Simulating the 2D Ising Model

Connor Tracy

Supervisor: Dr D Proment

September 2020

Contents

Abstract

This text will document the basic theory of the Ising model and numerical simulations of the Ising model on the two dimensional regular square lattice.[1] These simulations will be performed with the Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm. The accuracy of these simulations will be tested by estimating a critical exponent of the system. All proofs and tables will be found in the relevant appendices, along with an additional section on universality classes. The theory and simulation script developed during this project will serve as a first step towards a planned paper by this project's supervisor Dr D Proment of the University of East Anglia and myself as co-authors, in which we will apply the simulation techniques to answer questions on the Ising model on general graphs. The Python code for this project can be found on my GitHub at `https://github.com/Connor-Tracy/Ising-Simulation`.

1 The Ising Model

The Ising model is a model of magnetism which considers neighbouring sites, each of which being in one of two states, namely dipole atomic spins oriented either up or down. Adjacent sites interact according to their interaction potentials depending on their relative orientation. Moreover, we implement a mathematical trick to subject all sites to an external field to break the symmetry of the system. In the limit as the external magnetic field tends to zero, spontaneous magnetisation may occur[1], thus allowing the Ising model to be as Giovanni Gallavotti says in [20] "the simplest nontrivial example of a system undergoing phase transitions", in two or more dimensions. When this model is exactly solvable, one can compute thermodynamic quantities ("observables"), while numerical simulations provide a powerful tool for understanding the behaviours in more complicated scenarios, such as on a general graph. The simplicity of the model provides universality[2]; this model and its properties will apply to phase transitions being studied in a wide range of real-world applications from neuroscience[11], solids and gases[1, 5], operations research[24], finance[25] and bioinformatics[26] to name[9] just a few.

The *principle of minimum energy* states that a physical system will favour configurations with lower total energy[12]. A material modelled by a graph of spin sites is called *ferromagnetic* if interacting (nearest-neighbour) sites with agreeing spin directions have a lower potential interaction energy than those that disagree, or *antiferromagnetic* if this inequality is reversed. Hence by the principle of minimum energy, the model favours local alignment/anti-alignment of adjacent spins[45], since this results in a lower contribution to the total energy

[1] The author aims to strike a balance between completeness, brevity and personal learning. As such, the References section serves to provide an honest account of my sources and so do not reflect the author's understanding of proper academic referencing practices.

[2] A note on universality: Models are categorised by universality classes, according to their behaviour about criticality in the thermodynamic limit, in which properties including a critical exponent are the same. The Ising model's family of universality classes includes a wide range of other models from this ferromagnetic phase transition to liquid-gas transitions[36].

of the system. These microscopic effects have consequences on a statistical mechanical macroscopic scale, for example in magnetisation. Heat disturbs the configuration of the system and this allows[6] the neighbouring sites to flip their spin orientation, thus the material has the ability to tend to configurations in accordance with the principle of minimum energy. The study of this system is therefore statistical mechanical in nature and if if on a macroscopic scale the sites' spins are oriented identically, then the system represents a ferromagnetic material.

In 1924, Ernst Ising solved the one-dimensional model[15], proving using periodic boundary conditions that no phase transitions can occur, while the two-dimensional case was proven to include phase transitions by Lars Onsager[16] in 1944. This report will show the analytical proof of the one-dimensional case and then use numerical simulations to study the model in first two-dimensions and then in a graph more generally. Thermal agitation allows for spins to flip spontaneously, with the tendency towards neighbouring spins to orient in parallel or in anti-parallel being a counteracting force to bring long-range correlations of spins. The paper [45] explains these two forces lead to a critical *Curie temperature* at which there is large-scale disorder above this temperature and ferromagnetic order below this temperature.

1.1 Definitions and Heuristics

The Ising model imagines a $d-$dimensional regular lattice of N vertices called *sites*, the i^{th} state taking value $\sigma_i \in \{-1, +1\}$ to represent the spin orientation of an atomic dipole at this site[3]. There is an *interaction potential* for the spin-spin interactions which occurs between each of the site's $2d$ adjacent *nearest-neighbours* governed by a *coupling constant* J. When sites i, j are nearest-neighbours, we denote this by $\langle i, j \rangle$ and the potential energy is determined by the value of $\sigma_i \sigma_j$, i.e. whether the neighbouring sites' spins are parallel. A *configuration* is simply one of the 2^N possible arrangements of spins for every site. When a configuration has its statistical average magnetisation of the sites' spins non-zero, it is said to be a *ferromagnetic* configuration, or *antiferromagnetic* if the nearest neighbours' spins are sufficiently disordered (so neighbouring sites are mostly antiparallel). The nearest neighbours prefer to align in parallel when $J > 0$ and antiparallel when $J < 0$, called *ferromagnetic* and *antiferromagnetic* respectively, with this effect strongest for larger absolute values of the coupling constant J. In addition to spin-spin interactions with *coupling constant* J, there is also a constant external magnetic field[4] μ interacting with the sites. A positive or negative value for the magnetic field μ encourages sites to take the positive or negative spin orientations respectively. This is a useful tool to break the symmetry of the system and allow one to determine whether spontaneous magnetisation is possible as μ is made to go to zero.

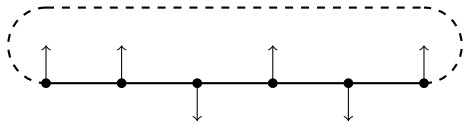
[3]Some authors use S_i instead of σ_i

[4]Some authors use H for the coupling constant and other authors also use H for the magnetic field.

The Hamiltonian[5] $\mathcal{H}$ describes the model's (free) energy for a given configuration. This is an important part of constructing the *partition function* (which is really like a generating function on exponential bases) from which one can calculate *observables* such as energy, magnetisation, heat capacity etc. Observables are macroscopic statistical expectation quantities arising from the microscopic arrangements. These macroscopic averages are denoted $\langle f \rangle$ for the expectation of f, calculated as a sum over the possible configurations σ with respect to *configuration probabilities* $P(\sigma)$. The *thermodynamic limit* refers to taking the limit as N (and the volume) tends to infinity in our derived formulae for some observable. Of particular interest is the critical *Curie temperature* at which the system spontaneously breaks its symmetry to enter a *phase transition* to a *ferromagnetic* material, when on a macroscopic level the spins align in the same direction. *Transition matrices* are another important tool which will be useful in deriving the free energy of the system to then calculate the partition function.

1.2 One-Dimensional Results

In one dimension, the Ising model concerns a one-dimensional lattice of N sites in a line. We assume 'periodic boundary conditions' (pbc), as did Ernst Ising in his original thesis[15], so we allow the two endpoints of the line to interact which joins the ends by setting $\sigma_{N+1} = \sigma_1$ to form a cycle or loop. This does not change the thermodynamic limit since it simply adds one additional interaction between the 1^{st} and N^{th} sites, though it does simplify the calculations. Boundary conditions do matter in general. We will later generalize the below theory in terms of nearest neighbours in higher dimensions.[6]



The *energy* of the system is given by the *Hamiltonian*[7]

$$\mathcal{H}(\sigma) := -J \sum_{i=1}^{N} \sigma_i \sigma_{i+1} - \mu \sum_{i=1}^{N} \sigma_i. \tag{1}$$

This sums over the energy of each (nearest-neighbour) pair's interaction potentials and each site's interaction with the field.

[5] Some authors use $E(\sigma)$ to free μ or H for the magnetic field

[6] The correlation length grows exponentially[23] near the critical temperature until it exceeds the size of the lattice, when the model fails to represent an infinite lattice. The paper [45] explained that conformal field theory helps describe how boundary conditions (bc's) affect the Ising model. For other explorations into the effect of bc's, see [21, 22, 45]. The paper also demonstrates that there are various techniques available to study the scaling limit, each applicable to different bc's. It was recently proven the scaling limit is conformally invariant.

Denoting the set of all possible configurations of a d-dimensional[7] lattice with N sites by $\mathcal{S}_d(N)$, the *partition function* Z is defined

$$Z := \sum_{\sigma \in \mathcal{S}_d(N)} \exp\{-\beta \mathcal{H}(\sigma)\} \tag{2}$$

$$= \sum_{\sigma \in \mathcal{S}_d(N)} \exp\left\{\beta J \sum_{i=1}^{N} \sigma_i \sigma_{i+1} + \beta\mu \sum_{i=1}^{N} \sigma_i\right\} \tag{3}$$

$$= \sum_{\sigma \in \mathcal{S}_d(N)} \exp\left\{\beta J \sum_{i=1}^{N} \sigma_i \sigma_{i+1} + \frac{\beta\mu}{2} \sum_{i=1}^{N} (\sigma_i + \sigma_{i+1})\right\}, \tag{4}$$

where the *inverse temperature* $\beta \geq 0$ is defined $(kT)^{-1}$ for k the Boltzmann constant[8].

