# Lecture 5: An FPRAS for the Ferromagnetic Ising Model

September 21, 2023

In this lecture, we use the canonical paths method from the previous lecture to build an FPRAS for the ferromagnetic Ising partition function.

## 1 The Ferromagnetic Ising Model

Let $G = (V, E)$ be a finite undirected graph, and let $A$ be its adjacency matrix. For a nonnegative parameter $\beta \geq 0$ and a vector $h \in \mathbb{R}$, define the Gibbs measure of the *ferromagnetic Ising model* on $G$ with *external field* $h$ as the following probability distribution over $\{\pm 1\}^V$:

$$\mu(\sigma) \propto \exp\left(\frac{\beta}{2}\sigma^\top A\sigma + h \cdot \langle \sigma, \mathbf{1}\rangle\right), \qquad \forall \sigma \in \{\pm 1\}^V. \tag{1}$$

Its associated partition function is given by

$$Z_G(\beta, h) \overset{\mathsf{def}}{=} \sum_{\sigma \in \{\pm 1\}^V} \exp\left(\frac{\beta}{2}\sigma^\top A\sigma + h \cdot \langle \sigma, \mathbf{1}\rangle\right).$$

This is one of the most famous and well-studied models in statistical mechanics dating back to the early 1900s [Len20; Isi25]. It was developed to mathematically model phase transition behavior in magnets, where we imagine a block of magnetic material as consisting of particles which themselves behave like tiny magnets.[1] As we can see from Eq. (1), particles (represented as vertices) interact with their nearest neighbors in a way which encourages alignment of their $\pm 1$ spins. The parameter $\beta \geq 0$ captures the *inverse temperature*; small $\beta$ corresponds to high temperature, which intuitively means weaker interactions since the particles "fluctuate more". The vector $h \in \mathbb{R}^V$ simulates the effect of external forces and effects.

Our goal is to algorithmically sample from this Gibbs distribution. Intuitively, if the temperature is high, then the interactions are weak and we expect $\mu$ to behave in some sense like the uniform measure over $\{\pm 1\}^V$. Therefore, in this regime of small $\beta$, we expect local Markov chains like Glauber dynamics to mix rapidly. On the other hand, if $\beta$ is large, then the distribution will concentrate mass around the two diametrically opposite maximizers $\mathbf{1}, -\mathbf{1} \in \{\pm 1\}^V$ of the Hamiltonian $H(\sigma) = \frac{\beta}{2}\sigma^\top A\sigma$. Since these maximizers differ in all coordinates, we expect that Glauber dynamics will mix slowly in the regime of large $\beta$.

For most interesting families of graphs, this is the case. Furthermore, there typically is a precise critical threshold $\beta_c$ (depending on the family of graphs under consideration) such that Glauber mixes rapidly if $\beta < \beta_c$, and mixes slowly if $\beta > \beta_c$. In this lecture, we show that despite this algorithmic phase transition, there is an FPRAS nonetheless.

**Theorem 1.1** (Jerrum–Sinclair [JS93])**.** *There is an* FPRAS *for the problem of estimating $Z_G(\beta, h)$, for any input graph $G$ and any $\beta, h \geq 0$.*

*Remark* 1. This result extends much more generally to any nonnegative matrix $A$. We can further replace $h \cdot \mathbf{1}$ with any vector $h \in \mathbb{R}^V$ as long as all entries have the same sign. Such external fields are sometimes called *consistent*, since they bias all entries of $\sigma \sim \mu$ towards the same spin in $\{\pm 1\}$.

[1] This was motivated by a real experiment performed by Pierre Curie, where a magnetic block of iron was gradually heated. It was observed that at some critical temperature, now called the Curie Temperature, the block of material spontaneously lost its magnetization.

## 1.1 The Even Subgraphs Representation

The first step in the proof of Theorem 1.1 is to construct a new distribution on a new state space which has the same partition function as $Z_G(\beta)$ (up to some easy-to-compute factor). For $F \subseteq E$, let $\mathsf{odd}(F) = \{v \in V : \deg_F(v) \text{ odd}\}$ denote the set of odd-degree vertices in the subgraph $(V, F)$; note the number $|\mathsf{odd}(F)|$ of odd-degree vertices must always be even. If $A, B \subseteq V$ are two sets of vertices, we write $A \oplus B$ for their symmetric difference[2]; we use the same notation if $A, B \subseteq E$ are two subsets of edges.

Let $0 \le \rho \le 1$ and $\lambda \ge 0$ be parameters, and define a probability distribution $\widehat{\mu}_{\rho,\lambda}$ over subsets of edges by

$$\widehat{\mu}_{\rho,\lambda}(F) \propto \rho^{|\mathsf{odd}(F)|}\lambda^{|F|}, \qquad \forall F \subseteq E,$$

which has partition function

$$\widehat{Z}_G(\rho,\lambda) \stackrel{\mathsf{def}}{=} \sum_{F \subseteq E} \rho^{|\mathsf{odd}(F)|}\lambda^{|F|}.$$

In the physics literature, this is sometimes called the *high-temperature expansion* for the ferromagnetic Ising model.[3] We will call this the *even subgraphs model*, since the distribution penalizes the presence of odd-degree vertices. The following tells us that to compute the ferromagnetic Ising partition function, it suffices to compute the partition function for the even subgraphs model.

