SIAM J. Discrete Math.
Vol. 27, No. 3, pp. 1544–1558

COMPUTING BOUNDS FOR ENTROPY OF STATIONARY $\mathbb{Z}^d$ MARKOV RANDOM FIELDS*

BRIAN MARCUS† AND RONNIE PAVLOV‡

Abstract. For any stationary $\mathbb{Z}^d$ Gibbs measure that satisfies strong spatial mixing, we obtain sequences of upper and lower approximations that converge to its entropy. In the case $d = 2$, these approximations are efficient in the sense that they are accurate to within ϵ and can be computed in time polynomial in $1/\epsilon$.

Key words. Markov random fields, Gibbs measures, entropy, disagreement percolation

AMS subject classifications. Primary, 37D35, 37B50; Secondary, 37B10, 37B40

DOI. 10.1137/120887382

1. Introduction. The entropy of a stationary $\mathbb{Z}^d$ Markov random field (MRF) is notoriously difficult to compute. Recently, Gamarnik and Katz [4] developed a technique for estimating entropy, and more generally pressure, for certain MRFs. Their approach built on earlier work of Weitz [20], who gave an algorithm for efficiently counting the number of independent sets in finite graphs. The algorithm was based on the construction of a computation tree, and the proof of efficiency relied on the concept of strong spatial mixing (SSM) [11, Part 2, section 2]. Coming from the direction of ergodic theory, we showed that a variant of the transfer matrix method provides efficient algorithms for estimating entropy for certain $\mathbb{Z}^2$ MRFs [14], [10]. Our argument relied on a version of SSM implied by a disagreement percolation condition developed in [18] (see Proposition 2.4 below). We regard an algorithm as "efficient" if it computes upper and lower bounds accurate to within ϵ in time polynomial in $1/\epsilon$.

While both approaches made use of SSM, they both required other assumptions as well, some involving the existence of certain kinds of periodic configurations. The purpose of this paper is to give approximations, using only SSM as a hypothesis, which estimate the entropy of $\mathbb{Z}^d$ MRFs (and do so efficiently in the case $d = 2$). General sufficient conditions for SSM can be found in the literature, e.g., [2] and [18]; one of these is reviewed in section 2.

Assuming a standard version of SSM (at exponential rate), we obtain upper and lower bounds that are exponentially tight (see Lemma 3.1 and Theorem 3.2). While these bounds are not explicitly computable in all cases, we believe them to be of independent interest. Of special interest are nearest-neighbor stationary $\mathbb{Z}^d$ Gibbs measures, which are MRFs given by nearest-neighbor interactions. For these measures, assuming SSM, we obtain an algorithm that approximates our bounds with specified precision (Theorem 4.5). Combining all of these results, we obtain an algorithm for approximating entropy of a nearest-neighbor stationary $\mathbb{Z}^d$ Gibbs measure that is accurate to within ϵ in time $e^{O((\log(1/\epsilon))^{(d-1)^2})}$ (see Corollary 4.7). Specializing to $d = 2$, the algorithm runs in time polynomial in $1/\epsilon$. We also show how to modify

*Received by the editors August 8, 2012; accepted for publication (in revised form) July 8, 2013; published electronically September 19, 2013.
http://www.siam.org/journals/sidma/27-3/88738.html
†Department of Mathematics, University of British Columbia, Vancouver, BC V6N 2N9, Canada (marcus@math.ubc.ca).
‡Department of Mathematics, University of Denver, Denver, CO 80208 (rpavlov@du.edu).

the algorithm to approximate the pressure of the interaction that defines the Gibbs measure (Corollary 4.13).

We emphasize that our algorithms are deterministic and establish rigorous estimates, as opposed to randomized algorithms, based on Monte Carlo simulation, which establish estimates which are frequently better but are only guaranteed with prescribed high degree of probability (for instance, see [8]). Though our algorithm requires SSM as a hypothesis, it does not depend on knowledge of the rate of correlation decay in the definition of SSM.

Classical examples of Gibbs measures include the Ising model and Potts model; see [12, Chapter 2] for an introduction to models of interest in statistical mechanics.

In section 2 we introduce many of the concepts used in the paper. In section 3, we establish the upper and lower bounds for MRFs. In section 4, we give algorithms to approximate these bounds for Gibbs measures.

2. Background. We focus on MRFs on the *d-dimensional cubic lattice*, the graph defined by vertex set $\mathbb{Z}^d$ and edge set $\{\{u,v\} \ : \sum_{i=1}^d |u_i - v_i| = 1\}$. The *boundary* of a set S, which is denoted by ∂S, is the set of $v \in \mathbb{Z}^d \setminus S$ which are adjacent to some element of S.

An *alphabet* $\mathcal{A}$ is a finite set with at least two elements. For a nonempty subset $S \subset \mathbb{Z}^d$, an element $u \in \mathcal{A}^S$ is called a *configuration*; here, 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 disjoint sets, $x \in \mathcal{A}^S$ 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. We will sometimes informally identify a configuration x on a shape S with the corresponding configuration on a translate $S + v$, namely, the configuration y on $S + v$ defined by $y_u = x_{u-v}$.

We use σ to denote the $\mathbb{Z}^d$ *shift action* on $\mathcal{A}^{\mathbb{Z}^d}$ defined by $(\sigma_v(x))_u = x_{u+v}$. The set $\mathcal{A}^{\mathbb{Z}^d}$ is a topological space when endowed with the product topology (where $\mathcal{A}$ has the discrete topology), and any subset inherits the induced topology. By a $\mathbb{Z}^d$-*measure*, we mean a Borel probability measure on $\mathcal{A}^{\mathbb{Z}^d}$. This means that any μ is determined by its values on the sets $[w] := \{x \in \mathcal{A}^{\mathbb{Z}^d} \ : \ x|_S = w\}$, where w is a configuration with arbitrary finite shape $S \subseteq \mathbb{Z}^d$. Such sets are called *cylinder sets*. Note that for configurations x and y on disjoint sets, we have $[xy] = [x] \cap [y]$. 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, $\mu(w, v \mid u)$ represents the conditional measure $\mu([w] \cap [v] \mid [u])$. A $\mathbb{Z}^d$-measure μ is *translation-invariant* (or *stationary*) if $\mu(A) = \mu(\sigma_v A)$ for all measurable sets A and $v \in \mathbb{Z}^d$. A $\mathbb{Z}^d$-measure is *fully supported* if it assigns strictly positive measure to every cylinder set in $\mathcal{A}^{\mathbb{Z}^d}$.

DEFINITION 2.1. *A $\mathbb{Z}^d$-measure μ is a $\mathbb{Z}^d$ MRF if, for any finite $S \subset \mathbb{Z}^d$, any $\eta \in \mathcal{A}^S$, any finite $T \subset \mathbb{Z}^d$ such that $\partial S \subseteq T \subseteq \mathbb{Z}^d \setminus S$, and any $\delta \in \mathcal{A}^T$ with $\mu(\delta) > 0$,*

$$\mu(\eta \mid \delta|_{\partial S}) = \mu(\eta \mid \delta). \tag{1}$$

Informally, μ 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 μ-conditionally independent given a configuration on the sites in ∂S. In many papers, the MRF condition is defined in terms of a parameter r, and the set of all sites in $\mathbb{Z}^d \setminus S$ that are within distance r of S plays the role of ∂S. Obviously our definition corresponds to the case $r = 1$ (a "nearest-neighbor" MRF).

Another commonly used variant on our definition of MRF involves conditioning, in the right-hand side of (1), on an entire configuration on $\mathbb{Z}^d \setminus S$ a.e. rather than arbitrarily large finite configurations. However, the definitions are equivalent (one can just take weak limits) and the finite approach is a bit more concrete.

