Computation of Entropy in Statistical Mechanics and Information Theory

SHMUEL FRIEDLAND* and URI N. PELED†
Department of Mathematics, Statistics and Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607-7045, USA
(Dated: March 18, 2003)

We outline the most recent theory for the computation of the exponential growth rate of the number of configurations on a multi-dimensional grid. As a demonstration, we compute the monomer-dimer constant for the plane lattice correct to 8 decimal digits.

PACS numbers: 65.40.Gr, 02.70.Hm
Keywords: monomer-dimer, entropy, spectral methods

The exponential growth rate h (with respect to the natural logarithm) of the number of configurations on a multi-dimensional grid arises in the theory of various physical phenomena [1–3]. In physics e^h is viewed as the entropy (per atom) of the corresponding hard model, in mathematics h is called the topological entropy [4], and in information theory h (with respect to $\log_2$) is called the multi-dimensional capacity [5]. The one-dimensional case is easy, namely e^h is the spectral radius $\rho(A)$ of the corresponding transfer matrix. There are very few 2-dimensional models where the value of h is known in closed form [6–9]. In all other cases there are estimates based on (a) asymptotic expansions, e.g., [10, 11]; (b) Monte-Carlo methods, e.g., [12, 13]; (c) bounds, e.g., [14–18]. In what follows we give the complete up-to-date theory for computing h by lower and upper bounds. It refines the techniques described in [19] by using an automorphism subgroup of a certain graph. As a demonstration of these techniques, we compute the topological entropy of the monomer-dimer system in a plane grid with 8-digit precision. We compare our results to the known results.

Consider the integer-valued lattice $\mathbb{Z}^d$ in d-dimensional space $\mathbb{R}^d$. At each point of the lattice we can place one of n kinds of atoms, denoted by $\langle n\rangle := \{1,\ldots,n\}$. The *hard type* model is specified by a d-graph $\Gamma := (\Gamma_1,\ldots,\Gamma_d)$, where $\Gamma_k \subseteq \langle n\rangle \times \langle n\rangle$. It imposes the restrictions that two atoms of the kinds p and q are allowed to occupy respectively the adjacent points $\mathbf{i} = (i_1,\ldots,i_d)$ and $\mathbf{i}+\mathbf{e}_k$ — where $\mathbf{e}_k := (\delta_{1k},\ldots,\delta_{dk})$ — only if $(p,q) \in \Gamma_k$. Configurations obeying these restrictions are called Γ-allowable or Γ-tilings. This general model is unisotropic: Γ_k may depend on k. The graph Γ_k is called *symmetric* when $(p,q) \in \Gamma_k \iff (q,p) \in \Gamma_k$. Γ is called the *nearest neighbor graph* if $\Gamma_1 = \cdots = \Gamma_d = \Delta$, and Δ is a symmetric graph. Let $\mathbf{m} = (m_1,\ldots,m_d) \geq \mathbf{1} = (1,\ldots,1)$ be a point on the lattice with positive coordinates. We denote the box of dimensions $m_1,\ldots,m_d$ by $\langle\mathbf{m}\rangle := \langle m_1\rangle \times\cdots\times \langle m_d\rangle$. Let $W(\mathbf{m})$ be the set of all Γ-allowable configurations of $|\mathbf{m}|_{\mathrm{pr}} := m_1\cdots m_d$ atoms of n kinds in $\langle\mathbf{m}\rangle$. It is easy to show that the function $\log\#W(\mathbf{m})$ is subadditive in each coordinate, i.e., $\log\#W(\mathbf{m}+p\mathbf{e}_k) \leq \log\#W(\mathbf{m}) + \log\#W(\mathbf{m}+(p-m_k)\mathbf{e}_k)$ for all $p \geq 1$ and $k = 1,\ldots,d$. It follows that the limit $h = \lim_{m_1,\ldots,m_d\to\infty} \frac{\log\#W(\mathbf{m})}{|\mathbf{m}|_{\mathrm{pr}}}$ exists; h is the exponential growth rate of $\#W(\mathbf{m})$. Furthermore it follows that $h \leq \frac{\log\#W(\mathbf{m})}{|\mathbf{m}|_{\mathrm{pr}}}$ for all $\mathbf{m} \geq \mathbf{1}$. It follows from König's Lemma that $h = \log 0 = -\infty$ if it is impossible to tile $\mathbb{Z}^d$ with an allowable configuration. A periodic tiling of $\mathbb{Z}^d$ having period m_k in each direction $\mathbf{e}_k$ is equivalent to a Γ-tiling of the torus $T(\mathbf{m}) := (\mathbb{Z}/m_1\mathbb{Z})\times\cdots\times(\mathbb{Z}/m_d\mathbb{Z})$. For $d = 1$, $h = \log\rho(A)$, where A is the incidence matrix of the graph Γ_1 and $\rho(A)$ is the spectral radius of A. In that case, if $h \geq 0$, then h is also the exponential growth rate of the number of allowable periodic configurations in $\mathbb{Z}$. We say that h is *computable* when given $\epsilon > 0$, we can find in a finite number of steps — depending on ϵ — a rational number r satisfying $|h - r| < \epsilon$. For $d \geq 2$, h is not computable in general [20]; this is so because there is a graph Γ such that there are Γ-allowable tilings of $\mathbb{Z}^d$, but none is periodic; hence there are no lower bounds for h for this Γ. However, a fundamental result in [4] asserts that if $d-1$ graphs out of $\Gamma_1,\ldots,\Gamma_d$ are symmetric, then the growth rate of the number of periodic configurations is equal to h, and h is computable, i.e., in addition to the above upper bound for h that converges to h, we have lower bounds for h that converge to h. In particular, these results hold for the nearest neighbor graph.