**Proposition 1.2.** *For every $\beta, h \ge 0$, we have the identity $Z_G(\beta, h) = C(\beta, h) \cdot \widehat{Z}_G(\rho, \lambda)$, where $\rho = \tanh(h), \lambda = \tanh(\beta)$ and $C(\beta, h) \stackrel{\mathsf{def}}{=} 2^{|V|}\cosh(h)^{|V|}\cosh(\beta)^{|E|}$.*

*Remark* 2. This transformation is completely general, and works for arbitrary $A \in \mathbb{R}^{n\times n}$. The only catch is that when negative entries are allowed in $A$, then $\widehat{Z}_G$ may contain negative terms, and we no longer get a probability distribution $\widehat{\mu}_{\rho,\lambda}$. Nonetheless, it is still a useful transformation; as an example for going beyond the ferromagnetic case, see the following paper studying the *Sherrington–Kirkpatrick spin glass model* [ALR87].

A proof is provided in Appendix A. The key trick is writing $e^x$ as $\cosh(x)\cdot(1+\tanh(x))$, allowing one to "average out" the identities of the spins $\{\pm 1\}$ in a nice way which "cancels out" many terms. When we discuss the Lee–Yang Theorem, we'll see that this trick $e^x = \cosh(x) \cdot (1 + \tanh(x))$ has a nice complex-analytic interpretation using Möbius transformations (after rewriting $Z_G(\beta, h)$ as the *cut polynomial* of $G$). There is also an interpretation of this transformation as a special case of *cluster expansion*, another technique we'll see later in the course. Unfortunately, the proof doesn't provide a lot of intuition behind the "meaning" of the subgraphs $F \subseteq E$.

For the rest of the lecture, we focus on estimating $\widehat{Z}_G(\rho, \lambda)$ for every $0 \le \rho, \lambda \le 1$. Note that even if we have a sampler for $\widehat{\mu}_{\rho,\lambda}$, it is not clear that we immediately get an approximate counter for $\widehat{Z}_G(\rho, \lambda)$. This is because, at least as it is currently defined, the problem of estimating $\widehat{Z}_G(\rho, \lambda)$ isn't self-reducible in a natural way similar to matchings. For instance, if we condition an edge $e$ to be in the sample $F \sim \widehat{\mu}_{\rho,\lambda}$, then in the resulting conditional distribution for the graph $G - e$, we need to replace $\mathsf{odd}(F)$ with $\mathsf{odd}(F \cup \{e\})$. We could relax the problem definition to include such distributions, but this will make the proofs more cumbersome than necessary. Instead, we use a different *annealing scheme* to show that approximate sampling from $\widehat{\mu}_{\rho,\lambda}$ for all $0 \le \rho, \lambda \le 1$ implies approximate counting for $\widehat{Z}_G(\rho, \lambda)$ for all $0 \le \rho, \lambda \le 1$.

**Proposition 1.3.** *Suppose there exists a* $\mathsf{FPAS}$ *for approximately sampling from $\widehat{\mu}_{\rho,\lambda}$ for every $0 \le \rho, \lambda \le 1$. Then there exists an* $\mathsf{FPRAS}$ *for estimating $\widehat{Z}_G(\rho, \lambda)$ for all $0 \le \rho, \lambda \le 1$.*

Combined with Proposition 1.2, this shows that to prove Theorem 1.1, it suffices to design an $\mathsf{FPAS}$ for approximately sampling from $\widehat{\mu}_{\rho,\lambda}$. The proof is provided in Appendix A. Given Proposition 1.3, our goal is now to construct an approximate sampler for the distribution $\widehat{\mu}_{\rho,\lambda}$. The core algorithm is Glauber dynamics $\mathsf{P}$, which for $\widehat{\mu}_{\rho,\lambda}$, evolves as follows: Suppose at time $t$, we have a set of edges $F_t \subseteq E$. To transition to $F_{t+1} \subseteq E$, we

1. select a uniformly random edge $e \in E$,

[2] The $\oplus$ notation comes from identifying each set $A$ with its corresponding $\{0, 1\}$-indicator function, viewed as a vector in $\mathbb{F}_2^V$.

[3] There are many other useful expansions for the ferromagnetic Ising model (see e.g. [Dum17]).

2. set $F_{t+1} = F_t + e$ with probability
$$\frac{\rho^{|\mathsf{odd}(F_t+e)|}\lambda^{|F_t+e|}}{\rho^{|\mathsf{odd}(F_t-e)|}\lambda^{|F_t-e|} + \rho^{|\mathsf{odd}(F_t+e)|}\lambda^{|F_t+e|}},$$
and set $F_{t+1} = F_t - e$ otherwise.

**Theorem 1.4.** *For every graph $G = (V, E)$ with $n$ vertices and $m$ edges, and every $0 < \rho, \lambda \leq 1$, Glauber dynamics has spectral gap $\gamma \geq \rho^6/m^2$ and hence, mixing time*[4]
$$\mathrm{T}_{\mathsf{mix}}(\epsilon) \leq O\left(\frac{m^2}{\rho^6}(m + \log(1/\epsilon))\right).$$

**Corollary 1.5.** *There exists an* FPAS *for sampling from $\widehat{\mu}_{\rho,\lambda}$ for every graph $G = (V, E)$ and every $0 \leq \rho, \lambda \leq 1$.*

Since the mixing time has an inverse polynomial dependence on $\rho > 0$, Theorem 1.4 doesn't directly give us a sampler for $\rho = 0$. To remedy this, we can take $\rho$ sufficiently small and add an extra rejection sampling trick at the end. The details are provided in Appendix A. In the remainder of the lecture, we prove Theorem 1.4 by constructing a set of canonical paths and bounding the congestion. Towards this, we introduce one more technical ingredient which will make life easy.