For two configurations $y, z \in \mathcal{A}^T$ on a finite set T, let $D(y,z) = \{v \in \mathbb{Z}^d : \ y_v \neq z_v\}$. Let $d(\cdot,\cdot)$ denote the L^1 distance on $\mathbb{Z}^d$.

DEFINITION 2.2. *A stationary $\mathbb{Z}^d$ MRF μ satisfies SSM if there exist constants $C, \alpha > 0$, such that for any finite $V \subset \mathbb{Z}^d$, $u \in V$, $\partial V \subseteq T \subset V \cup \partial V$, $x \in \mathcal{A}^{\{u\}}$, and $y, z \in \mathcal{A}^T$ satisfying $\mu(y), \mu(z) > 0$,*

$$\left|\mu(x \mid y) - \mu(x \mid z)\right| \leq Ce^{-\alpha d(\{u\}, D(y,z))}.$$

This definition of SSM is actually equivalent to a more general condition where the single site u is replaced with an arbitrary finite subset of sites U. For completeness we give a proof.

LEMMA 2.3. *For any stationary $\mathbb{Z}^d$ MRF that satisfies SSM, there exist constants $C, \alpha > 0$, such that for any finite $V \subset \mathbb{Z}^d$, $U \subseteq V$, $\partial V \subseteq T \subset V \cup \partial V$, $x \in \mathcal{A}^U$, and $y, z \in \mathcal{A}^T$ with $\mu(y), \mu(z) > 0$,*

$$\left|\mu(x \mid y) - \mu(x \mid z)\right| \leq |U| Ce^{-\alpha d(U, D(y,z))}. \tag{2}$$

(The constants C, α can be taken to be those in the definition of SSM.)

Proof. Arbitrarily order the sites in U as $1, 2, \ldots, |U|$. Then

$$\begin{aligned}
&\left|\mu(x \mid y) - \mu(x \mid z)\right| \\
&= \left| \left(\prod_{i=1}^{|U|} \mu(x_i \mid y, x_1, \ldots x_{i-1})\right) - \left(\prod_{i=1}^{|U|} \mu(x_i \mid z, x_1, \ldots x_{i-1})\right) \right| \\
&\leq \left[\sum_{i=1}^{|U|} \left(\prod_{j=1}^{i-1} \mu(x_j \mid y, x_1, \ldots x_{j-1})\right) \left(\prod_{j=i+1}^{|U|} \mu(x_j \mid z, x_1, \ldots x_{j-1})\right) \right. \\
&\qquad \left. \left|\mu(x_i \mid y, x_1, \ldots x_{i-1}) - \mu(x_i \mid z, x_1, \ldots x_{i-1})\right| \right] \leq C|U|e^{-\alpha d(U, D(y,z))}. \qquad \square
\end{aligned}$$

We note that SSM can be defined for probability measures on fairly arbitrary undirected graphs. Sometimes SSM, as we have defined it, is called "SSM with exponential rate."

There are a variety of conditions in the literature which guarantee SSM of an MRF: for example, see [2], [3], [6], [16], [18], and [20]. We present the one from [18] here as one of the most general and easy to state.

Let μ be a stationary MRF. Let

$$q(\mu) = \max_{y,z \in \mathcal{A}^{\partial \mathbf{0}} : \ \mu(y), \mu(z) > 0} \rho(\mu(\cdot \mid y), \mu(\cdot \mid z)),$$

where ρ denotes total variation distance of distributions on $\mathcal{A}^{\mathbf{0}}$. Let $p_c = p_c(\mathbb{Z}^d)$ denote the critical probability for site percolation in $\mathbb{Z}^d$. (We will not define $p_c(\mathbb{Z}^d)$ or discuss percolation theory here; for a good introduction to the subject, see [7].)

PROPOSITION 2.4. *If $q(\mu) < p_c$, then μ satisfies SSM.*

This result is essentially contained in [18, Theorem 1]; see [10, Theorem 3.10] for more explanation.

The following is the standard notion, in ergodic theory and information theory, of entropy.

DEFINITION 2.5. *Given a $\mathbb{Z}^d$-measure μ and a finite set $S \subset \mathbb{Z}^d$, one defines the* entropy *of μ on S as*

$$H_\mu(S) = \sum_{w \in \mathcal{A}^S} -\mu(w) \log(\mu(w)),$$

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

We also have the notion of conditional entropy.

DEFINITION 2.6. *Given a $\mathbb{Z}^d$-measure μ and disjoint finite sets $S, T \subset \mathbb{Z}^d$, one defines the* conditional entropy *of μ on S, given T, as*

$$H_\mu(S \mid T) = \sum_{w \in \mathcal{A}^{S \cup T}:\ \mu(w|_T) > 0} -\mu(w) \log \mu(w|_S \mid w|_T),$$

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

Let μ be a stationary $\mathbb{Z}^d$-measure. The following monotonicity property is well known: if $S, T, T' \subset \mathbb{Z}^d$ are finite, $T' \subset T$, and $S \cap T = \varnothing$, then $H_\mu(S \mid T) \leq H_\mu(S \mid T')$. We can now extend Definition 2.6 to infinite T by defining

$$H_\mu(S \mid T) = \lim_n H_\mu(S \mid T_n)$$

for a nested sequence of finite sets $T_1 \subset T_2 \subset \cdots$ with $\cup_n T_n = T$; by the monotonicity property just mentioned, the limit exists and does not depend on the particular choice of sequence T_n. With this definition, it is clear that the previously mentioned monotonicity also holds for infinite T and T'.

LEMMA 2.7. *Let μ be a stationary $\mathbb{Z}^d$-measure. If $S, T, T' \subset \mathbb{Z}^d$, S is finite, $T' \subset T$, and $S \cap T = \varnothing$, then*

$$H_\mu(S \mid T) \leq H_\mu(S \mid T').$$

We will find the following notation useful later. Let S and T be disjoint finite sets. For a stationary $\mathbb{Z}^d$ MRF μ and a fixed configuration $y \in \mathcal{A}^T$ with $\mu(y) > 0$, we define

$$H_\mu(S \mid y) = \sum_{x \in \mathcal{A}^S} -\mu(x \mid y) \log \mu(x \mid y).$$

Thus, we can write

$$H_\mu(S \mid T) = \sum_{y \in \mathcal{A}^T,\ \ \mu(y) > 0} \mu(y) H_\mu(S \mid y). \tag{3}$$

If T is the disjoint union of T_1 and T_2, we can write

$$H_\mu(S \mid T_1 \cup T_2) = \sum_{y \in \mathcal{A}^{T_1}:\ \mu(y) > 0} \mu(y) \sum_{w \in \mathcal{A}^{T_2}:\ \ \mu(wy) > 0} \mu(w \mid y) H_\mu(S \mid wy). \tag{4}$$

We can also define the entropy of a stationary $\mathbb{Z}^d$-measure itself, also known as entropy rate in information theory.

Definition 2.8. *The* measure-theoretic entropy *of a stationary* $\mathbb{Z}^d$*-measure* μ *on* $\mathcal{A}^{\mathbb{Z}^d}$ *is defined by*

$$h(\mu) = \lim_{j_1, j_2, \ldots, j_d \to \infty} \frac{H_\mu(S_{j_1 \ldots j_d})}{j_1 j_2 \cdots j_d},$$

where $S_{j_1 j_2 \ldots j_d}$ *denotes the* $j_1 \times j_2 \times \cdots \times j_d$ *rectangular prism* $\prod_{i=1}^d [1, j_i]$.

