PHASE TRANSITION IN THE ISING MODEL

1. The Ising Model

The Ising model is a crude but extremely important mathematical model of a ferromagnetic metal introduced by Ising about 70 years ago. Its importance stems from the fact that it is the one of the simplest mathematical models to exhibit a *phase transition*: at high temperature, there is a unique equilibrium state for the system, but at temperatures below a certain critical temperature, there are several distinct equilibrium states. This corresponds to the physical phenomenon of spontaneous magnetization: If unmagnetized iron is cooled to a very low temperature, it will magnetize; and if a magnet is heated to a sufficiently high temperature, it will demagnetize. The latter may be verified easily by experiment, using only a floppy disk and a household stove.

1.1. **Gibbs States.** Let $\mathcal{X}$ be a finite set and $H : \mathcal{X} \to \mathbb{R}$ a function, called the *Hamiltonian* of the system. In physical applications $H(x)$ represents the *energy* of the system when it is in state x. The *Gibbs state* $\mu = \mu_\beta$ for $\beta = 1/(kT)$, where $k =$ Boltzman's constant and $T =$ temperature, is the probability measure on $\mathcal{X}$ defined by

$$\mu_\beta(x) = e^{-\beta H(x)}/Z(\beta), \qquad \text{where} \tag{1.1}$$

$$Z(\beta) = \sum_{x\in\mathcal{X}} e^{-\beta H(x)}. \tag{1.2}$$

The normalizing constant $Z(\beta)$ is called the *partition function*. The family $\{\mu_\beta\}_{\beta>0}$ is a one-paramter *exponential family* of probability measures on $\mathcal{X}$, with $-\beta$ playing the role of the *natural parameter*, $H(x)$ the *sufficient statistic*, and $\log Z(\beta)$ the role of the $\psi-$function. Observe that, since the sum in (1.2) is finite, the partition function is well-defined and (real-)analytic in the domain $\beta > 0$.

1.2. **The Ising Hamiltonian.** In condensed-matter physics, field theory, and various other parts of statistical physics, the state space $\mathcal{X}$ is often of the form

$$\mathcal{X} = A^V \tag{1.3}$$

where V is a set of *sites* (which we will also call *vertices*) and A is a finite set. Elements of V usually represent spatial locations, and elements of A may represent atomic elements (in models of alloys), presence (+1) or absence (0) of particles (in models of gases), *spins* (in models of magnetism and in quantum field theory), and so on. In the Ising model, A is the two-element set $A = \{\pm 1\}$, and V is the set of vertices of a graph G; the most interesting case, from the standpoint of the physicist, is that where V is a subset of the d-dimensional integer lattice $\mathbb{Z}^d$. The *Ising Hamiltonian* is defined as follows: for any configuration $x \in \mathcal{X} := \{-1, +1\}^V$

$$H(x) = J \sum_{\substack{i,j \in V: \\ i \sim j}} x_i x_j \tag{1.4}$$

Here $i \sim j$ means that vertices i, j are *nearest neighbors*, that is, there is an edge of the graph G connecting i and j; each edge counts only once in the sum. The constant J is called the *coupling constant*: if $J < 0$ the model is called *ferromagnetic*, and if $J > 0$ it is *anti-ferromagnetic*. Unless otherwise specified, it is henceforth assumed that $J < 0$. Observe that in this case, the system "prefers" configurations in which neighboring spins are aligned, as these have lower energy. The degree to which this is true depends, of course, on the inverse temperature β — for larger values of β the preference for low-energy states is stronger.

1.3. **The Thermodynamic Limit.** The sort of magnet that you might carry around in your pocket would have on the order of 10^{25} iron atoms. The exact number isn't important — what is important is that the number is *big* (even to a computer scientist). Thus, it makes sense to study the behavior of the Ising model when the vertex set V is large, and in fact to inquire about the limiting behavior as V becomes infinite. There are two ways to go about this: (1) Look at the limiting behavior of the Gibbs states μ_β for finite V as V becomes larger; or (2) Try to extend the definition of Gibbs state to configuration spaces on infinite graphs. Program (2) is the theory of *DLR states* (for Dobrushin, Lanford, and Ruelle); it requires more mathematical machinery than I wish to invest in now, and so I shall only discuss program (1).