The *configuration probability* $P_\beta(\sigma)$ of a configuration $\sigma \in \mathcal{S}_d(N)$ is given by

$$P_\beta(\sigma) = e^{-\beta \mathcal{H}(\sigma)}/Z,$$

where the partition function Z is the normalising constant to obtain a valid probability distribution, as Z is the sum of the $e^{-\beta\mu(\sigma)}$ over all configurations $\sigma \in \mathcal{S}_d(N)$.[9] This is the probability mass function of the *Boltzmann distribution*. The expectation for an observable f is then given by

$$\langle f(\beta) \rangle := \sum_{\sigma} f(\sigma) P_\beta(\sigma).$$

Since many observables will be calculated by applying operators to the partition function, it will be useful to find a more closed form expression.

Proposition 1.1. The partition function Z is equal to

$$Z = \lambda_1^N + \lambda_2^N,$$

where

$$\lambda_1, \lambda_2 = e^{\beta J}\left(\cosh(\beta\mu) \pm \sqrt{\sinh^2(\beta\mu) + e^{-4\beta J}}\right). \tag{5}$$

Remark 1 ([2, 4]). In the case where $\mu = 0$, the partition function immediately factorises to give $Z = (2\cosh(H))^N$, avoiding any need for the transition matrix method[10].

[7] When the dimension and number of vertices is unambiguous by context, we may instead write $\mathcal{S}$. When we later encounter the more general case of a graph G, it will be natural to denote the set of possible configurations on the graph by $\mathcal{S}_G$.

[8] The last equality simply expresses the partition function in a more symmetrical form which will be useful later.

[9] For the Monte Carlo simulations which will be discussed later, the technique only require a function proportional to the true probability mass function.

[10] Since we are working towards results using $\mu \to 0$ as a mathematical trick, it may be noteworthy this specific result can be derived trivially when $H \to 0$.

Definition 1.2. The *(Helmholtz) free energy* F is the energy available to do work and can be defined[41] to be[11]

$$F := -\frac{1}{\beta}\ln(Z). \tag{6}$$

Proposition 1.3. The *mean energy*[12] (*internal energy*) is the expected value of the Hamiltonian over all possible configurations, since the Hamiltonian describes the total energy of a given configuration. The mean energy can be obtained[3] from the partition function Z via

$$\langle \mathcal{H} \rangle = \sum_{\sigma \in \mathcal{S}} \mathcal{H}(\sigma) P_\beta(\sigma) = -\frac{\partial \ln(Z)}{\partial \beta}. \tag{7}$$

Proposition 1.4. For $\lambda_1 > \lambda_2$ the larger of the two eigenvalues of the transition matrix $\mathbf{T}$, the *free energy per site* $f := \lim_{N\to\infty} F/N$ in the thermodynamic limit is

$$f = -\frac{1}{\beta}\ln(\lambda_1) = -J - \frac{1}{\beta} ln\left(\cosh(\beta\mu) + \sqrt{\sinh^2(\beta\mu) + e^{-4\beta J}}\right).$$

Proposition 1.5. Given the free energy F, the *magnetization* M and specific magnetization m satisfy

$$m := \frac{M}{N} := \frac{1}{N}\left\langle \sum_{i=1}^{N} \sigma_i \right\rangle = -\frac{1}{N}\frac{\partial F}{\partial \mu}. \tag{8}$$

Corollary 1.6. The specific magnetization is

$$m = \frac{\sinh(\beta\mu)}{\sqrt{\sinh^2(\beta\mu) + e^{-4\beta J}}}$$

in the thermodynamic limit as $N \to \infty$.

Theorem 1. *The one-dimensional Ising model possesses no phase transitions.*

Remark 2. Note the continuity of m contradicts the non-differentiability of the magnetization at the critical Curie temperature so this also shows no phase transition can occur in one dimension.[13]

[11] Equivalently, $F = E - ST$ for S the entropy as discussed in [44].

[12] Derivation of our definition of the free energy is given by [44], as well as the formulae for both the free energy (p19) and the average energy (p15) in one source. Another source presenting both the free energy and the average energy is on page 5.

[13] This is because we know from the earlier theory that magnetization is a second order phase transition, so its first derivative should be discontinuous, unlike infinite-order phase transitions in which this argument would not apply.

1.3 Critical Exponents

Important objects of study within models are their *critical exponents*, which describe the behaviour of the system about criticality (phase transitions). This report will use the estimation of one of these critical exponents to assess the accuracy of the simulations by comparing an estimated value to the analytical solution. The wider significance of critical exponents is due to universality classes, which can be understood by a consideration of the Renormalization Group, a discussion of which can be found in Appendix A.

An *order parameter* is a measure of orderedness across the boundaries of a system during a phase transition in which the function goes from being zero to non-zero[28]. The order parameter in the Ising model is magnetization, since parallel spins give rise to a ferromagnetic material. At most phase transitions, order parameters follow functions which are non-analytic at the critical point, and are of a particular *order*[14]. A *first order* phase transition is characterised by involving *latent heat*[29] causing a discontinuity in the energy since a fixed amount of heat is released or absorbed with no increase in temperature. A *second order* phase transition is continuous (the order parameter goes to zero continuously) but instead characterised by an infinite correlation length (fluctuations are correlated over all distances), power law behaviour about the critical point, and divergent specific heat (susceptibility). It is these decaying power laws which are described by critical exponents, classifying the second order phase transitions. In general, the operator applied to the partition function to derive an n^{th} order parameter is an n^{th} order derivative. Examples of a first order phase transition include the melting of ice and the boiling of water, while the superfluid transition and some superconductor transitions are second order phase transitions, alongside the Ising model's ferromagnetic transition[29].

The critical exponents classify the second order phase transitions as determine the decaying power laws about the critical point. We write $\tau := |T - T_c|/T_c$ for the *reduced temperature*, since the power law occurs about the critical temperature T_c. Then most critical exponents are x in the power law of some observable X, $X \propto \tau^x := (|T - T_c|/T_c)^x$. The main 6 critical exponents are denoted by the Greek letters α, β, γ, δ, η and ν, although δ instead gives a relation between J and another critical exponent. Table 1 [5, 28, 30] in Appendix A presents the observables following the power law of each exponent, the interpretation of each property in the Ising model and the value[15] of each critical exponent in each dimension.

[14] Infinite order phase transitions exist and instead involve a smooth function becoming non-zero[35]. An example of an infinite-order phase transition is the Berezinskii–Kosterlitz–Thouless transition (BKT transition) of the 2-dimensional *XY model*. The 2016 Nobel Prize for physics was awarded[43] for work on this phase transition.

[15] We give the values to 3 significant figures for the 3-dimensional case because theoretical and numerical results for the values have changed[32] over time as well as becoming more precise, though the best estimates for ω are not sufficiently precise.

2 Monte Carlo Theory

It is often difficult to sample from a given distribution, but Monte Carlo simulations provide important tools to do this. To simulate the configurations of our Ising model over time, we use the Metropolis-Hasting algorithm[47, 48] which is an example of an acceptance-rejection algorithm.

2.1 Acceptance-Rejection Algorithms

An acceptance-rejection algorithm takes a given *proposal distribution* from which we know how to sample, and generates samples from a *target distribution* by choosing whether to accepting a *candidate value* y from the proposal distribution with probability dependent on the ratio of the two distribution laws at this candidate value. This subsection will state the general acceptance-rejection algorithm, prove a sufficient condition for the algorithm to indeed sample from the target distribution, as well as proving a basic result concerning the efficiency of the algorithm. Convergence is best when the proposal distribution is of the same shape as the target distribution.

Remark 3. It is always assumed that it is possible to generate a sample from the uniform distribution $U[0,1]$ on the unit interval.

Algorithm 1 (Acceptance-Rejection)**.** Generate a sample x from the continuous or discrete random variable X with target law f_X, given a continuous or discrete proposal distribution Y with law h_Y and the same support $\mathcal{X}$ as X.

Step 1 Find a constant $C \in \mathbb{R}$ s.t. $Ch_Y(x) \geq f_X(x)\ \forall\ x \in \mathbb{R}$; optimal when $C \approx 1$.