It is well known that the limit exists independent of the rates at which each $j_1, j_2, \ldots, j_d$ approaches infinity [5, Theorem 15.12].

There is also a useful conditional entropy formula for $h(\mu)$. For this, we consider the usual lexicographic order on $\mathbb{Z}^d$: $x \prec y$ if for some $1 \leq k \leq d$, $x_i = y_i$ for $i = 1, \ldots, k-1$ and $x_k < y_k$. Let $\mathcal{P}^- = \{z \in \mathbb{Z}^d : z \prec \mathbf{0}\}$, where $\mathbf{0}$ denotes the origin.

Theorem 2.9 ([5, equation 15.18]). *Let* μ *be a stationary* $\mathbb{Z}^d$*-measure. Then*

$$h(\mu) = H_\mu(\mathbf{0} \mid \mathcal{P}^-).$$

When $d = 1$, $\mathbb{Z}^d = \mathbb{Z}$ can represent time, with the site $\mathbf{0}$ representing the present and $\mathcal{P}^-$ representing the past. The preceding result generalizes the interpretation of $h(\mu)$ as the average uncertainty of the present, conditioned on the past [19, Chapter 4].

Finally, we state a simple technical lemma that we will need.

Lemma 2.10. *Given constants* $C, \alpha > 0$*, there exists a constant* $C' > 0$ *such that if* $0 < a, b < 1$ *and* $|a - b| \leq Ce^{-\alpha n}$ *for some* $n \in \mathbb{N}$*, then*

$$|a \log a - b \log b| \leq C' n e^{-\alpha n}.$$

Proof. Clearly, without loss of generality, $0 < b \leq a < 1$. We first show that this implies

$$|a \log a - b \log b| \leq (a - b)(1 - \log(a - b)). \tag{5}$$

To see this, observe

$$\begin{aligned} |a \log a - b \log b| &\leq |a \log a - b \log a| + |b \log a - b \log b| \\ &= (a - b)(-\log a) + b \log\left(1 + \frac{a - b}{b}\right) \\ &\leq (a - b)(-\log(a - b)) + (a - b) = (a - b)(1 - \log(a - b)). \end{aligned}$$

Using (5) and the monotonicity of the function $x(1 - \log x)$ on $(0, 1]$, we see that if $|a - b| \leq Ce^{-\alpha n}$, then

$$|a \log a - b \log b| \leq Ce^{-\alpha n}(1 - \log(Ce^{-\alpha n})) = Ce^{-\alpha n}(1 - \log C + \alpha n) \leq C' n e^{-\alpha n}$$

for some C' depending only on C and α. ◻

3. Entropy bounds for stationary MRFs. Let $\mathcal{P}^+ = \{z \in \mathbb{Z}^d : z \succeq \mathbf{0}\}$. Then $\mathcal{P}^+ = \mathbb{Z}^d \setminus \mathcal{P}^-$. Let B_n denote the d-dimensional cube of side length $2n + 1$ centered at $\mathbf{0}$. Let $S_n = B_n \cap \mathcal{P}^+$ and $U_n = B_n \cap \partial \mathcal{P}^+$.

We claim that $U_n \subset \partial S_n$. To see this, note that by definition, if $x \in \partial \mathcal{P}^+$, then $x \in \mathcal{P}^-$ and x has a nearest neighbor $y \in \mathcal{P}^+$. It follows that for some $1 \leq k \leq d$, we have $x_i = y_i$ for all $i \neq k$ and either ($x_k = -1$ and $y_k = 0$) or ($x_k = 0$ and $y_k = 1$). In either case, if $x \in U_n = B_n \cap \partial \mathcal{P}^+$, then $y \in B_n$ and so $y \in S_n$. Thus, $x \in \partial S_n$. Figure 1 shows these sets for $d = 2$.

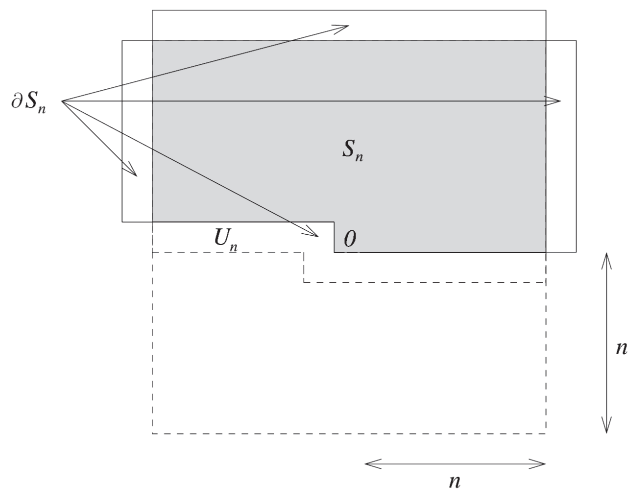

FIG. 1. *U_n, S_n, and ∂S_n. Here, the vertical axis represents the first coordinate and the horizontal axis represents the second coordinate.*

LEMMA 3.1. *Let μ be a stationary $\mathbb{Z}^d$ MRF. Then*

$$H_\mu(\mathbf{0} \mid \partial S_n) \le h(\mu) \le H_\mu(\mathbf{0} \mid U_n). \tag{6}$$

Proof. Since $h(\mu) = H_\mu(\mathbf{0} \mid \mathcal{P}^-)$ and $U_n \subset \mathcal{P}^-$, it follows from Lemma 2.7 that

$$H_\mu(\mathbf{0} \mid \partial S_n \cup \mathcal{P}^-) \le h(\mu) \le H_\mu(\mathbf{0} \mid U_n). \tag{7}$$

But since $\mathbf{0} \in S_n$, $S_n \cap \mathcal{P}^- = \varnothing$ and μ is a $\mathbb{Z}^d$ MRF, it follows that the left-hand sides of (6) and (7) agree. ∎

We remind the reader of standard notational conventions. For a function f on the integers, we write $f = O(n)$ to mean there exists a constant $C > 0$ such that for sufficiently large n, $|f(n)| \le Cn$, and $f = \Omega(n)$ to mean there exists a constant $C > 0$ such that for sufficiently large n, $f(n) \ge Cn$.

THEOREM 3.2. *Let μ be a stationary $\mathbb{Z}^d$ MRF that satisfies SSM. Then $|H_\mu(\mathbf{0}|U_n) - H_\mu(\mathbf{0}|\partial S_n)| = O(n)e^{-\alpha n}$, where α is the exponent in the definition of SSM.*

Proof. The main idea is that the SSM condition forces the probability of a letter on $\mathbf{0}$, conditioned on a configuration y on U_n, to be approximately the same as when conditioned on any extension of y to a configuration on ∂S_n.