Let $G = (V, \mathcal{E})$ be a countably infinite locally finite graph (such as the integer lattices $\mathbb{Z}^d$; *locally finite* means that for each vertex i there are only finitely many edges incident to i). Let $\mathcal{X} = \{\pm 1\}^V$ be the space of configurations on the vertex set V, and For any vertex $i \in V$, let $X_i : \mathcal{X} \to \{\pm 1\}$ be the ith coordinate evaluation map $X_i(x) = x_i$. The *Borel $\sigma-$algebra* $\mathcal{B}$ on the space $\mathcal{X}$ is the smallest $\sigma-$algebra that makes all of the random variables X_i measurable; equivalently, $\mathcal{B}$ is the $\sigma-$algebra generated by the open sets of the product topology on $\mathcal{X}$. For each *finite* subset $\Lambda \subset V$ define

$$H_\Lambda(x) = J \sum_{\substack{i \in \Lambda, j \in V \\ i \sim j}} x_i x_j; \tag{1.5}$$

here the sum is over all edges of the graph G with at least one endpoint in Λ. For each $z \in \mathcal{X}$ and each finite subset $\Lambda \subset V$ define the Gibbs state $\mu^z_\Lambda = \mu^z_{\Lambda,\beta}$ on Λ with external boundary condition z to be the discrete Borel probability measure on $\mathcal{X}$ determined by the rule

$$\begin{aligned} \mu^z_\Lambda(x) &= \exp\{-\beta H_\Lambda(x)\}/Z_{\Lambda,z}(\beta) \qquad \text{if } x_{\Lambda^c} = z_{\Lambda^c} \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{1.6}$$

where $Z_{\Lambda,z}(\beta)$ is the appropriate normalizing constant. Observe that μ^z_Λ is concentrated on a finite set of configurations, namely, those that agree with z outside of Λ. Also, if $z, w \in \mathcal{X}$ are two configurations that agree outside Λ then$\mu^z_\Lambda = \mu^w_\Lambda$.

The two most interesting boundary conditions (at least for now) are $z^+ \equiv 1$ and $z^- \equiv -1$: we shall denote by μ_Λ^+ and μ_Λ^- the Gibbs states with these external boundary conditions. It is important to note that these two Gibbs states are mirror images, in the following sense: if $\rho : \mathcal{X} \to \mathcal{X}$ is the mapping that flips every spin, that is, $\rho(x) = -x$, then for every finite $\Lambda \subset V$,

$$\mu_\Lambda^- = \mu_\Lambda^+ \circ \rho. \tag{1.7}$$

1.4. **Phase Transition in Dimension 2.** The observable physical phenomenon of spontaneous magnetization (and demagnetization) has a mathematical analogue in the Ising model in dimensions two and higher, a fact discovered by R. PEIERLS in the 1930s, some years after Ising introduced his model.[1] Let $G = (\mathbb{Z}^2, \mathcal{E})$ be the standard two-dimensional lattice (the edges $e \in \mathcal{E}$ connect points of $\mathbb{Z}^2$ that differ by $(1,0)$ or $(0,1)$), and let Λ_n be the square of side $2n+1$ centered at the origin o. Denote by μ_n^+ and μ_n^- the Gibbs states with external boundary conditions z^+ and z^- on the square Λ_n.

Theorem 1. *There exists β_c satisfying $0 < \beta_c < \infty$ such that the following is true: For each vertex $i \in \mathbb{Z}^2$*

$$\lim_{n\to\infty} \mu_n^+\{X_i = +1\} > 1/2 \qquad \text{if } \beta > \beta_c \tag{1.8}$$

$$\lim_{n\to\infty} \mu_n^+\{X_i = +1\} = 1/2 \qquad \text{if } \beta \le \beta_c. \tag{1.9}$$

The fact that $\beta_c < \infty$ is, in essence, Peierls discovery. I do not know who first proved that $\beta_c > 0$, but I consider this just as important.

Theorem 2. *For each β, as $n \to \infty$, the measures μ_n^+ converge in distribution to a probability measure μ^+, and the measures μ_n^- converge in distribution to a probability measure μ^-. These limiting measures are translation invariant, and μ^- is stochastically dominated by μ^+.*

It follows that the limits in Theorem 1 equal $\mu^+\{X_i = +1\}$. Since μ^+ is translation-invariant, it also follows that $\mu^+\{X_i = +1\} = \mu^+\{X_o = +1\}$ for all vertices i. By the symmetry relation (1.7), $\mu^-\{X_o = -1\} = \mu^-\{X_o = +1\}$.