Step 2 Generate a candidate sample y from the proposal distribution Y.

Step 3 Generate a sample u from $U[0,1]$.

Step 4 Calculate the acceptance rate[16] $a(y) := f_X(y)/Ch_Y(y)$.

Step 5 If $u \leq a(y)$, return $x = y$. Otherwise go to Step 2.

Theorem 2 (Acceptance-Rejection)**.** *If $\exists\ C \in \mathbb{R}$ constant s.t. $f_X(x)/h_Y(x) \leq C\ \forall\ X \in \mathcal{X}$, then Algorithm 1 samples once from a random variable X with law f_X, given a proposal law h_Y with acceptance rate $a(y) = f_X(y)/Ch_Y(y)$.*

Corollary 2.1. The number of loops of the acceptance-rejection algorithm to generate one sample will itself be distributed $Geo\,(1/C)$, thus the average number of loops will be C. The optimal choice of C will then be as close as possible to 1, i.e. when the proposal law is as close as possible to being equal to the target distribution: $h_Y(x) \approx f_X(x)\ \forall\ x \in \mathbb{R}$.

[16] This $a(y)$ is the rate at which the algorithm accepts a candidate value and thereby describes the efficiency as the proportion of candidate samples which are discarded.

2.2 The Metropolis-Hastings Algorithm

The Metropolis-Hastings is a specific type of acceptance-rejection algorithm originally proposed by Metropolis in 1953 [47] and revised by Hastings in 1970[48]. The Metropolis algorithm creates a Markov chain of configurations with stationary distribution equal to a target distribution π and accepts a transition from configuration a to b with some probability (in the case of a symmetric proposal distribution) $\mathbb{P}(a \to b) = min\,(1, \pi(b)/\pi(a))$. We will see in this section's proofs that the probability of jumping between two configurations depends on the energy change as $\curvearrowleft \to \curvearrowleft' = \max 1, \exp^{-E/kT}$. The Hastings revision multiplies the ratio by another ratio (seen in the full algorithm below) to allow for non-symmetric proposal distributions. Here the proposal distribution in fact only needs to be some function proportional to the target distribution, thereby eliminating the potentially difficult task of calculating the normalisation constant.

The usual method to estimate an observable is *direct sampling*: sampling values from the state space uniformly at random and taking the simple mean value of the observable. In the case of the Ising model however, the Markov steps constituting each successive simulation only differ in at most one spin site, a small change in N sites, moreover the probability of acceptance is small for larger energy increases. Thus once the chain finds itself on a probable configuration, it is very unlikely to jump to a configuration which happens to be particularly unlikely. Additionally, this construction also leads to nearby steps being correlated. Unfortunately, rare configurations can contribute significantly to the long-run average[3] which may be used to estimate an observable's expectation value. This rare configuration can mean the long-run average does not fluctuate about the true mean value and instead requires an extraordinarily rare configuration to bring the average close to the true expected value. An analogy given by Krauth in [3] is that of finding a needle by random selection from a haystack or hoping for a hole-in-one by driving golf balls at random. The golf analogy is particularly fitting as one could relate direct sampling from the space of possible configurations with dropping golf balls in random locations, while Markov-chain sampling is more akin to the driving of the balls across the green from each place they land. The rare event in which the hole is made or the needle is found massively changes the long-run average despite consistent previous results suggesting some other average value (zero in these cases). In fact, the expectation calculation means only configurations with value $X \approx \langle X \rangle$ contribute, but that this is an exponentially small fraction of the space of possible configurations. It is for this reason that direct sampling is not an appropriate method for Monte Carlo simulations and instead we require *importance sampling* to sample the rare configurations sufficiently often.

The Metropolis-Hastings algorithm addresses this issue of rare events by using importance sampling in which the probability laws are conditionally reweighted[17].

[17] Technically, this method addresses an issue of infinite variance of a stochastic integral by a finite sum approximation[3]. The product of the configuration's probability and the value of the observable remains unchanged, but the variance of the observable is reduced via Gaussian error propagation.

In direct sampling we choose configurations uniformly and estimate the observable weighting them with the probability $\exp^{-E/kT}$, while in importance sampling we choose configurations of energy E with probability $\exp^{-E/kT}$ and weight the observable equally. Thus the same long-run estimate is reached but the Markov chain is now able to preferentially transition to configurations with large values for the product of the configuration's probability and the value of the observable, rather than simply the most likely configurations. Now that the products are of the same order of magnitude, all configurations will contribute to the simple mean calculation, thereby eliminating the overlap problem from direct sampling.

Algorithm 2 (Metropolis-Hastings Algorithm)**.** [49] Generate the next configuration in the Markov Chain of samples from a target distribution with law π, given a current configuration x_i and a function $f \propto \pi$. Assume we can sample from the unnormalised proposal distribution with law $g(x'|x_t)$.

Step 1 Generate a candidate sample x' according to the law $g(x'|x_t)$.

Step 2 Generate a sample u from $U[0,1]$.

Step 3 Calculate[18] the acceptance rate $\alpha(x', x_t) := \min\left\{1, \dfrac{\pi(x')g(x', x_t)}{\pi(x_t)g(x_t, x')}\right\}$.

Step 4 If $u < \alpha(x', x_t)$, then set $x_{t+1} = x'$, otherwise set $x_{t+1} = x_t$.

Remark 4. This algorithm works because the resulting Markov chain has the target distribution as its stationary distribution. Notice it is the presence of the $\dfrac{g(x', x_t)}{g(x_t, x')}$ factor which accounts for the non-symmetric distributions, while the original Metropolis algorithm on symmetric distributions has this ratio equal to 1. Notice also that a candidate sample more likely than the current sample (i.e. if it has lower energy) configuration will give $\dfrac{\pi(x')g(x', x_t)}{\pi(x_t)g(x_t, x')} > 1$, so the algorithm always accepts a more likely configuration.

We now introduce the concept of ergodicity of a Markov process and the detailed balance property of transition probabilities, as these will be necessary to prove Theorem 3 (Metropolis-Hastings). It is not essential to understand these concepts to apply the Metropolis-Hastings algorithm.

Definition 2.2. A Markov process's *transition probabilities* $P(x_{t+1}|x_t)$ give the probability of the process jumping from state x_t at time t to state x_{t+1}. A distribution $\pi(x)$ is in *detailed balance* when every transition is *reversible*, that is when

$$\pi(x_t)\pi(x_{t+1}|x_t) = \pi(x_{t+1})\pi(x_t|x_{t+1}). \tag{9}$$

A Markov process is *ergodic* if every state is aperiodic and positive recurrent (or aperiodic and irreducible), or equivalently if

$$\exists\, T > 0 :\ P(x_t = j | x_1 = i) > 0 \ \forall\ i, j > 0 \ \forall\ t > T.$$

This is to say it is always possible to reach any state from any other state in some finite number of steps.

Theorem 3. *The Metropolis-Hastings algorithm samples from the target distribution in the limit.*

[18] Calculate the ratio $\pi(x')/\pi(x_t)$ via the ratio of $f \propto \pi$.

2.3 Optimization

There are several ways to improve a MCMC beyond the standard Metropolis-Hastings algorithm: parallel programming improves efficiency[19], while discarding some steps can help reduce the effect of autocorrelation. These considerations are particularly important for our analysis of critical exponents because it is precisely at the critical phase transition point where the correlation length diverges thus diverging the autocorrelation time, this phenomenon being known as *critical slowing down*. In one dimension, there are usually alternative methods available to sample from a distribution without the issue of autocorrelation, however in multivariate cases, a Markov Chain Monte Carlo approach is often necessary.

Parallel programming is a powerful approach to improving the efficiency of the simulations by running multiple streams of computations simultaneously. For example, this can be used to run simulations for different temperatures simultaneously to be used to generate plots for visually identifying the critical temperature. An explanation of the *checkerboard algorithm* is deferred to Appendix A. The method of applying parallization by exploiting the nature of nearest-neighbour interactions in a lattice is not used in my simulations in favour of instead parallelizing over multiple temperatures. This reduces overhead computations and better generalizes to our follow-on paper applying my code to the Ising model on more general graphs than lattices.