Let $L_n = \partial S_n \setminus U_n$. Note that ∂S_n is the disjoint union of U_n and L_n. For every configuration $y \in \mathcal{A}^{U_n}$ such that $\mu(y) > 0$, let

$$E(y) = \{w \in \mathcal{A}^{L_n} : \mu(yw) > 0\}.$$

By (3) and (4), we can write

$$H_\mu(\mathbf{0} \mid U_n) = \sum_{y \in \mathcal{A}^{U_n}:\ \mu(y)>0} \mu(y) H_\mu(\mathbf{0} \mid y) \text{ and} \tag{8}$$

$$H_\mu(\mathbf{0} \mid \partial S_n) = \sum_{y \in \mathcal{A}^{U_n}:\ \mu(y)>0} \mu(y) \sum_{w \in E(y)} \mu(w \mid y) H_\mu(\mathbf{0} \mid yw). \tag{9}$$

Fix y as above. Let C and α be the positive constants for SSM. For any configuration y on U_n and $w, w' \in E(y)$, we have $d(\{\mathbf{0}\}, D(w, w')) \geq n$. By SSM applied to $V = S_n$, $T = \partial S_n$, we have that for all $x \in \mathcal{A}^{\mathbf{0}}$, $y \in \mathcal{A}^{U_n}$, and $w, w' \in E(y)$,

$$\big|\mu(x \mid yw) - \mu(x \mid yw')\big| \leq Ce^{-\alpha n}.$$

Now,

$$\mu(x \mid y) = \sum_{w \in E(y)} \mu(w \mid y)\mu(x \mid yw),$$

and so for all $w \in E(y)$,

$$\begin{aligned}
\big|\mu(x \mid y) - \mu(x \mid yw)\big| &= \left| \left(\sum_{w' \in E(y)} \mu(w' \mid y)\mu(x \mid yw') \right) - \mu(x \mid yw) \right| \\
&= \left| \sum_{w' \in E(y)} \mu(w' \mid y)(\mu(x \mid yw') - \mu(x \mid yw)) \right| \\
&\leq \sum_{w' \in E(y)} \mu(w' \mid y)\big|\mu(x \mid yw') - \mu(x \mid yw)\big| \leq Ce^{-\alpha n}.
\end{aligned}$$

By Lemma 2.10, there is a constant C' depending only on C and α such that

$$\big|\mu(x \mid y)\log\mu(x \mid y) - \mu(x \mid yw)\log\mu(x \mid yw)\big| \leq C'ne^{-\alpha n}$$

for some $C' > 0$.

Thus,

$$\begin{aligned}
&\left| H_\mu(\mathbf{0} \mid y) - \sum_{w \in E(y)} \mu(w \mid y) H_\mu(\mathbf{0} \mid yw) \right| \\
&\quad = \left| \sum_{w \in E(y)} \mu(w \mid y)\big(H_\mu(\mathbf{0} \mid y) - H_\mu(\mathbf{0} \mid yw)\big) \right| \\
&\quad \leq \sum_{w \in E(y)} \mu(w \mid y)\big|H_\mu(\mathbf{0} \mid y) - H_\mu(\mathbf{0} \mid yw))\big| \\
&\quad \leq \sum_{x \in \mathcal{A}^{\mathbf{0}}} \sum_{w \in E(y)} \mu(w \mid y)\big|\mu(x \mid y)\log\mu(x \mid y) - \mu(x \mid yw)\log\mu(x \mid yw)\big| \\
&\quad \leq \sum_{x \in \mathcal{A}^{\mathbf{0}}} \sum_{w \in E(y)} \mu(w \mid y)C'ne^{-\alpha n} \leq |\mathcal{A}|C'ne^{-\alpha n}.
\end{aligned}$$

Applying (8) and (9), we get

$$\big|H_\mu(\mathbf{0} \mid U_n) - H_\mu(\mathbf{0} \mid \partial S_n)\big| \leq |\mathcal{A}|C'ne^{-\alpha n}. \qquad \square$$

By combining Lemma 3.1 and Theorem 3.2, we obtain exponentially accurate upper and lower approximations to $h(\mu)$ for any stationary $\mathbb{Z}^d$ MRF μ which satisfies SSM. In the following section, we show that when the MRF is a (nearest-neighbor) Gibbs measure (defined in the next section) and $d = 2$, there is an efficient algorithm to approximate these bounds. A version of the well-known Hammersley–Clifford theorem [15] shows that any fully supported (nearest-neighbor) MRF is a (nearest-neighbor) Gibbs measure. However, that result can fail in general; see [1] for an example based on a construction for finite graphs given in [13].

4. Computation of entropy bounds for stationary Gibbs measures. Let $\gamma : \mathcal{A} \to (0,\infty)$, $\beta_i : \mathcal{A} \times A \to [0,\infty)$, $i = 1,\dots,d$. For a finite $V \subset \mathbb{Z}^d$ and $w \in \mathcal{A}^V$, let

$$I(w) = \left(\prod_{v \in V} \gamma(v)\right) \prod_{i=1}^{d} \prod_{\{v \in V : v+e_i \in V\}} \beta_i(v, v+e_i).$$

In statistical physics, often $\log I(w)$ is referred to as the energy of the configuration w, and $\log \gamma$ and $\log \beta_i$ correspond to external fields and interaction strengths, respectively.

A configuration $\delta \in \mathcal{A}^{\partial V}$ is called V*-admissible* if there exists at least one $w \in \mathcal{A}^V$ such that $I(w\delta) > 0$.

DEFINITION 4.1. *Given γ, β_i as above, for all $|V| < \infty$ and V-admissible δ, define for all $w \in \mathcal{A}^V$,*

$$\Lambda^\delta(w) = \frac{I(w\delta)}{\sum_{x \in \mathcal{A}^V} I(x\delta)}.$$

The collection $\{\Lambda^\delta\}_{V,\delta}$ is called a stationary $\mathbb{Z}^d$ Gibbs specification *for the* local interactions γ, β_i.

Note that each Λ^δ is a probability measure on $\mathcal{A}^V$, and for $U \subset V$ and $w \in \mathcal{A}^U$,

$$\Lambda^\delta(w) = \sum_{c \in \mathcal{A}^{V \setminus U}} \Lambda^\delta(wc).$$

Also, we can regard Λ^δ as a probability measure on configurations $y \in \mathcal{A}^{V \cup \partial V}$ that agree with δ on ∂V.

Many classical models can be expressed using this framework (see [12, Chapter 2]):

- Ising model: $\mathcal{A} = \{\pm 1\}$, $\gamma(a) = e^{Ea}$, $\beta_i(a,b) = e^{Jab}$ for constants E (external magnetic field) and J (coupling strength).
- n-state Potts model: $\mathcal{A} = \{1,\dots,n\}$, $\gamma(a) = 1$, $\beta_i(a,b) = e^{J\delta_{ab}}$, where δ_{ab} is the Kronecker delta.
- n-coloring shift: $\mathcal{A} = \{1,\dots n\}$, $\gamma(a) = 1$, $\beta_i(a,b) = 1 - \delta_{ab}$; this can be thought of as the limiting case of the n-state Potts model as $J \to -\infty$.
- Hard square model: $\mathcal{A} = \{0,1\}$, $\gamma(a) = 1$, $\beta_i(a,b) = 1 - ab$. In other words, the hard square model weights nearest-neighbor configurations equally, subject to the "hard constraint" that nearest neighbor sites cannot both be occupied by 1.

DEFINITION 4.2. *A* stationary $\mathbb{Z}^d$ Gibbs measure *for a stationary $\mathbb{Z}^d$ Gibbs specification is a stationary $\mathbb{Z}^d$ MRF μ on $\mathcal{A}^{\mathbb{Z}^d}$ such that for any finite set V and $\delta \in \mathcal{A}^{\partial V}$, if $\mu(\delta) > 0$, then δ is V-admissible and for all $x \in \mathcal{A}^V$*

$$\mu(x \mid \delta) = \Lambda^\delta(x).$$

Specifications can be used to define MRFs, not just Gibbs measures (see [5]). However, we find the concept of specification most useful for Gibbs measures.

Gibbs measures, as defined here, are often referred to as nearest-neighbor Gibbs measures in the literature. Note that since the β_i are allowed to take on the value 0, a Gibbs measure need not be fully supported. Also, note that by definition, a necessary condition for $\mu(\delta) > 0$ is V-admissibility of δ. While there may be no finite procedure for determining if a configuration δ has positive measure, there is a finite procedure for determining if δ is V-admissible. For this reason, we impose an SSM condition on the specification that defines a Gibbs measure, rather than the Gibbs measure itself.