Corollary 3.

$$\mu^+ \ne \mu^- \qquad \text{if } \beta > \beta_c; \tag{1.10}$$

$$\mu^+ = \mu^- \qquad \text{if } \beta \le \beta_c. \tag{1.11}$$

Proof. The measure μ^+ stochastically dominates the measure μ^- for any value of β. By the extension of Strassen's monotone coupling theorem to infinite configuration spaces (HW set 2), there exist, on some probability space, random configurations Y^+, Y^- with distributions μ^+, μ^-, respectively, such that $Y^- \le Y^+$ almost surely. But by Theorem 2 and relation (1.9), if $\beta \le \beta_c$ then for every vertex i,

$$P\{Y_i^+ = +1\} = 1/2 = P\{Y_i^- = +1\},$$

[1] Ising's Ph. D. thesis supervisor LENZ had suggested to Ising that a phase transition might exist in the Ising model; Ising was able to prove that there is *no* phase transition in one dimension, but proved nothing about the behavior in higher dimensions.

and so it must be that $Y_i^+ = Y_i^-$ almost surely. □

In section 4 below, we shall prove that the limiting relation (1.8) holds for sufficiently large β (the low-temperature limit), by Peierls' original and elegant argument. In section 5 we shall prove that (1.9) holds for sufficiently small β by relating the construction of random fields with the distributions $\mu_\Lambda^\pm$ to *site percolation*.

2. The Markov Property for Gibbs States

For any configuration $x \in \mathcal{X} = \{\pm 1\}^V$ and any subset $\Lambda \subset V$, denote by x_Λ the restriction of the configuration x to the set Λ, that is, $x_\Lambda = (x_i)_{i\in\Lambda}$. Similarly, denote by X_Λ the vector $(X_i)_{i\in\Lambda}$ of coordinate evaluation mappings for sites in Λ.

Proposition 4. *Let Λ and Σ be disjoint, finite subsets of V. For any configuration $x \in \mathcal{X}$ the following is true:*

$$\mu^x_{\Lambda\cup\Sigma}(X_\Lambda = x_\Lambda \,|\, X_\Sigma = x_\Sigma) = \mu^x_\Lambda(X_\Lambda = x_\Lambda). \tag{2.1}$$

Thus, μ^x_Λ is the conditional distribution under $\mu^x_{\Lambda\cup\Sigma}$ of X_Λ given that $X_\Sigma = x_\Sigma$. It is this mutual consistency property that allows the possibility of extending the definition of a Gibbs state to infinite configuration spaces.

For any subset $\Lambda \subset V$, define the *outer boundary* $\partial\Lambda$ of Λ to be the set of all vertices at distance one from Λ.

Proposition 5. *For any configuration x and any finite subset $\Lambda \subset V$, the probability $\mu^x_\Lambda(X_\Lambda = x_\Lambda)$ depends only on $x_{\Lambda\cup\partial\Lambda}$.*

The proofs of Propositions 4 and 5 are left as exercises. Proposition 5 is quite easy. Proposition 4 is slightly more subtle: You will find it easiest to begin by showing that it suffices to consider the case where Σ is a singleton.

Corollary 6. *Let $\Sigma \subset \Lambda \subset V$ be finite subsets of V such that $\partial\Sigma \subset \Lambda$, and let x, y be configurations such that $x_\Lambda = y_\Lambda$. Then*

$$\mu^x_\Lambda(X_\Sigma = x_\Sigma \,|\, X_{\Lambda-\Sigma} = x_{\Lambda-\Sigma}) = \mu^y_\Lambda(X_\Sigma = x_\Sigma \,|\, X_{\Lambda-\Sigma} = x_{\Lambda-\Sigma}). \tag{2.2}$$

Let's make it a hat trick – this one is also an exercise.

3. Stochastic Monotonicity Results

Proposition 7. *For any finite subset $\Lambda \subset V$ and any two configurations $z, y \in \mathcal{X}$ such that $z \le y$, the probability measure μ^z_Λ is stochastically dominated by μ^y_Λ.*

This will be proved by appeal to a theorem of Holley. Let Λ be a finite set and $\mathcal{X}_\Lambda = \{-1, +1\}^\Lambda$ be the space of spin configurations on Λ. For any site $i \in \Lambda$, define the spin operators $\sigma_i^+, \sigma_i^- : \mathcal{X}_\Lambda \to \mathcal{X}_\Lambda$ by

$$(\sigma_i^\pm(x))_j = \begin{cases} x_j, & \text{if } j \neq i \\ \pm 1, & \text{if } j = i. \end{cases}$$

Theorem 8. *Let Λ be a finite set. Let μ and ν be probability distributions on $\{-1,+\}^\Lambda$ such that $\mu(x) > 0$ and $\nu(x) > 0$ for each $x \in \mathcal{X}$. If for every site i and every pair x, y of configurations such that $x \leq y$ it is the case that*

$$\frac{\mu(\sigma_i^+(x))}{\mu(\sigma_i^-(x))} \leq \frac{\nu(\sigma_i^+(y))}{\nu(\sigma_i^-(y))}, \tag{3.1}$$

then μ is stochastically dominated by ν.