By construction, nearby configurations (and the value of the observable on these configurations) on the Markov chain will be heavily correlated and the earlier steps in the chain will also be heavily dependent on its initial conditions. To address these issues, we will *couple from the past* by discarding the first results called the *burn-in* until the chain becomes uncorrelated from the initial state. We also iterate the algorithm multiple times before using the output configuration to reduce nearby correlation. It is also possible to compare the effect of initial conditions by running simulations starting from the two extremes[20] and [3] explains that if these two configurations converge to the same behaviour, then this squeezes all other initial configurations to agree. The important quantity in these discussions is the *(integrated) autocorrelation time* τ, defined in terms of the *autocorrelation function* $C(t)$, a function measuring the correlation of steps of the Markov Chain with the chain lagged by some number of steps t. In [39], the author recommends discarding a *burn-in* of at least the first $n >> \tau$ samples, with $n = 5\tau, 10\tau$ common choices, since then the new starting configuration is statistically uncorrelated from the initial conditions. Similarly, only after discarding τ samples should the next sample be considered statistically independent from the last.

We now define the integrated autocorrelation time, used to couple from the initial conditions and generate statistically independent samples.

[19] According to [39], one should compare efficiencies by the (wall-clock) time, rather than the number of updates.

[20] A *hot start* is a random/disordered initial configuration, while a *cold start* is a completely ordered initial configuration.

Definition 2.3 ([50])**.** Let $(x_i)_{0\leq n}$ be a finite of sequence of values of a random variable X_i indexed by times $0 \leq i \leq n$. Define the *autocorrelation function* $C(t)$ of $(X_i)_{0\leq n}$ by

$$C(t) = \frac{\sum_{i\leq n-t}(X_i X_{i+t}) - \langle X\rangle^2}{\langle X^2\rangle - \langle X\rangle^2} = \frac{\langle X_i X_{i+t}\rangle - \langle X\rangle^2}{\langle X^2\rangle - \langle X\rangle^2} = \frac{\langle X_i X_{i+t}\rangle - \langle X\rangle^2}{\mathbb{V}\mathrm{ar}(X)}.$$

Then the *(integrated) autocorrelation time* is defined

$$\tau = \frac{1}{2} + \sum_{t=1}^{\infty} C(t).$$

However this sum is often divergent so we should identify a *window* W to instead take the finite sum of $C(t)$ for $0 \leq t \leq W$. This introduces some error but after some time, $C(t)$ often deteriorates into nothing but noise around some non-zero value. Hence, continuing the sum beyond an appropriate window would increase the estimate of *tau* approximately linearly, but using noise and not towards a more meaningful value. As such, W should be taken such that $C(t)$ has plateaued for $T > W$. Although some research exists for how to best choose the window W, it is common to choose the window empirically according to the given problem.

Once an estimate for the integrated autocorrelation time has been evaluated, we discard at least 5τ initial values as the burn-in and discard τ values of our Markov chain of simulations between each recorded value. As such, one would expect approximately just $(m - 5\tau)/\tau$ statistically independent samples from an m-step simulation. Recall from Table 1 that the correlation length ζ goes to infinity as we approach the critical temperature $T_C \approx 2.27$, and so our samples become correlated over longer and longer distances, hence requiring many more Markov steps to generate as many statistically independent samples for temperatures close to T_C. This presents one limitation of these simulations since it is computationally inefficient to produce large sets of independent samples near the critical temperature, which is in fact even more important as this is when it takes the system the longest to equilibriate.

In addition to using the integrated autocorrelation time τ to produce statistically independent samples, it is also used in the formula for the expected standard error from the true mean of a Monte Carlo simulation, which is useful for producing confidence intervals for each observable estimate. We take directly from [39] the following definition.

Definition 2.4. In Monte Carlo analysis with correlated measurements $(X_i)_{1\leq i\leq n}$, the expected standard error from the true mean is

$$\delta_X = \sqrt{2\tau \frac{\sum_{i=1}^{n}(X_i - \langle X\rangle)^2}{n(n-1)}}.$$

It is beyond the scope of this report to justify this definition, but notice this error accounts for the autocorrelation by a simple $\sqrt{2\tau}$ multiple of the standard error of independent data without autocorrelation.

Another approach is to use *bins* of blocklength m much larger than the autocorrelation time τ so that we may assume the bins are statistically independent. Then we may use the standard error formula for uncorrelated measurements on the averages $\bar{X}^{(j)},\ j = 1, \cdots, M$ of the m values of the observable for each configuration in the j^{th} bin:

$$\sqrt{\frac{\sum_{j=1}^{M} \left(\bar{X}^{(j)} - \langle X \rangle\right)^2}{M(M-1)}}. \tag{10}$$

For simplicity we shall instead in our code use δ_X for our error estimate of the Monte Carlo measurements[21]. The $100(1-\alpha)\%$ confidence interval for a large-sample mean is then given by $x \pm z_{\alpha/2}\frac{\delta_X}{\sqrt{n}}$. Observe that the precision of a Monte Carlo estimate for the mean of an observable then increases as the inverse square root of the sample size n. This observation helps frame the perspective for the trade-off between the execution time for our code against estimate precision, whist affording a tool to help inform our code on when to terminate simulations upon 'sufficient' convergence.

Unfortunately due to time constraints, I have been unable thus far to get the code to implement integrated autocorrelation time calculations to work predictably. This means that no confidence intervals or convergence testing functionality have been included. I will continue to think about how to fix my autocorrelation time calculations and so this may be added to the code at a later date. In particular, it could be useful to implement the autocorrelation time before progressing on to applying these techniques to more general graphs. Instead of directly calling the function to calculate the autocorrelation time, manual 'sweeps' and 'skips' parameters are included to determine the number of Markov steps to discard before storing a sample and how large the burn-in should be as a proportion of the total Markov steps to be run.

3 Simulation Results

The script for this project's simulations can be found at `https://github.com/Connor-Tracy/Ising-Simulation`. This includes a function plot_ising2D to animate the evolution of a 2D Ising model lattice over time, and another function parallel_temps to estimate values for the following four observables over a range of temperatures: energy density, specific heat capacity, specific magnetization, specific magnetic susceptibility. An example GIF of an animation for the Ising model on a small regular square lattice will be included in the repository. The main focus however is the output of the parallel_temps function, which has been optimized in several ways, including using parallel programming to run simulations for multiple temperatures simultaneously. This function is highly configurable and has an option to plot the estimates arising from the simulations

[21] Perhaps in the follow-on research it may be worthwhile to exploit the independence of equation 10 and τ to provide an indicator for chain convergence.

for each temperature in the specified range. This will allow us to visually identify the critical temperature.

Figure X presents the plots estimates of the four observables over a range of 120 temperatures in the interval $[0.5T_c, 1.5T_c]$, computing the converged averages over $500,000$ samples for each of the 120 Markov chains of length over $5,000,000$.[22] With the exception of a small number of samples' specific heat capacity estimates, the convergence was very clear on such large simulations, with the trendlines very clear and narrow. Moreover, the analytical solutions for $T < T_c$ are displayed in each plot for reference. The analytical solution represents the result expected in the thermodynamic limit, whereas finite-size effects lead to the specific patterns observed, which agree with the literature on finite lattices. Due to time constraints, some desirable features have not been implemented such as critical exponent estimation, error bars, multidimensional simulations, automatic autocorrelation time estimation and implementation.

The method and code of this project will translate naturally to general graphs: at a high level a graph may be considered as a list of vertices together with a list of connecting edges. Then it is possible to accommodate general graphs into this project's pre-existing simulation script and thereby attempt to approach open questions numerically in more general settings. This will be the basis of a planned research paper between this project's supervisor Dr D Proment and myself which we hope to publish as co-authors if we are able to identify and work towards a better understanding of an open problem. One interesting application could be to better understand how the Ising model applies to graphs of non-integral dimension.

References

[1] David Chandler.
Introduction to Modern Statistical Mechanics, University of California, Berkeley.

[2] R. P. Feynman.
Statistical Mechanics: A Set of Lectures – California Institute of Technology.

[3] Werner Krauth.
Statistical Mechanics: Algorithms and Computations.

[4] Ising model
http://web.mit.edu/ceder/publications/Ising%20Model.pdf.

[5] Satya Pal Singh (2020).
The Ising Model: Brief Introduction and Its Application, Solid State Physics - Metastable, Spintronics Materials and Mechanics of Deformable Bodies - Recent Progress, Subbarayan Sivasankaran, Pramoda Kumar Nayak and Ezgi Günay, IntechOpen.

[22] On my personal (outdated) laptop, this took a little over 6.5 hours to run. The example provided at the end of the script is smaller and takes my laptop 13 minutes.

