Polynomial-time approximation algorithms for the Ising model*

Mark Jerrum *and* **Alistair Sinclair**

Department of Computer Science, University of Edinburgh

The King's Buildings, Edinburgh EH9 3JZ, Scotland

November 1989

(revised December 1991)

Abbreviated title: Approximation algorithms for the Ising model

*An Extended Abstract of this paper appeared in the "Proceedings of the International Colloquium on Automata, Languages and Programming," Warwick, UK, July 1990; published by Springer-Verlag as Lecture Notes in Computer Science Vol. 443.

Abstract

The paper presents a randomised algorithm which evaluates the partition function of an arbitrary ferromagnetic Ising system to any specified degree of accuracy. The running time of the algorithm increases only polynomially with the size of the system (i.e., the number of sites) and a parameter which controls the accuracy of the result. Further approximation algorithms are presented for the mean energy and the mean magnetic moment of ferromagnetic Ising systems.

The algorithms are based on Monte Carlo simulation of a suitably defined ergodic Markov chain. The states of the chain are not, as is customary, Ising spin configurations, but spanning subgraphs of the interaction graph of the system. It is shown that the expectations of simple operators on these configurations give numerical information about the partition function and related quantities.

The performance guarantees for the algorithms are rigorously derived, and rest on the fact that the Markov chain in question is rapidly mixing, i.e., converges to its equilibrium distribution in a polynomial number of steps. This is apparently the first time that rapid mixing has been demonstrated at all temperatures for a Markov chain related to the Ising model.

Keywords: The Ising model, statistical physics, ferromagnetism, spin-glasses, partition function, #P-completeness, approximation algorithms, Markov chains, rapid mixing, Monte Carlo simulation.

AMS subject classifications: 05C85, 60J10, 60J20, 60K35, 68Q20, 68Q25, 82B20, 82B31, 82B80.

1 Summary

This paper is concerned with computational solutions to a classical combinatorial problem of statistical physics. Generally known as the *Ising model*, the problem has been the focus of much attention in the physics and mathematics communities since it was first introduced by Lenz [24] and Ising [14] in the early 1920s. We will not present a detailed historical account here: a very readable survey is given by Cipra [6], while Welsh [30] sets the Ising model in the context of other combinatorial problems in statistical physics.

The problem is easily stated. Consider a collection of sites $[n] = \{0, 1, \ldots, n-1\}$, each pair i, j of which has an associated *interaction energy* V_{ij}. In most cases of physical interest, the set E of pairs with non-zero interaction energies forms a regular lattice graph $([n], E)$. A *configuration* is an assignment of positive ($\sigma_i = +1$) and negative ($\sigma_i = -1$) spins to each site $i \in [n]$. The *energy* of a configuration $\boldsymbol{\sigma} = (\sigma_i)$ is given by the Hamiltonian

$$H(\boldsymbol{\sigma}) \;=\; -\sum_{\{i,j\}\in E} V_{ij}\sigma_i\sigma_j - B\sum_{k\in[n]}\sigma_k,$$

where B is an external field.

In the case where all interaction energies are non-negative, such a system models the behaviour of a *ferromagnet*; in fact, it was towards an understanding of spontaneous magnetization that the model was first conceived. However, the Ising model has since become a powerful paradigm for the investigation of more general cooperative systems in which short-range interactions between elements can give rise to long-range order.

The central problem is to compute the *partition function*

$$Z \;=\; Z(V_{ij}, B, \beta) \;=\; \sum_{\boldsymbol{\sigma}} \exp(-\beta H(\boldsymbol{\sigma})),$$

where $\beta > 0$ is related to the temperature and the sum is over all possible configurations $\boldsymbol{\sigma}$. Almost all the physical properties of the system can be computed from knowledge of Z. Essentially, Z is the normalising factor in the calculation of probabilities: according to the fundamental theory of statistical mechanics, the probability that the system in equilibrium is found in configuration $\boldsymbol{\sigma}$ is $\exp(-\beta H(\boldsymbol{\sigma}))/Z$. Moreover, certain logarithmic derivatives of Z correspond to quantities such as the mean energy and the mean magnetic moment. Singularities in these derivatives generally correspond to *phase transitions*, when a small change in a parameter has an observable effect on the macroscopic properties of the system.

The search for efficient computational solutions to these problems has proved extremely hard and has generated a vast body of literature. A major breakthrough was achieved in the early 1960s by Kasteleyn [19] and Fisher [11], who reduced the problem of computing Z for any *planar* Ising system (i.e., one whose graph $([n], E)$ of non-zero interactions is planar) to the evaluation of a certain determinant. This must rank as one of the highlights in the field of combinatorial algorithms. It remains the state of the art as far as exact solutions are concerned; in particular, it does not appear to generalise to non-planar systems. On the other hand, a huge amount of computational effort is poured into *numerical* solutions of the Ising model for three-dimensional regular lattices and other non-planar systems. The problem is that the methods used here, while ingenious, generally lack a rigorous theoretical base, and rely for their validity largely on physical intuition.

In this paper, we exhibit what we believe to be the first provably efficient approximation algorithm for the partition function of an arbitrary ferromagnetic Ising system. By "efficient" here we mean that the algorithm is guaranteed to run in time polynomial in the number of sites n. The algorithm is a *fully polynomial randomised approximation scheme* (fpras), i.e., it will produce solutions which, with very high probability, fall within arbitrarily small error bounds specified by the user, the price of greater accuracy being a modest increase in runtime. We also show that such an algorithm is essentially the best one can hope for, in the sense that the existence of an efficient *exact* algorithm for the problem, or even of an efficient approximation algorithm for the non-ferromagnetic case, would have devastating and far-reaching consequences in the theory of computation.

From the point of view of theoretical computer science, our result provides a new example of a significant combinatorial enumeration problem which is #P-complete, and hence apparently intractable in exact form, but for which an efficient approximation algorithm exists. This is an intriguing class of problems, and includes the problems of computing the volume of a convex body [9], the partition function of a monomer-dimer system [16] and the permanent of a large class of 0–1 matrices [16]. Our algorithm is also of interest in its own right as a further application of the general technique of simulating an ergodic stochastic process whose rate of convergence can be analysed. This approach has recently attracted much attention, and its full algorithmic potential is only now becoming apparent.

The idea is the following. In order to compute weighted combinatorial sums, such as the Ising partition function, it is often enough to be able to sample configurations $\boldsymbol{\sigma}$ at random with probabilities proportional to their weights, in this case $\exp(-\beta H(\boldsymbol{\sigma}))$. This can be achieved by setting up an ergodic Markov chain whose states are configurations and whose transitions correspond to small local perturbations. If the chain is designed so that the equilibrium distribution to which it converges is the desired weighted distribution over configurations, then we get a random sampling procedure by simulating the chain for some number of steps and outputting the final state. For such a procedure to be efficient, the chain must be *rapidly mixing* in the sense that it gets very close to equilibrium after a small (i.e., polynomial) number of steps. This is a highly non-trivial requirement, since the number of states is exponentially large. Recent developments have provided appropriate analytical tools for establishing the rapid mixing property for chains of this kind [27, 29, 7, 28].

The Markov chain simulation approach to the Ising model is far from new: under the name of the Monte Carlo method, this technique has been applied extensively to a whole range of problems in statistical physics (see, e.g., [4]). The problem with the approach, however, is that it appears very difficult to define a Markov chain on Ising spin configurations $\boldsymbol{\sigma}$ which is rapidly mixing; indeed, the chains which are frequently used in practical simulation studies clearly do not have this property.

We overcome this obstacle by transforming the problem to an entirely new domain, where the configurations are spanning subgraphs of the interaction graph $([n], E)$. Each subgraph has an "energy" which is determined by weights attached to its edges and vertices. Although there is no direct correspondence between configurations in the two domains, and the subgraph configurations have no obvious physical significance, the two partition functions are, remarkably, very closely related. Moreover, and crucially, there is a natural Markov chain on the subgraphs with the appropriate equilibrium distribution which *is* rapidly mixing. Thus the Markov chain approach can be made to work efficiently in the new domain.

The above transformation is a classical result [26], often known as the "high-temperature expansion" of the Ising model partition function. However, the idea of viewing the graphs in this expansion as a statistical mechanical system which forms the basis of a Monte Carlo simulation appears to be new. To the best of our knowledge, our results represent the first rigorous proof of rapid mixing at all temperatures for a Markov chain related to the Ising model. Moreover, this property is entirely independent of the interaction topology and relies on no assumptions of any kind. We therefore believe that the chain deserves further investigation as a potentially powerful experimental tool.

The mechanism by which we use sampling of subgraph configurations to compute the partition function is perhaps of independent interest. This is achieved by subjecting an Ising system with fixed interactions and at a fixed temperature to varying external fields. By observing a small number of configurations, randomly selected at appropriately chosen values of the field, we are able to get an accurate estimate of Z. It is significant that this idea is motivated by combinatorial considerations and does not correspond to any obvious physical intuition.

As mentioned earlier, it is often derivatives of the partition function, rather than the function itself, which are of primary interest. For example, two important quantities are the *mean energy* $\mathcal{E} = -\partial(\ln Z)/\partial\beta$, and the *mean magnetic moment* $\mathcal{M} = \beta^{-1}\,\partial(\ln Z)/\partial B$. Our approximation algorithm for Z says nothing about our ability to compute these quantities accurately. However, it turns out that both $\mathcal{E}$ and $\mathcal{M}$ can be expressed in terms of expectations of certain simple operators on configurations in the subgraphs domain. Thus estimates of $\mathcal{E}$ and $\mathcal{M}$ can be read off from our configuration sampling algorithm, though again we may have to vary the external field in order to maximise the accuracy of the statistical experiment. As a result, we get a fpras for both $\mathcal{E}$ and $\mathcal{M}$ as well. We regard this as confirmation that our approach to the Ising model is robust and computationally effective.

The remainder of the paper is organised as follows. In Section 2 we describe the transformation of the Ising model to the new domain in which configurations are spanning subgraphs of the interaction graph. Section 4 is devoted to a discussion of the Markov chain on these configurations, and in particular to a proof that it is rapidly mixing. This fact is used in Section 3 to construct a fpras for the partition function of an arbitrary ferromagnetic Ising system, and in Section 5 to construct efficient approximation algorithms for the mean energy and the mean magnetic moment. Finally, in Section 6 we present strong evidence that our results are, in a precise theoretical sense, best possible.

2 The spins-world and subgraphs-world

Recall that our primary aim is to construct an algorithm for the following problem:

> INSTANCE: A real symmetric matrix $(V_{ij} : i, j \in [n])$ of interaction energies, a real number B (the external field), and a positive real number β.
>
> OUTPUT: The Ising partition function
>
> $$Z = Z(V_{ij}, B, \beta) = \sum_{\boldsymbol{\sigma} \in \{-1,+1\}^n} \exp(-\beta H(\boldsymbol{\sigma})), \tag{1}$$
>
> where the Hamiltonian $H(\boldsymbol{\sigma})$ is given by $H(\boldsymbol{\sigma}) = -\sum_{\{i,j\}\in E} V_{ij}\sigma_i\sigma_j - B\sum_{k\in[n]}\sigma_k$, and E is the set of unordered pairs $\{i,j\}$ with $V_{ij} \neq 0$.

Our algorithm will address the *ferromagnetic* case of the Ising model, which is characterised by the interaction energies V_{ij} being non-negative. Furthermore, rather than attempting to evaluate the partition function *exactly*, we shall content ourselves with a close approximation. The phrase "close approximation" will be given a precise meaning in the next section.

One strategy which has been applied successfully to problems of this type, and has, for example, been used to estimate the partition function of a general monomer-dimer system [27, 16], involves the simulation of an appropriately defined Markov chain. A direct application of this strategy to the Ising partition function would proceed as follows. View the configurations of the Ising system, namely the 2^n possible spin vectors $\boldsymbol{\sigma} \in \{-1,+1\}^n$, as the states of a Markov chain. Choose transition probabilities between states so that the Markov chain is ergodic and so that, in the stationary distribution, the probability of being in state $\boldsymbol{\sigma}$ is $Z^{-1}\exp(-\beta H(\boldsymbol{\sigma}))$. A reasonable way to achieve this, and one which is often used in practice, is to allow transitions to occur between spin configurations which differ in just one component, and choose transition probabilities according to the Metropolis rule [20]. If the resulting Markov chain is *rapidly mixing*, that is, if it converges rapidly to the stationary distribution regardless of the choice of initial state, then it can be used effectively to sample configurations $\boldsymbol{\sigma}$ from a distribution which is close to the stationary distribution. By collecting enough sample configurations, using different values of B and β, it should then be possible to estimate the partition function Z with good accuracy.

Unfortunately, it transpires that the Markov chain described above, and which we refer to as the *spins-world* process, is *not* rapidly mixing. It is well known that ferromagnetic Ising systems typically exhibit a phase transition at a certain value of the parameter β; for values of β above this critical value, the system settles into a state in which there is a preponderance of spins of one or other sign. Transitions between the majority $+1$ states and majority -1 states occur very infrequently, simply because the stationary distribution assigns small total weight to the configurations with balanced spins. (Informally, the state space has a constriction separating the majority $+1$ states and the majority -1 states.) Although it could be argued that the barrier to rapid mixing just described is somewhat trivial, there exist other more subtle barriers that apparently cannot be surmounted.

The problem caused by the absence of rapid mixing in the spins-world process can be circumvented by simulating a different Markov chain, which we refer to as the *subgraphs-world* process. The two Markov chains are structurally very different; furthermore, the subgraphs-world process has, as far as we are aware, no direct physical significance. However, the subgraphs-world process has a close connection with the Ising partition function and, crucially in the current application, is rapidly mixing. For the time being, we content ourselves with describing the subgraphs-world configurations and associated partition function. The description of the subgraphs-world process itself is deferred to Section 4.

We say that a subgraph is *spanning* if it includes all the vertices of the parent graph. (Note that spanning subgraphs are not in general connected.) The subgraphs-world configurations are spanning subgraphs of the interaction graph $([n], E)$. In the sequel we shall drop the adjective "spanning" where it seems safe to do so, and frequently identify a spanning subgraph $([n], X)$ with the set X of edges which define it. To simplify notation, let

$$\lambda_{ij} = \tanh \beta V_{ij} \quad \text{and} \quad \mu = \tanh \beta B. \tag{2}$$

Each configuration $X \subseteq E$ is assigned a *weight* according to the formula

$$w(X) = \mu^{|\mathrm{odd}(X)|} \prod_{\{i,j\} \in X} \lambda_{ij}, \tag{3}$$

where the notation $\mathrm{odd}(X)$ stands for the set of all odd-degree vertices in the graph X. The subgraphs-world partition function is simply

$$Z' = \sum_{X \subseteq E} w(X). \tag{4}$$

The above sum is generally known as the "high-temperature expansion."