Proof. The strategy is to build a discrete-time Markov chain $(X_n, Y_n)_{n\geq 0}$ on the space $\mathcal{X}_\Lambda \times \mathcal{X}_\Lambda$ of configuration pairs in such a way that (a) $X_n \leq Y_n$ for all n; (b) $(X_n)_{n\geq 0}$ is an aperiodic irreducible Markov chain with stationary distribution μ; and (c) $(Y_n)_{n\geq 0}$ is an aperiodic irreducible Markov chain with stationary distribution ν. The reader should convince himself/herself that this will imply $\mu \leq \nu$.

To build the (X_n, Y_n) chain, it is enough to specify the transition rules and then check that (a)–(c) hold. The transition rule goes like this: Assume that the current state is (x, y), where $x \leq y$. Choose $I \in \Lambda$ at random, uniformly on Λ; the configurations x, y will only be modified at site I, if at all. Given that $I = i$, update the spins (x_i, y_i) at site i with probabilities as follows:

$$
\begin{aligned}
(--) &\longrightarrow (++) && \text{with probability } \epsilon \\
(--) &\longrightarrow (--) && \text{with probability } 1-\epsilon \\[1ex]
(-+) &\longrightarrow (++) && \text{with probability } \epsilon \\
(-+) &\longrightarrow (--) && \text{with probability } \epsilon\frac{\nu(\sigma_i^- y)}{\nu(\sigma_i^+ y)} \\
(-+) &\longrightarrow (-+) && \text{with probability } 1-\epsilon-\epsilon\frac{\nu(\sigma_i^- y)}{\nu(\sigma_i^+ y)} \\[2ex]
(++) &\longrightarrow (--) && \text{with probability } \epsilon\frac{\nu(\sigma_i^- y)}{\nu(\sigma_i^+ y)} \\
(++) &\longrightarrow (-+) && \text{with probability } \epsilon\frac{\mu(\sigma_i^- x)}{\mu(\sigma_i^+ x)}-\epsilon\frac{\nu(\sigma_i^- y)}{\nu(\sigma_i^+ y)} \\
(++) &\longrightarrow (++) && \text{with probability } 1-\epsilon\frac{\mu(\sigma_i^- x)}{\mu(\sigma_i^+ x)},
\end{aligned}
$$

where $\epsilon > 0$ is chosen sufficiently small that all of the probabilities are positive and less than 1. Note that the hypothesis (3.1) guarantees that

$$\frac{\mu(\sigma_i^- x)}{\mu(\sigma_i^+ x)} - \frac{\nu(\sigma_i^- y)}{\nu(\sigma_i^+ y)} \geq 0.$$

It is clear that the transition probabilities specified above are such that the (X_n, Y_n) chain will only visit states (x, y) such that $x \leq y$. Therefore, to complete the proof, it suffices to show that the marginal processes X_n and Y_n are aperiodic irreducible Markov chains with stationary distributions μ and ν, respectively. Consider X_n. Given that $X_n = x$, $Y_n = y$, $I_n = i$ and any specification of the past, the conditional distribution of X_{n+1} satisfies

$$X_{n+1} = \sigma_i^+(x) \qquad \text{with probability } \epsilon \qquad \text{if } x = \sigma_i^-(x),$$

$$X_{n+1} = \sigma_i^-(x) \qquad \text{with probability } \epsilon\frac{\mu(\sigma_i^-(x))}{\mu(\sigma_i^+(x))} \qquad \text{if } x = \sigma_i^+(x).$$

Since these conditional probabilities don't depend on y or the past, it follows that the process $(X_n)_{n\geq 0}$ is Markov and aperiodic irredicible. Moreover, the transition probabilities $p(\cdot,\cdot)$ for this Markov chain satisfy the detailed balance equations

$$\begin{aligned}\mu(\sigma_i^+(x))\,p(\sigma_i^+(x),\sigma_i^-(x)) &= \mu(\sigma_i^+(x))\frac{\epsilon}{|\Lambda|}\frac{\mu(\sigma_i^-(x))}{\mu(\sigma_i^+(x))})\\ &= \mu(\sigma_i^-(x))\frac{\epsilon}{|\Lambda|}\\ &= \mu(\sigma_i^-(x))\,p(\sigma_i^-(x),\sigma_i^+(x)),\end{aligned}$$

where $|\Lambda|$ denotes the cardinality of Λ. Thus μ is the stationary distribution of the Markov chain $(X_n)_{n\geq 0}$. A similar calculation shows that $(Y_n)_{n\geq 0}$ is Markov, and has stationary distribution ν. □