## 2 Bounding Congestion via Flow Encodings

The challenge with bounding the congestion $\max_{a\to b} C_{\mathcal{P}}(a \to b)$ is adequately controlling the quantity $\sum_{x,y\in\Omega:(a,b)\in\mathcal{P}_{x\to y}} \mu(x)\mu(y)$. For instance, if $\mu$ is uniform, then $C_{\mathcal{P}}(a \to b) \leq \mathsf{poly}(n)$ essentially boils down to controlling the total number of paths
$$|\mathcal{P}_{x\to y}| \leq \mathsf{poly}(n) \cdot |\Omega|,$$
as we saw in the hypercube example from the previous lecture. But in general, the quantity $|\Omega|$ in the right-hand side is exactly the unknown quantity we wish to compute, so somehow we need to certify this inequality in some "implicit" manner. The following ingenious technique of Jerrum–Sinclair gives such a certificate. An important feature of this certificate is that we can verify it by making "local checks" to it; this will be more clear when we discuss a concrete application to even subgraphs.

**Definition 1** (Flow Encoding)**.** *Let $\mathcal{P} = \{\mathcal{P}_{x\to y}\}_{x,y\in\Omega}$ be a collection of canonical paths. For each transition $a \to b$ w.r.t. the Markov chain* P*, write $\mathsf{CP}_{a,b} = \{(x, y) \in \Omega^2 : \mathcal{P}_{x\to y} \ni (a \to b)\}$.*[5] *A* flow encoding *for $\mathcal{P}$ is a collection of injective maps $\boldsymbol{\eta} = \{\eta_{a\to b} : \mathsf{CP}_{a,b} \to \Omega\}_{a,b}$.*

**Lemma 2.1.** *Let $\boldsymbol{\eta} = \{\eta_{a\to b} : \mathsf{CP}_{a,b} \to \Omega\}_{a,b}$ be a flow encoding for a collection of canonical paths $\mathcal{P} = \{\mathcal{P}_{x\to y}\}_{x,y\in\Omega}$ such that for some $\alpha > 0$, we have the inequality*
$$\mu(x)\mu(y) \leq \alpha \cdot \mu(a)\mu(\eta_{a\to b}(x, y))\mathsf{P}(a \to b) \tag{2}$$
*uniformly for all transitions $a \to b$ in* P *and all $(x, y) \in \mathsf{CP}_{a\to b}$. Then we have the congestion bound*
$$\max_{a\to b} C_{\mathcal{P}}(a \to b) \leq \alpha.$$

*Proof.* Fix an arbitrary transition $a \to b$. Then
$$\begin{aligned}
C_{\mathcal{P}}(a \to b) &= \frac{1}{\mu(a)\mathsf{P}(a \to b)} \sum_{x,y\in\Omega:(a,b)\in\mathcal{P}_{x\to y}} \mu(x)\mu(y) && \text{(Definition of congestion)}\\
&\leq \alpha \sum_{(x,y)\in\mathsf{CP}_{a\to b}} \mu(\eta_{a\to b}(x, y)) && \text{(Flow encoding)}\\
&\leq \alpha. && \text{(Injectivity)}
\end{aligned}$$
□

*Remark* 3. One can relax the injectivity requirement, e.g. allowing for additional "side information", at some cost in the bound $\alpha$.

[4]The $\rho^6$ is for convenience and simplicity; with a little more care, the analysis can be sharpened to make this factor $\rho^4$.

[5]One can think of CP as being shorthand for "customer–producer" or "canonical paths".

## 2.1 Canonical Paths for Even Subgraphs

*Proof of Theorem 1.4.* We construct a collection of canonical paths to certify an upper bound on the inverse spectral gap. To bound the congestion, we then design a flow encoding and apply Lemma 2.1. Let $I, F \subseteq E$ be two subgraphs; we wish to define a canonical path $\mathcal{P}_{I \to F}$ from the initial edge set $I$ to the final edge set $F$. We use Glauber moves to flip the status of each edge of $I \oplus F$ one at a time, similar to the hypercube example from the previous lecture.

However, to achieve good congestion, the order in which we do this matters: We decompose $I \oplus F$ into a pairwise edge-disjoint collection of paths $P_1, \ldots, P_k$ and cycles $C_1, \ldots, C_\ell$, where $2k = |\mathsf{odd}(I \oplus F)|$. This is always possible via an inductive argument, since every odd-degree vertex in $I \oplus F$ must be connected by a path to some other odd-degree vertex. Note that such a decomposition is not unique in general. To each $P_i$, we arbitrarily designate one of the two ending vertices as the "start" vertex. Similarly, to each cycle $C_j$, we arbitrarily designate one of the vertices as the "start" vertex; we also need to specify a neighboring vertex to determine the "direction" we wish to traverse the cycle. These choices give rise to a fixed ordering of the edges $e_1, \ldots, e_t$ of the edges of $I \oplus F$; going in this order, we traverse from start to finish each path $P_1, \ldots, P_k$ one by one, and then traverse each cycle starting with the designated "start" vertex and walking in the designated direction. This then defines a canonical path $\mathcal{P}_{I \to F}$, where we "flip" each edge $e_i$ in order.

Clearly, the lengths of the canonical paths are at most $m$. We now define a flow encoding to bound the congestion. For a transition $A \to B = A \oplus e_i$ and $(I, F) \in \mathsf{CP}_{A \to B}$, define

$$\eta_{A \to B}(I, F) \stackrel{\mathsf{def}}{=} I \oplus F \oplus (A \cup B).$$

We prove the following claims.

- **Injectivity:** We show how to invert $\eta_{A \to B}$. Since $\eta_{A \to B}(I, F) \oplus (A \cup B) = I \oplus F$, we can recover the ordering of the edges $e_1, \ldots, e_t$ used to flip from $I$ to $F$. Furthermore, we can recover the exact $1 \le i \le t$ at which the transition $A \to B$ occurred in this flipping process, since $e_i = A \oplus B$. Hence, we can recover $I$ by starting with $A$ and applying flips to $e_1, \ldots, e_{i-1}$; similarly, we can recover $F$ by starting with $B$ and applying flips to $e_{i+1}, \ldots, e_t$.