We now elaborate our results. Let us fix $\mathbf{m}^{\sim d} := (m_1,\ldots,m_{d-1}) \geq \mathbf{1}$. Let $A(\mathbf{m}^{\sim d})$ be the transfer matrix between two layers ϕ and ψ of $(\Gamma_1,\ldots,\Gamma_{d-1})$-allowable configurations in the box $\langle\mathbf{m}^{\sim d}\rangle$ of the $(d-1)$-dimensional lattice. That is, the entry $a_{\phi\psi}$ of $A(\mathbf{m}^{\sim d})$ is equal to 1 if the combined configuration of ψ on top of ϕ is a Γ-allowable configuration of $\mathbf{m} = (\mathbf{m}^{\sim d}, 2)$. Otherwise $a_{\phi\psi} = 0$. Let $\rho(A(\mathbf{m}^{\sim d}))$ be the spectral radius of $A(\mathbf{m}^{\sim d})$. The Perron-Frobenius Theorem yields that $\rho(A(\mathbf{m}^{\sim d}))$ is an eigenvalue of $A(\mathbf{m}^{\sim d})$. From this it is possible to deduce that $\lim_{m_d\to\infty} \frac{\log\#W(\mathbf{m})}{m_d} = \log\rho(A(\mathbf{m}^{\sim d}))$. In particular, we have the upper bound

$$h \leq \frac{\log\rho(A(\mathbf{m}^{\sim d}))}{|\mathbf{m}^{\sim d}|_{\mathrm{pr}}}. \tag{1}$$

Assume now that Γ_1 is symmetric. Let $W_{\mathrm{per},\{1\}}(\mathbf{m}) \subseteq W(\mathbf{m})$ be the set of all Γ-allowable configurations in $\langle \mathbf{m} \rangle$ that can be extended periodically in the direction $\mathbf{e}_1$. Define $\bar{h}(m_1) := \lim_{m_2,\ldots,m_d \to \infty} \frac{\log \# W_{\mathrm{per},\{1\}}(\mathbf{m})}{m_2 \ldots m_d}$ and $\bar{h}(0) := \lim_{m_2,\ldots,m_d \to \infty} \frac{\log \# W((1,m_2,\ldots,m_d))}{m_2 \ldots m_d}$. Then for all positive integers p, r and nonnegative integers q, we have the lower and upper bounds

$$\frac{\bar{h}(p+2q) - \bar{h}(2q)}{p} \le h \le \frac{\bar{h}(2r)}{2r}. \tag{2}$$