[6] *https://en.wikipedia.org/wiki/Ising_model*.

[7] Unofficial.
https://jeffjar.me/statmech2/intro4.html.

[8] Marco Pangallo, Alberto Sala, Cecilia Panigutti (2014).
Monte Carlo simulations of the Ising model on lattices and power law networks.

[9] Lots of people (Kun Yang, Yi-Fan Chen, George Roumpos, Chris Colby, John Anderson) (2019).
High Performance MC Sims of Ising Model.

[10] *Wikipedia – Metropolis-Hastings Algorithm.*

[11] J. J. Hopfield (1982) *Neural networks and physical systems with emergent collective computational abilities.*

[12] Herbert B. Callen (1985). (... is the reference given by wikipedia) *Thermodynamics and an Introduction to Thermostatistics (2nd ed.).*

[13] Silvio R. A. Salinas (2001). `http://www.lps.ens.fr/~krzakala/ISINGMODEL.pdf` *Introduction to Statistical Physics*

[14] EDOUARD BRÉZIN *Introduction to Statistical Field Theory*

[15] Ernst Ising (1924). `https://link.springer.com/content/pdf/10.1007/BF02980577.pdf` *Beitrag zur Theorie des Ferromagnetismus*

[16] Lars Onsager (1944). `https://journals.aps.org/pr/abstract/10.1103/PhysRev.65.117` *Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition*

[17] Stephen G. Brush (1967).
Rev. Mod. Phys. 39, 883 *History of the Lenz-Ising Model*

[18] Giovanni Gallavotti (1972). *Instabilities and phase transitions in the ising model. A review*

[19] Giovanni Gallavotti (1999). *Statistical Mechanics: A Short Treastise*

[20] *https://en.wikipedia.org/wiki/Ising_model*.

[21] G. G. Cabrera (1987). *Role of boundary conditions in the finite-size Ising model*

[22] M. Achilles, J. Bendisch, H. v. Trotha (2000). *Effect on the groundstate threshold p c by different boundary conditions in* $2D \pm J$ *Ising models*

[23] User: M Quack (2012). `https://www.physicsforums.com/threads/effect-of-sample-size-when-using-periodic-boundary-conditions-in-2d-ising-model.590446/` *Effect of sample size when using periodic boundary conditions in 2D Ising model*

[24] Francisco Prieto-Castrillo, Amin Shokri Gazafroudi, Javier Prieto, and Juan Manuel Corchado (2018) *An ising spin-based model to explore efficient flexibility in distributed power systems.*

[25] Tobias Preis, Wolfgang Paul, and Johannes J. Schneider (2008) *Fluctuation patterns in high frequency financial asset returns.*

[26] Norm Jouppi (2017). *Quantifying the performance of the tpu, our first machine learning chip.*

[27] *http://www.maths.nuigalway.ie/~rquinlan/linearalgebra/section2-2.pdf*

[28] Kari Rummukainen. *https://www.mv.helsinki.fi/home/rummukai/simu/simulation.pdf*

[29] *https://en.m.wikipedia.org/wiki/Phase_transition*

[30] *https://en.m.wikipedia.org/wiki/Ising_critical_exponents*

[31] W. Heisenberg (1927). *The Physical Content of Quantum Kinematics and Mechanics*

[32] Andrea Pelissetto, Ettore Vicari (2002). *Critical Phenomena and Renormalization-Group Theory*

[33] *http://users.ox.ac.uk/~phys1116/msm-rg2.pdf*

[34] *https://en.m.wikipedia.org/wiki/Renormalization_group*

[35] Finite-size scaling at infinite-order phase transitions (2016). *G. T. Barkema, Rick Keesman, Jules Lamers, R. A. Duine,*

[36] *https://www.tcm.phy.cam.ac.uk/~bds10/phase/rg.pdf*

[37] Kadanoff, Leo P (1966). *Scaling laws for Ising models near* T_c

[38] *An Introduction to the Theory of Critical Phenomena and the Renormalization Group (1976).*

[39] *https://www.mv.helsinki.fi/home/rummukai/simu/fss.pdf*

[40] *https://www.mv.helsinki.fi/home/rummukai/lectures/montecarlo_oulu/lectures/mc_notes2.pdf*

[41] Annick LESNE (1996). *A comparative introduction to the renormalization methods used in statistical mechanics and for dynamical systems*

[42] *http://www.scholarpedia.org/article/Scaling_laws*

[43] *https://www.nobelprize.org/prizes/lists/all-nobel-prizes-in-physics/*

[44] *http://www.damtp.cam.ac.uk/user/us248/Lectures/Notes/statphys.pdf*

[45] Clément Hongler And Kalle Kytölä (2011). *Ising Interfaces And Free Boundary Conditions*

[46] Kris Hauser (2013). *Why does the Metropolis-Hastings procedure satisfy the detailed balance criterion?*

[47] Nicholas Metropolis (1953). *Equation of State Calculations by Fast Computing Machines.*

[48] W.K. Hastings (1970). *Monte Carlo Sampling Methods Using Markov Chains and Their Applications.*

[49] Anna Mikusheva (2007). *Time Series Analysis, Lecture 25; MIT.*

[50] Anders W. Sandvik (Fall 2018). *Monte Carlo simulations in classical statistical physics; Lecture 5*
PY 502 Computational Physics, Boston University Department of Physics.

[51] Emanuel Schmidt (2011) *Monte Carlo Simulation of the 2D Ising model*

A Checkerboard Algorithm

One can exploit a property of the nearest-neighbour interactions in a lattice to paralellize the update steps of the algorithm via *checkerboard ordering*. The order in which our algorithm traverses the sites of the graph can be chosen without materially affecting our results[39]. There are three main orderings: *typewriter*, random and checkerboard. In typewriter ordering, the sites are considered sequentially: row-by-row and column-by-column. The code can be written naturally, especially for the case of a regular square lattice, such that this is particularly fast. A slower but less artificial method is that of a random ordering in which sites are chosen at random. Finally, the approach of checkerboard ordering leverages the efficiency gains of [39] the code by exploiting a property [symmetry??] of the Ising model on a regular square lattice. Namely, since sites only act on their nearest neighbours, labelling the sites into black and white types in a checkerboard pattern leads to sites of a particular colour only interacting with sites of the opposite colour. Since the change of energy of a configuration is determined by the changed site and its $2d$ nearest-neighbours, one can update all of the sites of one colour while fixing the sites of the other colour. The same can then be done for the other colour and so on, making use of parallel programming to execute the independent updates simultaneously. The author compares in [39] the execution times of algorithms using the different orderings and showed that indeed checkerboard ordering is the fastest[23].

[23]The author of [39] also explains that although the typewriter ordering is faster than random ordering, it breaks the Ising model's detailed balance equations. The consequence of this is that the update step for the algorithm must be chosen 'carefully'

B Universality Classes

A *universality class* is a collection of models which may differ on finite scales but each tend to the exact same behaviour in the thermodynamic limit, which is scale invariant. Thus any study into the behaviour of one model in the thermodynamic limit will apply exactly to any other model in the same universality class, in that any two models in the same universality class will have the same critical exponents; as we will see later, critical exponents describe the behaviour of observables near criticality. There is a *lower critical dimension* below which there is no phase transition, namely dimension 2 for the Ising model (since the 2-dimensional case exhibits a phase transition while the 1-dimensional case does not), so therefore this does not belong to a universality class. Similarly, the Ising model has a universality class for each higher dimension up to degeneracy from the 4-dimensional (*upper critical dimension*) case, from which point the critical exponents will be exactly the same for each higher dimension. One may use mean field theory to calculate the critical exponents for the upper critical dimension. Hence the Ising models of dimensions 2, 3 and 4+ are members of three distinct universality classes. In the 2-dimensional Ising model, one can calculate the critical exponents using the 'minimal model'. Mean field theory is required for the 4-dimensional model while the 3-dimensional model is not yet exactly solved, though it has primarily been studied using Renormalization Group methods and Monte Carlo simulations. The $2-$dimensional Ising model shares[28] its universality class with the liquid-vapour transition of a lattice gas. We will briefly discuss the Renormalization Group approach as it also gives a natural explanation for the "miracle"[33] of universality classes.