- **Eq. (2) holds with $\alpha = m/\rho^6$:** By reversibility of $\mathsf{P}$, since $B = A \oplus e_i$,

$$\mu(A)\mu(\eta_{A \to B}(I, F))\mathsf{P}(A \to B) = \mu(A \cup B)\mu(I \oplus F \oplus (A \cup B))\mathsf{P}(A \cup B \to A \cap B).$$

In other words, we only need to consider the transition from the larger set $A \cup B$ to the smaller set $A \cap B$. In particular, since $|A \cap B| \le |A \cup B|$, we can lower bound the transition probability as

$$\mathsf{P}(A \cup B \to A \cap B) \ge \frac{\rho^2}{m},$$

independent of $\lambda$. Hence, to prove the desired bound on $\alpha$, it suffices to show that

$$\mu(I)\mu(F) \le \rho^{-4} \cdot \mu(\eta_{A \to B}(I, F))\mu(A \cup B).$$

We can cancel $\widehat{Z}_G(\rho, \lambda)^2$ from the denominator in both sides; similarly, the factors of $\lambda$ cancel. Since $0 \le \rho \le 1$, it thus suffices to show that for any set of edges $T$ we encounter in the canonical path $\mathcal{P}_{I \to F}$ (e.g. $A \cup B$), then

$$4 + |\mathsf{odd}(I)| + |\mathsf{odd}(F)| \ge |\mathsf{odd}(T)| + |\mathsf{odd}(I \oplus F \oplus T)|.$$

Let us first consider the following special case.

**Claim 2.2.** *Suppose $T \subseteq E$ is obtained from $I$ by completely traversing some of the paths and cycles, and hasn't started traversing the next path/cycle. In other words, $T$ is of the form $T = I \oplus P_1 \oplus \cdots \oplus P_j$ or $T = I \oplus P_1 \oplus \cdots \oplus P_k \oplus C_1 \oplus \cdots \oplus C_j$ for some $j$. Then*

$$|\mathsf{odd}(I)| + |\mathsf{odd}(F)| = |\mathsf{odd}(T)| + |\mathsf{odd}(I \oplus F \oplus T)|.$$

*Proof.* Let us start with the simplest nontrivial case $T = I \oplus P_1$ and $I \oplus F \oplus T = F \oplus P_1$. Let $u, v$ be the two endpoints of the path $P_1$; note that $u, v \in \mathsf{odd}(I \oplus F) = \mathsf{odd}(I) \oplus \mathsf{odd}(F)$. If $u \in \mathsf{odd}(I) \setminus \mathsf{odd}(F)$ and $v \in \mathsf{odd}(F) \setminus \mathsf{odd}(I)$, then both $\mathsf{odd}(T) = \mathsf{odd}(I) - u + v$ and $\mathsf{odd}(I \oplus F \oplus T) = \mathsf{odd}(F) + u - v$ hold, which immediately implies the desired equality. The same reasoning applies for the other three cases

- $u \in \mathsf{odd}(F) \setminus \mathsf{odd}(I), v \in \mathsf{odd}(I) \setminus \mathsf{odd}(F)$
- $u, v \in \mathsf{odd}(I) \setminus \mathsf{odd}(F)$
- and $u, v \in \mathsf{odd}(F) \setminus \mathsf{odd}(I)$.

It follows that for any $1 \le j \le k$, letting $T_j = T_{j-1} \oplus P_j$ and $T_0 = I$, we get

$$\begin{aligned} |\mathsf{odd}(T_j)| + |\mathsf{odd}(I \oplus F \oplus T_j)| &= |\mathsf{odd}(T_{j-1})| + |\mathsf{odd}(I \oplus F \oplus T_{j-1})| \\ &= \cdots \\ &= |\mathsf{odd}(I)| + |\mathsf{odd}(F)| \,. \end{aligned}$$

The case of flipping all edges in a cycle is easier, since the set of odd-degree vertices doesn't change at all. □

Now, let us consider the fully general case, where $T$ could be in the middle of traversing one of the paths $P_j$ (with a similar analysis if instead we're in the middle of traversing one of the cycles $C_j$). Let $T_{\mathsf{start}}$ denote the edge set we had right before starting to traverse $P_j$; similarly, let $T_{\mathsf{end}}$ denote the edge we would get as soon as we finish traversing $P_j$. We claim that

$$\left| |\mathsf{odd}(T)| - \frac{|\mathsf{odd}(T_{\mathsf{start}})| + |\mathsf{odd}(T_{\mathsf{end}})|}{2} \right| \le 2.$$

To see this, observe that for any $A \subseteq E$ and $e \in E$, we have $||\mathsf{odd}(A)| - |\mathsf{odd}(A \oplus e)|| \le 2$ since flipping $e$ only changes the membership of its two endpoints in the set of odd-degree vertices. Furthermore, since we are flipping *consecutive* edges along a traversal of a path/cycle, this discrepancy never increases above 2. The same reasoning also establishes that

$$\left| |\mathsf{odd}(I \oplus F \oplus T)| - \frac{|\mathsf{odd}(I \oplus F \oplus T_{\mathsf{start}})| + |\mathsf{odd}(I \oplus F \oplus T_{\mathsf{end}})|}{2} \right| \le 2.$$

Once we have these two inequalities, then we're done, since

$$\begin{aligned} |\mathsf{odd}(I)| + |\mathsf{odd}(F)| &= |\mathsf{odd}(T_{\mathsf{start}})| + |\mathsf{odd}(I \oplus F \oplus T_{\mathsf{start}})| \\ |\mathsf{odd}(I)| + |\mathsf{odd}(F)| &= |\mathsf{odd}(T_{\mathsf{end}})| + |\mathsf{odd}(I \oplus F \oplus T_{\mathsf{end}})| \end{aligned}$$

from Claim 2.2.