DEFINITION 4.3. *A stationary $\mathbb{Z}^d$ Gibbs specification Λ satisfies SSM if there exist constants $C, \alpha > 0$, such that for all finite $V \subset \mathbb{Z}^d$, $u \in V$, $\partial V \subseteq T \subset V \cup \partial V$, $x \in \mathcal{A}^{\{u\}}$, $y, z \in \mathcal{A}^T$, such that $\delta = y|_{\partial V}$ and $\eta = z|_{\partial V}$ are V-admissible and $\Lambda^\delta(y), \Lambda^\eta(z) > 0$, and then*

$$\left|\Lambda^\delta(x \mid y) - \Lambda^\eta(x \mid z)\right| \le Ce^{-\alpha d(\{u\}, D(y,z))}.$$

Note that if the specification of a Gibbs measure μ satisfies SSM, then the measure μ itself satisfies SSM as an MRF. It is well known that when the specification satisfies SSM there is a unique Gibbs measure corresponding to the specification. In fact, a weaker notion of spatial mixing, known as weak spatial mixing [11], is sufficient.

Just as in the case of MRFs, the single-site version of SSM for Gibbs specifications implies a finite-set version, and the proof, which we omit, is very similar to that of Lemma 2.3.

LEMMA 4.4. *For any stationary $\mathbb{Z}^d$ Gibbs specification that satisfies SSM, there exist constants $C, \alpha > 0$, such that for any finite $V \subset \mathbb{Z}^d$, $U \subseteq V$, $\partial V \subseteq T \subset V \cup \partial V$, $x \in \mathcal{A}^U$, $y, z \in \mathcal{A}^T$, such that $\delta = y|_{\partial V}$ and $\eta = z|_{\partial V}$ are V-admissible and $\Lambda^\delta(y), \Lambda^\eta(z) > 0$, and then*

$$\left|\Lambda^\delta(x \mid y) - \Lambda^\eta(x \mid z)\right| \le |U| Ce^{-\alpha d(U, D(y,z))}. \tag{10}$$

(The constants C, α can be taken to be those in the definition of SSM.)

We note that there are conditions, such as one analogous to Proposition 2.4, that imply SSM for stationary Gibbs specifications.

The following is the main result of this section.

THEOREM 4.5. *Let μ be a stationary $\mathbb{Z}^d$ Gibbs measure whose specification satisfies SSM. Let (K_n), $n \in \mathbb{N}$, be a sequence of sets satisfying $K_n \subset B_n$ and $|K_n| = O(n^{d-1})$. Then there is an algorithm which, on input n, computes upper and lower bounds to $H_\mu(\mathbf{0} \mid K_n)$ in time $e^{O(n^{(d-1)^2})}$ to within tolerance $e^{-n^{d-1}}$.*

Remark 4.6. For this and all subsequent results involving running time of algorithms involving μ, we do not count computation of the Gibbs parameters γ and β_i towards the claimed running time. (In other words, we assume that we are given approximations to γ and β_i with arbitrarily good precision before performing any computation.) We also note that the algorithms here do not depend on knowledge of specific values of the parameters C and α of SSM.

As an immediate consequence of Lemma 3.1, Theorem 3.2, and Theorem 4.5 (applied to $K_n = \partial S_{n-1}$ and $K_n = U_n$), we have the following.

COROLLARY 4.7. *Let μ be a stationary $\mathbb{Z}^d$ Gibbs measure whose specification satisfies SSM. Then there is an algorithm which, on input n, computes upper and lower bounds to $h(\mu)$ in time $e^{O(n^{(d-1)^2})}$ to within tolerance $e^{-\Omega(n)}$.*

Note that for $d = 2$ this gives an algorithm to compute $h(\mu)$ to within $O(1/n)$ in polynomial time (in n).

For the proof of Theorem 4.5, we will need the following result.

LEMMA 4.8. *Let μ be a stationary $\mathbb{Z}^d$ Gibbs measure. Let (K_n), $n \in \mathbb{N}$, be a sequence of sets satisfying $K_n \subset B_n$ and $|K_n| = O(n^{d-1})$. Then for any sequence $(m_n)_{n\in\mathbb{N}}$ of positive integers, there is an algorithm which, on input n, determines which $\delta \in \mathcal{A}^{\partial B_{n+m_n}}$ are B_{n+m_n}-admissible and, for those which are, computes $\Lambda^\delta(w)$ for all $w \in \mathcal{A}^{K_n}$, in running time $e^{O((n+m_n)^{d-1})}$.*

Proof. For simplicity, we prove this only for $d = 2$. The general case follows along similar lines.

Fix sequences (K_n) and (m_n) as in the statement of the lemma, a particular value of n, a B_{n+m_n}-admissible δ, and $w \in \mathcal{A}^{K_n}$. Define

$$I^\delta(w) := \sum_{c \in \mathcal{A}^{B_{n+m_n} \setminus K_n}} I(wc\delta).$$

We will show that

$$I^\delta(w) = I(\delta)x^T \overline{M}_{-n-m_n} \overline{M}_{-n-m_n+1} \cdots \overline{M}_{n+m_n-1} y, \tag{11}$$

where each $\overline{M}_i$ is a square matrix and x, y are vectors, all indexed by $\mathcal{A}^{[-n-m_n,n+m_n]}$.

For $a, b \in \mathcal{A}^{[-n-m_n,n+m_n]}$, we write $a = a_{-n-m_n}, \dots a_{n+m_n}$, $b = b_{-n-m_n}, \dots b_{n+m_n}$, and

$$\gamma(a) = \prod_{j=-n-m_n}^{n+m_n} \gamma(a_j),\ \ \beta_1(a,b) = \prod_{j=-n-m_n}^{n+m_n} \beta_1(a_j,b_j),\ \ \beta_2(a) = \prod_{j=-n-m_n}^{n+m_n-1} \beta_2(a_j,a_{j+1}).$$

For $i = -n - m_n, \dots, n + m_n$, define the transfer matrix M_i by

$$(M_i)_{(a,b)} = \gamma(a)\beta_1(a,b)\beta_2(a)\beta_2(\delta_{i,-n-m_n-1}, a_{-n-m_n})\beta_2(a_{n+m_n}, \delta_{i,n+m_n+1})$$

for $a, b \in A^{[-n-m_n,n+m_n]}$. Let $V_i = \{i\} \times [-n - m_n + 1, \dots n + m_n - 1]$ and let

$$(\overline{M}_i)_{(a,b)} = (M_i)_{(a,b)}$$

except if a, b are such that $V_i \cap K_n \neq \varnothing$ and $a|_{\{j:(i,j)\in K_n\}} \neq w|_{V_i \cap K_n}$, in which case we set $(\overline{M}_i)_{(a,b)} = 0$. Let $\delta_L = \delta|_{\{-n-m_n-1\}\times\{[-n-m_n,n+m_n]\}}$ and $\delta_R = \delta|_{\{n+m_n+1\}\times\{[-n-m_n,n+m_n]\}}$ be the restrictions of δ to the left and right sides of ∂B_{n+m_n}. We define vectors x, y by

$$x_a = \beta_1(\delta_L, a), \quad y_a = (M_{n+m_n})_{(a,\delta_R)} \text{ for } a \in A^{[-n-m_n,n+m_n]}.$$

To see that the expression (11) holds, observe that the interactions that make up $I^\delta(w)$ are assembled one column at a time by each matrix $\overline{M}_i$, with adjustments on the border of B_{n+m_n} contributed by $I(\delta)$, x, and y.

