note that although the volume of our partitioned shape is in fact $(n+2)^{d-1}(n - \ell + 2)$, since $\frac{(n+2)^{d-1}(n-\ell+2)}{n^d} \to 1$, replacing this volume by $n^d$ affects nothing.)

The reader may check that the $d = 2$ construction above has properties (i)-(v), and that for any $d$, if these five properties are satisfied, then the conclusion of Theorem 3.4 holds. We then only must describe how to get such a partition for $d+1$ from one for $d$. To that end, assume that $[0, n+1]^{d-1} \times [0, n - \ell + 1]$ can be partitioned into $\{C_i\}_{i=0}^M$, that the $C_i$ are broken into two classes $B$ and $S$, and that sets $S_i$ have been chosen, for which (i)-(v) above hold. There is a small technical point in the induction; we actually need to assume that (ii) and (iii) above hold for $2k$ for the $d$-dimensional partition in order to get (ii) and (iii) to hold for $k$ for the $(d+1)$-dimensional partition.

We wish to construct a partition $\{C'_i\}_{i=0}^{M'}$ for $[0, n+1]^d \times [0, n - \ell' + 1]$, $B'$ and $S'$, and $S'_i$ with the same properties. We first define $C'_0 = \underline{\partial}([0, n+1]^d \times [0, n - \ell' + 1])$ and assign $C'_0 \in S'$. Define a parameter $m := m_n$ such that $\frac{m}{n} \to 0$ and $\frac{m}{M} \to \infty$ as $n \to \infty$, which is possible by (i). We only consider values of $n$ large enough that $m > 2R, k$. Similarly to the $d = 2$ case, where $C_1$, $C_2$, and $C_3$ partitioned $[1, n] \times [1, m]$, we begin by making a partition of $[1, n]^d \times [1, m]$. Retroactively define $\ell$ above to be equal to $km$, which satisfies $\frac{\ell}{n} \to 0$ since $\frac{m}{n} \to 0$.

Define $C'_1 = \bigcup_{h=1}^m ([1, n]^{d-1} \times [1, k(m-h)] \times \{h\})$ and assign $C'_1 \in S'$. Then, for $1 \leq i \leq M$, define $C'_{i+1} = \bigcup_{h=1}^m (C_i + kme_d + h(-ke_d + e_{d+1}))$, assigning $C'_{i+1} \in B'$ if and only if $C_i \in B$. Finally, define $C'_{M+2} = ([1, n]^d \times [1, m]) \setminus \bigcup_{q=1}^{M+1} C'_q = \bigcup_{h=1}^m ([1, n]^d \times [n - kh, n] \times \{h\})$, and assign $C'_{M+2} \in S'$.

We now partition the prisms $[1, n]^d \times [jm + 1, (j+1)m]$, $1 \leq j < \lceil \frac{n-\ell'}{m} \rceil - 1$, via translates of $C'_i$, $1 \leq i \leq M+2$: for any such $j$ and $i$, $C'_{j(M+2)+i}$ is defined to be $C'_i + mje_{d+1}$, and is in $B'$ if and only if $C'_i$ was.

Finally, we define $M' = (\lceil \frac{n-\ell'}{m} \rceil - 1)(M+2) + 1$ and $C'_{M'} = [1, n]^d \times [m \lceil \frac{n-\ell'}{m} \rceil - m + 1, n - \ell' + 1]$, and assign it to $S'$. We still must define the sets $S_i$ required for (iv) and (v):

- Define $S'_0 = [R+1, n-R]^d \times [R+1, n - \ell' - R]$ as for $d = 2$.
- For any $0 \leq j < \lfloor \frac{n}{m} \rfloor$, define

$$S'_{j(M+2)+1} = ([R+1, n-R]^{d-1} \times [km + R, n - R] \times [mj + R + 1, m(j+1)]) \\ \cup ([R+1, n-R]^d \times [m(j+1) + R + 1, n - \ell' - R]).$$