See [19] for other lower bounds. Note that $\bar{h}(q)$ is the entropy of the corresponding configurations on a $(d-1)$-dimensional lattice, which are given by corresponding graphs $(\Gamma_{2,q}, \ldots, \Gamma_{d,q})$. If Γ_j is symmetric, then $\Gamma_{j,q}$ is symmetric. It can be shown that the upper bound in (2) is an improvement of (1) for $m_1 = 2r$. If $p = 2r$ and $q = 0$, then the upper and the lower bounds in (2) converge to h when $r \to \infty$. Hence h is computable if $\bar{h}(2r)$ is computable for each r.

In what follows we assume that $\Gamma_1, \ldots, \Gamma_{d-1}$ are symmetric. Then $\mathbb{Z}^d$ can be tiled by a Γ-allowable configuration if and only of $A((2,\ldots,2))$ is not nilpotent, and in what follows we assume that $A((2,\ldots,2))$ is not nilpotent. We apply (2) recursively to obtain computable upper and lower bounds for h that converge to h, thereby showing that h is computable. The upper bound for h obtained from this process has the following form. Let $A_{\mathrm{per}}(\mathbf{m}^{\sim d})$ be the submatrix of $A(\mathbf{m}^{\sim d})$ whose entries are the $a_{\phi\psi}$ such that ϕ and ψ can be extended to periodic $(\Gamma_1, \ldots, \Gamma_{d-1})$-allowable configurations in $\mathbb{Z}^{d-1}$. Then [19]

$$h = \limsup_{m_1,\ldots,m_{d-1} \to \infty} \frac{\log \rho(A_{\mathrm{per}}(\mathbf{m}^{\sim d}))}{|\mathbf{m}^{\sim d}|_{\mathrm{pr}}} \tag{3}$$

$$h \le \frac{\log \rho(A_{\mathrm{per}}(\mathbf{m}^{\sim d}))}{|\mathbf{m}^{\sim d}|_{\mathrm{pr}}} \quad \text{if } m_1, \ldots, m_{d-1} \text{ are even.} \tag{4}$$

The right-hand side of (3) is by definition the growth rate of the number of periodic configurations. Inequality (4) can fail if at least one m_i is odd. Note that (4) is an improvement of (1).

For $d = 2$, (2) yields that for all integers $p, r \ge 1,\ q \ge 0$

$$\frac{\log \rho(A_{\mathrm{per}}(p+2q)) - \log \rho(A_{\mathrm{per}}(2q))}{p} \le h \le \le \frac{\log \rho(A_{\mathrm{per}}(2r))}{2r}. \tag{5}$$

Our numerical experiments show that the best lower bounds in (5) are obtained when $p = 1$ or $p = 2$. It follows from (3) and the upper bound in (5) that $h = \lim_{r \to \infty} \frac{\log \rho(A_{\mathrm{per}}(2r))}{2r}$.

To summarize, our lower and upper bounds involve computing spectral radii of some nonnegative $N \times N$ matrices $A = (a_{ij})_1^N$. The problem is that for $d = 2, 3$, in all physical models that we know of where an analytical expression for h is not available, the dimension N of the matrix $A = A_{\mathrm{per}}(\mathbf{m}^{\sim d})$ grows exponentially with $|\mathbf{m}^{\sim d}|_{\mathrm{pr}}$. For example, for the monomer-dimer problem discussed later, $N = 2^{|\mathbf{m}^{\sim d}|_{\mathrm{pr}}}$. Thus, if $d = 2$ and $\mathbf{m}^{\sim 2} = 16$, then $N = 2^{16}$, and finding the spectral radius of an arbitrary $2^{16} \times 2^{16}$ nonnegative matrix is beyond the power of modern computers. However, there is a way to reduce the computations for $\log \rho(A)$ if A has a nontrivial automorphism group.