The *Renormalization Group*[24] (RG) is an algebraic tool used to analyse changing systems over *scale transformations* when investigating changes in force laws arising from different resolutions or *scales*. Note that a related idea is that of the Uncertainty Principle[31] which implies that energy/momentum and resolution scales are inversely related. This RG tool is used to investigate the critical exponents the 3-dimensional case of the Ising model, which is not yet exactly solved[34] unlike the 2-dimensional and 4-dimensional cases, as well as to provide the key insight that universality classes in fact correspond exactly to fixed points of the Renormalization Group. Each renormalization step iteratively transforms the coupling parameters[36] until reaching a fixed point, hence the emergence of universality classes for the limiting behaviour of different models, when the parameters are invariant under this rescaling transformation.

Fixed points can[36] have a critical point at $T = T_c,\ J = J_c$ called the *Curie temperature* where the parameters are non-zero and *scale invariant*. Alternatively, two other fixed points are possible with either

$$T = 0,\ J \to \infty,\ \text{or } T \to \infty,\ J = 0.$$

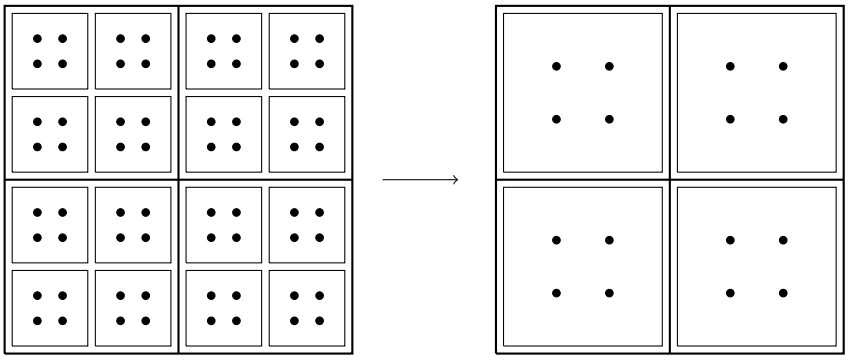
The importance of an observable or parameter is determined to be *relevant* or *not relevant* if its magnitude always increases or always decreases respectively,

[24] It will not be needed for this project to understand the Renormalization Group but a short and basic exposition of this important technique is given for interest.

as the resolution decreases, equivalently as the *scale length* increases. If neither of these happen, then it is undetermined whether the observable is relevant or not. A relevant parameter is necessary to describe the macroscopic scaling law while a non-relevant observable is not needed. The block spin pedagogical model gives an intuition on how complex systems might be reduced such that the macroscopic system is described by very few relevant parameters. The redundancy of many microscopic properties in this way as non-relevant parameters gives rise to the universality classes which group models with identical limiting behaviour.

It is the Renormalizing Group which provides a clear explanation for the surprising connections made between disparate models via universality, explaining critical exponents as some of the relevant variables, while microscopic differences in the systems are determined by irrelevant variables, which are much fewer in number than the degrees of freedom that arise from considering the microscopic model directly. The block spin RG is an approach introduced by Leo P. Kadanoff in [37] and will serve as a useful pedagogical tool for building intuition of the Renormalization Group and its connection to the Ising model, universality classes and therefore critical exponents.

Consider a square lattice and *renormalize* it by replacing each vertex with a block of four vertices. This block variable now behaves as the average of its four constituent sites and one now has a new lattice with $1/4th$ the number of vertices. This is called the *RG step* or *RG transformation*. Repeating this process increases the scale until we effectively find the long-range behaviour between sites[34].



At criticality there is long-range correlation of spins[38] as introduced by the block-spin approach, but there is always correlation of near spins so consider the system consisting of larger blocks of spins which can behave as one large spin. This removes many of the degrees of freedom in this new scaling and yet will have the same critical behaviour as the standard nearest-neighbour Ising model. This self-similarity suggests that many degrees of freedom are not necessary to describe the critical behaviour, hence disparate models with different microscopic properties can be described by the same few relevant variables to

belong in the same universality class and thus share the same critical exponents describing their limiting behaviour near the critical points.

Critical Exponent	Observable	Ising Property	Ising 2D	Ising 3D[25]	Ising 4+
α	$C \propto \tau^{-\alpha}$	Specific heat capacity C_V $= \frac{1}{VkT}\left(\langle E^2\rangle - \langle E\rangle^2\right)_{V\to\infty}$ $= \frac{1}{V}\frac{\partial\langle E\rangle}{\partial T}_{V\to\infty}$	0	0.110	0
β	Order parameter $\Psi \propto (-\tau)^{-\beta}$	Magnetisation $\Psi = M = \frac{1}{V}\frac{\partial F}{\partial \mu}_{\substack{\mu\to 0\\ V\to\infty}}$	1/8	0.326	1/2
γ	$\frac{\partial\Psi}{\partial J} \propto \tau^{-\gamma}$	Specific magnetic susceptibility χ_M $= V\left(\langle M^2\rangle - \langle M\rangle^2\right)_{\substack{\mu\to 0\\ V\to\infty}}$ $= \frac{1}{V}\left(\frac{\partial^2 F}{\partial\mu^2}\right)_{\substack{\mu\to 0\\ V\to\infty}}$	7/4	1.24	1
δ	δ s.t. $J = \Psi^\delta$	Critical isotherm	15	4.79	3
η	Correlation function $\Gamma(x) = \frac{1}{V}\sum_i \sigma_i\sigma_{i+x}$	η s.t. correlation function scales[26] as $r^{-d+2-\eta}$	1/4	0.036	0
ν	Correlation length $\xi \propto \tau^{-\nu}$	Correlation length ξ s.t. $\Gamma(x) \propto e^{-x/\xi}$	1	0.630	1/2

Table 1: Critical Exponents of the Ising Model.

[25] To 3 sf's

[26] In the sense of [42].

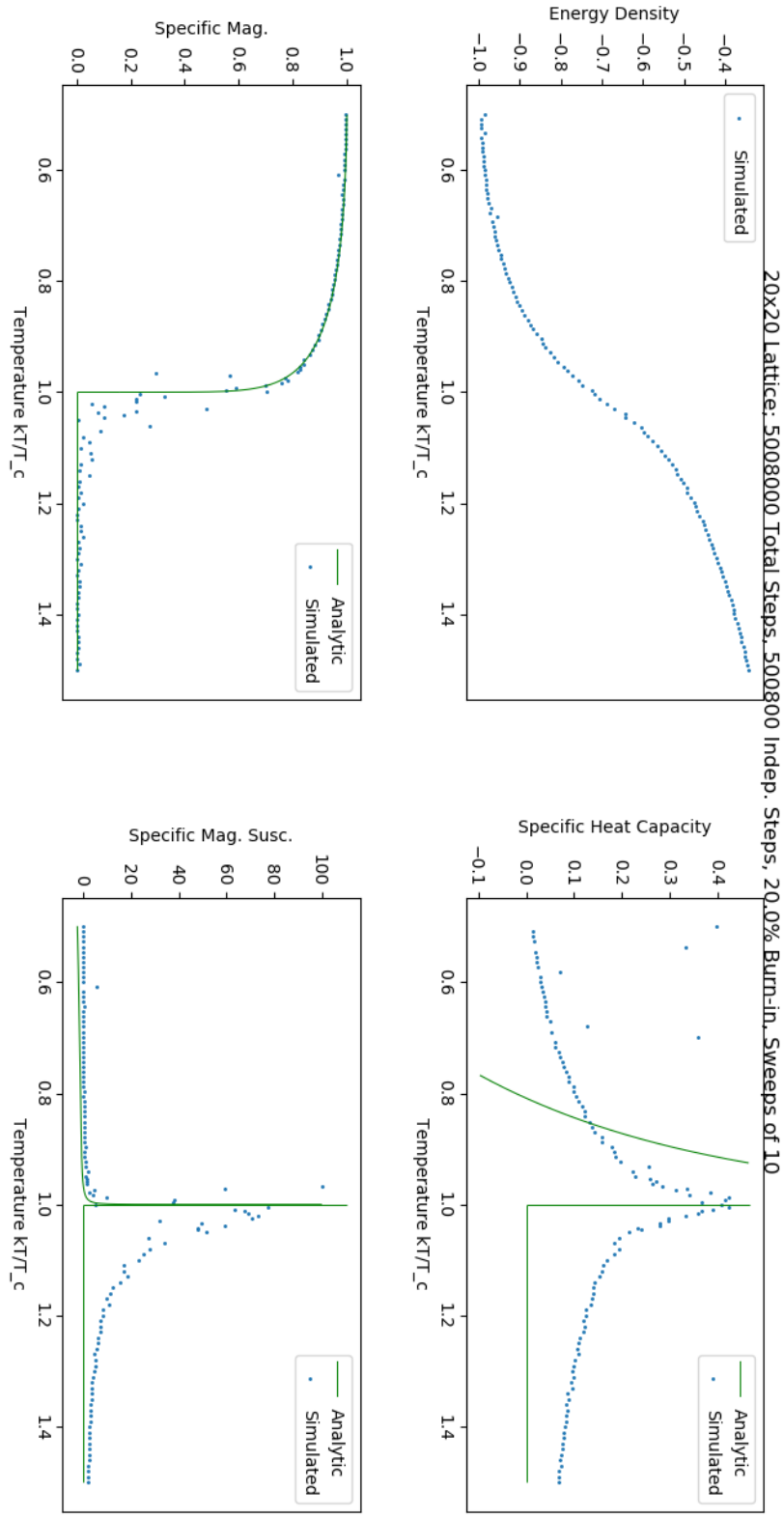


Figure 1: Large simulation plotting estimates for four observables against temperature