- For any $0 \le j < \lfloor \frac{n}{m} \rfloor$, define
$$S'_{(j+1)(M+2)} = [R+1, n-R]^d \times [m(j+1)+R+1, n-\ell'-R].$$
- For any $0 \le j < \lfloor \frac{n}{m} \rfloor$ and $2 \le i \le M+1$, define
$$S'_{j(M+2)+i} = ((S_{i-1} + kme_d) \times [mj+R+1, m(j+1)]) \\ \cup ([R+1, n-R]^d \times [m(j+1)+R+1, n-\ell'-R]).$$
- Finally, define $S'_{M'} = \varnothing$. (Again, this is analogous to the $\mathbb{Z}^2$ proof.)

Due to the quite long and technical proof, we leave it to the reader to check that (i)-(v) are satisfied for this partition, completing the proof by induction.

□

## 4. Pressure Approximation schemes

In this section, we derive, as a consequence of our pressure representation results, an algorithm for approximating the pressure of a shift-invariant nearest-neighbor Gibbs interaction $\Phi$. For this, we assume that the exact values of $\Phi$ are known; it would be impossible to hope for a bound on computation time for $P(\Phi)$ if $\Phi$ were uncomputable or very hard to compute. The main idea of this section comes from Gamarnik and Katz [5, Corollary 1].

We would like to apply our main results to shift-invariant measures $\nu$ which are as easy as possible to integrate against, for instance the atomic measure supported on a periodic orbit. In general, a $\mathbb{Z}^d$ SFT $X$ need not have any periodic points ([1]). However, any nearest neighbor SFT $X$ that satisfies SSF must have a periodic point: choose any $b \in \mathcal{A}$ and let $a \in \mathcal{A}$ such that $b^{N_0} a^{\{0\}}$ is locally admissible in $X$, then the point $z$ defined by $z_v = a$ if $\sum_i v_i$ is even and $z_v = b$ if $\sum_i v_i$ is odd is periodic and in $X$.

**Proposition 4.1.** *Let $\Phi$ be a nearest neighbor $\mathbb{Z}^d$ interaction and $X = X_\Phi$. Assume that*

(i) *$X$ satisfies SSF,*
(ii) *$\Phi$ satisfies SSM at exponential rate.*

*Then there is an algorithm to compute $P(\Phi)$ to within $\epsilon$ in time $e^{O((\log \frac{1}{\epsilon})^{d-1})}$.*

Note that in the case $d = 2$ this gives a polynomial-time approximation scheme.

*Proof.* Let $\mu$ be the unique Gibbs state corresponding to $\Phi$.

Let $z$ be a periodic point, which exists by SSF, and $\nu$ the shift-invariant atomic measure supported on the orbit of $z$. The assumptions of Theorem 3.1 are satisfied: assumptions (A1) and (A3) follow from SSF by Propositions 2.24 and 2.17, and assumption (A2) follows from SSM and Proposition 2.14.

We conclude from Theorem 3.1 that
$$P(\Phi) = \int (I_\mu + A_\Phi) d\nu = (1/|D|)(\sum_{v \in D} - \log p_\mu(\sigma^v(z)) - \sum_{v \in D,\ v' \sim v} \Phi(z(\{v, v'\})))$$
(here, $D \subset \mathbb{Z}^2$ is a fundamental domain for $z$).

Since we assume that the exact values of the interaction $\Phi$ are known, it suffices to compute the desired approximations to $p_\mu(x)$ for all $x = \sigma^v(z),\ v \in D$. We may assume $v = 0$ (the proof is the same for all $v$).