□

# 3 On Random Initializations

Arguably, a more "natural" solution to the bottleneck from applying Glauber dynamics directly to $\mu$ is instead to "average out" the identities of the spins $\{\pm 1\}$ at the algorithmic level. More specifically, we choose a special initial distribution $\nu$ and hope that $\|\nu \mathsf{P}^t - \mu\|_{\mathsf{TV}} \le \epsilon$ for $t \le \mathsf{poly}(n, \log(1/\epsilon))$, even though the mixing time from a worst-case initial distribution $\delta_x$ can be exponentially large. This is probably a more practical solution to sampling from the ferromagnetic Ising model, but there are much fewer available tools for analyzing such methods.

Given that at low temperatures, we expect $\mu$ to concentrate around the $\mathbf{1}$ and $-\mathbf{1}$ configurations, one natural choice could be $\nu = \frac{1}{2}\delta_{\mathbf{1}} + \frac{1}{2}\delta_{-\mathbf{1}}$. This was recently studied in a paper of Gheissari–Sinclair [GS22], where the case of random $d$-regular graphs and the $d$-dimensional torus are considered. However, this initialization definitely fails for more general graphs with sparse cuts. For instance, imagine the graph $G$ consists of $\sqrt{n}$ copies of disjoint complete graphs $K_{\sqrt{n}}$, each on $\sqrt{n}$ vertices, and with no edges between them. Then at low temperature, $\mu$ actually concentrates on $2^{\sqrt{n}}$ configurations; the only constraint is that all vertices in the same copy of $K_{\sqrt{n}}$ must have the same $\{\pm 1\}$-spin.

Another natural initial distribution is to take $\mathsf{Unif}\{\pm 1\}^n$. As far as we are aware, it open to prove or disprove the following.

**Question 1.** *Is it true that for any graph $G = (V, E)$ and every $\beta \geq 0$, if we take the initial distribution of Glauber dynamics to be $\nu = \mathsf{Unif}\{\pm 1\}^n$, then $\|\nu \mathsf{P}^t - \mu\|_{\mathsf{TV}} \leq \epsilon$ for $t \leq \mathsf{poly}(n, \log(1/\epsilon))$.*

From a mathematical perspective, methods based on bounding worst-case mixing for various expansions of the Ising model (e.g. even subgraphs) seem to be more "robust" to perturbations in the input (e.g. introducing consistent but nonuniform external fields $h \in \mathbb{R}^V_{\geq 0}$). However, at least for the even subgraphs analysis we did in this lecture, the ultimate mixing time scales as $O(m^3/\rho^6)$. Although this has been improved drastically to nearly-linear mixing time for constant $\rho$ (see [CLV21a; CLV21b] for the bounded-degree case, and [CZ23] more generally), in the regime of $\rho \leq 1/n$ (corresponding to small external field $h \leq O(1/n)$), the running time scales at least like $O(n^6 m^3)$ from the proof of Theorem 1.4 (and the proof of Corollary 1.5); the techniques in more recent works [CLV21b; CZ23] break down completely in this regime of $\rho$. This raises the following question.

**Question 2.** *Does there exist a nearly-linear time sampler for the ferromagnetic Ising model for all graphs $G = (V, E)$, and all $\beta \in \mathbb{R}_{\geq 0}, h \in \mathbb{R}$?*

## References


[ALR87] M. Aizenman, J. L. Lebowitz, and D. Ruelle. "Some rigorous results on the Sherrington-Kirkpatrick spin glass model". In: *Communications in Mathematical Physics* 112.1 (1987), pp. 3–20 (cit. on p. 2).

[CLV21a] Zongchen Chen, Kuikui Liu, and Eric Vigoda. "Optimal Mixing of Glauber Dynamics: Entropy Factorization via High-Dimensional Expansion". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2021. Virtual, Italy: Association for Computing Machinery, 2021, pp. 1537–1550 (cit. on p. 6).

[CLV21b] Zongchen Chen, Kuikui Liu, and Eric Vigoda. "Spectral Independence via Stability and Applications to Holant-Type Problems". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. FOCS 2021. 2021, pp. 149–160 (cit. on p. 6).

[CZ23] Xiaoyu Chen and Xinyuan Zhang. "A Near-Linear Time Sampler for the Ising Model with External Field". In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2023, pp. 4478–4503. DOI: 10.1137/1.9781611977554.ch170. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611977554.ch170 (cit. on p. 6).

[Dum17] Hugo Duminil-Copin. "Random currents expansion of the Ising model". In: *arXiv preprint arXiv:1607.06933* (2017) (cit. on p. 2).

[GS22] Reza Gheissari and Alistair Sinclair. "Low-Temperature Ising Dynamics with Random Initializations". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. Rome, Italy: Association for Computing Machinery, 2022, pp. 1445–1458. ISBN: 9781450392648. DOI: 10.1145/3519935.3519964 (cit. on p. 5).

[Isi25] Ernst Ising. "Beitrag zur Theorie des Ferromagnetismus". In: *Zeitschrift für Physik* 31 (1925), pp. 253–258 (cit. on p. 1).

[JS93] Mark Jerrum and Alistair Sinclair. "Polynomial-time approximation algorithms for the Ising model". In: *SIAM Journal on Computing* 22.5 (1993), pp. 1087–1116 (cit. on pp. 1, 8).

[Len20] W. Lenz. "Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern". In: *Physikalische Zeitschrift* 21 (1920), pp. 613–615 (cit. on p. 1).