Recall that the automorphism group $\mathrm{Aut}(A)$ is a subgroup of the symmetric group S_N on $\langle N \rangle$, defined by $\mathrm{Aut}(A) := \{\pi \in S_N \mid a_{\pi(i)\pi(j)} = a_{ij} \text{ for all } i, j \in \langle N \rangle\}$. Let G be a subgroup of $\mathrm{Aut}(A)$. Let $\mathcal{O}(G) := \langle N \rangle / G$ be the orbit space, and let $M = \#\mathcal{O}(G)$. Define $\hat{A} = (\hat{a}_{\alpha\beta})_{\alpha,\beta \in \mathcal{O}(G)}$ to be the $M \times M$ matrix with $\hat{a}_{\alpha\beta} = \sum_{j \in \beta} a_{ij}$ for any $i \in \alpha$. Then $\rho(A) = \rho(\hat{A})$. Furthermore, if A is symmetric, then $\hat{A}$ is symmetric with respect to the inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{\alpha \in \mathcal{O}(G)} w_\alpha x_\alpha y_\alpha$, where w_α is the cardinality of the orbit α; in particular, $\hat{A}$ has real eigenvalues. Clearly, $M \ge N/\#G$, but in our computations M is not much larger than $N/\#G$. Assume that $\hat{A}$ is a nonnegative matrix, symmetric with respect to the inner product $\langle \cdot, \cdot \rangle$. Then the computation of $\rho(\hat{A})$ can be done by the power method as follows. Let $\mathbf{x}_0 = \mathbf{1}$ and $\mathbf{x}_\ell = (x_{\alpha,\ell})_{\alpha \in \mathcal{O}(G)} = \mathbf{x}_{\ell-1}\hat{A}$ for all positive integers ℓ. Set $\underline{\rho}_\ell := \frac{\langle \mathbf{x}_\ell \hat{A}, \mathbf{x}_\ell \rangle}{\langle \mathbf{x}_\ell, \mathbf{x}_\ell \rangle}$ and $\bar{\rho}_\ell := \max_\alpha \frac{(\mathbf{x}_\ell \hat{A})_\alpha}{\mathbf{x}_{\alpha,\ell}}$. Then $\underline{\rho}_\ell \le \rho(\hat{A}) \le \bar{\rho}_\ell$, the sequence $\{\bar{\rho}_\ell\}$ is nonincreasing, and $\lim_{\ell \to \infty} \bar{\rho}_\ell = \lim_{\ell \to \infty} \underline{\rho}_\ell = \rho(\hat{A})$.

Assume now that Γ is the nearest neighbor graph. Let $G(\mathbf{m}^{\sim d})$ the automorphism group of the adjacency graph of the torus $T(\mathbf{m}^{\sim d})$. It is not difficult to see that $G(\mathbf{m}^{\sim d})$ is isomorphic to a subgroup of $\mathrm{Aut}(A_{\mathrm{per}}(\mathbf{m}^{\sim d}))$. Clearly for $d = 2$, the automorphism group $\mathrm{Aut}(G(m))$ is of order $2m$ for $m \ge 3$, and of order 2 for $m = 2$. The product $\mathrm{Aut}(G(m_1)) \times \cdots \times \mathrm{Aut}(G(m_{d-1}))$ is a subgroup of $G(\mathbf{m}^{\sim d})$. If $m_i = m_j$ for some $1 \le i < j < d$, we can also permute the coordinates i and j in $T(\mathbf{m}^{\sim d})$. Thus if $m_1 = \cdots = m_{d-1} = m \ge 3$, then $\#G(\mathbf{m}^{\sim d}) \ge (2m)^{d-1}(d-1)!$ [in the special case of $m = 4$, even more is true, namely $\#G(\mathbf{m}^{\sim d}) \ge 2^{2(d-1)}(2(d-1))! > 8^{d-1}(d-1)!$; this follows from the observation that $T(4) \simeq T((2,2))$, and so $T(\mathbf{m}^{\sim d}) \simeq T(\mathbf{n})$, where $\mathbf{n} \in \mathbb{Z}^{2(d-1)}$ with $n_i = 2$; hence $\#G((4,4)) \ge 2^4 4! > 8^2 2!$]. Thus for $d = 2$ and $\mathbf{m}^{\sim 2} = 16$, we reduce the dimension N of $A_{\mathrm{per}}(16)$ roughly by the factor 32 by considering the matrix $\hat{A}_{\mathrm{per}}(16)$ induced by the action of $G(16)$.