Note that each $\overline{M}_i$ can be constructed in time $(e^{O(n+m_n)})^2 = e^{O(n+m_n)}$, x and y can be constructed in time $e^{O(n+m_n)}$, and $I(\delta)$ can be constucted in time $O(n+m_n)$. Each matrix multiplication takes time at most $(e^{O(n+m_n)})^3 = e^{O(n+m_n)}$. Thus, $I^\delta(w)$ can be constructed in time $e^{O(n+m_n)}$. This can be done for all $w \in \mathcal{A}^{K_n}$ in time $e^{O(n+m_n)}e^{O(n)} = e^{O(n+m_n)}$. (This is the only part of the proof where we use the assumption on the size of K_n.)

Since

$$\Lambda^\delta(w) = \frac{I^\delta(w)}{\sum_{x\in\mathcal{A}^{K_n}} I^\delta(x)},$$

we can compute $\Lambda^\delta(w)$ for all $w \in \mathcal{A}^{K_n}$ and all B_{n+m_n}-admissible $\delta \in \mathcal{A}^{\partial B_{n+m_n}}$ in time $(e^{O(n+m_n)})^2 = e^{O(n+m_n)}$.

For $d > 2$, the proof follows along similar lines using transfer matrices indexed by configurations on $(d-1)$-dimensional arrays. $\square$

Proposition 4.9. *Let μ be a stationary $\mathbb{Z}^d$ Gibbs measure whose specification satisfies SSM with constants C and α. Let (K_n), $n \in \mathbb{N}$, be a sequence of sets satisfying $K_n \subset B_n$ and $|K_n| = O(n^{d-1})$. Then for any sequence (m_n) of positive integers, there is an algorithm which, on input n, computes upper and lower bounds $\mu^+(w)$ and $\mu^-(w)$ to $\mu(w)$, for all $w \in \mathcal{A}^{K_n}$, in time $e^{O((n+m_n)^{d-1})}$, such that*

$$\mu^+(w) - \mu^-(w) \le Ce^{-\alpha m_n}|K_n|.$$

Proof. Fix sequences (K_n) and (m_n) as in the statement of the proposition, a particular value of n, and $w \in \mathcal{A}^{K_n}$. Observe that

$$\mu(w) = \sum_{\delta \in \mathcal{A}^{\partial B_{m_n+n}}:\ \mu(\delta)>0} \mu(w \mid \delta)\mu(\delta).$$

Let δ^w be a configuration δ which achieves $\max_{\{B_{n+m_n}\text{-admissible } \delta\}} \Lambda^\delta(w)$ and let δ_w be a configuration δ which achieves $\min_{\{B_{n+m_n}\text{-admissible } \delta\}} \Lambda^\delta(w)$. Since strict positivity of $\mu(\delta)$ implies B_{n+m_n}-admissibility, it follows that

$$\Lambda^{\delta_w}(w) \le \mu(w) \le \Lambda^{\delta^w}(w).$$

Since μ satisfies SSM, it follows by Lemma 2.3 (applied to $V = B_{n+m_n}, T = \partial V$ and $U = K_n$) that

$$0 \le \Lambda^{\delta^w}(w) - \Lambda^{\delta_w}(w) \le Ce^{-\alpha m_n}|K_n|. \tag{12}$$

By Lemma 4.8, we can identify all B_{m_n+n}-admissible δ and compute $\Lambda^\delta(w)$ for all such δ and all $w \in \mathcal{A}^{K_n}$ in time $e^{O((n+m_n)^{d-1})}$. Thus in time $e^{O((n+m_n)^{d-1})}$ we can identify, for all $w \in \mathcal{A}^{K_n}$, δ_w and δ^w, and compute the upper and lower bounds $\Lambda^{\delta_w}(w)$ and $\Lambda^{\delta^w}(w)$.

This, together with (12), completes the proof. ∎

Similarly, we have the following.

Proposition 4.10. *Let μ be a stationary $\mathbb{Z}^d$ Gibbs measure whose specification satisfies SSM with constants C and α. Let (K_n), $n \in \mathbb{N}$, be a sequence of sets satisfying $K_n \subset B_n \setminus \{\mathbf{0}\}$ and $|K_n| = O(n^{d-1})$. Then for any sequence (m_n) of positive integers, there is an algorithm which, on input n, computes upper and lower bounds $\mu^+(x_0 \mid w)$ and $\mu^-(x_0 \mid w)$ to $\mu(x_0 \mid w)$ for all $x_0 \in \mathcal{A}$ and $w \in \mathcal{A}^{K_n}$ with $\mu(w) > 0$ in time $e^{O((n+m_n)^{d-1})}$ such that*

$$\mu^+(x_0 \mid w) - \mu^-(x_0 \mid w) \le Ce^{-\alpha m_n}.$$

Proof. Fix sequences (K_n) and (m_n) as in the statement of the proposition, a particular value of n, and $w \in \mathcal{A}^{K_n}$. Write

$$\mu(x_0 \mid w) = \sum_{\delta \in \mathcal{A}^{\partial B_{m_n+n}}:\ \mu(w\delta)>0} \mu(x_0 \mid w, \delta)\mu(\delta \mid w).$$

As in the proof of Proposition 4.9, we can find B_{n+m_n}-admissible $\delta^{x_0,w}$ and $\delta_{x_0,w}$ such that

$$\Lambda^{\delta_{x_0,w}}(x_0|w) \le \mu(x_0 \mid w) \le \Lambda^{\delta^{x_0,w}}(x_0|w) \text{ and}$$
$$0 \le \Lambda^{\delta^{x_0,w}}(x_0|w) - \Lambda^{\delta_{x_0,w}}(x_0|w) \le Ce^{-\alpha m_n}.$$

(Here, we apply SSM to $V = B_{n+m_n}$, $T = (\partial V) \cup K_n$, $U = \{\mathbf{0}\}$.) Then apply Lemma 4.8 to compute these bounds, i.e., compute $\Lambda^{\delta^{x_0,w}}(x_0 w)$, $\Lambda^{\delta^{x_0,w}}(w)$, $\Lambda^{\delta_{x_0,w}}(x_0 w)$, and $\Lambda^{\delta_{x_0,w}}(w)$. $\square$

Proof of Theorem 4.5. Let (K_n), $n \in \mathbb{N}$, be a sequence of sets satisfying $K_n \subset B_n$ and $|K_n| = O(n^{d-1})$.

Fix any sequence (m_n), a particular value of n, and $w \in \mathcal{A}^{K_n}$. Let $\mu^+(w), \mu^-(w)$, $\mu^+(x_0|w), \mu^-(x_0|w)$ be as in Propositions 4.9 and 4.10. We will use these to obtain upper and lower estimates to $H_\mu(\mathbf{0} \mid K_n)$ and bound the error. Later, we will describe how to select values of (m_n) that will yield our algorithm with the asserted properties.

Let $f(x) = -x \log x$. Let $\mu^{--}(x_0 \mid w)$ denote whichever of $\mu^+(x_0 \mid w), \mu^-(x_0 \mid w)$ achieves $\min(f(\mu^+(x_0 \mid w)), f(\mu^-(x_0 \mid w)))$. Using concavity of f and Lemma 2.10, there exists $C' > 0$ (independent of n and m_n) such that

$$0 \le f(\mu(x_0 \mid w)) - f(\mu^{--}(x_0 \mid w)) \le C' m_n e^{-\alpha m_n}. \tag{13}$$

Recall that

$$H_\mu(\mathbf{0} \mid K_n) = \sum_{w \in \mathcal{A}^{K_n}} \mu(w) \sum_{x_0 \in \mathcal{A}^{\mathbf{0}}} f(\mu(x_0 \mid w)).$$

Let $H_\mu^-(\mathbf{0} \mid K_n)$ denote the expression obtained by substituting $\mu^-(w)$ for $\mu(w)$ and $\mu^{--}(x_0 \mid w)$ for $\mu(x_0 \mid w)$:

$$H_\mu^-(\mathbf{0} \mid K_n) = \sum_{w \in \mathcal{A}^{K_n}} \mu^-(w) \sum_{x_0 \in \mathcal{A}^{\mathbf{0}}} f(\mu^{--}(x_0 \mid w)).$$

Then $H_\mu^-(\mathbf{0} \mid K_n) \le H_\mu(\mathbf{0} \mid K_n)$.