## A Unfinished Proofs for Even Subgraphs

*Proof of Proposition 1.2.* The key is the identity $e^x = \cosh(x)\cdot(1+\tanh(x))$. Then each term of $Z_G(\beta,h)$ is given by

$$\begin{aligned}
&\exp\left(\frac{\beta}{2}\sigma^\top A\sigma + h\cdot\langle\sigma,\mathbf{1}\rangle\right)\\
&= \prod_{uv\in E}\exp(\beta\sigma_u\sigma_v)\prod_{v\in V}\exp(h\sigma_v)\\
&= \prod_{uv\in E}\cosh(\beta\sigma_u\sigma_v)\prod_{uv\in E}(1+\tanh(\beta\sigma_u\sigma_v))\prod_{v\in V}\cosh(h\sigma_v)\prod_{v\in V}(1+\tanh(h\sigma_v))\\
&= \cosh(\beta)^{|E|}\cosh(h)^{|V|}\cdot\prod_{uv\in V}(1+\sigma_u\sigma_v\tanh(\beta))\prod_{v\in V}(1+\sigma_v\tanh(h))\\
&\hfill \text{(cosh is even and tanh is odd)}\\
&= \cosh(\beta)^{|E|}\cosh(h)^{|V|}\cdot\left(\sum_{F\subseteq E}\tanh(\beta)^{|F|}\prod_{uv\in F}\sigma_u\sigma_v\right)\left(\sum_{S\subseteq V}\tanh(h)^{|S|}\prod_{v\in S}\sigma_v\right)\\
&= \cosh(\beta)^{|E|}\cosh(h)^{|V|}\cdot\sum_{\substack{F\subseteq E\\ S\subseteq V}}\tanh(\beta)^{|F|}\tanh(h)^{|S|}\underbrace{\left(\prod_{v\in\mathsf{odd}(F)}\sigma_v\right)\left(\prod_{v\in S}\sigma_v\right)}_{=\prod_{v\in\mathsf{odd}(F)\oplus S}\sigma_v}. \qquad \text{(Using } \sigma_v^2=1\text{)}
\end{aligned}$$

It follows that

$$Z_G(\beta,h) = \cosh(\beta)^{|E|}\cosh(h)^{|V|}\cdot\sum_{\substack{F\subseteq E\\ S\subseteq V}}\tanh(\beta)^{|F|}\tanh(h)^{|S|}\sum_{\sigma\in\{\pm1\}^V}\prod_{v\in\mathsf{odd}(F)\oplus S}\sigma_v.$$

Since $\sum_{\sigma\in\{\pm1\}^V}\prod_{v\in A}\sigma_v$ equals $2^{|V|}$ if $A=\emptyset$ and 0 otherwise, the only terms which survive the above summations are those such that $S=\mathsf{odd}(F)$. The claim immediately follows. $\square$

*Proof of Proposition 1.3.* Let $(\rho_t)_{t=0}^T$ be a decreasing sequence such that $\rho_T=\rho$ and $\rho_0=1$; we instantiate this sequence later. Then

$$\widehat{Z}_G(\rho,\lambda) = \widehat{Z}_G(\rho_0,\lambda)\cdot\prod_{t=1}^T\frac{\widehat{Z}_G(\rho_t,\lambda)}{\widehat{Z}_G(\rho_{t-1},\lambda)} = \widehat{Z}_G(1,\lambda)\cdot\prod_{t=1}^T\mathbb{E}_{F\sim\widehat{\mu}_{\rho_{t-1},\lambda}}\left[\left(\frac{\rho_t}{\rho_{t-1}}\right)^{|\mathsf{odd}(F)|}\right].$$

Note that $\widehat{Z}_G(1,\lambda)=(1+\lambda)^{|E|}$, which is trivial to compute. We combine the given FPAS with the Monte Carlo method to estimate each of the terms $\mathbb{E}_{F\sim\widehat{\mu}_{\rho_{t-1},\lambda}}\left[\left(\frac{\rho_t}{\rho_{t-1}}\right)^{|\mathsf{odd}(F)|}\right]$. Since $(\rho_t)_{t=0}^T$ is decreasing, $F\mapsto\left(\frac{\rho_t}{\rho_{t-1}}\right)^{|\mathsf{odd}(F)|}$ is a 1-bounded function. We prove the following claim, which tells us we can take $T\approx(1-\rho)n$ and $\rho_t=\rho_{t-1}-\frac{1}{n}$ for all $t$; for this interpolation, the Monte Carlo method efficiently estimates the desired expectations.

**Claim A.1.** *Suppose $\rho_t\le\rho_{t-1}\le\rho_t+\frac{1}{n}$. Then*

$$\mathbb{E}_{F\sim\widehat{\mu}_{\rho_{t-1},\lambda}}\left[\left(\frac{\rho_t}{\rho_{t-1}}\right)^{|\mathsf{odd}(F)|}\right]\ge\Omega(1).$$

*Proof.* This lower bound holds for two different reasons depending on the order of $\rho_{t-1},\rho_t$. If $C\in(0,1)$ is some absolute constant, then in the regime $\rho_{t-1}\ge C$, we have $\rho_{t-1}=\left(1\pm O\left(\frac{1}{n}\right)\right)\cdot\rho_t$, whence $\left(\frac{\rho_t}{\rho_{t-1}}\right)^{|\mathsf{odd}(F)|}\ge\Omega(1)$ uniformly for all $F\subseteq E$ just from the fact that $|\mathsf{odd}(F)|\le|V|=n$.