A Proofs for Section 1 (The Ising Model)

Proposition 1.1. The partition function Z is equal to

$$Z = \lambda_1^N + \lambda_2^N,$$

where

$$\lambda_1, \lambda_2 = e^{\beta J}\left(\cosh(\beta\mu) \pm \sqrt{\sinh^2(\beta\mu) + e^{-4\beta J}}\right). \tag{5}$$

Proof. [13] Rewriting the exponential sum in the symmetric form of the partition function as in equation (4), one may rewrite the partition function as

$$Z = \sum_{\sigma \in \mathcal{S}_d(N)} \prod_{i=1}^{N} T(\sigma_i, \sigma_{i+1}), \tag{11}$$

where

$$T(\sigma_i, \sigma_{i+1}) = \exp\left\{\beta J \sigma_i \sigma_{i+1} + \frac{\beta\mu}{2}(\sigma_i + \sigma_{i+1})\right\}. \tag{12}$$

This function T has four possible inputs corresponding to the possible pairs of spin values of the i^{th} and $(i+1)^{st}$ sites so define the transfer matrix

$$\mathbf{T} = \begin{pmatrix} T(+1,+1) & T(+1,-1) \\ T(-1,+1) & T(-1,-1) \end{pmatrix} = \begin{pmatrix} e^{\beta(J+\mu)} & e^{\beta(-J)} \\ e^{\beta(-J)} & e^{\beta(J-\mu)} \end{pmatrix}.$$

Let us now notice that equation (5) for the partition function can be rewritten as the trace of the matrix T^N. Since the real transition matrix $\mathbf{T}$ is symmetric, it is diagonalizable with distinct real eigenvalues[27]. Denote the eigenvalues $\lambda_1 > \lambda_2$, then

$$\begin{aligned} Z &= Tr\left(\mathbf{T}^N\right) \\ &= Tr\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^N = Tr\begin{pmatrix} \lambda_1^N & 0 \\ 0 & \lambda_2^N \end{pmatrix} \\ &= \lambda_1^N + \lambda_2^N. \end{aligned} \tag{13}$$

We now set the characteristic equation of $\mathbf{T}$ to equal 0,

$$\begin{aligned} 0 = \det(\mathbf{T} - \lambda I) &= \begin{pmatrix} e^{\beta(J+\mu)} - \lambda & e^{\beta(-J)} \\ e^{\beta(-J)} & e^{\beta(J-\mu)} - \lambda \end{pmatrix} \\ &= 2\sinh(2\beta J) - 2\lambda e^{\beta J}\cosh(\beta\mu) + \lambda^2. \end{aligned}$$

Finally the quadratic formula gives

$$\begin{aligned}\lambda_1, \lambda_2 &= \frac{1}{2}\left(2e^{\beta J}\cosh(\beta\mu) \pm \sqrt{\left(2e^{\beta J}\cosh(\beta\mu)\right)^2 - 4\left(e^{2\beta J} - e^{-2\beta J}\right)}\right) \\ &= e^{\beta J}\cosh(\beta\mu) \pm \sqrt{e^{2\beta J}\cosh^2(\beta\mu) - 2\sinh(2\beta J)} \\ &= e^{\beta J}\cosh(\beta\mu) \pm \sqrt{e^{2\beta J}\sinh^2(\beta\mu) + e^{-2\beta J}} \\ &= e^{\beta J}\left(\cosh(\beta\mu) \pm \sqrt{\sinh^2(\beta\mu) + e^{-4\beta J}}\right). \end{aligned} \tag{14}$$

□

Proposition 1.3. The *mean energy*[27] (*internal energy*) is the expected value of the Hamiltonian over all possible configurations, since the Hamiltonian describes the total energy of a given configuration. The mean energy can be obtained[3] from the partition function Z via

$$\langle \mathcal{H} \rangle = \sum_{\sigma \in \mathcal{S}} \mathcal{H}(\sigma) P_\beta(\sigma) = -\frac{\partial \ln(Z)}{\partial \beta}. \tag{7}$$

Proof. The chain rule gives $\dfrac{\partial \ln(Z)}{\partial \beta} = \dfrac{1}{Z} \cdot \dfrac{\partial Z}{\partial \beta}$, so

$$\begin{aligned} -\frac{\partial \ln(Z)}{\partial \beta} &= -\frac{1}{Z} \cdot \frac{\partial}{\partial \beta}\left(\sum_{\sigma \in \mathcal{S}} \exp\{-\beta \mathcal{H}(\sigma)\}\right) \\ &= \sum_{\sigma \in \mathcal{S}} \mathcal{H}(\sigma) \cdot \frac{1}{Z}\exp\{-\beta \mathcal{H}(\sigma)\} \\ &= \sum_{\sigma \in \mathcal{S}} \mathcal{H}(\sigma) P_\beta(\sigma) = \langle \mathcal{H} \rangle. \end{aligned}$$

□

Proposition 1.4. For $\lambda_1 > \lambda_2$ the larger of the two eigenvalues of the transition matrix $\mathbf{T}$, the *free energy per site* $f := \lim_{N\to\infty} F/N$ in the thermodynamic limit is

$$f = -\frac{1}{\beta}\ln(\lambda_1) = -J - \frac{1}{\beta} ln\left(\cosh(\beta\mu) + \sqrt{\sinh^2(\beta\mu) + e^{-4\beta J}}\right).$$

Proof. Proposition 1 says the partition function can be written $Z = \lambda_1^N + \lambda_2^N$, which can be rewritten as

$$Z = \lambda_1^N\left(1 + \left(\frac{\lambda_2}{\lambda_1}\right)^N\right). \tag{15}$$

[27] Derivation of our definition of the free energy is given by [44], as well as the formulae for both the free energy (p19) and the average energy (p15) in one source. Another source presenting both the free energy and the average energy is on page 5.

The Helmholtz free energy F is then

$$F = -\frac{1}{\beta}\ln(Z) = -\frac{1}{\beta}N\ln(\lambda_1) - \frac{1}{\beta}\ln\left(1 + \left(\frac{\lambda_2}{\lambda_1}\right)^N\right), \tag{16}$$

thus the free energy per site in the thermodynamic limit is

$$f = \lim_{N\to\infty}\frac{F}{N} = -\frac{1}{\beta}\ln(\lambda_1), \tag{17}$$

where

$$\lambda_1 = e^{-\beta J}\left(\cosh(\beta\mu) + \sqrt{\sinh^2(\beta\mu) + e^{-4\beta J}}\right).$$

□

Proposition 1.5. Given the free energy F, the *magnetization* M and specific magnetization m satisfy

$$m := \frac{M}{N} := \frac{1}{N}\left\langle \sum_{i=1}^{N}\sigma_i \right\rangle = -\frac{1}{N}\frac{\partial F}{\partial \mu}. \tag{8}$$

Proof.

$$\begin{aligned}
-\frac{\partial F}{\partial \mu} &= \frac{\partial}{\partial \mu}\left(\frac{1}{\beta}\ln\left(\sum_{\sigma\in\mathcal{S}}\exp\left\{\beta J\sum_{i=1}^{N}\sigma_i\sigma_{i+1} + \beta\mu\sum_{i=1}^{N}\sigma_i\right\}\right)\right) \\
&= \frac{1}{\beta}\sum_{\sigma\in\mathcal{S}}\left(\beta\left(\sum_{i=1}^{N}\sigma_i\right)\sum_{i=1}^{N}e^{-\beta\mathcal{H}(\sigma)}\cdot\frac{1}{Z}\right) \\
&= \sum_{\sigma\in\mathcal{S}}\left(\sum_{i=1}^{N}\sigma_i\right)P_\beta(\sigma) = \left\langle\sum_{i=1}^{N}\sigma_i\right\rangle =: M.
\end{aligned}$$

□

Corollary 1.6. The specific magnetization is

$$m = \frac{\sinh(\beta\mu)}{\sqrt{\sinh^2(\beta\mu) + e^{-4\beta J}}}$$

in the thermodynamic limit as $N \to \infty$.