It is a surprising fact that the spins- and subgraphs-world partition functions Z and Z' are related in a simple way. Define

$$A = (2 \cosh \beta B)^n \prod_{\{i,j\} \in E} \cosh \beta V_{ij}, \tag{5}$$

and note that A is an easily computed function of the parameters that specify the Ising system. The following classical result [26] relates the two partition functions.

Theorem 1 $Z = AZ'$.

In recognition of the central role it plays in our algorithm, we present a full proof of this result below.

Theorem 1 prompts us to consider a statistical mechanical system whose configurations are spanning subgraphs of $([n], E)$. We shall define a Markov chain whose states are these configurations, and whose stationary distribution assigns probability $\pi(X) = w(X)/Z'$ to configuration X. This subgraphs-world process will be analysed in detail in Section 4, and shown to be rapidly mixing. Hence it will provide us with an efficient means of sampling subgraphs-world configurations with probabilities roughly proportional to their weights. Since Z' is a weighted sum of the configurations, we might expect such a procedure to give us useful information about Z' itself, and hence about the original spins-world partition function Z. The next section confirms that this is indeed the case.

Proof of Theorem 1 Taking equation (1) as a starting point, apply the identity $\mathrm{e}^x = \cosh x\,(1 + \tanh x)$ to recast the partition function in the form

$$Z = 2^{-n} A \sum_{\boldsymbol{\sigma} \in \{-1,+1\}^n} \prod_{\{i,j\} \in E} \{1 + \tanh(\beta V_{ij} \sigma_i \sigma_j)\} \prod_{k \in [n]} \{1 + \tanh(\beta B \sigma_k)\},$$

where A is defined in (5). Note that the spin variables σ_k disappear from the expression for A because $\sigma_k = \pm 1$ for all k, and $\cosh x$ is an even function. Similarly, since $\tanh x$

is an odd function, the spin variables may be brought outside allowing Z to be rewritten as

$$Z = 2^{-n} A \sum_{\boldsymbol{\sigma}} \prod_{\{i,j\}\in E} \{1 + \sigma_i \sigma_j \tanh \beta V_{ij}\} \prod_{k\in[n]} \{1 + \sigma_k \tanh \beta B\}.$$

Expanding the two products, and changing variables according to (2), we obtain

$$Z = 2^{-n} A \sum_{\boldsymbol{\sigma}} \sum_{X\subseteq E} \Bigg(\prod_{\{i,j\}\in X} \lambda_{ij} \sigma_i \sigma_j \Bigg) \sum_{U\subseteq[n]} \Bigg(\prod_{k\in U} \mu \sigma_k \Bigg),$$

which on interchanging the order of summation yields

$$Z = 2^{-n} A \sum_{X\subseteq E} \sum_{U\subseteq[n]} \sum_{\boldsymbol{\sigma}} W(U, X, \boldsymbol{\sigma}), \tag{6}$$

where

$$W(U, X, \boldsymbol{\sigma}) = \prod_{k\in U} \mu \sigma_k \prod_{\{i,j\}\in X} \lambda_{ij} \sigma_i \sigma_j.$$

Now we claim that $\sum_{\boldsymbol{\sigma}} W(U, X, \boldsymbol{\sigma}) = 0$ unless X is a graph in which all vertices in U have odd degree, and all vertices in $[n] - U$ have even degree. To see this, fix U and X, and let $k \in [n]$ be such that *either* $k \in U$ and has even degree in X *or* $k \in [n] - U$ and has odd degree in X. For any vector $\boldsymbol{\sigma} \in \{-1, +1\}^n$, let $\boldsymbol{\sigma}^{(k)}$ denote the vector derived from $\boldsymbol{\sigma}$ by inverting the sign of the kth component. Then the terms $W(U, X, \boldsymbol{\sigma})$ and $W(U, X, \boldsymbol{\sigma}^{(k)})$ are equal in size but opposite in sign. Hence the terms of the sum $\sum_{\boldsymbol{\sigma}} W(U, X, \boldsymbol{\sigma})$ cancel out in pairs.

Conversely, suppose that X *is* a graph in which all vertices in U have odd degree, and all vertices in $[n] - U$ have even degree. Then, for all $\boldsymbol{\sigma} \in \{-1, +1\}^n$ and $k \in [n]$, the terms $W(U, X, \boldsymbol{\sigma})$ and $W(U, X, \boldsymbol{\sigma}^{(k)})$ are equal. Thus the value of $W(U, X, \boldsymbol{\sigma})$ is independent of $\boldsymbol{\sigma}$ and

$$\sum_{\boldsymbol{\sigma}} W(U, X, \boldsymbol{\sigma}) = 2^n \mu^{|U|} \prod_{\{i,j\}\in X} \lambda_{ij} = 2^n w(X).$$

Finally, substituting for $\sum_{\boldsymbol{\sigma}} W(U, X, \boldsymbol{\sigma})$ in equation (6) we obtain the identity $Z = A \sum_{X\subseteq E} w(X)$, as required. □

3 Estimating the partition function

The aim of this section is to present an efficient approximation algorithm for computing the partition function Z of a ferromagnetic Ising system. The section is structured as follows. First, we define precisely what we mean by an efficient approximation algorithm.

Then we state, without proof, the properties of the sampling procedure for subgraphs-world configurations which plays a key role in our algorithm: the construction and analysis of this procedure, based on a suitably defined Markov chain, is left to the next section. Finally, we explain how to use samples produced by this procedure to obtain a reliable approximation of Z.

Our definition of efficient approximation algorithm is a very demanding one, following Karp and Luby [18] and others. For non-negative real numbers a, $\tilde{a}$, ϵ, we say that $\tilde{a}$ *approximates* a *within ratio* $1+\epsilon$ if $a(1+\epsilon)^{-1} \le \tilde{a} \le a(1+\epsilon)$. Let f be any function from problem instances to real numbers. (The Ising partition function is an example of such a function.) A *randomised approximation scheme* for f is a probabilistic algorithm which, when presented with an instance x and a real number $\epsilon \in (0,1]$, outputs a number which, with high probability, approximates $f(x)$ within ratio $(1+\epsilon)$. We shall take the phrase "with high probability" to mean with probability at least $3/4$. This is because a failure probability of $1/4$ can be reduced to any desired value $\delta > 0$ by performing only $\mathrm{O}(\log \delta^{-1})$ trials and taking the median of the results [17]. (This claim is also justified in the proof of Lemma 3 below.) Of course, it is not enough just to obtain an accurate result with high reliability; the result must also be obtained efficiently. Accordingly, we call an approximation scheme *fully polynomial* if it runs in time polynomial in ϵ^{-1} and the size of the problem instance x. The reader will appreciate that a fully polynomial randomised approximation scheme, or *fpras*, embodies a strong notion of efficient approximation.

With an eye to simplicity of presentation, we shall not concern ourselves with the errors which arise through the inexact nature of computer arithmetic. Instead, we shall assume a computational model in which real arithmetic is performed with perfect accuracy, and in which arithmetic operations and standard functions, such as exp, are charged at unit cost. After all, we are aiming only at an *approximate* evaluation of the partition function, and it will become apparent that our technique does not rely on intermediate computations being carried out to untoward accuracy. Again with simplicity in mind, we will take n, the number of sites, as the size of the problem instance, even though the number of parameters to the model would be a more reasonable measure from an information-theoretic point of view.

As we have already mentioned, our approximation algorithm for Z is based on a sampling procedure for subgraphs-world configurations. We must now be more precise about the properties of the sampling procedure. For a ferromagnetic Ising system $\langle \lambda_{ij}, \mu \rangle$, with λ_{ij} and μ as defined in equation (2) of the previous section, let Ω denote the set of subgraphs-world configurations, i.e., the set of spanning subgraphs of the interaction graph $([n], E)$, and define the probability distribution π over Ω by $\pi(X) = w(X)/\sum_{X'} w(X') = w(X)/Z'$, where w is the weight function defined in equation (3). (Note that, since the system is ferromagnetic, $w(X) \ge 0$ for all $X \in \Omega$, so π *is* a probability distribution.) We wish to formalise the notion of an algorithm which, given a ferromagnetic system, selects a configuration from a distribution which is "close to" π. A *generator* for subgraphs-world configurations is a probabilistic algorithm which takes as input a ferromagnetic Ising system in the form $\langle\lambda_{ij},\mu\rangle$, plus a positive real tolerance δ, and outputs an element of Ω drawn from a distribution p satisfying

$$\|p-\pi\| \le \delta.$$

Here $\|\cdot\|$ denotes variation distance, i.e.,

$$\|p-\pi\| = \frac{1}{2}\sum_{X\in\Omega}|p(X)-\pi(X)| = \max_{A\subseteq\Omega}|p(A)-\pi(A)|.$$

It turns out to be possible to construct an efficient generator for subgraphs-world configurations, as the following theorem states.

Theorem 2 *There exists a generator for subgraphs-world configurations which, on inputs $\langle\lambda_{ij},\mu\rangle$ and δ, runs in time bounded by a polynomial in n, μ^{-1} and $\log\delta^{-1}$. Specifically, the runtime of the generator is $\mathrm{O}(m^2\mu^{-8}(\log\delta^{-1}+m))$, where $m=|E|$ is the number of non-zero interactions.*

Remarks (i) The presence of μ^{-1} in the time bound implies that the generator is inefficient for systems with a very small external field. This dependence on the field is inessential and can be removed with a little extra work (see Theorem 10 of Section 5).

(ii) Efficient generators for combinatorial structures, of which the above is a particular example, are discussed in a general framework in [27, 29].

The construction of a generator with the above properties, based on simulation of a suitably defined Markov chain, is described and justified in detail in the next section. For the moment we will simply assume Theorem 2 and concentrate on showing how samples produced by the generator can be used to obtain an efficient approximation algorithm for the partition function $Z(V_{ij},B,\beta)$. Our approach, which we now describe, is an instance of a computational technique which will be employed repeatedly in this paper.

Suppose we want to estimate the value of some physical quantity associated with a ferromagnetic Ising system. The first step is to express the quantity as the expectation of a suitably defined random variable over configurations in the subgraphs-world. Then we can estimate the quantity by sampling configurations at random, with the aid of the generator of Theorem 2, and computing the sample mean.

More precisely, let f be a non-negative real-valued function defined on the set Ω of subgraphs-world configurations of a ferromagnetic Ising system. Viewing Ω as a sample space with probability distribution $\pi(X) = w(X)/Z'$, the function f becomes a random variable with expectation

$$\mathbf{E}(f) = \frac{1}{Z'} \sum_{X \in \Omega} w(X) f(X).$$

It is a simple matter to get an estimate of $\mathbf{E}(f)$ using the generator of Theorem 2: construct an independent sample $\{X_i\}$ of configurations, of size s, and compute the sample mean $s^{-1} \sum_i f(X_i)$. Provided the tolerance input to the generator is small, this will be an almost unbiased estimator of $\mathbf{E}(f)$. By making the sample size s large enough, we can achieve any desired degree of accuracy with reasonable confidence. Moreover, we may drastically reduce the probability that the estimator falls outside the acceptable range of accuracy by repeating the entire process t times and taking the median of the t results.

The efficiency of such an experiment depends on how large the numbers s and t must be in order to achieve a specified accuracy with specified confidence. This in turn depends on the variance of the random variable f, or more precisely on the quantity $\max(f)/\mathbf{E}(f)$, where $\max(f)$ denotes the maximum value of f on Ω. The next lemma quantifies these effects; the proof is straightforward and is left till the end of the section.

Lemma 3 *Let f be a non-negative real-valued random variable defined on the set Ω of subgraphs-world configurations of a ferromagnetic Ising system, and let ξ, η be real numbers with $0 < \xi \leq 1$ and $0 < \eta \leq 1/2$. Then there is an experiment of the form described above which uses a total of $504\xi^{-2}\lceil \lg \eta^{-1} \rceil \max(f)/\mathbf{E}(f)$ samples from the generator, each with input $\langle \lambda_{ij}, \mu \rangle$ and tolerance $\delta = \xi\,\mathbf{E}(f)/8\max(f)$, and produces an output Y satisfying*

$$\Pr\Big(Y \text{ approximates } \mathbf{E}(f) \text{ within ratio } 1+\xi\Big) \geq 1-\eta.$$

Lemma 3 makes it clear that, whenever we employ the above technique, we will need to ensure that the ratio $\max(f)/\mathbf{E}(f)$ is not too large for the random variable f under consideration. In particular, our criterion for efficiency demands that the ratio be bounded by a polynomial function of n, the size of the system.

We turn now to an explanation of how the technique can be applied to compute the partition function $Z(V_{ij}, B, \beta)$. Recall from Theorem 1 of the previous section that $Z = AZ'$, where A is simple to evaluate directly. We therefore concentrate on computing Z'. Our first step is to write Z' explicitly as a function of μ as follows:

$$Z' \equiv Z'(\mu) = \sum_{X \subseteq E} \mu^{|\mathrm{odd}(X)|} \prod_{\{i,j\} \in X} \lambda_{ij} = \sum_{k=0}^{\lfloor n/2 \rfloor} c_k \mu^{2k}. \tag{7}$$

Note that only even powers of μ need be included in the sum since the number of odd-degree vertices in a subgraph X is necessarily even. We are thus viewing Z' as a polynomial in μ^2 with coefficients

$$c_k \;=\; \sum_{X\,:\,|\mathrm{odd}(X)|=2k} \;\prod_{\{i,j\}\in X} \lambda_{ij}. \tag{8}$$

In the ferromagnetic case all the c_k are positive, so $Z'(\mu)$ is an increasing function of μ.

Clearly the coefficients c_k actually depend on the λ_{ij}, and hence on the interactions V_{ij} of the system and on the parameter β. However, in what follows we will regard these quantities, and therefore also the coefficients, as fixed, and consider what happens when μ is varied. In spins-world terminology, this corresponds to subjecting a system with fixed interactions and at a fixed temperature to a varying external field. Our task is to evaluate the partition function at a specified external field value B. By the above discussion, this is reduced to evaluating the polynomial $Z'(\mu) = \sum c_k \mu^{2k}$ at the point $\mu = \tanh \beta B$.