Recall the notation from the proof of Proposition 2.14:

$$S_n = B_n \setminus \mathcal{P}_n,\ \partial S_n = U_n \cup C_n$$

where

$$U_n = (\partial S_n) \cap \mathcal{P}, C_n = \partial S_n \setminus U_n.$$

Note that no site in $C_n$ neighbors one in $U_n$. So, for any locally admissible configurations on $C_n$ and $U_n$, their concatentation is locally admissible, and therefore globally admissible by SSF. Then, by Propositions 2.23 and 2.24, any such concatenation has positive $\mu$-measure as well. Therefore, we may use the fact that $\mu$ is an MRF to represent $p_\mu$ as a weighted average:

$$p_\mu(z) = \sum_{\text{locally admissible } \delta \in \mathcal{A}^{C_n}} \mu(z(0) \mid z(U_n)\delta)\mu(\delta).$$

Let $\delta^{z(0),n}$ achieve $\max \mu(z(0) \mid z(U_n)\delta)$ and $\delta_{z(0),n}$ achieve $\min \mu(z(0) \mid z(U_n)\delta)$ over all locally admissible $\delta \in \mathcal{A}^{C_n}$. Clearly,

$$\mu(z(0) \mid z(U_n)\delta_{z(0),n}) \le p_\mu(z) \le \mu(z(0) \mid z(U_n)\delta^{z(0),n}).$$

By SSM at exponential rate, there are constants $C, \alpha > 0$ such that these upper and lower bounds on $p_\mu(z)$ differ by at most $Ce^{-\alpha n}$.

This gives sequences of upper and lower bounds on $p_\mu(z)$ with accuracy $e^{-\Omega(n)}$. For $\delta \in \mathcal{A}^{C_n}$, the time to compute $\mu(z(0) \mid z(U_n)\delta)$ is $e^{O(n^{d-1})}$ since this is the ratio of two probabilities of configurations of size $O(n^{d-1})$, each of which can be computed using the transfer matrix method from [11, Lemma 4.8] in time $e^{O(n^{d-1})}$. Since $|\mathcal{A}^{C_n}| = e^{O(n^{d-1})}$, the total time to compute the upper and lower bounds is $e^{O(n^{d-1})}e^{O(n^{d-1})} = e^{O(n^{d-1})}$. □

There are many sufficient conditions for SSM at exponential rate (for instance, see the discussion in [11]).

We remark that Corollary 4.13 of [11] gives an algorithm to compute $P(\Phi)$ that is less efficient, in that the approximation to within $\epsilon$ requires time $e^{O((\log \frac{1}{\epsilon})^{(d-1)^2})}$. However, it applies more generally than Proposition 4.1 in that it requires only assumption (ii) above and not assumption (i).

Finally, we note that, for a nearest-neighbor interaction $\Phi$, the algorithm here to compute $P(\Phi)$ also computes $h(\mu)$ for any Gibbs measure corresponding to $\Phi$.

## 5. Connections with Thermodynamic Formalism

In this section, we connect Corollary 3.2 with results from Ruelle's thermodynamic formalism. We begin by taking a new look at Corollary 3.2.

Consider the following probability-vector-valued functions:

$$\hat{p}_{\mu,n}(x) := \mu(y(0) = \cdot \mid y(\mathcal{P}_n) = x(\mathcal{P}_n))$$

and

$$\hat{p}_\mu(x) := \mu(y(0) = \cdot \mid y(\mathcal{P}) = x(\mathcal{P})).$$

Note that these functions do not depend on $x_0$ ($\hat{p}_{\mu,n}$ is similar in spirit to the function $\hat{p}^n_\mu$, which was introduced and used in the proof of Proposition 2.14).

Note that $\hat{p}_\mu$ is defined only $\mu$-a.e. By Martingale convergence, $\hat{p}_{\mu,n}$ converges to $\hat{p}_\mu$ for $\mu$-a.e. $x \in \text{supp}(\mu)$.

The following result relates unform convergence of $\hat{p}_{\mu,n}$ with a continuity property of $\hat{p}_\mu$.