The interesting case is when $\rho_{t-1}\le C$. In this case, we can't hope to uniformly lower bound $\left(\frac{\rho_t}{\rho_{t-1}}\right)^{|\mathsf{odd}(F)|}$ for all $F$. However, because $\rho_{t-1}$ is not too large, the distribution $\widehat{\mu}_{\rho_{t-1},\lambda}$ places more mass on those $F\subseteq E$ with small $|\mathsf{odd}(F)|$; for such $F$, we can afford a larger multiplicative gap between $\rho_{t-1}$ and $\rho_t$. This is the intuition. To formally prove the claim, we use the original

Ising model formulation. Let $h_t = \operatorname{arctanh} \rho_t, h_{t-1} = \operatorname{arctanh} \rho_{t-1}$ and $\beta = \operatorname{arctanh} \lambda$. Since $\operatorname{arctanh}(x)$ blows up to $\pm\infty$ as $x \to \pm 1$, but remains $O(1)$-Lipschitz for $x$ bounded away from $\pm 1$, the restriction $\rho_{t-1} \leq C$ ensures that $|\rho_{t-1} - \rho_t| \leq 1/n$ implies $|h_{t-1} - h_t| \leq O(1/n)$. This is the reason it is advantageous to use $Z_G(\cdot, \cdot)$ from the "spin world". Using this, we have

$$
\begin{aligned}
\mathbb{E}_{F \sim \widehat{\mu}_{\rho_{t-1},\lambda}} \left[ \left( \frac{\rho_t}{\rho_{t-1}} \right)^{|\mathsf{odd}(F)|} \right] &= \frac{\widehat{Z}_G(\rho_t, \lambda)}{\widehat{Z}_G(\rho_{t-1}, \lambda)} \\
&= \left( \frac{\cosh(h_{t-1})}{\cosh(h_t)} \right)^{|V|} \cdot \frac{Z_G(\beta, h_t)}{Z_G(\beta, h_{t-1})} && \text{(Proposition 1.2)} \\
&\geq \left( \frac{\cosh(h_{t-1})}{\cosh(h_t)} \right)^{|V|} \cdot \min_{\sigma \in \{\pm 1\}^V} \exp\left( (h_t - h_{t-1}) \cdot \langle \sigma, \mathbf{1} \rangle \right) \\
& && \left(\text{Using } \tfrac{a_1 + \cdots + a_n}{b_1 + \cdots + b_n} \geq \min_{i \in [n]} \tfrac{a_i}{b_i}\right) \\
&= \exp\left( n \cdot (\phi(h_{t-1}) - \phi(h_t)) \right). && (\text{Setting } \phi(x) = \log\cosh(x) - x)
\end{aligned}
$$

Again, since $\rho_{t-1} \leq C$, we have $|h_{t-1} - h_t| \leq O(1/n)$. Hence, the $\Omega(1)$ lower bound follows provided $\phi$ is $O(1)$-Lipschitz on $\mathbb{R}$; one can show this by observing that $\phi'(x) = \tanh(x) - 1$ is bounded in absolute value by 2 for all $x \in \mathbb{R}$. □

□

*Proof of Corollary 1.5.* If $\rho \geq 1/n$, then running Glauber dynamics and applying Theorem 1.4 already furnishes an FPAS. If $\rho \leq 1/n$, we add an extra rejection sampling step on top. Suppose we want to sample from $\widehat{\mu}_{\rho,\lambda}$ to within $0 < \delta < 1$ total variation error, where $\rho \leq 1/n$. Let $0 < C < 1$ be the universal constant lower bound in Claim A.1.[6] We run Glauber dynamics to generate an approximate sample $F \sim \nu$ where $\left\|\nu - \widehat{\mu}_{1/n,\lambda}\right\|_{\mathsf{TV}} \leq C^2\delta/4$, and accept this as a valid sample with probability

$$\Pr[\mathsf{Accept} \mid \text{Proposed } F] = (\rho n)^{|\mathsf{odd}(F)|} < 1.$$

Otherwise, we reject and try again. If we are aiming for $\delta$ total variation error, then we make $O(\log(1/\delta))$ independent attempts; if we reject all such proposals, then we output an arbitrarily chosen $F_* \subseteq E$ (e.g. $\emptyset$).

If it were the case that $\nu = \widehat{\mu}_{1/n,\mu}$, then any accepted proposal is automatically distributed perfectly according to $\widehat{\mu}_{\rho,\lambda}$. However, our proposals are only approximately distributed according to $\widehat{\mu}_{1/n,\lambda}$; furthermore, we don't have access to the quantity $\nu(F)$, so it isn't clear how to "algorithmically fix" the output distribution of an accepted proposal. Instead, we argue that both $\Pr_\nu[\mathsf{Accept}] \geq \Omega(1)$ and $\left\|\nu - \widehat{\mu}_{1/n,\lambda}\right\|_{\mathsf{TV}} \leq C^2\delta/4$ together imply that the distribution of any accepted proposal is close to $\widehat{\mu}_{\rho,\lambda}$. We begin by lower bounding the overall acceptance probability, which is also necessary for the efficiency of the algorithm.

**Claim A.2.** *We have the lower bound $\Pr_\nu[\mathsf{Accept}] \geq C/2$, which is a universal constant independent of $\delta$. Here, the subscript $\nu$ indicates w.r.t. which distribution the proposals are being generated from.*

*Proof.*

$$
\begin{aligned}
\Pr_\nu[\mathsf{Accept}] &= \sum_{F \subseteq E} \Pr[\mathsf{Accept} \mid \text{Proposed } F] \cdot \Pr_\nu[\text{Proposed } F] \\
&= \mathbb{E}_{F \sim \nu} \left[ (\rho n)^{|\mathsf{odd}(F)|} \right] && (\text{Proposal distribution is } \nu) \\
&= -C^2\delta/4 + \mathbb{E}_{F \sim \widehat{\mu}_{1/n,\lambda}} \left[ (\rho n)^{|\mathsf{odd}(F)|} \right] && \left(\text{Using } \left\|\nu - \widehat{\mu}_{1/n,\lambda}\right\|_{\mathsf{TV}} \leq C\delta\right) \\
&\geq C \cdot \left( 1 - \frac{C\delta}{4} \right) && \text{(Claim A.1)} \\
&\geq C/2. && (0 \leq \delta, C \leq 1)
\end{aligned}
$$

□

[6] In the Jerrum–Sinclair paper, they show that one can take $C = 1/10$ [JS93].