Our starting point is the observation that the value of $Z'(\mu)$ at $\mu = 1$ can be computed directly. To see this, note from (7) and (8) that

$$Z'(1) \;=\; \sum_{k=0}^{\lfloor n/2 \rfloor} c_k \;=\; \sum_{X\subseteq E} \;\prod_{\{i,j\}\in X} \lambda_{ij} \;=\; \prod_{\{i,j\}\in E} (1+\lambda_{ij}). \tag{9}$$

We are now going to relate the desired value $Z'(\tanh\beta B)$ to $Z'(1)$ using the values of $Z'(\mu)$ at certain intermediate points $\tanh \beta B < \mu < 1$.

The mechanism for relating the values of Z' at two points $\mu = \mu_0$ and $\mu = \mu_1$, with $1 \geq \mu_0 > \mu_1 \geq 0$, is the following. Consider the random variable $f(X) = (\mu_1/\mu_0)^{|\mathrm{odd}(X)|}$ over configurations of the system at $\mu = \mu_0$. The expectation of f is given by

$$\mathbf{E}_{\mu_0}(f) \;=\; \frac{1}{Z'(\mu_0)} \sum_{k=0}^{\lfloor n/2 \rfloor} c_k \mu_0^{2k} \left(\frac{\mu_1}{\mu_0}\right)^{2k} \;=\; \frac{Z'(\mu_1)}{Z'(\mu_0)}.$$

(Here and in the sequel we will use notation such as $\mathbf{E}_{\mu_0}(f)$ to indicate the fact that the expectation is with respect to a particular value of μ, in this case μ_0.) Hence we can estimate the quantity $Z'(\mu_1)/Z'(\mu_0)$ using the sampling technique discussed earlier. By Lemma 3, this process will be efficient provided the ratio $\max(f)/\,\mathbf{E}_{\mu_0}(f)$ is not too large. Clearly, this cannot be guaranteed for arbitrary values of μ_0 and μ_1. However, if the values are reasonably close together then the ratio is bounded rather tightly, as we now show. First, note that certainly $\max(f) \leq 1$. It is therefore enough to obtain a lower bound on the expectation $\mathbf{E}_{\mu_0}(f)$. Such a bound is provided by the next lemma, whose proof we defer to the end of the section.

Lemma 4 *Let μ_0 and μ_1 be arbitrary real numbers in the range $[0,1]$ satisfying $\mu_1 < \mu_0 \leq \mu_1 + n^{-1}$. Then the ratio $Z'(\mu_1)/Z'(\mu_0)$ is bounded below by $1/10$.*

Lemma 4 suggests that we should be able to bootstrap the known value $Z'(1)$ to the desired value $Z'(\tanh \beta B)$ by performing statistical experiments at a sequence of intermediate values of μ which are a distance n^{-1} apart. Specifically, let $r < n$ be the natural number satisfying

$$\frac{n-r}{n} > \tanh \beta B \geq \frac{n-r-1}{n}, \tag{10}$$

and define the sequence (μ_k) for $0 \leq k \leq r+1$ by

$$\mu_k = \begin{cases} (n-k)/n & \text{for } 0 \leq k \leq r; \\ \tanh \beta B & \text{for } k = r+1. \end{cases} \tag{11}$$

Note that $\mu_k \in [0,1]$ and $\mu_{k+1} < \mu_k \leq \mu_{k+1} + n^{-1}$. Hence by the above discussion we may estimate the ratio $Z'(\mu_{k+1})/Z'(\mu_k)$ efficiently for each k. This is enough to yield an estimate of $Z'(\tanh \beta B)$, since we have

$$Z'(\tanh \beta B) = Z'(1) \times \prod_{k=0}^{r} \frac{Z'(\mu_{k+1})}{Z'(\mu_k)}. \tag{12}$$

We are now in a position to write down our approximation algorithm for the partition function Z. We assume that the input consists of a ferromagnetic Ising system in the form $\langle V_{ij}, B, \beta \rangle$, and a positive real $\epsilon \in (0,1]$ which specifies the desired accuracy. As usual, we set $\lambda_{ij} = \tanh \beta V_{ij}$.

Step 1. Compute $A = (2\cosh \beta B)^n \prod_{\{i,j\}\in E} \cosh \beta V_{ij}$, and $Z'(1) = \prod_{\{i,j\}\in E}(1+\lambda_{ij})$.

Step 2. Define the sequence (μ_k) for $0 \leq k \leq r+1$ as in (10) and (11) above. For each $k = 0, 1, \ldots, r$ in turn, do the following:

> Let $f(X) = (\mu_{k+1}/\mu_k)^{|\mathrm{odd}(X)|}$ for each subgraphs-world configuration X, so that $\mathbf{E}_{\mu_k}(f) = Z'(\mu_{k+1})/Z'(\mu_k)$. Using the technique of Lemma 3 applied to the system at $\mu = \mu_k$, with $\xi = \epsilon/2n$ and $\eta = 1/4n$, compute a quantity Y_k satisfying
>
> $$\Pr\Big(Y_k \text{ approximates } Z'(\mu_{k+1})/Z'(\mu_k) \text{ within ratio } 1+\epsilon/2n\Big) \geq 1 - 1/4n.$$

Step 3. Output the product

$$A \times Z'(1) \times \prod_{k=0}^{r} Y_k.$$

Theorem 5 *The above algorithm is an fpras for the partition function Z of a ferromagnetic Ising system.*

Proof The output of the algorithm is the product of the quantities A and $Z'(1)$, computed exactly in Step 1, together with $r+1 \leq n$ random variables Y_k arising from experiments in Step 2. From (12) and the property of the Y_k expressed in Step 2, it is immediate that the product approximates $Z(V_{ij}, B, \beta)$ within ratio $(1+\epsilon/2n)^n \leq 1+\epsilon$ with probability at least $(1-1/4n)^n \geq 3/4$. It remains only to show that the runtime of the algorithm is bounded by a polynomial in n and ϵ^{-1}.

Steps 1 and 3 can clearly be executed in time $\mathrm{O}(n^2)$. Now consider the operation of Step 2 for a particular value of k. Appealing to Lemmas 3 and 4, we see that the process of computing the estimate Y_k requires $N = 20160\epsilon^{-2}n^2\lceil \lg 4n \rceil$ calls to the generator of Theorem 2. Moreover, the tolerance supplied on each call is $\delta = \epsilon/160n$, and the value of μ is never less than n^{-1}. It follows from Theorem 2 that the runtime of each call is bounded by $q(n, \epsilon^{-1})$ for some polynomial $q(\cdot, \cdot)$. The total execution time of Step 2 is therefore $\mathrm{O}(nN\, q(n, \epsilon^{-1}))$, which is a polynomial function of n and ϵ^{-1}. The algorithm therefore satisfies all the requirements of an fpras. $\square$

Remarks (i) The statement of Theorem 2 actually gives an upper bound on the polynomial q appearing at the end of the above proof. From this, it is easily seen that the overall runtime of the fpras of Theorem 5 is $\mathrm{O}\big(\epsilon^{-2}m^2n^{11}\log n\,(\log(\epsilon^{-1}n)+m)\big)$. Now we may assume without loss of generality that $\epsilon \geq 2^{-m}$, since otherwise we can evaluate Z exactly by brute force in time $\mathrm{O}(m\epsilon^{-1})$. Hence the expression for the runtime simplifies to $\mathrm{O}(\epsilon^{-2}m^3n^{11}\log n)$. (See also the Remark following the proof of Theorem 7 at the end of the next section.)

(ii) Closer analysis reveals that the sequence of coefficients (c_k) of the polynomial expression (7) for Z' is *log-concave*, i.e.,

$$c_{k+1}c_{k-1} \leq c_k^2 \qquad \text{for } k = 1, 2, \ldots, \lfloor n/2 \rfloor - 1.$$

(The proof makes use of the ideas introduced in the proof of Theorem 7 of the next section.) This is a surprising result in its own right, since naturally occurring log-concave sequences are quite rare in combinatorics. It also suggests an alternative method for approximating Z: by log-concavity, for each k it is possible to choose a value of μ which assigns to configurations with precisely k pairs of odd-degree vertices the largest aggregated weight. This in turn means that we can read off all significant coefficients c_k by sampling at appropriate values of μ, again using $Z'(1)$ as a reference value. (An analogous approach was used in [27, 16] to obtain the coefficients of a polynomial related to matchings, or monomer-dimer configurations, in a graph.) This method is both more complex and rather less efficient than the one presented in Theorem 5. However, it does supply more detailed information about Z, in the form of the coefficients of Z'. We have been unable to determine whether these quantities have any inherent physical significance, so we will not present the alternative algorithm in detail here.

We close the section by providing the missing proofs of Lemmas 3 and 4.

Proof of Lemma 3 Let $\mathrm{Var}(f)$ denote the variance of f, i.e., $\mathrm{Var}(f) = \mathbf{E}(f^2) - \mathbf{E}(f)^2$. The generator of Theorem 2 selects elements of Ω from a distribution p which is slightly different from π. Accordingly, define the mean and variance of f with respect to this distribution by

$$\begin{aligned} \mathbf{E}'(f) &= \sum_{X\in\Omega} p(X)f(X); \\ \mathrm{Var}'(f) &= \sum_{X\in\Omega} p(X)f(X)^2 - \mathbf{E}'(f)^2. \end{aligned}$$

Since the variation distance satisfies $\|p-\pi\| \le \delta$, we have

$$\begin{aligned} |\,\mathbf{E}(f) - \mathbf{E}'(f)| &\le \delta \max(f) = \xi\,\mathbf{E}(f)/8; \\ |\mathrm{Var}(f) - \mathrm{Var}'(f)| &\le 3\delta \max(f)^2 = 3\xi\,\mathbf{E}(f)\max(f)/8. \end{aligned} \tag{13}$$

Now let $\{X_i\}$ be an independent sample of size s produced by the generator, and let $Y_0 = s^{-1}\sum_i f(X_i)$ be the sample mean. Clearly Y_0 has expectation $\mathbf{E}'(f)$ and variance $s^{-1}\mathrm{Var}'(f)$. Therefore, by Chebyshev's inequality we have

$$\Pr\Big(|Y_0 - \mathbf{E}'(f)| > \tfrac{\xi}{3}\,\mathbf{E}'(f)\Big) \le \frac{9}{\xi^2}\,\frac{\mathrm{Var}'(f)}{s\,\mathbf{E}'(f)^2}. \tag{14}$$

But if $|Y_0 - \mathbf{E}'(f)| \le \frac{\xi}{3}\,\mathbf{E}'(f)$ then, from (13),

$$\begin{aligned} |Y_0 - \mathbf{E}(f)| &\le |Y_0 - \mathbf{E}'(f)| + |\,\mathbf{E}'(f) - \mathbf{E}(f)| \\ &\le \tfrac{\xi}{3}\,\mathbf{E}'(f) + \tfrac{\xi}{8}\,\mathbf{E}(f) \\ &\le \tfrac{\xi}{3}\left(1+\tfrac{\xi}{8}\right)\mathbf{E}(f) + \tfrac{\xi}{8}\,\mathbf{E}(f) \\ &\le \tfrac{\xi}{2}\,\mathbf{E}(f). \end{aligned} \tag{15}$$

Note that this in turn implies that Y_0 approximates $\mathbf{E}(f)$ within ratio $1+\xi$. Moreover, applying (13) again we have

$$\frac{\mathrm{Var}'(f)}{\mathbf{E}'(f)^2} \le \frac{\mathrm{Var}(f) + \frac{3}{8}\,\mathbf{E}(f)\max(f)}{\left(\frac{7}{8}\,\mathbf{E}(f)\right)^2} \le \frac{\frac{11}{8}\,\mathbf{E}(f)\max(f)}{\left(\frac{7}{8}\,\mathbf{E}(f)\right)^2} < \frac{2\max(f)}{\mathbf{E}(f)}, \tag{16}$$

where in the second inequality we have used the distribution-independent bound $\mathrm{Var}(f) \le \mathbf{E}(f)\max(f)$, valid for any non-negative random variable f. Combining (15) and (16) with (14), and choosing sample size $s = 72\xi^{-2}\max(f)/\,\mathbf{E}(f)$, gives

$$\Pr\Big(Y_0 \text{ approximates } \mathbf{E}(f) \text{ within ratio } 1+\xi\Big) \le \frac{18}{\xi^2 s}\,\frac{\max(f)}{\mathbf{E}(f)} = \frac{1}{4}. \tag{17}$$

Now consider performing the above experiment an odd number t times, independently, and let Y denote the median of the resulting t values of Y_0. In view of (17), the probability that Y fails to approximate $\mathbf{E}(f)$ within ratio $1+\xi$ is at most

$$\sum_{i=(t+1)/2}^{t} \binom{t}{i} \left(\frac{1}{4}\right)^{i} \left(\frac{3}{4}\right)^{t-i} \leq \left(\frac{1}{4}\right)^{t/2} \left(\frac{3}{4}\right)^{t/2} \sum_{i=(t+1)/2}^{t} \binom{t}{i} \leq \left(\frac{3}{16}\right)^{t/2} 2^t = \left(\frac{3}{4}\right)^{t/2}.$$

Taking $t = 6\lceil \lg \eta^{-1} \rceil + 1$, this probability is bounded above by $\eta^{3 \lg(4/3)} < \eta$. The random variable Y therefore satisfies the requirement of the lemma. The total number of samples required from the generator is st, which is bounded above by $504\xi^{-2}\lceil \lg \eta^{-1} \rceil \max(f)/\mathbf{E}(f)$ as claimed. $\square$

Proof of Lemma 4 We split the argument into two cases.

Case I: $\mu_0 \geq 3/4$. In this case, we have

$$\frac{\mu_1}{\mu_0} \geq 1 - \frac{1}{n\mu_0} \geq 1 - \frac{4}{3n}.$$

Therefore, since $Z'(\mu) = \sum_{k=0}^{\lfloor n/2 \rfloor} c_k \mu^{2k}$, and all coefficients c_k are positive,

$$\frac{Z'(\mu_1)}{Z'(\mu_0)} \geq \left(\frac{\mu_1}{\mu_0}\right)^{2\lfloor n/2 \rfloor} \geq \left(1 - \frac{4}{3n}\right)^{n} \geq \frac{1}{9},$$

assuming as we may that $n \geq 2$. (The problem is trivial otherwise.)