**Definition 5.1.** *A function $g$ is* **past-continuous** *on a shift space $X$ if it is continuous on $X$ and, for all $x \in X$, $g(x)$ depends only on $x(\mathcal{P})$: if $x, y \in X$ and $x(\mathcal{P}) = y(\mathcal{P})$, then $g(x) = g(y)$.*

**Proposition 5.2.** *Let $\mu$ be a stationary measure.*

(i) *$\hat{p}_{\mu,n}$ converges uniformly on* $\mathrm{supp}(\mu)$ *iff $\hat{p}_\mu$ is past-continuous on* $\mathrm{supp}(\mu)$ *(i.e., $\hat{p}_\mu$ agrees with a past-continuous function ($\mu$-a.e.)).*

(ii) *If $\hat{p}_{\mu,n}$ converges uniformly on* $\mathrm{supp}(\mu)$, *then* $\lim_{S \to \mathcal{P}} \; p_{\mu,S} = p_\mu(x)$, *uniformly on* $\mathrm{supp}(\mu)$.

*Proof.* For finite $S \subset \mathcal{P}$, $a \in \mathcal{A}$ and $x \in \mathrm{supp}(\mu)$, we can write

(18)
$$\mu(y(0) = a \mid y(S) = x(S)) = \frac{1}{\mu([x(S)])} \int_{[x(S)]} p_\mu(y(0) = a \mid y(\mathcal{P}) = x(\mathcal{P})) d\mu(y).$$

(i), $\Leftarrow$: If $\hat{p}_\mu$ is past-continuous on $\mathrm{supp}(\mu)$, then we can take the integrand in (18) to be a continuous, and therefore uniformly continuous, function of $y(\mathcal{P}),\ y \in [x(S)]$. Thus, taking $S = \mathcal{P}_n$ we get uniform convergence of $\hat{p}_{\mu,n}$.

(i), $\Rightarrow$: If $\hat{p}_{\mu,n}$ converges uniformly on $\mathrm{supp}(\mu)$, then its limit is a uniform limit of past-continuous functions on $\mathrm{supp}(\mu)$ and thus is past-continuous on $\mathrm{supp}(\mu)$. By Martingale convergence, this limit agrees with $\hat{p}_\mu$ ($\mu$-a.e.).

(ii): By (i), we may assume that $\hat{p}_\mu$ is past-continuous. Take $a = x(0)$ in (18). Given $\epsilon > 0$, for sufficiently large $n$, if $S$ is a finite set satisfying $\mathcal{P}_n \subset S \subset \mathcal{P}$, then for all $x \in \mathrm{supp}(\mu)$, the integrand in (18) is within $\epsilon$ of $p_\mu(x)$, and so $p_{\mu,S}(x)$ is within $\epsilon$ of $p_\mu(x)$.

□

**Corollary 5.3.** *If $\Phi$ is a nearest-neighbor interaction with underlying SFT $X$, $\mu$ is a Gibbs measure for $\Phi$,*

(D1) *$X$ satisfies the classical D-condition,*
(D2) *$\hat{p}_\mu$ is past-continuous on $X$, and*
(D3) *$c_\mu > 0$,*

*then*

$$P(\Phi) = \int I_\mu(x) + A_\Phi(x)\ d\nu = \int I_\mu(x) - \sum_{i=1}^{d} \Phi(x(\{0, e_i\}))\ d\nu$$

*for every shift-invariant measure $\nu$ with* $\mathrm{supp}(\nu) \subseteq X$.

*Proof.* This follows immediately from Corollary 3.2 and Proposition 5.2. □

Next, we will show how Ruelle's thermodynamic formalism [14] can be applied to obtain a result similar to Corollary 5.3. For this, we need to give a definition of interactions more general than nearest-neighbor.

