in terms of the dual variables $\widetilde{s}_n$ that we're now working with. It's easy to see that

$$C_p(\beta) = (\cosh\beta)\exp\left[p\log\tanh\beta\right] = (\cosh\beta)\exp\left[\frac{1-\widetilde{s}_j\widetilde{s}_k}{2}\log\tanh\beta\right],$$

substituting $p = (1-\widetilde{s}_j\widetilde{s}_k)/2$. Breaking up the exponential gives us

$$C_p(\beta) = (\cosh\beta\sinh\beta)^{1/2}\exp\left[-\frac{1}{2}\widetilde{s}_j\widetilde{s}_k\log\tanh\beta\right].$$

Inserting this into the partition function, the product over all links just provides $2N$ factors of the $\widetilde{s}_n$-independent first term, and we can then convert the product of exponentials into an exponential of the sum, producing

$$Z = \frac{1}{2}(2\cosh\beta\sinh\beta)^N \sum_{\{\widetilde{s}\}}\exp\left[-\frac{\log\tanh\beta}{2}\sum_{(jk)}\widetilde{s}_j\widetilde{s}_k\right].$$

$Z(\beta) \propto Z(\widetilde{\beta})$

If we define

$$\widetilde{\beta} \equiv -\frac{\log\tanh\beta}{2},$$

then we can recognize the sum over $\{\widetilde{s}\}$ configurations as simply a zero-field Ising model partition function $Z(\widetilde{\beta})$ as in Eq. 112. We can also recognize this definition of $\widetilde{\beta}$ as equivalent to Eq. 122:

$$\tanh\beta = e^{-2\widetilde{\beta}}$$

$$e^{2\widetilde{\beta}} - e^{-2\widetilde{\beta}} = \frac{1}{\tanh\beta} - \tanh\beta = \frac{\cosh\beta}{\sinh\beta} - \frac{\sinh\beta}{\cosh\beta} = \frac{1}{\cosh\beta\sinh\beta}$$

$$\cosh\beta\sinh\beta = \frac{1}{4}\left(e^{\beta}+e^{-\beta}\right)\left(e^{\beta}-e^{-\beta}\right) = \frac{1}{4}\left(e^{2\beta}-e^{-2\beta}\right) = \frac{1}{2}\sinh(2\beta)$$

$$\sinh(2\widetilde{\beta}) = \frac{1}{\sinh(2\beta)}$$

8 May

10 May

Using similar manipulations, we can express the spin-independent prefactor in terms of either β or $\widetilde{\beta}$,

$$2\cosh\beta\sinh\beta = \sinh(2\beta) = \frac{1}{\sinh(2\widetilde{\beta})},$$

or in the mixed form that reproduces Eq. 121:

$$2\cosh\beta\sinh\beta = 2\cosh^2\beta\tanh\beta = \frac{2\cosh^2\beta}{e^{2\widetilde{\beta}}} \quad\Longrightarrow\quad \frac{Z(\beta)}{(2\cosh^2\beta)^N} = \frac{Z(\widetilde{\beta})}{2e^{2N\widetilde{\beta}}}.$$

We have successfully derived Kramers–Wannier duality! Now let's briefly interpret what it means. It's a worthwhile **exercise** to show that multiplying a partition function by an overall spin-independent factor, $Z(\beta) \to c(\beta)Z(\beta)$, has no

effect on expectation values. (Try it!) Therefore the relation above identifies a $d = 2$ Ising system at temperature β with another such system at temperature $\widetilde{\beta}$, where small β corresponds to large $\widetilde{\beta}$ and vice versa.

However, this does *not* mean that $d = 2$ Ising model behaves the same at low and high temperatures. Indeed, we saw already in Section 9.2 that it changes between qualitatively different ordered and disordered phases in these two regimes. What Kramers–Wannier duality is telling us is that the ordered phase of Ising spins in two dimensions, characterized by their order parameter, is secretly equivalent to the disordered phase of the *dual* spins — a different set of degrees of freedom, which can be characterized by a 'disorder parameter'. Similarly, the disordered phase of the original system maps onto the ordered phase of the dual system.