Proof of Proposition 7. Under μ_Λ^x, the configuration $X_{V-\Lambda} = x_{V-\Lambda}$ almost surely. Thus, it suffices to show that if $z \leq y$ then the distribution of X_Λ under μ_Λ^z is stochastically dominated by its distribution under μ_Λ^y. For this, we use the sufficient condition provided by Theorem 8. Let $\tilde{z},\tilde{y}$ be configurations that coincide with z,y, respectively, outside Λ, and such that $\tilde{z} \leq \tilde{y}$. Then

$$\begin{aligned}\frac{\mu_\Lambda^z(\sigma_i^+(\tilde{z}))}{\mu_\Lambda^z(\sigma_i^-(\tilde{z}))} &= \exp\{-2\beta J\sum_{j:j\sim i}\tilde{z}_j\}\\ &\leq \exp\{-2\beta J\sum_{j:j\sim i}\tilde{y}_j\}\\ &= \frac{\mu_\Lambda^y(\sigma_i^+(\tilde{y}))}{\mu_\Lambda^y(\sigma_i^-(\tilde{y}))}.\end{aligned}$$

Consequently, by Theorem 8, μ_Λ^z is stochastically dominated by μ_Λ^y. □

Corollary 9. *For each integer* $n \geq 1$,

$$\mu_n^+ \geq \mu_{n+1}^+, \tag{3.2}$$

$$\mu_n^- \leq \mu_{n+1}^-, \textit{ and} \tag{3.3}$$

$$\mu_n^- \leq \mu_n^+. \tag{3.4}$$

Proof. Set $\Lambda = \Lambda_n$, $\Lambda^* = \Lambda_{n+1}$, and $\Sigma = \Lambda_{n+1} - \Lambda_n$. By Proposition 4, μ_n^+ is the conditional distribution, under μ_{n+1}^+, of X_Λ given the event $X_\Sigma = z_\Sigma^+$ (recall that $z^+ =$ all pluses). By Proposition 7, this dominates the conditional distribution of X_Λ under any other event $X_\Sigma = x_\Sigma$. Therefore, μ_n^+ stochastically dominates the *unconditional* distribution of X_Λ

under μ^+_{n+1} (Exercise: Explain why.) This proves that $\mu^+_n \geq \mu^+_{n+1}$. The second inequality is similar, and the third follows directly from Proposition 7. □

Proof of Theorem 2. Since $\mu^+_n \geq \mu^+_{n+1}$, on some probability space there exist $\mathcal{X}$-valued random variables $X^{(n)}$ with marginal distributions μ^+_n such that $X^{(n)} \geq X^{(n+1)}$ for all n. (This is a consequence of Strassen's Monotone Coupling Theorem for measures on $\mathcal{X}$.) Since the components $X^{(n)}_i$ are $\pm$, it follows that the configurations $X^{(n)}$ converge coordinatewise, monotonically, to a limit X. The distribution μ of X must be the weak limit of μ^+_n. A similar argument shows that the measures μ^-_n converge to a limit μ^-, and that $\mu^- = \mu^+ \circ \rho$. That $\mu^+ \geq \mu^-$ follows from the fact that $\mu^+_n \geq \mu^-_n$. The transition invariance of the measures μ^+, μ^- can alos be proved by a stochastic comparison argument (the details are omitted for now). □

4. Peierls' Contour Argument

Peierls' argument is based on the observation that the Ising Hamiltonian H_Λ defined by (1.5) depends only on the number of $+/-$ nearest neighbor pairs in the configuration:

$$H_\Lambda(x) = -2JL_\Lambda(x) + C_\Lambda \qquad \text{where} \tag{4.1}$$

$$L_\Lambda(x) = \sum_{\substack{i\in\Lambda, j\in V\\ i\sim j}} \delta(x_i, -x_j), \tag{4.2}$$

with $\delta(\cdot,\cdot)$ being the Kronecker delta function. Evaluation of $L_\Lambda(x)$ can be accomplished by partitioning the vertices of $\Lambda \cup \partial\Lambda$ into (maximal) connected clusters of $+$ spins and $-$ spins in x, as in Figure ; $L_\Lambda(x)$ is the number of edges in $\Lambda \cup \partial\Lambda$ connecting $+$ clusters to $-$ clusters. For two-dimensional graphs, $L_\Lambda(x)$ may be evaluated by drawing *boundary contours* around the connected clusters, as shown in the following lemma. For the remainder of this section, assume that G is the standard two-dimensional integer lattice $\mathbb{Z}^2$.