Case II: $\mu_0 < 3/4$. This case is handled by appealing to the original spins-world expansion of Z. First note from the definition (4) of Z' that

$$\frac{Z'(\mu_1)}{Z'(\mu_0)} = \frac{Z(V_{ij}, B_1, \beta)}{Z(V_{ij}, B_0, \beta)} \times \left(\frac{2\cosh \beta B_0}{2\cosh \beta B_1}\right)^{n} \geq \frac{Z(V_{ij}, B_1, \beta)}{Z(V_{ij}, B_0, \beta)}, \tag{18}$$

where $\beta > 0$ is arbitrary and B_0, B_1 are defined by $\mu_i = \tanh \beta B_i$. Note that $B_0 \geq B_1$. Moreover, the upper bound $\mu_0 - \mu_1 \leq n^{-1}$ translates to a bound on $B_0 - B_1$ via the inequality $\tanh x - \tanh y \geq (x-y)\operatorname{sech}^2 x$, valid for $x \geq y \geq 0$. We get

$$\beta(B_0 - B_1) \leq (\mu_0 - \mu_1)/\operatorname{sech}^2 \beta B_0 \leq 16/7n, \tag{19}$$

where we have used the fact that $\tanh \beta B_0 = \mu_0 < 3/4$ and $\operatorname{sech}^2 x = 1 - \tanh^2 x$.

But from the definition (1) of the partition function Z we have

$$\frac{Z(V_{ij}, B_1, \beta)}{Z(V_{ij}, B_0, \beta)} \geq \min_{\boldsymbol{\sigma}} \exp\Big(\beta(B_1 - B_0)\textstyle\sum_{k\in[n]} \sigma_k\Big) \geq \exp\Big(-n\beta(B_0 - B_1)\Big),$$

which by (19) is bounded below by $\mathrm{e}^{-16/7}$ and hence by $1/10$. Together with (18) this yields the desired bound on $Z'(\mu_1)/Z'(\mu_0)$. $\square$

4 An analysis of the subgraphs-world process

We shall assume that the reader is familiar with the elementary theory of finite Markov chains in discrete time: an introduction can be found, for example, in [10, Chapter XV].

Assume $\mu > 0$. Taking our cue from the form of equation (4), we define the subgraphs-world process, $\mathcal{MC}_{\text{Ising}}$, as follows. The state space, Ω, of the Markov chain $\mathcal{MC}_{\text{Ising}}$ is the set of all spanning subgraphs $X \subseteq E$; note that $|\Omega| = 2^m$ where $m = |E|$ is the number of unordered pairs $\{i,j\}$ with $\lambda_{ij} \neq 0$. For $X, X' \in \Omega$ with $X \neq X'$, the transition probability from X to X' is given by

$$p(X,X') = \begin{cases} 1/2m & \text{if } |X \oplus X'| = 1 \text{ and } w(X') \geq w(X); \\ w(X')/(2m\, w(X)) & \text{if } |X \oplus X'| = 1 \text{ and } w(X') < w(X); \\ 0 & \text{otherwise,} \end{cases}$$

where $X \oplus X'$ denotes the symmetric difference of X and X'. The self-loop probabilities $p(X,X)$ are defined implicitly by complementation, so that $p(X,X) = 1 - \sum_{X' \neq X} p(X,X')$. Thus, transitions in $\mathcal{MC}_{\text{Ising}}$ are perturbations in which a single edge is added to, or deleted from a subgraph. Note that exactly m transitions are available from any state, and all transition probabilities are bounded above by $1/2m$. Hence the transition probabilities are well defined, and the self-loop probabilities $p(X,X)$ for each state X are bounded below by $1/2$.

We pause to observe that the above chain is very easy to simulate. Suppose the current state of the chain is $X \in \Omega$. Then the transitions from X can be selected according to the following model:

1. with probability $1/2$ set $X' = X$, otherwise
2. select an edge $e \in E$ uniformly at random, and let $Y = X \oplus \{e\}$ (the symmetric difference of X and $\{e\}$);
3. if $w(Y) \geq w(X)$ then set $X' = Y$; if $w(Y) < w(X)$ then with probability $w(Y)/w(X)$ set $X' = Y$, otherwise set $X' = X$.

It will be seen that this procedure correctly models the transition probabilities specified earlier. It is worth remarking that there is no need to compute the weight functions $w(X)$ and $w(Y)$ from scratch at each iteration; since Y and X differ by a single edge, the quotient $w(Y)/w(X)$ can be computed using just two multiplications.

The Markov chain $\mathcal{MC}_{\text{Ising}}$ is irreducible (all states communicate via the empty state $\emptyset$) and aperiodic (the self-loop probabilities are non-zero). Thus there is a well defined stationary distribution on Ω which is independent of the initial state. Define $\pi : \Omega \to \mathbb{R}$ by $\pi(X) = w(X)/\sum_{X'} w(X') = w(X)/Z'$. We shall see presently that π is indeed the stationary distribution on Ω. For $X, X' \in \Omega$ define $q(X,X') = \pi(X)\, p(X,X')$.

We claim that q is symmetric in its two arguments. If $X = X'$ then there is nothing to prove. If $|X \oplus X'| > 1$ then $p(X, X') = 0$ and hence $q(X, X') = 0$. Finally, it is straightforward to verify from the definition of the transition probability $p(X, X')$ that

$$q(X, X') = (2m)^{-1} \min\{\pi(X), \pi(X')\}, \quad \text{if } |X \oplus X'| = 1. \tag{20}$$

Since q is symmetric, the so-called *detailed balance* condition holds:

$$\pi(X)p(X, X') = q(X, X') = \pi(X')p(X', X). \tag{21}$$

Suppose, as is the case here, that the function $p(\cdot, \cdot)$ describes the transition probabilities of an ergodic Markov chain. It is a fact [16, Lemma 2.1] that if there is *any* function $\pi : \Omega \to \mathbb{R}$ satisfying detailed balance together with the normalisation condition $\sum_{X \in \Omega} \pi(X) = 1$, then the Markov chain is (*time-*)*reversible*, and π is its stationary distribution. Thus the stationary distribution of the Markov chain $\mathcal{MC}_{\text{Ising}}$ is indeed given by $\pi(X) = w(X)/Z'$, as claimed above, and we can use the chain to sample configurations $X \in \Omega$ with probabilities approximately proportional to $w(X)$.

As explained informally earlier, if the Markov chain $\mathcal{MC}_{\text{Ising}}$ is to be the basis of an efficient sampling procedure for configurations then it must be *rapidly mixing*, in the sense that, if it is allowed to evolve from a suitable initial state, the distribution of its final state will be very close to the stationary distribution after only polynomially many steps. Note that this is a highly non-trivial requirement: since the number of states in the chain is exponentially large, we are demanding that it converges after visiting only a tiny fraction of its state space. Our argument that the chain is rapidly mixing is in two parts: first, in Theorem 6, we state a general characterisation of the rapid mixing property in terms of a measure known as the *conductance*; then, in Theorem 7, we estimate the conductance of $\mathcal{MC}_{\text{Ising}}$.

For an ergodic reversible Markov chain, the conductance [27, 29] is defined by

$$\Phi = \min \left\{ \sum_{\substack{X \in S \\ X' \notin S}} q(X, X') \Big/ \sum_{X \in S} \pi(X) \right\},$$

where the minimisation is over all subsets S of states with $0 < \sum_{X \in S} \pi(X) \leq 1/2$. (Note that $0 < \Phi \leq 1$.) The conductance in some sense measures the rate at which the process can flow around the state space: specifically, it provides a lower bound on the conditional probability that the stationary process escapes from a small subset S of the state space in a single step, given that it is initially in S. Thus a chain with large conductance is unlikely to "get stuck" in any small region of the state space, so we might expect it to converge fast. This intuition is captured in the following theorem.

Theorem 6 *Let Φ be the conductance of an ergodic, reversible Markov chain with stationary distribution π and* $\min_X p(X, X) \geq 1/2$. *Let $p^{(t)}$ denote the distribution of the* *state at time t given that the initial state is X_0. Then the variation distance $\|p^{(t)} - \pi\|$ satisfies*

$$\|p^{(t)} - \pi\| \ \leq \ \frac{(1-\Phi^2)^t}{\pi(X_0)}.$$

(The requirement that $\min_X p(X,X) \geq 1/2$, i.e, that every state has a self-loop probability of at least $1/2$, is a technical device which removes periodicity; note that $\mathcal{MC}_{\text{Ising}}$ satisfies this requirement by construction.)

Proof The theorem is essentially a restatement of Theorem 3.4 of [29], to which the interested reader is referred for details; we mention here only the necessary modifications. The main difference stems from the fact that in the former result we used the stronger *relative pointwise distance* (r.p.d.), rather than the variation distance, as a measure of deviation from the stationary distribution. In similar fashion to the r.p.d., the variation distance at time t may be related, by elementary linear algebra, to the second eigenvalue λ_1 of the Markov chain: we get

$$\|p^{(t)} - \pi\| \ \leq \ \frac{\lambda_1^t}{\pi(X_0)}. \tag{22}$$

(See, for example, Proposition 2 of [7], which presents a marginally stronger result, with $2\sqrt{\pi(X_0)}$ replacing $\pi(X_0)$ in the denominator. Note that the presence of a self-loop probability of $1/2$ on every state ensures that all eigenvalues are non-negative.) The bound in (22) differs from that on the r.p.d. in Lemma 3.1 of [29] only in that $\pi(X_0)$ replaces $\min_X \pi(X)$.

Now the main result of Section 3 of [29], Lemma 3.3, relates λ_1 to the conductance via the bound $\lambda_1 \leq 1 - \Phi^2/2$, valid for an arbitrary reversible chain. It is easily seen from the proof of the lemma that the marginally stronger bound

$$\lambda_1 \ \leq \ 1 - \Phi^2 \tag{23}$$

holds for chains in which all self-loop probabilities are at least $1/2$. Putting (22) and (23) together establishes the theorem. $\square$

Remarks (i) The heart of the above proof is the eigenvalue bound (23). This is a discrete analogue of Cheeger's inequality for Riemannian manifolds [5]. Related bounds have been observed by several authors: see, e.g., [8, 1, 23, 25].

(ii) Theorem 6 has a converse, which says that if a Markov chain is rapidly mixing then its conductance cannot be too small: see, e.g., [23, 27, 28]. Thus the conductance provides a *characterisation* of the rapid mixing property.

Theorem 6 allows us to investigate the rate of convergence of a reversible chain by examining its transition structure, as reflected in the conductance. In particular, if we wish to ensure a variation distance of at most δ then it is clear that $\Phi^{-2}(\ln \delta^{-1} + \ln \pi(X_0)^{-1})$ steps suffice. Thus the rapid mixing property will generally follow from an inverse polynomial lower bound on the conductance. Such a bound is available for the chain $\mathcal{MC}_{\text{Ising}}$ defined above. Specifically, we have

Theorem 7 *The conductance of the Markov chain $\mathcal{MC}_{\text{Ising}}$ is bounded below by $\mu^4/4m$.*

The proof of Theorem 7 is the main content of this section. Before proceeding with it, however, let us first use the result to verify our claim in Theorem 2 of the previous section that an efficient generator for subgraphs-world configurations exists. This will complete the validation of our approximation algorithm for the partition function.

Proof of Theorem 2 The generator operates as follows. Given as input a ferromagnetic Ising system in the form $\langle \lambda_{ij}, \mu \rangle$, with $0 < \mu \leq 1$, and a tolerance $\delta \in (0,1]$, simulate the associated Markov chain $\mathcal{MC}_{\text{Ising}}$ for $16m^2\mu^{-8}(\ln \delta^{-1} + m)$ steps, starting in state $X_0 = \varnothing$ (i.e, the empty graph on vertex set $[n]$). Since $\lambda_{ij} < 1$ for all i, j, and $\mu \leq 1$, it is clear that $w(X_0) \geq w(X)$ for all configurations X. Hence $\pi(X_0) \geq 2^{-m}$. Appealing to Theorem 6, we conclude that the specified number of simulation steps is enough to ensure a variation distance of at most δ. The theorem is therefore established. $\square$

Proof of Theorem 7 The proof rests on a path counting argument, similar to those employed in previous applications [27, 16] of the conductance bound. We present a preliminary sketch map of the proof technique before considering the technical details which arise when applying the path counting argument to the particular chain under consideration.

For each pair of states $I, F \in \Omega$, a *canonical path* from I (the initial state) to F (the final state) is specified. The canonical path proceeds via a number of intermediate states using only valid transitions of the Markov chain. Each canonical path is assigned a weight which is the product of the stationary probabilities at the initial and final states; thus the weight of the path from I to F is $\pi(I)\pi(F)$, independent of the intermediate states in the path. A careful choice of canonical paths is essential to secure a good bound on conductance.

Suppose it can be shown that, for each transition $T \to T'$, the aggregated weight of canonical paths which use transition $T \to T'$ is bounded above by $b\, q(T, T')$, where q is as in equation (21). Consider any partition of the state space Ω into two sets S and $\overline{S}$ with $\sum_{X \in S} \pi(X) \leq 1/2$. Then, on the one hand, the total weight of canonical paths which cross the cut defined by S and $\overline{S}$ is at least $\sum_{I\in S}\sum_{F\in\overline{S}}\pi(I)\pi(F) = \pi(S)\pi(\overline{S}) \geq \pi(S)/2$. On the other hand, summing over transitions $T \to T'$ with $T \in S$ and $T' \in \overline{S}$, we have that the total weight of canonical paths which cross the cut is bounded above by $b\sum_{T\in S}\sum_{T'\in\overline{S}} q(T,T')$. Since S and $\overline{S}$ represent a general partition of the state space Ω, it follows immediately that the conductance of the Markov chain is bounded below by $1/2b$.

It will be perceived that the principal barrier to applying the above idea is likely to lie in obtaining a good value for the bound b. This is achieved using a combinatorial encoding technique, as follows. For each transition $T \to T'$, let $\mathrm{cp}(T,T')$ denote the set of all pairs $(I,F) \in \Omega^2$ such that the canonical path from I to F includes the transition $T \to T'$. Fix any particular transition $T \to T'$. Then it turns out that we can define an *injective map* from $\mathrm{cp}(T,T')$ to the state space Ω. Since the map is injective, any state $U \in \Omega$ picks out at most one canonical path, from I to F say, which uses the transition $T \to T'$; the state U can be thought of as an *encoding* of the canonical path. Moreover, the injective map can be chosen so that the weight of the canonical path, namely $\pi(I)\pi(F)$, is roughly proportional to $\pi(U)$, the probability assigned to the encoding U by the stationary distribution. Since π is a probability distribution, the sum of $\pi(U)$ over all encodings U is at most 1; this upper bound translates to an upper bound on the total weight of paths using $T \to T'$, and hence to a value for b.