Proof. Define $c := \cosh(\beta\mu)$, $s := \sinh(\beta\mu)$, and $\sqrt{\cdot} := \sqrt{\sinh^2(\beta\mu) + e^{-4\beta J}}$ for

ease of reading. Now recall

$$\begin{aligned}\lambda_1, \lambda_2 &= e^{-\beta J}\left(c \pm \sqrt{\cdot}\right), \\ F &= -\frac{1}{\beta}\ln(Z), \\ Z &= \lambda_1^N\left(1 + \left(\frac{\lambda_2}{\lambda_1}\right)^N\right), \\ M &= -\frac{\partial F}{\partial \mu},\end{aligned}$$

from equation (5), (10), (12) and (15) respectively. First notice that $c^2 + s^2 \equiv 1$ gives

$$(c + \sqrt{\cdot})(c - \sqrt{\cdot}) = 1 - e^{4\beta J}, \tag{18}$$

and recall the derivative of the logarithm of a function is given by

$$\frac{d}{dx}\ln(f(x)) = \frac{f'(x)}{f(x)}. \tag{19}$$

Differentiating equation (10) after substituting equation (12) wrt μ yields

$$\begin{aligned}M &\overset{(15)}{=} -\frac{\partial F}{\partial \mu} = \frac{1}{\beta}\frac{\partial}{\partial \mu}\left[N\ln(\lambda_1) + \ln\left(1 + \left(\frac{\lambda_2}{\lambda_1}\right)^N\right)\right] \\ &\overset{(17)}{=} \frac{1}{\beta}\left[N\frac{\partial \lambda_1}{\partial \mu}/\lambda_1 + N\left(\frac{\lambda_2}{\lambda_1}\right)^{N-1}\frac{\partial}{\partial \mu}\left(\frac{\lambda_2}{\lambda_1}\right)\Big/\left(1 + \left(\frac{\lambda_2}{\lambda_1}\right)^N\right)\right].\end{aligned} \tag{20}$$

Hence one must compute

$$\frac{\partial \lambda_1}{\partial \mu}/\lambda_1 = \frac{\beta(s + sc/\sqrt{\cdot})}{c + \sqrt{\cdot}} = \frac{s\beta}{\sqrt{\cdot}}, \tag{21}$$

and

$$\frac{\lambda_2}{\lambda_1} = \frac{c - \sqrt{\cdot}}{c + \sqrt{\cdot}} \overset{(16)}{=} \frac{(c - \sqrt{\cdot})^2}{1 - e^{4\beta J}}, \tag{22}$$

so then

$$\frac{\partial}{\partial \mu}\left(\frac{\lambda_2}{\lambda_1}\right) \overset{(20)}{=} \frac{2\beta s(1 - c/\sqrt{\cdot})(c - \sqrt{\cdot})}{1 - e^{4\beta J}} = -\frac{2\beta s(c - \sqrt{\cdot})}{(c + \sqrt{\cdot})\sqrt{\cdot}}. \tag{23}$$

Now substituting equations (19) - (21) into equation (18) yields

$$\begin{aligned}M &= \frac{1}{\beta}\left[N\frac{s\beta}{\sqrt{\cdot}} - 2N\beta s\left(\frac{\lambda_2}{\lambda_1}\right)^{N-1}\frac{c - \sqrt{\cdot}}{(c + \sqrt{\cdot})\sqrt{\cdot}}\Big/\left(1 + \left(\frac{\lambda_2}{\lambda_1}\right)^N\right)\right] \\ &\overset{(20)}{=} \frac{Ns}{\sqrt{\cdot}}\left[1 + \left(\frac{\lambda_2}{\lambda_1}\right)^N - 2\left(\frac{\lambda_2}{\lambda_1}\right)^N\right]\Big/\left(1 + \left(\frac{\lambda_2}{\lambda_1}\right)^N\right) \\ &= \frac{Ns}{\sqrt{\cdot}}\left[\frac{1 - (\lambda_2/\lambda_1)^N}{1 + (\lambda_2/\lambda_1)^N}\right] = \frac{Ns}{\sqrt{\cdot}}\left[1 - \frac{2}{1 + (\lambda_1/\lambda_2)^N}\right].\end{aligned} \tag{24}$$

Finally, this gives the specific magnetisation

$$m := \frac{M}{N} = \frac{s}{\sqrt{\cdot}}\left[1 - \frac{2}{1 + (\lambda_1/\lambda_2)^N}\right] \to \frac{s}{\sqrt{\cdot}} = \frac{\sinh(\beta\mu)}{\sqrt{\sinh^2(\beta\mu) + e^{-4\beta J}}},$$

in the thermodynamic limit as $N \to \infty$, since $\lambda_1 > \lambda_2$ by construction. □

Theorem 1. *The one-dimensional Ising model possesses no phase transitions.*

Proof. [14] In the thermodynamic limit, Proposition 1.4 states

$$m = \frac{1}{\beta}\frac{\partial \lambda_1}{\partial \mu}.$$

The Maclaurin expansion $\cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$ for H small gives

$$\lambda_1 = 2\cosh(\beta J)\left(1 + \frac{\beta^2\mu^2}{2}e^{2\beta J} + \mathcal{O}\left(\mu^4\right)\right).$$

Now as $\mu \to 0$, we see $m \to 0$ implies the magnetization $m = \frac{1}{\beta}\frac{\partial \lambda_1}{\partial \mu}$ vanishes when the external magnetic field is removed independently of β, hence no phase transitions occur in the one-dimensional model at any temperature. □

B Proofs for Section 2 (Monte Carlo Theory)

Theorem 2 (Acceptance-Rejection)**.** *If $\exists\ C \in \mathbb{R}$ constant s.t. $f_X(x)/h_Y(x) \leq C\ \forall\ X \in \mathcal{X}$, then Algorithm 1 samples once from a random variable X with law f_X, given a proposal law h_Y with acceptance rate $a(y) = f_X(y)/Ch_Y(y)$.*

Proof. fghnj □

Theorem 3. *The Metropolis-Hastings algorithm samples from the target distribution in the limit.*

Proof. The Metropolis-Hastings algorithm generates a Markov chain of configurations so we shall prove that this Markov process asymptotically reaches a unique stationary distribution equal to the target distribution. A sufficient condition for the existence of a stationary distribution π for a Markov process is for its transition probabilities to be in detailed balance with π and for the Markov process to be ergodic. Hence we must prove that the Metropolis-Hastings algorithm satisfies these two properties and that the stationary distribution is in fact the target distribution.

A key feature of the algorithm is to first generate a candidate sample from the law $g(x'|x_t)$ and then choose whether to accept this as the next step according to the acceptance rate, thus the transition probability is

$$P(x'|x_t) = g(x'|x_t)\alpha(x', x_t),$$

which means we can rewrite equation (24) as

$$\frac{P(x'|x_t)}{P(x_t|x')} = \frac{P(x')}{P(x_t)} = \frac{g(x'|x_t)\alpha(x', x_t)}{g(x_t|x')\alpha(x_t, x')}$$
$$\implies$$
$$\frac{\alpha(x', x_t)}{\alpha(x_t, x')} = \frac{P(x')g(x_t|x')}{P(x_t)g(x'|x_t)}.$$

Since the Metropolis-Hastings algorithm's acceptance rate $\alpha(x', x_t) = \min\left\{1, \frac{\pi(x')g(x_t, x')}{\pi(x_t)g(x', x_t)}\right\}$ satisfies this equation, detailed balance is satisfied so this proves the existence of a stationary distribution for the Markov process. Moreover, since P is designed to satisfy the detailed balance equation with respect to the target distribution π, then

$$\sum_{x'} \pi(x')P(x', x_t) \overset{(eqn of detbal)}{=} \pi(x') \sum_{x'} P(x_t, x') = \pi(x'),$$

hence π is the stationary distribution of the Markov chain. Finally, [46] presents the "somewhat subtle" argument to prove ergodicity. □