Lemma 10. *For each vertex $i \in \Lambda \cup \partial\Lambda$, let $K_i = K_i(x)$ be the maximal connected set of vertices j such that sites i and j have the same spin in configuration x. Then for any two vertices i, j such that $K_i \neq K_j$ there is a simple closed curve $\gamma = \gamma_{i,j}$, called a* boundary contour *(possibly empty) separating K_i from K_j. The curve γ is a finite union of horizontal and vertical segments in the dual lattice. Each such segment bisects an edge connecting a vertex in K_i to a vertex in K_j.*

Proof. The curve γ may be constructed using a "maze-walking" algorithm. Begin by choosing an edge e connecting K_i to K_j (if there is one), and let the first segment γ_1 of γ be a perpendicular bisector of e. Define (oriented) segments γ_n, for $n = 2, 3, \ldots$, inductively, in such a way that if one traverses the segment γ_n then a vertex of K_i is on the right and a vertex of K_j is on the left. Eventually the sequence γ_n will enter a cycle. This cycle must include all of the segments γ_n because otherwise the right/left rule would be violated somewhere. Therefore, the cycle determines a closed curve. This closed curve must completely separate the regions K_i and K_j, because otherwise one of them could not be connected. Consult your local topologist for further details. □

Corollary 11. $L_\Lambda(x) = \sum_{i,j} |\gamma_{i,j}|$. □

Assume now that the region Λ is a square. Fix a vertex $i \in \Lambda$, and let $x \in \mathcal{X}$ be a configuration such that $x_{\Lambda^c} = z^+_{\Lambda^c}$. If $x_i = -1$, then it must be that the vertex i is completely surrounded by a contour that separates it from $\partial\Lambda$, as the vertices outside Λ all have + spins. In particular, the boundary of the connected cluster $K_i = K_i(x)$ of $-$ spins to which vertex i belongs contains a unique contour $\gamma := \gamma_{i,\infty}$ that separates K_i from the exterior Λ^c of the square Λ. (Note that this contour may in general surround other connected components K_j.) Define C_γ ot be the set of all vertices j that are surrounded by γ; define configuration $\tilde{x}$ to be the configuration obtained from x by flipping all spins inside $\gamma_{i,\infty}$

$$(\tilde{x})_j = \begin{cases} -x_j \text{ if } j \in C_\gamma \\ +x_j \text{ if } j \notin C_\gamma. \end{cases} \tag{4.3}$$

Lemma 12. *Let $x \in \mathcal{X}$ be any configuration such that $x_i = -1$, and let $\gamma = \gamma_{i,\infty}$ be the contour that separates K_i from Λ^c. If $\tilde{x}$ is the configuration defined by (4.3), then*

$$\frac{\mu^+_\Lambda(x)}{\mu^+_\Lambda(\tilde{x})} = \exp\{-2\beta J|\gamma|\}. \tag{4.4}$$

Proof. For all nearest neighbor pairs j, k, the spin products $x_j x_k$ and $\tilde{x}_j \tilde{x}_k$ are related as follows:

$$\begin{aligned} x_j x_k &= -\tilde{x}_j \tilde{x}_k && \text{if } j,k \text{ are on opposite sides of } \gamma; \\ &= +\tilde{x}_j \tilde{x}_k && \text{otherwise.} \end{aligned}$$

Consequently,

$$H_\Lambda(\tilde{x}) - H_\Lambda(x) = 2J|\gamma|.$$

□

Lemma 13. *The mapping $x \mapsto (\tilde{x}, \gamma)$ is one-to-one.*

Proof. Given $(\tilde{x}, \gamma)$, one can recover x by negating in the region C_γ surrounded by γ. □

Proposition 14. *For each $\beta > 0$ and each square Λ containing vertex i,*

$$\mu^+_\Lambda(X_i = -1) \le \sum_{n=4}^{\infty} n 3^n e^{-2\beta J n}. \tag{4.5}$$

Proof. On the event $X_i = -1$ the connected cluster K_i of $-$ spins containing the vertex i must be separated from the connected cluster K_∞ of + spins containing the vertices on the boundary $\partial\Lambda$. Let Γ be the boundary contour of $K_i = K_i(X)$ that separates K_i from K_∞. By Lemmas 13 and 12, the μ^+_Λ−probability that $X_i = -1$ and $\Gamma = \gamma$ satisfies

$$\mu^+_\Lambda(X_i = -1 \text{ and } \Gamma = \gamma) \le \exp\{-2\beta J|\gamma|\}.$$