All the above must now be specialised to the Markov chain $\mathcal{MC}_{\mathrm{Ising}}$. The first task is to specify a canonical path for each pair $I, F \in \Omega$. View I and F as graphs with vertex set $[n]$. Let $\Delta = I \oplus F$ be the symmetric difference of I and F, and suppose that the graph Δ has $2k$ vertices of odd degree. (The number of odd-degree vertices in a graph is necessarily even.) Cover Δ with a collection $C_1, C_2, \ldots, C_r$ of assorted trails and circuits which are pairwise edge-disjoint, imposing the condition that the first k objects, C_1, $C_2, \ldots$, C_k, are all open *trails* (walks with no repeated edges) while the remainder, C_{k+1}, $C_{k+2}, \ldots$, C_r, are all *circuits* (closed trails).

That this can be done with so few (open) trails follows from a simple induction on k. When $k = 0$, every vertex of Δ is of even degree, so each connected component of Δ is Eulerian and can be covered by a single circuit. Now suppose $k > 0$ and i is a vertex in Δ of odd degree. The connected component of Δ containing i must contain at least one other odd-degree vertex, say j. Connect i and j by a trail, letting this be one of the trails in the required decomposition of Δ. Deleting this trail from Δ yields a graph with $2(k-1)$ odd-degree vertices, which can be covered by $k-1$ trails (together with some number of circuits) by the induction hypothesis.

The covering of Δ by trails and circuits is not in general unique, and we assume that some rule is employed to pick out a particular choice of C_1, $C_2, \ldots$, C_r. We further assume that this rule also specifies a distinguished *initial vertex* for each trail or circuit, and a *direction* for each circuit. In the case of a trail the initial vertex must be an end-point of the trail; in the case of a circuit the initial vertex may be arbitrary. The canonical path from I to F is now obtained by *unwinding* the trails and circuits C_1, $C_2, \ldots$, C_r in sequence. Unwinding C_i involves processing each edge of C_i in sequence, starting at the initial vertex and, in the case of a circuit, following the assigned direction. Each processed edge, e, generates a single transition on the path from I to F. If e is in F (and hence not in I) the transition involves adding the edge e to the current state; if e is in I (and hence not in F) the transition involves deleting the edge e from the current state. It is clear that this process defines a canonical path of legal transitions from state I to state F.

The next task is to define the injective map (encoding) from the set of canonical paths using a given transition to Ω. Recall that $\mathrm{cp}(T,T')$ denotes the set of all pairs $(I,F)\in\Omega^2$ such that the canonical path from I to F employs the transition $T\to T'$. Define the map $\eta_{T\to T'}:\mathrm{cp}(T,T')\to\Omega$ by $\eta_{T\to T'}(I,F)=I\oplus F\oplus(T\cup T')$ for all $(I,F)\in\mathrm{cp}(T,T')$. The intention here is that the encoding should agree with I on the trails and circuits already processed, and with F elsewhere.

We verify that $\eta_{T\to T'}$ is injective by demonstrating that I and F are uniquely determined by $U=\eta_{T\to T'}(I,F)$. Indeed, given U, we can compute $U\oplus(T\cup T')=I\oplus F$ and hence the uniquely defined covering C_1, $C_2,\ldots$, C_r of $I\oplus F$ by trails and circuits. The edge $e=T\oplus T'$ which is added or deleted by the transition $T\to T'$ points out which trail or circuit, C_i, is being unwound, and how far the unwinding of C_i has progressed. Starting at state T', we may complete the unwinding of C_i and successive trails/circuits to discover the final state F; equally well, we may use the reverse process to recover the initial state I. So the map $\eta_{T\to T'}$ is injective, as claimed.

The other property we require of the encoding is that its weight should be roughly proportional to that of the encoded path. Precisely, we require of $U=\eta_{T\to T'}(I,F)$ that

$$\pi(U)q(T,T')\geq(2m)^{-1}\mu^4\pi(I)\pi(F), \tag{24}$$

or, equivalently, multiplying through by $(Z')^2$ and using assertion (20),

$$w(U)w(T)\geq\mu^4w(I)w(F)\quad\text{and}\quad w(U)w(T')\geq\mu^4w(I)w(F). \tag{25}$$

The verification of the left hand inequality will be treated in detail below; the right hand inequality will then be shown to follow by symmetry.

For $X\in\Omega$, write $\Lambda(X)=\prod_{\{i,j\}\in X}\lambda_{ij}$, so that $w(X)=\Lambda(X)\mu^{|\mathrm{odd}(X)|}$. To verify the left hand inequality of (25) it is enough to demonstrate separately that

$$\Lambda(U)\Lambda(T)\geq\Lambda(I)\Lambda(F), \tag{26}$$

and

$$|\mathrm{odd}(U)|+|\mathrm{odd}(T)|-|\mathrm{odd}(I)|-|\mathrm{odd}(F)|\leq 4. \tag{27}$$

(We have used here the fact that $0 < \mu \leq 1$.) We deal first with inequality (26), which is the more straightforward. From the construction of canonical paths we have $I \cap F \subseteq T \cup T' \subseteq I \cup F$, while the definition of U entails $U \oplus (T \cup T') = I \oplus F$. It follows by elementary set theory from these two observations that $U \cap (T \cup T') = I \cap F$ and $U \cup (T \cup T') = I \cup F$. Hence

$$\Lambda(U)\Lambda(T \cup T') = \Lambda(U \cup (T \cup T'))\,\Lambda(U \cap (T \cup T')) = \Lambda(I \cup F)\,\Lambda(I \cap F) = \Lambda(I)\Lambda(F),$$

which, together with $\Lambda(T \cup T') \leq \Lambda(T)$, implies inequality (26).

We now turn to inequality (27). For $i \in [n]$ define

$$\alpha(i) = \chi_{\mathrm{odd}(U)}(i) + \chi_{\mathrm{odd}(T)}(i) - \chi_{\mathrm{odd}(I)}(i) - \chi_{\mathrm{odd}(F)}(i),$$

where χ_S denotes the characteristic function of a set S. Note that $-2 \leq \alpha(i) \leq 2$. Inequality (27) can be re-expressed as

$$\sum_{i \in [n]} \alpha(i) \leq 4. \tag{28}$$

We shall argue that $\alpha(i) \leq 0$ for all i outside a small set of exceptions. In order to discuss these exceptions, we give names to three vertices which have special significance. Denote by s the initial vertex of the circuit which is in the process of being unwound when the transition $T \to T'$ is made (s is undefined if the transition occurs on a trail). Denote by u and v the endpoints of the edge $e = T \oplus T'$ which is added or subtracted during the transition $T \to T'$; vertex u is distinguished from v by being the first to be encountered in the direction of unwinding. We shall see that the vertices s, u, and v are the only ones which can provide a positive contribution to the sum in (28).

Consider first the conditions under which $\alpha(i) = 2$ can occur. It must be the case that i has even degree in both I and F, and odd degree in both U and T. Now it is a consequence of the way canonical paths are constructed that a vertex which has even degree in both I and F will generally have even degree in the intermediate configuration T; the only exceptions are the vertex s (whose degree became odd when the unwinding of the circuit commenced) and the vertex u (whose degree was made odd by the previous transition, and whose parity is restored by the transition $T \to T'$ itself). To summarise: the case $\alpha(i) = 2$ can only occur when $i = s$ or $i = u$.

Consider now the conditions under which $\alpha(i) = 1$ can occur. This case is ruled out, with two exceptions, by simple parity considerations. Since $I \oplus F = U \oplus (T \cup T')$, the value of $\alpha(i)$ must be even unless $i = u$ or $i = v$. (These are the only points at which the set $T \cup T'$ may differ from T.) Combining this observation with the previous analysis for the case $\alpha(i) = 2$, we see that only three terms of the sum occurring in (28) can possibly be strictly positive, and that the sum itself cannot exceed 5. (The worst case is achieved by setting $\alpha(s) = 2$, $\alpha(u) = 2$, and $\alpha(v) = 1$.) However, the sum cannot actually attain 5 on grounds of parity: each of the terms appearing on the left hand side of inequality (27) is necessarily even. This completes the verification of inequality (28), and hence of (27) and the left hand inequality of (25). The right hand inequality of (25) follows by a symmetrical argument, with the roles of T and T', and u and v, interchanged.

Finally, summing (24) over all canonical paths which employ the transition $T \to T'$ we obtain the following upper bound on the total weight of canonical paths which use $T \to T'$:

$$\sum_{(I,F)\in\mathrm{cp}(T,T')} \pi(I)\pi(F) \le \sum_{(I,F)\in\mathrm{cp}(T,T')} 2m\mu^{-4}\pi(\eta_{T\to T'}(I,F))\, q(T,T') \le 2m\mu^{-4}q(T,T'),$$

where the second inequality rests on the fact that $\eta_{T\to T'}$ is injective. In the notation of the sketch map presented at the beginning of the proof, $b = 2m\mu^{-4}$. Therefore the conductance of $\mathcal{MC}_{\mathrm{Ising}}$ is bounded below by $1/2b = \mu^4/4m$, as claimed at the outset. $\square$

Remark The main task in the proof of Theorem 7 is to estimate the "bottleneck" measure b; this is then used to get a bound on conductance, and hence on the rate of convergence of $\mathcal{MC}_{\mathrm{Ising}}$. In fact, b can be used *directly* to obtain a significantly sharper bound on the rate of convergence: for the details, see [28]. Specifically, the runtime of the generator quoted in Theorem 2 is reduced by a factor $\mathrm{O}(\mu^{-4})$, which improves the runtime of the approximation algorithm for the partition function (Theorem 5) by a factor of $\mathrm{O}(n^4)$. Similar improvements apply to the runtimes of our other algorithms that make use of the generator.

5 The mean energy and mean magnetic moment

Of greater immediate importance to physicists than the partition function Z itself are the partial derivatives of $\ln Z$ with respect to β and B. The key quantities of interest are the *mean energy* $\mathcal{E} = -\partial(\ln Z)/\partial\beta$, and the *mean magnetic moment* $\mathcal{M} = \beta^{-1}\,\partial(\ln Z)/\partial B$. As their names suggest, both of these can be viewed, in the spins-world, as expectations of the corresponding physical quantities.

There is no reason to suppose that an fpras for the partition function Z will directly yield an fpras for these derivatives of $\ln Z$. However, we demonstrate in this section that polynomial-time approximation algorithms for $\mathcal{E}$ and $\mathcal{M}$ do indeed exist. The construction of these algorithms relies on the surprising fact that $\mathcal{E}$ and $\mathcal{M}$ can be viewed as expectations of appropriately defined random variables in the subgraphs-world. Although some technical complications arise, it is possible to estimate these expectations more or less directly by simulating the subgraphs-world process for a polynomially bounded number of steps.

The mean magnetic moment is slightly easier to work with, and we treat it first. The main result is preceded by a technical lemma, whose proof is deferred to the end of the section. Recall that in the subgraphs-world distribution, each configuration X occurs with probability $w(X)/Z'$.

Lemma 8 *Suppose the configuration $X \in \Omega$ is randomly selected according to the subgraphs-world distribution. Then*

(i) $\Pr(|\mathrm{odd}(X)| > 0) \geq \mu^2/2$, *provided* $\sum \lambda_{ij} \geq 1$;

(ii) $\Pr(|\mathrm{odd}(X)| = 2) \geq \mu^2/10$, *provided* $\sum \lambda_{ij} \geq 1$ *and* $\mu \leq n^{-1}$.

Theorem 9 *There exists an fpras for the mean magnetic moment $\mathcal{M} = \beta^{-1}\, \partial(\ln Z)/\partial B$, where Z is the partition function of a ferromagnetic Ising system.*

Proof We shall express the quantity $\mathcal{M}$ as an expectation in the subgraphs-world by differentiating the logarithm of the expansion given in Theorem 1 with respect to B. Before doing this, it is convenient to perform some preparatory computations. Since $\mathcal{M} = 0$ when $B = 0$, we may assume that $B > 0$. Recall that $w(X) = \Lambda(X)\mu^{|\mathrm{odd}(X)|}$, where $\mu = \tanh \beta B$ by definition, and $\Lambda(X)$ is independent of B. Then

$$
\begin{aligned}
\frac{\partial}{\partial B} w(X) &= \Lambda(X)|\mathrm{odd}(X)|\mu^{|\mathrm{odd}(X)|-1}(\operatorname{sech} \beta B)^2 \beta \\
&= \beta\, w(X)|\mathrm{odd}(X)|(\tanh \beta B)^{-1}(\operatorname{sech} \beta B)^2 \\
&= w(X)\frac{2\beta|\mathrm{odd}(X)|}{\sinh 2\beta B}.
\end{aligned}
$$

Furthermore, from the definition of A in equation (5),

$$
\frac{\partial}{\partial B} \ln A = \frac{\partial}{\partial B} n \ln \cosh \beta B = n\beta \tanh \beta B.
$$

With these identities in mind, we compute $\mathcal{M}$ using the expansion of Theorem 1 as the starting point:

$$
\begin{aligned}
\mathcal{M} = \frac{1}{\beta}\frac{\partial}{\partial B} \ln Z &= \frac{1}{\beta}\frac{\partial}{\partial B} \ln A + \frac{1}{\beta}\frac{\partial}{\partial B} \ln Z' \\
&= n \tanh \beta B + \frac{1}{\beta Z'} \sum_X \frac{\partial}{\partial B} w(X) \\
&= n \tanh \beta B + \frac{1}{Z'} \sum_X w(X) \frac{2|\mathrm{odd}(X)|}{\sinh 2\beta B}.
\end{aligned}
$$

Using, as before, the notation $\mathbf{E}(f) = (Z')^{-1}\sum_X w(X)f(X)$ to express the expectation of a random variable f in the subgraphs-world, the above identity can be written more compactly as

$$\mathcal{M} = n\tanh\beta B + \frac{2}{\sinh 2\beta B}\,\mathbf{E}\,|\mathrm{odd}(X)|. \tag{29}$$

Note that to approximate $\mathcal{M}$ within ratio $1+\epsilon$, it is enough, since both terms of (29) are positive, to estimate $\mathbf{E}\,|\mathrm{odd}(X)|$ within ratio $1+\epsilon$. We propose to do this by using the Markov chain $\mathcal{MC}_{\mathrm{Ising}}$ analysed in Section 4 to provide a polynomial number of sample configurations X from the subgraphs-world distribution, and returning the average of $|\mathrm{odd}(X)|$ over the sample. As noted in the discussion preceding Lemma 3, this approach will yield an fpras for $\mathcal{M}$ provided the ratio $\max|\mathrm{odd}(X)|/\,\mathbf{E}\,|\mathrm{odd}(X)|$ is bounded by a polynomial in n. Although such a bound often holds, a more refined approach is necessary in certain circumstances. We proceed by case analysis.