Now, we estimate the difference between $H_\mu(\mathbf{0} \mid K_n)$ and $H_\mu^-(\mathbf{0} \mid K_n)$. We have

$$\begin{aligned}
&H_\mu(\mathbf{0} \mid K_n) - H_\mu^-(\mathbf{0} \mid K_n) \\
&\quad = \sum_{w \in \mathcal{A}^{K_n}} \mu(w) \sum_{x_0 \in \mathcal{A}} f(\mu(x_0 \mid w)) - \sum_{w \in \mathcal{A}^{K_n}} \mu^-(w) \sum_{x_0 \in \mathcal{A}^{\mathbf{0}}} f(\mu^{--}(x_0 \mid w)) \\
&\quad = \sum_{w \in \mathcal{A}^{K_n}} \mu(w) \sum_{x_0 \in \mathcal{A}^{\mathbf{0}}} (f(\mu(x_0 \mid w)) - f(\mu^{--}(x_0 \mid w))) \\
&\qquad + \sum_{w \in \mathcal{A}^{K_n}} (\mu(w) - \mu^-(w)) \sum_{x_0 \in \mathcal{A}^{\mathbf{0}}} f(\mu^{--}(x_0 \mid w)) \\
&\quad \le |\mathcal{A}| C' m_n e^{-\alpha m_n} + |\mathcal{A}|^{|K_n|} |K_n| e^{-1} |\mathcal{A}| C e^{-\alpha m_n} \\
&\quad \le |\mathcal{A}| C' m_n e^{-\alpha m_n} + e^{\eta n^{d-1}} C e^{-\alpha m_n}
\end{aligned} \tag{14}$$

for some constant η (depending on the upper bound of $\frac{|K_n|}{n^{d-1}}$); here, we have used (13) and Proposition 4.9 in the first inequality.

The reader can check that there then exists a constant L, depending on $|\mathcal{A}|, C', C, \alpha$, and η, so that for every n, if $m_n = L n^{d-1}$, then $H_\mu(\mathbf{0} \mid K_n) - H_\mu^-(\mathbf{0} \mid K_n) < 0.5 e^{-n^{d-1}}$.

We also note that the computation time of $H_\mu^-(\mathbf{0} \mid K_n)$ is $e^{O((n+m_n)^{d-1})}$ (the total amount of time to compute $\mu^-(w)$ and $f(\mu^{--}(x_0 \mid w))$ for all $w \in \mathcal{A}^{K_n}$ and $x_0 \in \mathcal{A}^{\mathbf{0}}$.)

For the upper bound, let $\mu^{++}(x_0 \mid w)$ be whichever of $\mu^+(x_0 \mid w), \mu^-(x_0 \mid w)$ achieves $\max(f(\mu^+(x_0 \mid w)), f(\mu^-(x_0 \mid w)))$ if $x, y \le 1/e$ or $x, y \ge 1/e$, and $1/e$

otherwise. Using Lemma 2.10 and the fact that $f(x)$ achieves its maximum at $x = 1/e$, we have

$$0 \le f(\mu^{++}(x_0 \mid w)) - f(\mu(x_0 \mid w)) \le C' m_n e^{-\alpha m_n}. \tag{15}$$

(The C' is the same as above.) Then

$$H^+_\mu(\mathbf{0} \mid K_n) = \sum_{w \in \mathcal{A}^{K_n}} \mu^+(w) \sum_{x_0 \in \mathcal{A}^{\mathbf{0}}} f(\mu^{++}(x_0 \mid w))$$

is an upper bound for $H_\mu(\mathbf{0} \mid K_n)$.

Using (15), we see that

$$\begin{aligned}
(16)\quad & H^+_\mu(\mathbf{0} \mid K_n) - H_\mu(\mathbf{0} \mid K_n) \\
&= \sum_{w \in \mathcal{A}^{K_n}} \mu^+(w) \sum_{x_0 \in \mathcal{A}^{\mathbf{0}}} f(\mu^{++}(x_0 \mid w)) - \sum_{w \in \mathcal{A}^{K_n}} \mu(w) \sum_{x_0 \in \mathcal{A}^{\mathbf{0}}} f(\mu(x_0 \mid w)) \\
&= \sum_{w \in \mathcal{A}^{K_n}} \mu(w) \sum_{x_0 \in \mathcal{A}^{\mathbf{0}}} (f(\mu^{++}(x_0 \mid w)) - f(\mu(x_0 \mid w))) \\
&\quad + \sum_{w \in \mathcal{A}^{K_n}} (\mu^+(w) - \mu(w)) \sum_{x_0 \in \mathcal{A}^{\mathbf{0}}} f(\mu^{++}(x_0 \mid w)) \\
&\le |\mathcal{A}| C' m_n e^{-\alpha m_n} + |\mathcal{A}|^{|K_n|} |K_n| e^{-1} |\mathcal{A}| C e^{-\alpha m_n} \\
&\le |\mathcal{A}| C' m_n e^{-\alpha m_n} + e^{\eta n^{d-1}} C e^{-\alpha m_n}.
\end{aligned}$$

For every n, if $m_n = Ln^{d-1}$ (the L is the same as for the lower bound), then $H^+_\mu(\mathbf{0} \mid K_n) - H_\mu(\mathbf{0} \mid K_n) < 0.5e^{-n^{d-1}}$. The time to compute $H^+_\mu(\mathbf{0} \mid K_n)$ is $e^{O((n+m_n)^{d-1})}$, the same as for $H^-_\mu(\mathbf{0} \mid K_n)$.

We now describe the algorithm for choosing the values (m_n). We note that without knowledge of the explicit constants C and α from the SSM of μ, we cannot explicitly compute the constant L. However, for our purposes, knowledge of L is unnecessary.

The algorithm uses parameters n and j which both start off equal to 1, though they will be incremented later. The algorithm consists of one main loop which is run repeatedly. At the beginning of the loop, the above bounds $H^-_\mu(\mathbf{0} \mid K_n)$ and $H^+_\mu(\mathbf{0} \mid K_n)$ are computed for $m_n = jn^{d-1}$. If the bounds are not within $e^{-n^{d-1}}$ of each other, then j is incremented by 1 and the algorithm returns to the beginning of the loop. When the bounds are within $e^{-n^{d-1}}$ of each other (which will happen for large enough j by the comments following (14) and (16)), then m_n is defined to be jn^{d-1}, the value of n is incremented by 1, and the algorithm returns to the beginning of the loop.

By the above discussion, there exists L so that j will never be incremented beyond L in this algorithm. This means that there exists J so that for all sufficiently large n, $m_n = Jn^{d-1}$. Therefore, for all n, the algorithm yields upper and lower bounds to within tolerance $e^{-n^{d-1}}$ in time $e^{O((n+Jn^{d-1})^{d-1})} = e^{O(n^{(d-1)^2})}$. ∎

Remark 4.11. We remark that the algorithms in Propositions 4.9 and 4.10 can be simplified if one uses knowledge of specific values of the constants C and α in the definition of SSM. Namely, one can compute $\Lambda^\delta(w)$ (or $\Lambda^\delta(x_0 \mid w)$) for any fixed B_{n+m_n}-admissible configuration δ and then set the upper and lower bounds $\mu^\pm(w)$ (or $\mu^\pm(x_0 \mid w)$) to be $\Lambda^\delta(w) \pm Ce^{-\alpha m_n}|K_n|$ (or $\Lambda^\delta(x_0 \mid w) \pm Ce^{-\alpha m_n}$).