We now discuss the monomer-dimer tiling of $\mathbb{Z}^d$; see [2]. A *monomer-dimer tiling* of $\langle \mathbf{m} \rangle$ is a partitioning of $\langle \mathbf{m} \rangle$ into s pairs of neighboring vertices called dimers, and $|\mathbf{m}|_{\mathrm{pr}} - 2s$ single vertices called monomers. Let $W_0(\mathbf{m}, s)$ be the set of all such monomer-dimer tiling for a fixed s, $0 \le s \le |\mathbf{m}|_{\mathrm{pr}}/2$, and $W_0(\mathbf{m}) := \cup_{0 \le s \le |\mathbf{m}|_{\mathrm{pr}}/2} W_0(\mathbf{m}, s)$. The pure dimer case corresponds to $W_0(\mathbf{m}, |\mathbf{m}|_{\mathrm{pr}}/2)$

if $|\mathbf{m}|_{\rm pr}$ is even. Let h_d and $\tilde{h}_d$ be the exponential growth rate of the number of monomer-dimer and dimer tiling, respectively. It is shown in [21] that if $\langle \mathbf{m}_q \rangle$ is an infinite sequence of boxes satisfying $m_{q,i} \to \infty$ for each $i \in \langle d \rangle$, and $s_q \in [0, |\mathbf{m}_q|_{\rm pr}/2]$ is an infinite integer sequence satisfying $\lim_{q\to\infty} 2s_q/|\mathbf{m}_q|_{\rm pr} = t$, then $\lim_{q\to\infty} \frac{\log \# W_0(\mathbf{m}_q, s_q)}{|\mathbf{m}_q|_{\rm pr}} = \lambda_d(t)$, where $\lambda_d(t)$ is a continuous and concave function of t. Therefore $\lambda_d(1) = \tilde{h}_d$, and it can be shown that $h_d = \max_{t\in[0,1]} \lambda_d(t)$. Moreover, h_d, $\tilde{h}_d$ and $\lambda_d(t)$ can be computed by using periodic monomer-dimer tiling [21]. It is known [3] that $\lambda_1(t) = (1-\frac{t}{2})\log(1-\frac{t}{2}) - \frac{t}{2}\log\frac{t}{2} - (1-t)\log(1-t)$. Hence $\tilde{h}_1 = 0$ and $h_1 = \lambda_1(1-\frac{1}{\sqrt{5}}) = \log\frac{1+\sqrt{5}}{2}$. From [6] and [7] we know that $\tilde{h}_2 = \frac{1}{\pi}\sum_{q=0}^{\infty} \frac{(-1)^q}{(2q+1)^2} = 0.29156090\ldots$ Other than that, only estimates for $\lambda_2(t)$ and $\lambda_3(t)$ are known [22]. It is shown in [15] that $\tilde{h}_3 \le 0.463107$, and the permanent inequality in [23] yields $0.440075 \le \tilde{h}_3$. The Monte-Carlo methods in [13] yield $\tilde{h}_3 = 0.4466 \pm 0.0006$. We recall the estimate $\tilde{h}_3 \sim 0.446$ from asymptotic expansions [10].