**Definition 5.4.** *An* ***interaction*** *is a shift-invariant function $\Phi$ from $\mathcal{A}^*$ $\mathbb{R} \cup \{\infty\}$ which takes on the value $\infty$ for only finitely many configurations on shapes* *containing* $0$. *An interaction is* ***finite-range*** *if there exists* $N$ *so that if* $\Phi(w) \neq 0$, *then* $w$ *has shape with diameter at most* $N$.

We remark, without proof, that the results of this paper extend to finite-range interactions, in particular to interactions which are non-zero only on vertices and edges (such interactions are often also called nearest-neighbor interactions, but form a slightly more general class than the interactions that we have called nearest-neighbor in this paper).

**Definition 5.5.** *An interaction is called* ***summable*** *if*

$$\sum_{\Lambda \subset \mathbb{Z}^d, \Lambda \ni 0, |\Lambda| < \infty} \frac{1}{|\Lambda|} \max_{x \in \mathcal{A}^\Lambda:\ \Phi(x) \neq \infty} |\Phi(x)| < \infty.$$

*An interaction is called* ***absolutely summable*** *if*

$$\sum_{\Lambda \subset \mathbb{Z}^d, \Lambda \ni 0, |\Lambda| < \infty} \max_{x \in \mathcal{A}^\Lambda:\ \Phi(x) \neq \infty} |\Phi(x)| < \infty.$$

The underlying SFT corresponding to an interaction $\Phi$ is defined as follows.

$$X_\Phi = \{x \in \mathcal{A}^{\mathbb{Z}^d}\ :\ \Phi(x(S)) \neq \infty \text{ for all finite } S \subseteq \mathbb{Z}^d\}.$$

The role of configurations on finite subsets with $\Phi(x) = \infty$ corresponds to forbidden configurations and defines the SFT $X_\Phi$. This is equivalent to Ruelle's setting where one starts with a background SFT $X$ and only considers finite-valued interactions defined on configurations that are not forbidden.

A continuous function $A_\Phi$ and underlying SFT $X_\Phi$ can be associated to any summable interaction $\Phi$, in a similar fashion as what was done for nearest-neighbor interactions in Section 2.2:

$$A_\Phi(x) := -\sum_{\Lambda \subseteq \mathcal{P} \cup \{0\}, \Lambda \ni 0, |\Lambda| < \infty,\ \Phi(x(\Lambda)) \neq \infty} \Phi(x(\Lambda)).$$

Let $X$ be a nonempty SFT and let $I(X)$ denote the set of summable interactions $\Phi$ with $X_\Phi = X$. Ruelle shows in [14, Section 3,2] that for any continuous function $f$ on $X$, there exists a summable interaction $\Phi$ such that $X_\Phi = X$ and $f = A_\Phi$. That is, the mapping $\Phi \mapsto A_\Phi$ from $I(X)$ to $C(X)$ is surjective. However, the restriction of this mapping to the set of absolutely summable interactions is not surjective.

For an absolutely summable interaction, the concepts of energy function, partition function and Gibbs measure are defined analogously to those in Section 2.2. For details, see [14]. For us, a Gibbs measure is shift-invariant by definition, as in the nearest-neighbor case.

Gibbs measures and equilibrium states are intimately connected by the following two standard theorems (these were mentioned at the end of Section 2.6 for the special case of nearest-neighbor interactions). Proofs of both can be found in [14].

**Theorem 5.6.** ([8]) *If* $\Phi$ *is an absolutely summable interaction, then any equilibrium state for* $A_\Phi$ *on* $X_\Phi$ *is a Gibbs measure for* $\Phi$.

**Theorem 5.7.** ([4]) *If* $\Phi$ *is an absolutely summable interaction whose underlying SFT* $X_\Phi$ *satisfies the D-condition, then any Gibbs measure for* $\Phi$ *is an equilibrium state for* $A_\Phi$ *on* $X_\Phi$.

**Definition 5.8.** ([14]) *Absolutely summable interactions $\Phi$ and $\Phi'$ with the same underlying SFT $X$ are called* ***physically equivalent*** *if $A_\Phi$ and $A_{\Phi'}$ have a common equilibrium state.*

The following result is Proposition 4.7(b) from [14].