Case I: $\sum\lambda_{ij} \geq 1$. We identify two subcases, according to the size of μ.

Case Ia: $\mu \geq n^{-1}$. In this range, we may estimate $\mathcal{M}$ by direct experiment. From Lemma 8,

$$\mathbf{E}\,|\mathrm{odd}(X)| \geq 2\Pr(|\mathrm{odd}(X)| > 0) \geq \mu^2 \geq n^{-2},$$

while, clearly, $\max|\mathrm{odd}(X)| \leq n$. Thus the ratio $\max|\mathrm{odd}(X)|/\,\mathbf{E}\,|\mathrm{odd}(X)|$ is bounded above by n^3.

Case Ib: $\mu < n^{-1}$. Intuitively, the problem when μ is small is that a randomly sampled configuration may, with high probability, satisfy $|\mathrm{odd}(X)| = 0$; in this situation, very many trials may be required to obtain a sufficiently accurate estimate of the expectation of $|\mathrm{odd}(X)|$. The solution is to perform experiments at an increased value of μ, say $\hat{\mu}$, at which the event $|\mathrm{odd}(X)| > 0$ occurs with sufficiently high probability. Since we shall be allowing μ to vary, it is convenient, as before, to refine our notation to make the dependence on μ explicit. In particular, $\mathbf{E}_{\hat{\mu}}(f)$ will denote the expectation of the random variable f when experiments are undertaken with μ set to some revised value $\hat{\mu}$. The unsubscripted notation $\mathbf{E}(f)$ will be reserved for expectations with respect to the original value of μ.

Set $\hat{\mu} = n^{-1}$ and define

$$f(X) = |\mathrm{odd}(X)|\left(\frac{\mu}{\hat{\mu}}\right)^{|\mathrm{odd}(X)|}.$$

Straightforward manipulations yield the identity

$$\mathbf{E}\,|\mathrm{odd}(X)| = \frac{Z'(\hat{\mu})}{Z'(\mu)}\,\mathbf{E}_{\hat{\mu}}(f), \tag{30}$$

which relates the quantity we are attempting to estimate to the expectation of f at the revised value of μ. Since $Z'(\hat{\mu})$ and $Z'(\mu)$ may be estimated by the techniques of

Section 3, we concentrate here on the estimation of $\mathsf{E}_{\hat\mu}(f)$. From part (ii) of Lemma 8 we have $\Pr_{\hat\mu}(|\mathrm{odd}(X)| = 2) \geq \hat\mu^2/10$; this inequality allows the expectation of f to be bounded below:

$$\mathsf{E}_{\hat\mu}(f) \geq 2\left(\frac{\mu}{\hat\mu}\right)^2 \Pr_{\hat\mu}(|\mathrm{odd}(X)| = 2) \geq \frac{\mu^2}{5}.$$

The maximum of f, meanwhile, satisfies the crude bound $\max(f) \leq (\mu/\hat\mu)^2 n$. Thus the ratio $\max(f)/\mathsf{E}_{\hat\mu}(f)$ is bounded above by $5n\hat\mu^{-2} = 5n^3$.

Case II: $\sum \lambda_{ij} < 1$. In this rather pathological case, the essential problem we face is that a randomly sampled configuration may, with high probability, be the empty set. As before, we shall conduct experiments at an artificially inflated value of μ and use equation (30) to relate the results of these experiments to the value we are attempting to estimate. This time, however, we choose to work with $\hat\mu = 1$.

Unfortunately, even with $\hat\mu$ set to 1, the highest possible value, it may still happen that the empty configuration $X = \varnothing$ occurs with very high probability. We overcome this problem by sampling from the distribution obtained by conditioning on the event $X \neq \varnothing$. With f as before, and noting that $\hat\mu = 1$ and $f(\varnothing) = 0$, we have

$$\begin{aligned}\mathsf{E}_{\hat\mu}(f) &= \sum_{k=0}^{\lfloor n/2\rfloor} 2k\mu^{2k} \Pr_{\hat\mu}(|\mathrm{odd}(X)| = 2k) \\ &= \Pr_{\hat\mu}(X \neq \varnothing) \sum_{k=1}^{\lfloor n/2\rfloor} 2k\mu^{2k} \Pr_{\hat\mu}\Big(|\mathrm{odd}(X)| = 2k \;\Big|\; X \neq \varnothing\Big).\end{aligned}$$

Now $Z'(\hat\mu) = \prod(1+\lambda_{ij})$, and hence

$$\Pr_{\hat\mu}(X \neq \varnothing) = \frac{Z'(\hat\mu) - 1}{Z'(\hat\mu)} \tag{31}$$

is easy to evaluate directly. It is enough, therefore, to estimate the expectation of f with respect to the distribution of *conditional* probabilities, in which each non-trivial configuration $X \neq \varnothing$ occurs with probability $(Z'(\hat\mu)-1)^{-1}\prod_{\{i,j\}\in X}\lambda_{ij}$. This conditional distribution may be sampled without recourse to Markov chain simulation, the direct method being as follows. Start with $X = \varnothing$ and perform a sequence of m trials, each trial determining whether a particular edge is to be added to X. The probability governing each trial has one of two values, depending on whether any of the previous trials have contributed an edge to X. Define $p_{ij} = \lambda_{ij}(1+\lambda_{ij})^{-1}$, and suppose that we are about to decide whether to add a certain edge $\{i,j\}$ to X. If $X \neq \varnothing$, the edge $\{i,j\}$ is added to X with probability p_{ij}; otherwise, the edge is added to X with probability $p_{ij}[1 - \prod(1-p_{hk})]^{-1}$, where the product is over all edges $\{h,k\}$ whose fate is yet to be decided, including edge $\{i,j\}$ itself. It is straightforward to verify that this procedure yields the required distribution.

It only remains to verify that a polynomially bounded number of sample configurations suffice to provide an accurate estimate of the expectation of f. Again, we do this by demonstrating an upper bound on the ratio between the maximum of f and the mean of f. Observe that

$$Z'(\hat{\mu}) = \prod(1+\lambda_{ij}) \le \exp\left(\sum \lambda_{ij}\right) \le 1 + 2\sum \lambda_{ij}, \tag{32}$$

and

$$\Pr_{\hat{\mu}}(|\mathrm{odd}(X)| = 2) \ge \frac{1}{Z'(\hat{\mu})} \sum \lambda_{ij}, \tag{33}$$

where in equation (32) we have used the fact that $\mathrm{e}^x \le 1 + 2x$ in the range $0 \le x \le 1$. Combining equations (31) and (33), we obtain a lower bound on the probability of $|\mathrm{odd}(X)| = 2$ conditional on $X \ne \varnothing$:

$$\Pr_{\hat{\mu}}\Big(|\mathrm{odd}(X)| = 2 \;\Big|\; X \ne \varnothing\Big) = \frac{\Pr_{\hat{\mu}}(|\mathrm{odd}(X)| = 2)}{\Pr_{\hat{\mu}}(X \ne \varnothing)} \ge \frac{Z'(\hat{\mu})}{Z'(\hat{\mu}) - 1} \frac{1}{Z'(\hat{\mu})} \sum \lambda_{ij} \ge \frac{1}{2};$$

the final step here relies on inequality (32). It follows immediately that the expectation of f with respect to the conditional distribution is at least μ^2. Using the crude bound $\max(f) \le n\mu^2$, we see that the ratio of the maximum to the mean of f is bounded above by n. This completes the analysis of Case II.

We conclude by analysing the time complexity of the proposed approximation scheme for $\mathcal{M}$. The worst case is realised by Case Ib, where our method demands that the three quantities appearing on the right hand side of identity (30) be known with sufficiently high accuracy. To obtain an fpras for $\mathcal{M}$, it is enough to estimate each of these three quantities within ratio $1 + \epsilon/4$ and with failure probability $1/12$. Setting $\xi = \epsilon/4$ and $\eta = 1/12$ in Lemma 3, we see that $\mathrm{O}(\epsilon^{-2}n^3)$ samples from the generator suffice to estimate $\mathbf{E}_{\hat{\mu}}(f)$ within ratio $1 + \epsilon/4$ and with failure probability $1/12$. The production of each sample requires time $\mathrm{O}(m^2\hat{\mu}^{-8}(\log\delta^{-1} + m))$, where $\delta^{-1} = \mathrm{O}(\epsilon^{-1}n^3)$. We may assume that $\epsilon \ge 2^{-m}$, otherwise there would be time enough to evaluate $\mathcal{M}$ exactly using a brute force algorithm. With this simplifying assumption, and noting that $\hat{\mu} = n^{-1}$, the time to produce each sample is seen to be $\mathrm{O}(m^3n^8)$, and the total time to estimate $\mathbf{E}_{\hat{\mu}}(f)$ is $\mathrm{O}(\epsilon^{-2}m^3n^{11})$. The overall execution time for the algorithm is thus dominated by the time taken to estimate $Z'(\mu)$ and $Z'(\hat{\mu})$ within ratio $1 + \epsilon/4$ and with failure probability $1/12$; from remark (i) following Theorem 5 this dominant term is seen to be $\mathrm{O}(\epsilon^{-2}m^3n^{11}\log n)$. □

We turn now to the mean energy $\mathcal{E}$. Up to this point, we have always sampled configurations with μ set to some value which is at least n^{-1}. In the sequel, we shall need to sample configurations at smaller values of μ; at these values Theorem 2 no longer guarantees an execution time for the sampling procedure which is polynomial in n. However, efficient sampling is possible, even at $\mu = 0$.

Theorem 10 *There exists a generator for subgraphs-world configurations which, on inputs* $\langle \lambda_{ij}, \mu \rangle$ *and* δ*, runs in* expected *time bounded by a polynomial in* n *and* $\log \delta^{-1}$*. Specifically, the expected execution time of the generator is* $\mathrm{O}(m^2 n^8 (\log \delta^{-1} + m))$*.*

Proof The result for $\mu \geq n^{-1}$ follows directly from Theorem 2. So assume that $\mu < n^{-1}$. We employ the generator of Theorem 2 but with μ artificially boosted to $\hat{\mu} = n^{-1}$, and δ decreased to $\hat{\delta} = \delta/10$. To sample a configuration X from the distribution corresponding to the *original* value of μ, perform a sequence of trials of the following form. First produce a random configuration X using the generator of Theorem 2 (with the modified parameters); then, with probability $(\mu/\hat{\mu})^{|\mathrm{odd}(X)|}$, declare the trial successful, otherwise declare the trial a failure. The sequence of trials is halted at the first successful trial, and the configuration X produced by that trial is returned as result.

The probability that a trial is declared successful is at least $\Pr_{\hat{\mu}}(|\mathrm{odd}(X)| = 0)$, which by Lemma 4 (setting $\mu_1 = 0$ and $\mu_0 = n^{-1}$) is at least $1/10$. Thus the expected number of trials required to generate a configuration is at most 10. It is straightforward to check that the procedure described above, viewed as a generator with respect to the *original* value of μ, has tolerance at most $10\hat{\delta} = \delta$. $\square$

As in the case of the mean magnetic moment, the main result rests on a technical lemma, whose proof is deferred.

Lemma 11 *Suppose* $B = 0$*, i.e., that there is no external field. If* $w(X) \leq 1/32nm^2$ *for all* $X \neq \emptyset$*, then* $Z' = \sum_X w(X) \leq m$*.*

Theorem 12 *There exists an fpras for the (negation of the) mean energy* $-\mathcal{E} = \partial(\ln Z)/\partial\beta$*, where* Z *is the partition function of a ferromagnetic Ising system.*

Proof We shall assume that $B = 0$, i.e., that there is no external field; the proof in the general case introduces extra technical complications, but requires no new ideas. At the end of the proof, we sketch the modifications required to deal with a non-zero external field.

When $B = 0$ the partition function, Z, simplifies to $Z = A \sum_X w(X)$, where

$$A = \prod_{\{i,j\} \in E} \cosh \beta V_{ij}, \quad w(X) = \prod_{\{i,j\} \in X} \tanh \beta V_{ij},$$

and the sum is over all closed $X \subseteq E$. (A graph X is said to be *closed* if every vertex of X has even degree.) Define

$$c = \sum_{\{i,j\} \in E} V_{ij} \tanh \beta V_{ij}, \quad g(X) = \sum_{\{i,j\} \in X} 2V_{ij}/\sinh 2\beta V_{ij},$$

and let $f(X) = c + g(X)$. Then

$$\begin{aligned}\frac{\partial}{\partial\beta} w(X) &= \frac{\partial}{\partial\beta} \prod_{\{i,j\}\in X} \tanh \beta V_{ij} \\ &= \prod_{\{i,j\}\in X} \tanh \beta V_{ij} \sum_{\{i,j\}\in X} 2V_{ij}/\sinh 2\beta V_{ij} \\ &= w(X)g(X).\end{aligned}$$

Using this identity, and starting from the expansion given in Theorem 1:

$$\begin{aligned}-\mathcal{E} = \frac{\partial}{\partial\beta}(\ln Z) &= \frac{\partial}{\partial\beta}\ln A + \frac{\partial}{\partial\beta}\ln Z' \\ &= \frac{\partial}{\partial\beta}\sum_{\{i,j\}\in E} \ln\cosh \beta V_{ij} + \frac{1}{Z'}\frac{\partial}{\partial\beta}\sum_X w(X) \\ &= \sum_{\{i,j\}\in E} V_{ij}\tanh\beta V_{ij} + \frac{1}{Z'}\sum_X w(X)g(X) \\ &= \frac{1}{Z'}\sum_X w(X)(c + g(X)).\end{aligned}$$

Thus, the mean energy can be expressed as an expectation: $-\mathcal{E} = \mathbf{E}(f)$. This expression for $-\mathcal{E}$ immediately suggests that we attempt to estimate the mean energy by taking an average of $f(X)$ over some polynomially bounded set of sample configurations X. This basic idea can be made to work, with a little refinement. As before, we proceed by case analysis.