Claim A.2 ensures that the probability of rejecting all $O(\log(1/\delta))$ independent attempts is at most $O(\delta)$. Again, if the true proposal distribution $\nu$ were actually the ideal proposal distribution $\widehat{\mu}_{1/n,\lambda}$, then we'd be done immediately. No further analysis would be needed. However, we need to correctly account for the case where we only have $\|\nu - \widehat{\mu}_{\rho,\lambda}\|_{\mathsf{TV}} \le C^2\delta/4$.

Let $\xi$ denote the law of the output of the overall sampling algorithm and let $L \le O(\log(1/\delta))$ denote the maximum allowed number of proposals. Then for all $F \subseteq E$,

$$\begin{aligned}
\xi(F) &= \sum_{\ell=0}^{L-1} \Pr[\mathsf{Accept} \mid \text{Proposed } F] \cdot \nu(F) \cdot \Pr_{\nu}[\mathsf{Reject}]^{\ell} + \Pr_{\nu}[\mathsf{Reject}]^{L} \cdot \mathbb{I}[F_* = F] \\
&= \frac{\Pr[\mathsf{Accept} \mid \text{Proposed } F] \cdot \nu(F)}{\Pr_{\nu}[\mathsf{Accept}]} \cdot \left(1 - \Pr_{\nu}[\mathsf{Reject}]^{L}\right) + \Pr_{\nu}[\mathsf{Reject}]^{L} \cdot \mathbb{I}[F_* = F].
\end{aligned}$$

It follows that

$$\begin{aligned}
&\|\xi - \widehat{\mu}_{\rho,\lambda}\|_{\mathsf{TV}} \\
&\le \Pr_{\nu}[\mathsf{Reject}]^{L} \cdot \|\delta_{F_*} - \widehat{\mu}_{\rho,\lambda}\|_{\mathsf{TV}} + \left(1 - \Pr_{\nu}[\mathsf{Reject}]^{L}\right) \cdot \frac{1}{2} \sum_{F \subseteq E} \left| \frac{\Pr[\mathsf{Accept} \mid \text{Proposed } F] \cdot \nu(F)}{\Pr_{\nu}[\mathsf{Accept}]} - \widehat{\mu}_{\rho,\lambda}(F) \right| \\
&\le \Pr_{\nu}[\mathsf{Reject}]^{L} + \underbrace{\frac{1}{2} \sum_{F \subseteq E} \left| \frac{\Pr[\mathsf{Accept} \mid \text{Proposed } F] \cdot \nu(F)}{\Pr_{\nu}[\mathsf{Accept}]} - \frac{\Pr[\mathsf{Accept} \mid \text{Proposed } F] \cdot \widehat{\mu}_{1/n,\lambda}(F)}{\Pr_{\widehat{\mu}_{1/n,\lambda}}[\mathsf{Accept}]} \right|}_{(*)}.
\end{aligned}$$

The main nontrivial step is controlling $(*)$, since we are comparing two different proposal distributions. Since $F \mapsto (\rho n)^{|\mathsf{odd}(F)|}$ is a 1-bounded function and $\left\|\nu - \widehat{\mu}_{1/n,\lambda}\right\|_{\mathsf{TV}} \le C^2\delta/4$, we have

$$\left| \Pr_{\nu}[\mathsf{Accept}] - \Pr_{\widehat{\mu}_{1/n,\lambda}}[\mathsf{Accept}] \right| = \left| \mathbb{E}_{F \sim \nu}\left[(\rho n)^{|\mathsf{odd}(F)|}\right] - \mathbb{E}_{F \sim \widehat{\mu}_{1/n,\lambda}}\left[(\rho n)^{|\mathsf{odd}(F)|}\right] \right| \le C^2\delta/4.$$

Furthermore, Claim A.2 says that $\Pr_{\nu}[\mathsf{Accept}] \ge C/2$, and so this $C^2\delta/4$ additive approximation is also a $(1 \pm C\delta/2)$-multiplicative approximation. It follows that

$$\begin{aligned}
(*) &\le \frac{1}{\Pr_{\nu}[\mathsf{Accept}]} \cdot \frac{1}{2} \sum_{F \subseteq E} (\rho n)^{|\mathsf{odd}(F)|} \cdot \left|(1 \pm C\delta/2)\nu(F) - \widehat{\mu}_{1/n,\lambda}(F)\right| \\
&\le \frac{1}{C} \cdot \left( \underbrace{\sum_{F \subseteq E} \left|\nu(F) - \widehat{\mu}_{1/n,\lambda}(F)\right|}_{\le C^2\delta/2} + \frac{C\delta}{2} \underbrace{\sum_{F \subseteq E} \nu(F)}_{=1} \right) \qquad (\text{Using } \rho n \le 1 \text{ and } \Pr_{\nu}[\mathsf{Accept}] \ge C/2) \\
&\le \frac{1+C}{2} \cdot \delta.
\end{aligned}$$

Taking $L \le O(\log(1/\delta))$ sufficiently large, we also obtain $\Pr_{\nu}[\mathsf{Reject}]^{L} \le \frac{1-C}{2} \cdot \delta$, and so putting everything together, we get $\|\xi - \widehat{\mu}_{\rho,\lambda}\|_{\mathsf{TV}} \le \delta$ as desired. $\square$