Consequently,

$$\mu^+_\Lambda(X_i = -1) \le \sum_\gamma \exp\{-2\beta J|\gamma|\},$$

where the sum is over all contours in the (dual) integer lattice surrounding i. To estimate the number of such surrounding contours of length k, observe that any such contour must intersect the vertical upward ray emanating from vertex i at some point within distance k of i. Starting from this intersection point, the contour is formed by attaching successive line segments, one at a time; at each stage, there are at most 3 such segments to choose from. Hence, the number of surrounding contours of length k is at most $k3^k$. The estimate (4.5) now follows easily. □

Since the sum on the right side of inequality (4.5) is less than 1/2 for all sufficiently large values of β Proposition 14, together with Theorem 2, implies that (1.8) holds at low temperature.

5. The High Temperature Limit

In this section we shall prove the following proposition, which implies that (1.9), and hence also (1.11), hold at high temperature.

Proposition 15.

$$\tanh(-4\beta J) < 1/4 \Longrightarrow \lim_{n\to\infty} \mu_n^+\{X_i = -1\} = 1/2. \tag{5.1}$$

The proof, unlike Peierls' argument, does not really depend on planarity of the underlying graph, and may be extended not only to the higher-dimensional integer lattices but to arbitrary vertex-regular graphs (graphs with the property that all vertices have the same number of incident edges). We shall only discuss the case $G = \mathbb{Z}^2$.

5.1. **Bernoulli-p Site Percolation.** The upper bound of 1/4 in (5.1) for $\tanh(-4\beta J)$ emerges from the world of *site percolation.* In its simplest incarnation, site percolation has to do wtih the connectivity properties of the random graph obtained from the two-dimensional integer lattice by tossing a $p-$coin at every vertex, then erasing the vertex, and all edges incident to it, if the coin toss results in a T. *Percolation* is the event that the resulting subgraph of $\mathbb{Z}^2$ has an infinite connected cluster of vertices, equivalently, that $\mathbb{Z}^2$ has an infinite connected cluster of $H-$vertices.

Proposition 16. *If $p < 1/4$ then percolation occurs with probability* 0.

Proof. It is enough to show that for any vertex i, the probability that i is part of an infinite connected cluster of Hs is zero. Denote by K the (maximal) connected cluster of vertices containing i at which the coin toss is H. Define sets $F_0, F_1, F_2, \dots$ inductively as follows: Let $F_0 = \{i\}$, and for each $n \geq 0$ define F_{n+1} to be the set of all vertices at which the coin toss is H that are nearest neighbors of vertices in F_n and that have not been listed in $\cup_{j=0}^n F_j$. I claim that

$$E|F_{n+1}| \leq 4pE|F_n|. \tag{5.2}$$

To see this, observe that, for each vertex $j \in F_n$ there are at most 4 vertices adjacent to j that can be included in F_{n+1}. For each of these, the conditional probability that it is included in F_{n+1}, given the coin tosses that have resulted in constructing $F_0, F_1, \dots, F_n$, is at most p; consequently, the expected number that are included is no more than $4p$.

The cluster K is the union of the sets $F_0, F_1, \dots,$ and so its expected cardinality is bounded by $\sum_n E|F_n|$. By inequality (5.2), if $4p < 1$ then $E|K| < \infty$, in which case K is finite with probability 1. □

Fix a site $i \in V = \mathbb{Z}^2$, and denote by Λ_n the square of side $2n+1$ centered at the origin. Say that i percolates to $\partial\Lambda_n$ if the connected cluster of Hs containing site i extends to the boundary of Λ_n, equivalently, if there is a path of $H-$vertices from i ot the boundary of Λ_n. Denote this event by $A(i,n)$.

Corollary 17. *If $p < 1/4$ then $\lim_{n\to\infty} P_p(A(i,n)) = 0$ for each site i.* □

5.2. **Monotone Coupling of Gibbs States.**

Proposition 18. *Fix $\beta > 0$, and set $p = \tanh(-4\beta J)$. On some probability space $(\Omega, \mathcal{F}, P)$, there exist $\mathcal{X}-$valued random variables $Z^{(n)} \leq Y^{(n)}$ with marginal distributions μ_n^- and μ_n^+, respectively, such that*

$$P(Z_i^{(n)} \neq Y_i^{(n)}) \leq P_p(A(i,n)), \tag{5.3}$$

where $P_p(A(i,n))$ is the probability that site i percolates to $\partial\Lambda_n$ in Bernoulli-p site percolation.