Case I: $\beta V_{ij} > 1$ for some pair i, j. The condition guarantees that the constant c is not too small in relation to $g(X)$, and hence that the ratio $\max(f)/\mathbf{E}(f)$ is not too large. This, as we have seen, implies that $\mathbf{E}(f)$ can be estimated by direct experiment. First note that the existence of a pair i, j with $\beta V_{ij} > 1$ entails

$$c \geq V_{ij}\tanh\beta V_{ij} > \frac{3V_{ij}}{4} > \frac{3}{4\beta}. \tag{34}$$

Then observe that the inequality $x/\sinh x < 1$, valid for $x > 0$, implies

$$g(X) = \sum_{\{i,j\}\in X} 2V_{ij}/\sinh 2\beta V_{ij} < \sum_{\{i,j\}\in X}\frac{1}{\beta} \leq \frac{m}{\beta}.$$

Thus $\max(f) = c + \max(g) \leq c + m/\beta \leq c(1 + 4m/3)$, where the final step relies on inequality (34). Since $\mathbf{E}(f)$ is certainly bounded below by c, it follows that $\max(f)/\mathbf{E}(f) \leq 1 + 4m/3 \leq 7m/3$.

Case II: $\beta V_{ij} \leq 1$ for all i, j. To estimate $\mathbf{E}(f)$ within ratio $1+\epsilon$ it is enough, since c is positive, to estimate $\mathbf{E}(g) = \mathbf{E}(f) - c$ within ratio $1+\epsilon$. For simplicity, we shall, in the sequel, work with g instead of f. Using the bounds $1/2 < x/\sinh x < 1$, valid for x in the range $0 < x \leq 2$, we have $1/2\beta < 2V_{ij}/\sinh 2\beta V_{ij} < 1/\beta$, implying

$$1/2\beta < g(X) < m/\beta, \quad \text{for } X \neq \varnothing. \tag{35}$$

Let $C \neq \varnothing$ be a closed subgraph which maximises $w(C)$. Note that C is necessarily a cycle, and can be found in polynomial time using a standard shortest paths algorithm. There are two subcases, depending on the magnitude of $w(C)$.

Case IIa: $w(C) \geq 1/32nm^2$. In this subcase we may estimate $\mathbf{E}(g)$ by direct experiment. Since $Z' \geq w(\varnothing) + w(C) = 1 + w(C)$, it follows that

$$\Pr(X = \varnothing) = \frac{w(\varnothing)}{Z'} = \frac{1}{Z'} \leq (1 + w(C))^{-1} \leq 1 - \frac{1}{64nm^2}.$$

Combining this inequality with inequality (35), we obtain $\mathbf{E}(g) > (1/2\beta)\Pr(X \neq \varnothing) \geq 1/128\beta nm^2$. Then a further application of inequality (35) yields the required bound: $\max(g)/\mathbf{E}(g) \leq 128nm^3$.

Case IIb: $w(C) < 1/32nm^2$. The essential problem in this case is that we have no lower bound on the expectation of g. The solution is to increase the expectation artificially, by adjusting the weight function w. Naturally, a new weight function induces a new probability distribution on configurations, which in turn alters the expectation of g. It is therefore important to adjust the weights systematically, so that it is possible to recover the expectation of g with respect to the *original* distribution from the knowledge of the expectation of g with respect to the *new* distribution.

The new weight function w_α is parameterised by a real number α in the range $0 \leq \alpha \leq 1$. We define adjusted edge weights $\lambda_{ij}^{(\alpha)} = \lambda_{ij}^{1-\alpha}$ which induce a new weight function w_α:

$$w_\alpha(X) = \prod_{\{i,j\}\in X} \lambda_{ij}^{(\alpha)} = \prod_{\{i,j\}\in X} \lambda_{ij}^{1-\alpha} = w(X)^{1-\alpha}.$$

Note that the original weight function corresponds to setting $\alpha = 0$, i.e., $w(X) = w_0(X)$. Note also that $0 < \lambda_{ij}^{(\alpha)} \leq 1$, so the new edge weights $\lambda_{ij}^{(\alpha)}$ correspond to a well defined subgraphs-world process; as a consequence, it is possible to sample configurations according to the distribution which assigns probability $w_\alpha(X)/Z'_\alpha$ to configuration X, where $Z'_\alpha = \sum_X w_\alpha(X)$. Let $\mathbf{E}_\alpha(\cdot)$ denote the expectation operator with respect to the new distribution, i.e., $\mathbf{E}_\alpha(h) = (Z'_\alpha)^{-1}\sum_X w_\alpha(X)h(X)$.

Now fix α so that $w_\alpha(C) = 1/32nm^2$; the required value of α is the solution to the equation $w(C)^{1-\alpha} = 1/32nm^2$, and lies in the range $0 < \alpha < 1$. For any $X \neq \varnothing$, maximality of C implies $w(C) \geq w(X)$, which in turn implies $w_\alpha(C) = w(C)^{1-\alpha} \geq w(X)^{1-\alpha} = w_\alpha(X)$; thus C remains a maximum weight non-trivial configuration under the new weight function w_α. Now the quantity we wish to evaluate, namely $\mathbf{E}(g)$, can be written as an expectation with respect to the new distribution:

$$\begin{aligned}\mathbf{E}(g) = \frac{1}{Z'}\sum_X w(X)g(X) &= \frac{1}{Z'}\sum_X w_\alpha(X)w(X)^\alpha g(X) \\ &= \frac{Z'_\alpha}{Z'}\,\mathbf{E}_\alpha(w(X)^\alpha g(X)). \end{aligned} \tag{36}$$

Since Z' and Z'_α can be computed by the fpras of Section 3, we merely have to show that $\mathbf{E}_\alpha(w(X)^\alpha g(X))$ can be approximated in polynomial time. As before, this can be achieved by bounding the ratio of the maximum to the mean.

By maximality of C and inequality (35),

$$\max_X(w(X)^\alpha g(X)) \le w(C)^\alpha\,\frac{m}{\beta}.$$

Also, by Lemma 11 and inequality (35),

$$\begin{aligned}\mathbf{E}_\alpha(w(X)^\alpha g(X)) &\ge \frac{w_\alpha(C)}{Z'_\alpha}(w(C)^\alpha g(C)) \\ &\ge \frac{1}{32nm^3}\,\frac{w(C)^\alpha}{2\beta} = \frac{w(C)^\alpha}{64\beta nm^3}.\end{aligned}$$

Putting these inequalities together, we obtain

$$\frac{\max(w(X)^\alpha g(X))}{\mathbf{E}_\alpha(w(X)^\alpha g(X))} \le 64nm^4.$$

This completes the analysis of Case IIb.

It is this final case which determines the execution time of the proposed fpras for $-\mathcal{E}$. Consider the cost of evaluating the three factors appearing in (36), within ratio $1+\epsilon/4$ and with failure probability $1/16$. From Section 3, $\mathrm{O}(\epsilon^{-2}m^3n^{11}\log n)$ time suffices to obtain satisfactory estimates for the factors Z' and Z'_α. Setting $\xi = \epsilon/4$ and $\eta = 1/16$ in Lemma 3, we see that $\mathrm{O}(\epsilon^{-2}nm^4)$ samples from the generator suffice to evaluate the remaining factor, $\mathbf{E}_\alpha(w(X)^\alpha g(X))$, within ratio $1+\epsilon/4$ and with failure probability $1/16$. By Theorem 10, the generation of each sample requires expected time $\mathrm{O}(m^3n^8)$. (As before, we are justified in assuming that $\epsilon \ge 2^{-m}$.) Thus the expected time required to obtain a sufficiently accurate estimate of $\mathbf{E}_\alpha(w(X)^\alpha g(X))$ is $\mathrm{O}(\epsilon^{-2}m^7n^9)$; it can be seen that this term dominates those which arise in the other steps of the algorithm. Now set a definite upper bound on the aggregated execution time of the generator which is 16 times the *expected* execution time. By this means the quoted *average case* time-complexity is converted into a *worst case* time complexity, at the cost of introducing an additional failure probability of $1/16$. This additional failure probability may be absorbed into the overall failure probability of $1/4$ which the definition of fpras allows.

Finally, we sketch the modifications required to handle the case of non-zero external field, i.e., $B > 0$. Starting with the general subgraphs-world expansion for Z, and differentiating $\ln Z$ with respect to β, the mean energy $-\mathcal{E}$ can again be expressed as the expectation of an appropriately defined random variable $f(X)$. Naturally, the form of f is now more complex. Lemma 11 continues to hold, but with $1/64n^2m^2$ replacing $1/32nm^2$ as the upper bound on $w(X)$. The proof of Lemma 11 increases in technical complexity, but the main idea remains as before. The complications arise from the fact that the maximum weight non-trivial subgraph may be either a cycle (as before) or a single edge (previously excluded). The case analysis in the proof of the theorem proceeds as before, but the upper bound on overall execution time rises to $O(\epsilon^{-2}m^7n^{10})$. $\square$

In this section, we have presented fully polynomial randomised approximation schemes for the first derivatives of $\ln Z$ with respect to β and B. The *second* derivatives of $\ln Z$ also have physical significance: $\mathcal{C} = k\beta^2\,\partial^2(\ln Z)/\partial\beta^2$ is the *specific heat*, and $\mathcal{X} = \beta^{-1}\,\partial^2(\ln Z)/\partial B^2$ the *magnetic susceptibility*. (Here, k denotes Boltzmann's constant.) It is natural to ask whether the techniques presented in this section can be extended to these quantities. With a certain amount of computational effort, it is possible to express $\mathcal{C}$ and $\mathcal{X}$ as expectations of appropriate random variables in the subgraphs-world. Unfortunately, however, these random variables are not necessarily positive, and the proof techniques of Theorems 9 and 12 are therefore not applicable. At present, the question of the existence of an fpras for $\mathcal{C}$ and $\mathcal{X}$ remains open.

We close the section by presenting proofs of the technical lemmas.

Proof of Lemma 8 We demonstrate, by a simple mapping argument, that

$$\Pr(|\mathrm{odd}(X)| = 2) \geq \mu^2 \Pr(|\mathrm{odd}(X)| = 0). \tag{37}$$

Let Ω_k denote the set $\{X \in \Omega : |\mathrm{odd}(X)| = 2k\}$. Associate with each configuration $X \in \Omega_0$ the set $S(X) = \{X' \in \Omega : |X' \oplus X| = 1\} \subseteq \Omega_1$. It is straightforward to verify that the the subsets $\{S(X) : X \in \Omega_0\}$ are pairwise disjoint, and that $\sum_{X' \in S(X)} w(X') \geq w(X)\mu^2$ for all $X \in \Omega_0$. (For $X = \varnothing$ we need the condition $\sum \lambda_{ij} \geq 1$.) Thus $\sum_{X\in\Omega_1} w(X) \geq \mu^2 \sum_{X\in\Omega_0} w(X)$, and inequality (37) follows by dividing through by Z'.

It follows from inequality (37) that $\Pr(|\mathrm{odd}(X)| > 0) \geq \mu^2/(1+\mu^2) \geq \mu^2/2$; this deals with the first part of the lemma. Furthermore, Lemma 4 assures us that $\Pr(|\mathrm{odd}(X)| = 0) \geq 1/10$ whenever $\mu \leq n^{-1}$. Combining this observation with inequality (37) yields the second part of the lemma. $\square$

Proof of Lemma 11 Since $B = 0$, it is only the closed subgraphs $X \subseteq E$ which have non-zero weight: $w(X) = \prod_{\{i,j\}\in X} \lambda_{ij}$. Let X_0, X_1, $X_2, \ldots,$ X_{s-1} be an enumeration of the closed subgraphs of E in order of non-increasing weight; thus $X_0 = \emptyset$, $w(X_0) = 1$, and $w(X_1) \leq 1/32nm^2$. For each edge $e = \{i,j\} \in E$ define $l(e)$, the *length* of e, to be $-\ln \lambda_{ij}$. Extend the length function to subsets of E by summation, so that $l(X) = -\ln w(X)$. (Clearly, these "lengths" are merely weights which combine additively rather than multiplicatively. Even so, they will prove to be a convenient notational and conceptual aid.)

Define $L = l(X_1) = -\ln w(X_1)$. Let $C = (e_1, e_2, \cdots, e_r)$ be any circuit in E; here, each $e_i = \{v_{i-1}, v_i\}$ is an edge, and $v_r = v_0$. Define $d_0 = 0$, and $d_k = \sum_{i=1}^{k} l(e_i)$ for $1 \leq k \leq r$. Call a directed edge $e_k = (v_{k-1}, v_k)$ of C a *pier* if there exists an integer h with $d_{k-1} < hL/2 \leq d_k$. We make two observations about piers. First, the circuit C is completely determined by the start vertex v_0 and the set of all piers. This is because the total length of edges in C which form a span between two consecutive piers is strictly less than $L/2$. Thus, the existence of two distinct spans between consecutive piers would imply the existence of a circuit of length less than L, and hence of a non-trivial closed subgraph of weight greater than $\mathrm{e}^{-L} = w(X_1)$; this would contradict the assumption that X_1 is maximal. Second, the total number of piers in C is at most $2d_r/L = 2l(C)/L$. Intuitively, the role of piers is to provide a compact encoding of circuits.

Now suppose $\alpha \geq 0$, and let $X \subseteq E$ be a general closed subgraph with $l(X) \leq \alpha L$. Decompose X into its connected components; each of these components is Eulerian and hence can be regarded as a circuit with specified start vertex. Encode each component of X as a sequence consisting of the start vertex of the circuit followed by the piers of the circuit in sequence. Encode X itself by concatenating the codes for the various components; note that X is completely determined by the code so formed. Since each connected component of X has length at least L, the total number of vertices in the code (which is equal to the number of components) is at most $l(X)/L = \alpha$. Furthermore, the total number of directed edges in the code (which is the total number of piers) is at most $2l(X)/L = 2\alpha$. These observations yield an upper bound on the number of distinct codes for closed X with $l(X) \leq \alpha L$. The number of ways of selecting a sequence of at most α distinct vertices is bounded by n^α; that of selecting a sequence of at most 2α distinct directed edges by $(2m)^{2\alpha}$; that of interleaving the vertex and edge sequences by $2^{3\alpha}$. Thus the number of distinct codes, and hence the number of closed X with $l(X) \leq \alpha L$, is bounded above by $(32nm^2)^\alpha$.