**Theorem 5.9.** *If $X$ is an SFT that satisfies the D-condition, and if $\Phi, \Phi'$ are physically equivalent absolutely summable interactions with underlying SFT $X$, then $\int (A_\Phi - A_{\Phi'})\, d\nu$ is constant for all shift-invariant measures $\nu$ on $X$.*

In order to make a connection with Corollary 5.3, we need the following.

**Lemma 5.10.** *If $\Phi$ is a shift-invariant nearest-neighbor interaction, $\mu$ is an equilibrium state for $A_\Phi$, and $I_\mu$ is continuous, then $\mu$ is also an equilibrium state for $-I_\mu$.*

*Proof.* Let $X^{\mathcal{P}} = \{x(\mathcal{P}) : x \in X\}$.

Since log is a concave function, we can use Jensen's inequality to see that for any shift-invariant measure $\nu$ on $X$,

$$\int_X I_\nu - I_\mu \, d\nu = \int_X \log\left(\frac{\mu(x(0) \mid x(\mathcal{P}))}{\nu(x(0) \mid x(\mathcal{P}))}\right) \, d\nu \le \log \int_X \frac{\mu(x(0) \mid x(\mathcal{P}))}{\nu(x(0) \mid x(\mathcal{P}))} \, d\nu$$
$$= \log \int_{X^{\mathcal{P}}} \sum_{a \in \mathcal{A}} \frac{\mu(x(0) = a \mid x(\mathcal{P}))}{\nu(x(0) = a \mid x(\mathcal{P}))} \, \nu(x(0) = a \mid x(\mathcal{P})) \, d\nu(\mathcal{P})$$
$$= \log \int_{X^{\mathcal{P}}} \sum_{a \in \mathcal{A}} \mu(x(0) = a \mid x(\mathcal{P}) \, d\nu(\mathcal{P})) = \log \int_{X^{\mathcal{P}}} \, d\nu(\mathcal{P}) = \log 1 = 0.$$

So, $\int I_\mu \, d\nu \ge \int I_\nu \, d\nu$. For any $\nu$, we have $h(\nu) = \int I_\nu \, d\nu$. Therefore,

$$h(\nu) - \int I_\mu \, d\nu \le h(\nu) - \int I_\nu \, d\nu = 0 = h(\mu) - \int I_\mu \, d\mu.$$

This means that the function $h(\rho) + \int -I_\mu \, d\rho$ is maximized at $\rho = \mu$, so $\mu$ is an equilibrium state for $-I_\mu$ by definition.

□

**Corollary 5.11.** *If $\Phi$ is a nearest-neighbor interaction with underlying SFT $X = X_\Phi$, $\mu$ is a Gibbs measure for $\Phi$,*

(E1) *$X$ satisfies the classical D-condition, and*
(E2) *$-I_\mu = A_{\Phi'}$ for some absolutely summable interaction $\Phi'$ with $X_{\Phi'} = X$,*

*then*

$$P(\Phi) = \int I_\mu(x) + A_\Phi(x) \, d\nu = \int I_\mu(x) - \sum_{i=1}^{d} \Phi(x(\{0, e_i\})) \, d\nu$$

*for every shift-invariant measure $\nu$ with* $\operatorname{supp}(\nu) \subseteq X$.