Observe that this proposition and Corollary 17 imply Proposition 15, because Corollary 17 implies that the probability that site i percolates to $\partial\Lambda_n$ converges to zero if $p = \tanh(-4\beta J) < 1/4$. The proof of Proposition 18 will use the following lemma, which explains the occurrence of the quantity $\tanh(-4\beta J)$.

Lemma 19. *For any two configurations z, y such that $z \leq y$, and for any finite regions $\Sigma \subset \Lambda$ and any site $i \in \Lambda - \Sigma$,*

$$\mu_\Lambda^+(X_i = +1 \,|\, X_\Sigma = y_\Sigma) - \mu_\Lambda^-(X_i = +1 \,|\, X_\Sigma = z_\Sigma) \leq \tanh(-4\beta J). \tag{5.4}$$

Proof. In view of the Markov property (Proposition 4), it suffices to show that for any two configurations x, y,

$$|\mu_{\Lambda-i}^x(X_i = +1) - \mu_{\Lambda-i}^y(X_i = +1)| \leq \tanh(-4\beta J). \tag{5.5}$$

The two probabilities in (5.5) are easily calculated, as they depend only on the spins x_j, y_j at the four nearest neighbors of i. The maximum discrepancy occurs when the $x-$spins are all $+1$ and the $y-$spins are all -1: it is $\tanh(-4\beta J)$. □

Proof of Proposition 18. Fix n, and abbreviate $\Lambda = \Lambda_n$, $Z = Z^{(n)}$, and $Y = Y^{(n)}$. There are $N := (2n+1)^2$ sites in the square Λ: label these $1, 2, \dots, N$ in order, starting from the sites at distance 1 from $\partial\Lambda$, then proceeding through the sites at distance 2 from $\partial\Lambda$, and so on, but omitting site i until the very end, so that it is listed as site N. We will construct Z, Y one site at a time, proceeding through the sites $1, 2, \dots, N$ in order, using independent uniform-$(0,1)$ random variables $U_1, U_2, \dots, U_N$. (The values $Z_i = -1$ and $Y_i = +1$ are determined by the requirement that the marginal distributions of Z and Y are μ_n^- and μ_n^+, respectively.)

To construct Z_1, Y_1, use the uniform U_1 to choose $\pm$ spins from the Gibbs distributions $\mu_n^-(X_1 \in dx)$ and $\mu_n^+(X_1 \in dx)$. By Proposition 7, these distributions are stochastically ordered, so the assignment of spins may be done in such a way that $Z_1 \leq Y_1$. Moreover,by Lemma 19,

$$|\mu_n^-(X_1 = +1) - \mu_n^+(X_1 = +1)| \leq p,$$

so the probability that $Z_1 < Y_1$ is no larger than p. Hence, the assignment of spins may be dome in such a way that the event $Z_1 < Y_1$ is contained in the event $U_1 < p$.

Now suppose that Z_j, Y_j, for $1 \leq j < m$, are defined. Conditional on the event $Z_j = z_j$, $Y_j = y_j$, with $z_j \leq y_j$, use the uniform random variable U_m to assign the spins $\pm$ at Z_m and Y_m using the conditional distributions

$$Z_m \sim \mu_n^-(X_m \in dx \,|\, X_j = z_j \;\forall\, 1 \leq j < m) \text{ and}$$
$$Y_m \sim \mu_n^+(X_m \in dx \,|\, X_j = y_j \;\forall\, 1 \leq j < m).$$

Since $z_j \leq y_j$, these conditional distributions are again stochastically ordered, by Proposition 7 and Proposition 4; consequently, the assignment of spins may be done in such a way that $Z_m \leq Y_m$. Moreover, by Lemma 19, the conditional probability that $Z_m < Y_m$, given the assignments $Y_j = y_j$ and $Z_j = z_j$ for $1 \leq j < m$, is, once again, no larger than p; consequently, the assignments may be done in such a way that the event $Z_m < Y_m$ is contained in the event $U_m < p$.

It remains to show that the inequality (5.3) holds. By construction, $Z_j < Y_j$ can only occur if $U_j < p$. Moreover, by Corollary 6, if in the course of the construction it develops that $Z_j = Y_j$ for all sites j in a contour that surrounds site i, then it must be the case that $Z_i = Y_i$, as the conditional distributions of the spins Z_k and Y_k will coincide for all sites k after completion of the contour. Thus, $Z_i \neq Y_i$ can only occur if there is a connected path of sites j leading from site i to $\partial\Lambda$ along which $U_j < p$. But this is precisely the event that site i percolates to $\partial\Lambda$ in Bernoulli-p percolation. □