We apply the inequalities (5) to estimate h_2. We were able to compute $\rho(A_{\rm per}(n))$ up to $n = 15$, and it turns out that the sequence $\frac{\log \rho(A_{\rm per}(2r))}{2r}$ is decreasing for $r = 1, \ldots, 7$ whereas the sequence $\frac{\log \rho(A_{\rm per}(2r-1))}{2r-1}$ is increasing for $r = 1, \ldots, 8$. So the best upper bound we could obtain from (5) is achieved for $r = 7$. Our best lower bound was obtained when $p = 1$ and $q = 7$. These bounds give $h_2 = 0.66279897$ correct to 8 digits. [It follows from (3) that $h_2 \ge \limsup_{r\to\infty} \frac{\log \rho(A_{\rm per}(2r-1))}{2r-1}$. If we assume that $\frac{\log \rho(A_{\rm per}(2r-1))}{2r-1}$ is nondecreasing for *all* r, then the above lim sup is a lim and we could improve the lower bound in (5) to $h_2 \ge \frac{\log \rho(A_{\rm per}(2r-1))}{2r-1}$, which would yield $h_2 = 0.662798973$ correct to 9 digits.] The maximal value of the Monte-Carlo estimates of $\lambda_2(t)$ in [22], given in t increments of 0.05, is 0.6676 (for $t = 0.65$). This value is about 1% higher than the true maximal value of $\lambda_2(t)$, namely h_2, according to our computations above. This contradicts the statement in [22] that the Monte-Carlo estimates of $\lambda_2(t)$ are "probably 1%–2% lower" than the actual values. We also used (5) to see how well it estimates the known dimer entropy $\tilde{h}_2$. We computed $\rho(A_{\rm per}(n))$ up to $n = 15$. Again we found that the sequence $\frac{\log \rho(A_{\rm per}(2r))}{2r}$ is decreasing for $r = 1, \ldots, 7$ whereas the sequence $\frac{\log \rho(A_{\rm per}(2r-1))}{2r-1}$ is increasing for $r = 1, \ldots, 8$. The values of the increasing sequence we computed were always lower bounds for $\tilde{h}_2$. However, the estimates for $\tilde{h}_2$ we could obtain gave only 2 correct decimal digits of $\tilde{h}_2$.

Finally it is worth mentioning the theoretical work of [24], which shows that the general monomer-dimer problem in arbitrary planar graphs (as opposed to planar grids) is computationally intractable.

---

* `friedlan@uic.edu`

† `uripeled@uic.edu`

[1] L. Pauling, J. Amer. Chem. Soc. **57**, 2680 (1935).

[2] R. Fowler and G. Rushbrooke, Trans. Faraday Soc. **33**, 1272 (1937).

[3] J. D. B. E.G.D. Cohen and Z. Salsburg, Physica **21**, 137 (1955).

[4] S. Friedland, Linear Algebra Appl. **252**, 199 (1997).

[5] W. Weeks and R. Blahut, IEEE Trans. Info. Theory **44**, 1193 (1998).

[6] M. Fisher, Phys. Rev. **124**, 1664 (1961).

[7] P. Kasteleyn, Physica **27**, 1209 (1961).

[8] E. Lieb, Phys. Rev. **162**, 162 (1967).

[9] R. Baxter, J. Phys. A: Math. Gen. **13**, L61 (1980).

[10] J. Nagle, Phys. Rev. **152**, 190 (1966).

[11] D. Gaunt, Phys. Rev. **179**, 174 (1969).

[12] J. Hammersley, in *Proc. Comput. Physics Con.* (Inst. of Phys. & Phys. Soc., London, 1970).

[13] I. Beichl and F. Sullivan, J. Comp. Phys. **149**, 128 (1999).

[14] J. Hammersley, Proc. Camb. Phil. Soc. **64**, 455 (1966).

[15] M. Ciucu, Duke Math. J. **94**, 1 (1998).

[16] N. Calkin and H. Wilf, SIAM J. Discr. Math. **11**, 54 (1998).

[17] S. Forschhammer and J. Justesen, IEEE Trans. Info. Theory **45**, 118 (1999).

[18] Z. Nagy and K. Zeger, IEEE Trans. Info. Theory **46**, 1030 (2000).

[19] S. Friedland, in *Mathematical System Theory in Biology, Communication, Computation and Finance*, edited by D. Gilliam and J. Rosenthal (Springer, 2002), to appear.

[20] R. Berger, Mem. Amer. Math. Soc. **66** (1966).

[21] J. Hammersley, in *Reseach Papers in Statistics: Festschrift for J. Neyman*, edited by F. David (Wiley, London, 1966), pp. 125–146.

[22] J. Hammersley and V. Menon, J. Inst. Maths. Applics **6**, 341 (1970).

[23] A. Schrijver, J. Comb. Theory **B 72**, 122 (1998).

[24] M. Jerrum, J. Stat. Phys. **48**, 121 (1987).