In theory, we can also dispense with the auxiliary sequence (m_n) in Proposition 4.10: we could instead bound $\mu(x_0 \mid w)$ by the minimum and maximum possible values of $\mu(x_0 \mid w, \delta)$ for configurations δ on ∂B_n, which would give approximations of tolerance $Ce^{-\alpha n}$ in time $O(e^{n^{d-1}})$. A similar simplification could be done for Proposition 4.9 as well, but it would not be useful for our proof of Theorem 4.5: note that in formula (14), the upper bound on $\mu(w) - \mu^-(w)$ is multiplied by $|A|^{|K_n|}$, and so this upper bound must be at most $e^{-\Omega(n^{d-1})}$. Therefore, the described simplification for Proposition 4.10 would not reduce the overall order of computation time for Theorem 4.5, since the algorithm from Proposition 4.9 would still require time $e^{O(n^{(d-1)^2})}$.

Finally, we note that in Proposition 4.10 when $K_n = \partial S_{n-1}$, there is no need to bound the conditional probabilities $\mu(x_0 \mid w)$, as they can be computed exactly (by using the methods of Lemma 4.8).

We will now describe how to extend Theorem 4.5 and Corollary 4.7 to give bounds for pressure in addition to entropy. Given local interactions γ, β_i, define

$$X = \{x \in \mathcal{A}^{\mathbb{Z}^d} : \text{ for all } v \in \mathbb{Z}^d \text{ and } 1 \le i \le d, \quad \beta_i(x_v, x_{v+e_i}) > 0\}.$$

X is the set of configurations on $\mathbb{Z}^d$ defined by nearest-neighbor constraints and so belongs to the class of nearest-neighbor (or 1-step) shifts of finite type [9].

Let $f : X \to \mathbb{R}$ be defined by

$$f(x) = \log \gamma(x_0) + \sum_{i=1}^{d} \log \beta_i(x_0, x_{e_i}). \tag{17}$$

Definition 4.12. *Let X and f be as above. Define the* pressure *of f by*

$$P(f) = \max_{\nu} \ h(\nu) + \int f d\nu,$$

where the max *is taken over all stationary measures ν with support contained in X. A measure which achieves the* max *is called an* equilibrium state *for f.*

Alternatively, pressure can be defined directly in terms of X and f, without reference to stationary measures. The definition of pressure which we have used is a corollary of the well-known variational principle [19, Chapter 9]. For general dynamical systems, the max is merely a sup; however, in our context, the sup is always achieved.

It is well known that any equilibrium state for f is a Gibbs measure for the specification defined by the interactions γ, β_i [17, Chapter 4]. As mentioned earlier, when the specification satisfies SSM, there is only one Gibbs measure μ that satisfies that specification, and so μ is an (unique) equilibrium state for f.

Corollary 4.13. *Let γ, β_i be local interactions which define a stationary $\mathbb{Z}^d$ Gibbs specification that satisfies SSM. Let f be as in* (17)*. Then there is an algorithm to compute upper and lower bounds to $P(f)$ in time $e^{O(n^{(d-1)^2})}$ to within tolerance $e^{-\Omega(n)}$.*

Proof. Let μ be the unique Gibbs measure that satisfies the specification. Then Corollary 4.7 applies to compute such bounds for $h(\mu)$.

It follows from Proposition 4.9 that for any configuration w on a single site or edge of $\mathbb{Z}^d$, one can compute upper and lower bounds to $\mu(w)$ in time $e^{O(n^{d-1})}$ to within tolerance $e^{-\Omega(n)}$. (In fact, this follows easily from weak spatial mixing.) Thus one can compute upper and lower bounds to $\int f d\mu$ in the same time with the same tolerance.

Finally, recall that μ is an equilibrium state since its specification satisfies SSM, and so we can compute the desired bounds for $h(\mu) + \int f d\mu = P(f)$. $\square$

Acknowledgments. The authors thank David Gamarnik for several helpful discussions and Andrea Montanari for pointing out the connection between our work [10], [14] and work of Gamarnik and Katz [4]. We are also grateful to the anonymous referees for valuable comments that improved this paper.

REFERENCES

[1] N. Chandgotia, *Markov Random Fields and Measures with Nearest Neighbour Gibbs Potentials*, MSc thesis, University of British Columbia, Vancouver, BC, 2011; also available online from http://www.math.ubc.ca/∼nishant/Masters thesis.pdf.

[2] R. Dobrushin, *Description of a random field by means of conditional probabilities and conditions for its regularity*, Theory Probab. Appl., 13 (1968), pp. 197–224.

[3] D. Gamarnik, D. Katz, and S. Misra, *Strong spatial mixing for list coloring of graphs*, arXiv:1207.1223, 2012.

[4] D. Gamarnik and D. Katz, *Sequential cavity method for computing free energy and surface pressure*, J. Statist. Phys., 137 (2009), pp. 205–232.

[5] H. Georgii, *Gibbs Measures and Phase Transitions*, de Gruyter Stud. Math., Walter de Gruyter, Berlin, 1988.

[6] L. Goldberg, R. Martin, and M. Paterson, *Strong spatial mixing with fewer colours for lattice graphs*, SIAM J. Comput., 35 (2005), pp. 486–517.

[7] G. Grimmett, *Percolation*, 2nd ed., Springer, New York, 1999.

[8] C. Kenyon, D. Randall, and A. Sinclair, *Approximating the number of monomer-dimer coverings of a lattice*, J. Statist. Phys., 83 (1996), pp. 637–659.

[9] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge, 1995.

[10] 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 Dynam. Systems, 33 (2013), pp. 186–220.

[11] F. Martinelli, *Lectures on Glauber dynamics for discrete spin models*, in Lectures on Probability Theory and Statistics, Lecture Notes in Math. 1717, Springer-Verlag, New York, pp. 93–191.

[12] M. Mezard and A. Montanari, *Information, Physics and Computation*, Oxford Graduate Texts, Oxford University Press, Oxford, 2009.

[13] J. Moussouris, *Gibbs and Markov random systems with constraints*, J. Statist. Physics, 10 (1974), pp. 11–33.

[14] R. Pavlov, *Approximating the hard square entropy constant with probabilistic methods*, Ann. Probab., 40 (2012), pp. 2362–2399.

[15] C. Preston, *Gibbs States on Countable Sets*, Cambridge University Press, Cambridge, 1974.

[16] R. Restrepo, J. Shin, P. Tetali, E. Vigoda, and L. Yang, *Improved mixing condition on the grid for counting and sampling independent sets*, arXiv:1105.0914, 2011.

[17] D. Ruelle, *Thermodynamic Formalism*, Cambridge University Press, Cambridge, 1978.

[18] J. van den Berg and C. Maes, *Disagreement percolation in the study of Markov random fields*, Ann. Probab., 22 (1994), pp. 749–763.

[19] P. Walters, *An Introduction to Ergodic Theory*, Grad. Texts in Math. 79, Springer, New York, 1975.

[20] D. Weitz, *Computing independent sets up to the tree threshold*, in Proceedings of the 38th Symposium on the Theory of Computing, 2006, pp. 140–149.