Now consider the subsequence X_0, $X_1, \ldots,$ X_{k-1}, consisting of the k closed subgraphs of greatest weight (i.e., shortest length), and let $\alpha = l(X_{k-1})/L$. Then the coding argument implies $(32nm^2)^\alpha \geq k$. On the other hand, from the definition of α, and using the bound on $w(X_1)$ guaranteed in the statement of the lemma,

$$(32nm^2)^\alpha \le \frac{1}{w(X_1)^\alpha} = \frac{1}{w(X_{k-1})}.$$

Combining these two inequalities we obtain $w(X_{k-1}) \le k^{-1}$, and hence

$$Z' = \sum_{k=1}^{s} w(X_{k-1}) \le \sum_{k=1}^{s} \frac{1}{k} \le \sum_{k=1}^{2^m} \frac{1}{k}.$$

By a well known asymptotic result [21, p. 74], the latter sum is bounded by m for all sufficiently large m; indeed it is enough that $m \ge 3$. The lemma holds trivially for $m < 3$. $\square$

6 Completeness Results

In this paper we have restricted our attention to the *ferromagnetic* case of the Ising model; moreover we have contented ourselves with solutions which are *approximate* only. The results of this section aim to justify these apparently limited goals. Since we are concerned here with *negative* results, it will be an advantage to work with a simplified version of the Ising problem.

> INSTANCE: A symmetric matrix $(V_{ij} : i, j \in [n])$ with entries in $\{-1, 0, +1\}$, and a natural number, β, presented in unary notation.
>
> OUTPUT: The partition function $Z = Z(V_{ij}, \beta) = \sum_{\boldsymbol{\sigma}} 2^{-\beta H(\boldsymbol{\sigma})}$, where $H(\boldsymbol{\sigma}) = -\sum_{\{i,j\}\in E} V_{ij}\sigma_i\sigma_j$, and the sum is over $\boldsymbol{\sigma} \in \{-1, +1\}^n$. (As usual, E is the set of pairs $\{i, j\}$ with $V_{ij} \ne 0$.)

We refer to the problem in this form as ISING. The main points to note are that the external field is zero, and that powers of e have been replaced by powers of 2. The latter modification merely amounts to a scaling of β, and is made to avoid problems which would arise from the introduction of real arithmetic. The restrictions imposed on the various quantities appearing in an instance of ISING ensure that the output is a rational number whose numerator is a binary integer with a polynomially bounded number of digits, and whose denominator is a certain power of two. The output can thus be considered as a fixed-point binary number with an explicit "binary point". Adopting this viewpoint, it is not difficult to locate ISING within the class #P of combinatorial enumeration problems. (See Garey and Johnson's book [12, p. 167] for an explanation of #P and its completeness class.)

The two problems which form the starting point for the intractability results of this section are MAXCUT:

INSTANCE: An undirected graph G and a positive integer b.

QUESTION: Is there a cut-set in G of size b? I.e., is there a partition of the vertex set of G into two subsets such that the number of edges which span the two subsets is at least b?

and the related #MAXCUT:

INSTANCE: An undirected graph G.

OUTPUT: The number of cut-sets in G of maximum size.

The following is a slight extension of a known result.

Lemma 13 MAXCUT *is* NP-*complete, and* #MAXCUT *is* #P-*complete.*

Proof NP-completeness of MAXCUT is proved in [13]. The reductions used there are not "parsimonious" [12, p. 169], and hence do not immediately imply #P-completeness of #MAXCUT. As usual, however, the reductions (given in the proofs of Theorems 1.1 and 1.2 of that paper) can be modified, without great difficulty, to yield parsimonious versions. For those who wish to follow the argument in detail, the necessary modifications are presented below.

In [13, Theorem 1.1], new variables $\{e_i : 1 \le i \le m\}$ should be introduced, and the definition of the clause set S' amended to read

$$\begin{aligned} S' = \bigcup_{i=1}^{m} \Big\{ &(d_i \vee \neg e_i), \\ &(a_i), (b_i), (c_i), (d_i), (e_i), \\ &(a_i \vee b_i), (a_i \vee c_i), \ldots, (d_i \vee e_i), \\ &(\neg a_i \vee \neg b_i), (\neg a_i \vee \neg c_i), \ldots, (\neg d_i \vee \neg e_i) \Big\}, \end{aligned}$$

where each ellipsis stands for seven omitted disjuctions. Note that there are 26 clauses in S' arising from the ith clause, $(a_i \vee b_i \vee c_i)$, of S. If the ith clause of S is satisfied, then there is precisely one way to choose truth values for the variables d_i and e_i so that 20 of these 26 clauses of S' are satisfied. Conversely, if the ith clause of S is not satisfied then, however d_i and e_i are chosen, at most 19 of the 26 clauses can be satisfied. Thus, setting $k = 20m$, the original proof goes through. Note that the reduction is now parsimonious.

In [13, Theorem 1.2], duplicate clauses should be removed by replacing each clause $C_i' = (u_i \vee v_i)$ by the seven clauses

$$(u_i \vee \neg c_i), (\neg u_i \vee c_i), (u_i \vee \neg d_i), (\neg u_i \vee d_i), (c_i \vee \neg d_i), (\neg c_i \vee d_i), \text{ and } (c_i \vee v_i),$$

where c_i and d_i are new variables. For a given assignment to u_i and v_i, one must set $c_i = d_i = u_i$ in order to maximise the total number of satisfied clauses within these seven. Now, if C_i' is satisfied then all seven clauses may be satisfied; however, if C_i' is not satisfied then at most six of the clauses may be satisfied. The existing proof goes through with k set to $k' + 6q$. Again, the modified reduction is parsimonious. $\square$

The first theorem of the section presents evidence that our restriction to the ferromagnetic case of the Ising model cannot be relaxed.

Theorem 14 *There can be no fpras for* Ising *unless* RP = NP.

RP is the class of decision problems which can be solved in polynomial time by a certain type of randomised algorithm which is allowed one-sided errors. (See [2, p. 138] for a precise definition.) It is widely conjectured that RP is strictly contained in NP. Thus Theorem 14 can be interpreted as strong evidence that no approximation algorithm exists for the Ising partition function in the non-ferromagnetic, or "spin-glass" case. Indeed, the existence of such an algorithm would imply, by virtue of Theorem 14, the existence of efficient randomised algorithms for such hard problems as testing satisfiability of a Boolean formula and the Travelling Salesman Problem.

Proof of Theorem 14 Let $G = ([n], E)$ be a graph and b a positive integer defining an instance of MaxCut. Construct an instance of Ising by setting $\beta = n$, and $V_{ij} = -1$ when $\{i,j\} \in E$. (As usual, $V_{ij} = 0$ when $\{i,j\} \notin E$.) Each spin-vector $\boldsymbol{\sigma}$ partitions $[n]$ into two subsets, and hence defines a cut-set of G: $\mathrm{cut}(\boldsymbol{\sigma}) = \{\{i,j\} \in E : \sigma_i\sigma_j = -1\}$. Note that $H(\boldsymbol{\sigma}) = m - 2|\mathrm{cut}(\boldsymbol{\sigma})|$, where $m = |E|$. Let N_k denote the number of spin-vectors $\boldsymbol{\sigma}$ for which $|\mathrm{cut}(\boldsymbol{\sigma})| = k$. Then the simplified partition function can be re-expressed as $Z = \sum_{k=0}^{m} N_k 2^{\beta(2k-m)}$.

Note that if a cut-set of size b exists in G then $Z \geq 2^{\beta(2b-m)} = 2^{n(2b-m)}$; in contrast, if no such cut-set exists, $Z \leq 2^n 2^{\beta(2b-2-m)} = 2^{-n} 2^{n(2b-m)}$. Now the existence of an fpras for Ising would imply that these two cases could be distinguished in polynomial time, with failure probability at most $\frac{1}{4}$; in other words, MaxCut $\in$ BPP. From this it would follow—since MaxCut is NP-complete and BPP is closed under polynomial time reductions—that NP $\subseteq$ BPP. However, the inclusion NP $\subseteq$ BPP entails RP = NP [22]. $\square$

Our final theorem states that Ising is a complete problem for the class #P. Thus a polynomial time algorithm which solved it *exactly* would yield similar algorithms for a range of presumably intractable problems, such as counting the number of satisfying assignments of a Boolean formula and counting optimal Travelling Salesman tours. We should therefore not be too disappointed that we have obtained only *approximation* algorithms for the Ising problem.

Theorem 15 ISING *is* #P*-complete even when the matrix* V_{ij} *is non-negative (i.e., even in the ferromagnetic case).*

Proof We present an easy polynomial-time (Turing) reduction from #MAXCUT. Let $G = ([n], E)$ be an instance of #MAXCUT. Set $V_{ij} = +1$ when $\{i,j\} \in E$, and $V_{ij} = 0$ otherwise. Note that $H(\boldsymbol{\sigma}) = 2|\mathrm{cut}(\boldsymbol{\sigma})| - m$. With N_k as before we have $Z = \sum_{k=0}^{m} N_k 2^{\beta(m-2k)} = 2^{\beta m} p(4^{-\beta})$, where $p(x) = \sum_k N_k x^k$ is a polynomial of degree m. Suppose that the value of p is known at the points $\beta = 0, 1, \ldots, m$, i.e, at $x = 1, 4^{-1}, 4^{-2}, \ldots, 4^{-m}$. Then the coefficients of p can be recovered in polynomial time from these values by interpolation. Using Newton's formula, this process can be carried out using only rational arithmetic; moreover, the lengths of the numerators and denominators all remain polynomially bounded. The leading (non-zero) coefficient of p is twice the number of maximum cut-sets in G. (Note that each bipartition of $[n]$ corresponds to a pair of opposite spin-vectors.) $\square$

Further completeness results related to the ones in this section can be found in [3, 15].

References

[1] Alon, N., Eigenvalues and expanders, *Combinatorica* **6** (1986), pp. 83–96.

[2] Balcázar, J. L., Díaz, J. and Gabarró, J., *Structural Complexity, Volume I,* Springer-Verlag, Berlin, 1988.

[3] Barahona, F., On the computational complexity of Ising spin glass models, *Journal of Physics A: Mathematical & General* **15** (1982), pp. 3241–3253.

[4] Binder, K., Monte Carlo investigations of phase transitions and critical phenomena, in *Phase Transitions and Critical Phenomena, Volume 5b* (C. Domb and M. S. Green eds.), Academic Press, London, 1976, pp. 1–105.

[5] Cheeger, J., A lower bound for the smallest eigenvalue of the Laplacian, In *Problems in Analysis* (R. C. Gunning ed.), Princeton University Press, New Jersey, 1970, pp. 195–199.

[6] Cipra, B., An introduction to the Ising model, *American Mathematical Monthly* **94** (1987), pp. 937–959.

[7] Diaconis, P., and Stroock, D., Geometric bounds for eigenvalues of Markov chains, *Annals of Applied Probability* **1** (1991), pp. 36–61.

[8] Dodziuk, J., Difference equations, isoperimetric inequality and transience of certain random walks, *Transactions of the American Mathematical Society* **284** (1984), pp. 787–794.

[9] Dyer, M., Frieze, A., and Kannan, R., A random polynomial time algorithm for approximating the volume of convex bodies, *Proceedings of the 21st ACM Symposium on Theory of Computing,* 1989, pp. 375–381.

[10] Feller, W., *An introduction to probability theory and its applications, Volume I* (3rd edition), John Wiley, New York, 1968.

[11] Fisher, M. E., On the dimer solution of planar Ising models, *Journal of Mathematical Physics* **7** (1966), pp. 1776–1781.

[12] Garey, M. R. and Johnson, D. S., *Computers and Intractability: a Guide to the Theory of NP-completeness,* Freeman, San Francisco, 1979.

[13] Garey, M. R., Johnson, D. S., and Stockmeyer, L., Some simplified NP-complete problems, *Theoretical Computer Science* **1** (1976), pp. 237–267.

[14] Ising, E., Beitrag zur Theorie des Ferromagnetismus, *Zeitschrift für Physik* **31** (1925), pp. 253–258.

[15] Jerrum, M. R., 2-dimensional monomer-dimer systems are computationally intractable, *Journal of Statistical Physics* **48** (1987), pp. 121–134.

[16] Jerrum, M. R. and Sinclair, A. J., Approximating the permanent, *SIAM Journal on Computing* **18** (1989), pp. 1149–1178.

[17] Jerrum, M. R., Valiant, L. G., and Vazirani, V. V., Random generation of combinatorial structues from a uniform distribution, *Theoretical Computer Science* **43** (1986), pp. 169–188.

[18] Karp, R. M. and Luby, M., Monte-Carlo algorithms for enumeration and reliability problems, *Proceedings of the 24th IEEE Symposium on Foundations of Computer Science,* 1983, pp. 56–64.

[19] Kasteleyn, P. W., Dimer statistics and phase transitions, *Journal of Mathematical Physics* **4** (1963), pp. 287–293.

[20] Kirkpatrick, S., Gelatt, C., and Vecchi, M., Optimisation by simulated annealing, *Science* **220** (May 1983), pp. 671–680.

[21] Knuth, D. E., *The Art of Computer Programming, Volume 1: Fundamental Algorithms,* Addison-Wesley, 1975.

[22] Ker-I Ko, Some observations on probabilistic algorithms and NP-hard problems, *Information Processing Letters* **14** (1982), pp. 39–43.

[23] Lawler, G. F. and Sokal, A. D., Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality, *Transactions of the American Mathematical Society* **309** (1988), pp. 557–580.

[24] Lenz, W., Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern, *Zeitschrift für Physik* **21** (1920), pp. 613–615.

[25] Mohar, B., Isoperimetric numbers of graphs, *Journal of Combinatorial Theory, Series B* **47** (1989), pp. 274–291.

[26] Newell, G. F. and Montroll, E. W., On the theory of the Ising model of ferromagnetism, *Reviews of Modern Physics* **25** (1953), pp. 353–389.

[27] Sinclair, A. J., Randomised Algorithms for Counting and Generating Combinatorial Structures, Ph.D. Thesis, Dept. of Computer Science, University of Edinburgh, June 1988.

[28] Sinclair, A. J., Improved bounds for mixing rates of Markov chains and multicommodity flow, Technical Report ECS-LFCS-91-178, Dept. of Computer Science, University of Edinburgh, October 1991.

[29] Sinclair, A. J. and Jerrum, M. R., Approximate counting, uniform generation, and rapidly mixing Markov chains, *Information and Computation* **82** (1989), pp. 93–133.

[30] Welsh, D. J. A., The computational complexity of some classical problems from statistical physics, in *Disorder in Physical Systems*, Oxford University Press, February 1990, pp. 307–321.