*Proof.* Since $X$ satisfies the $D$-condition, $\mu$ is an equilibrium state for $A_\Phi$. By Lemma 5.10, $A_\Phi$ and $-I_\mu$ are physically equivalent. Since $-I_\mu = A_{\Phi'}$ for an absolutely summable interaction $\Phi'$ and the D-condition holds, by Theorem 5.9, $\int A_\Phi + I_\mu \, d\nu$ is a constant over all shift-invariant measures $\nu$ on $X$. But, for $\nu = \mu$, this integral is $\int A_\Phi + I_\mu \, d\mu = P(A_\Phi) = P(\Phi)$. Therefore, $P(\Phi) = \int I_\mu + A_\Phi \, d\nu$ for all shift-invariant $\nu$ on $X$. □

Corollary 5.3 and Corollary 5.11 give the same integral representation for $P(\Phi)$ for every shift-invariant measure $\nu$. The classical D-condition is assumed for both, but the other hypotheses relate to different types of continuity: (D2) and (D3) of Corollary 5.3 imply past-continuity of $I_\mu$ (by Proposition 2.16), while (E2) of Corollary 5.11 is a strong form of continuity of $I_\mu$ (as mentioned earlier, continuity of $I_\mu$ implies that $I_\mu = A_\Phi$ for some summable, but not necessarily absolutely summable, interaction $\Phi$).

## Acknowledgments

We are grateful to Nishant Chandgotia and David Gamarnik for several helpful discussions.

## References

[1] R. Berger. The undecidability of the domino problem. *Mem. Amer. Math. Soc.*, 66, 1966.
[2] R. Bradley. A caution on mixing conditions for random fields. *Statist. Probab. Lett.*, 8:489–491, 1989.
[3] N. Chandgotia and T. Meyerovitch. Markov random fields, Markov cocycles and the 3-colored chessboard. Preprint, ArXiv: 1305.0808v1, 2013.
[4] R.L. Dobrushin. Description of a random field by means of conditional probabilities and conditions for its regularity. *Theory Probab. Appl.*, 13:197–224, 1968.
[5] D. Gamarnik and D. Katz. Sequential cavity method for computing free energy and surface pressure. *J. Stat. Physics*, 137:205 – 232, 2009.
[6] H. Georgii. *Gibbs measures and phase transitions.* de Gruyter Studies in Mathematics, Walter de Gruyter & Co., Berlin, 1988.
[7] M. Hochman and T. Meyerovitch. A characterization of the entropies of multidimensiona shifts of finite type. *Ann. Math*, 171(3):2011–2038, 2012.
[8] O. E. Lanford III and D. Ruelle. Observables at infinity and states with short range correlations in statistical mechanics. *Comm. Math. Phys.*, 13:194 –215, 1969.
[9] U. Krengel. *Ergodic Theorems.* De Gruyter, 1985.
[10] D. Lind and B. Marcus. *An introduction to symbolic dynamics and coding.* Cambridge University Press, 1995, reprinted 1999.
[11] B. Marcus and R. Pavlov. Computing bounds on entropy of $\mathbb{Z}^d$ stationary Markov random fields. *SIAM J. Discrete Math*, to appear, 2012.
[12] B. Marcus and R. Pavlov. Approximating entropy for a class of Markov random fields and pressure for a class of functions on shifts of finite type. *Ergodic Theory and Dynamical Systems*, 33:186–220, 2013.
[13] T. Meyerovitch. Gibbs and equilibrium measures for some families of subshifts. *Ergodic Theory and Dynamical Systems*, 33:934 – 953, 2013.
[14] D. Ruelle. *Thermodynamic Formalism.* Cambridge University Press, 1978.
[15] P. Walters. *An Introduction to Ergodic Theory*, volume 79 of *Graduate Texts in Mathematics.* Springer, 1975.
[16] P. Walters. A variational principle for the pressure of continuous transformations. *Amer. J. Math.*, 97(4):937–971, 1975.
[17] D. Weitz. Combinatorial criteria for uniqueness of Gibbs measures. *Random Structures and Algorithms*, 27:445–475, 2005.

Brian Marcus, Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, BC V6T 1Z2
*E-mail address*: `marcus@math.ubc.ca`

Ronnie Pavlov, Department of Mathematics, University of Denver, 2360 S. Gaylord St., Denver, CO 80208
*E-mail address*: `rpavlov@du.edu`