If we assume there is a single phase transition where the ordered and disorederd phases coincide, then Kramers–Wannier duality implies this must occur when $\beta = \widetilde{\beta}$. In other words,

$$\sinh^2(2\beta_c) = 1 \quad \Longrightarrow \quad \beta_c = \frac{1}{2}\operatorname{arcsinh}(1) = \frac{\log\left(1+\sqrt{2}\right)}{2} = 0.440686\ldots,$$

recalling (or looking up) that $\operatorname{arcsinh}(x) = \log\left(x + \sqrt{x^2+1}\right)$. Because this exact critical temperature $T_c = 2/\log\left(1+\sqrt{2}\right) = 2.269185\ldots$ was predicted three years before Onsager analytically solved the $d = 2$ Ising model, the fact that his solution correctly reproduced this T_c was a significant check of its correctness.

As mentioned at the start of this subsection, dualities of this sort are a pillar of theoretical physics in the 21st century. In general, these dualities are much more complicated than Kramers–Wannier duality, in two main ways. First, the dual degrees of freedom are typically different — and have different interactions — than the original degrees of freedom. For example, if we were to carry out a similar analysis of the three-dimensional Ising model, we would find that the dual system is *not* just another Ising model. The $d = 2$ Ising model is a special case of a **self-dual** system, which we exploited to determine T_c.

Second, the Ising model is special in that we were able to explicitly derive the duality it exhibits, which is typically not (yet) possible. Instead, most dualities have to be conjectured and then checked by subjecting them to as many tests as possible. For example, this is the case for *holographic dualities* that are conjectured to relate certain theories of quantum gravity to non-gravitational quantum systems that can be much easier to analyze mathematically. As an aside, the fully connected lattice we encountereed in recent tutorials also makes an appearance in holography — the behaviour of interacting fermions on this fully connected lattice (known as the SYK model, named after Subir Sachdev, Jinwu Ye and Alexei Kitaev) is conjected to be dual to the gravitational dynamics of quantum black holes. Roughly one thousand scientific studies of the SYK model and its conjectured holographic duality have been published since 2016!

Unit 10: Synthesis and broader applications

10.1 Monte Carlo importance sampling

Although we were able to derive some exact results for the Ising model in one and two dimensions, it's worth recalling that for $3 \leq d < \infty$ no exact solution is known even for this simple system. In general, interacting statistical systems are not exactly solvable. In order to explore their broad applications throughout the mathematical sciences and beyond, we therefore need to analyze them either through systematic approximation schemes (such as *perturbation theory*) or by numerical computations. Numerical methods have become increasingly important over the past fifty years, and in this section we'll outline the general methods they employ.

Our goal is to compute expectation values of interest, which are formally defined by sums over all micro-states. Considering the canonical ensemble for simplicity,

$$\langle \mathcal{O} \rangle = \sum_{i=1}^{M} \mathcal{O}_i \, p_i = \frac{1}{Z} \sum_{i=1}^{M} \mathcal{O}_i \, e^{-\beta E_i} = \frac{\sum_{i=1}^{M} \mathcal{O}_i \, e^{-\beta E_i}}{\sum_{i=1}^{M} e^{-\beta E_i}}.$$

We already saw, at the end of Section 9.1, that enormous computational resources would be required to carry out such sums over micro-states. Even for tiny Ising systems with $N \sim 1000$ spins, the largest existing or foreseeable supercomputers would have to run for far longer than the age of the universe in order to evaluate the roughly $2^{1000} \sim 10^{300}$ terms in the partition function. To quantify 'tiny', consider that $N \sim 1000$ would correspond to a $10 \times 10 \times 10$ lattice in three dimensions or a $6 \times 6 \times 6 \times 6$ lattice in four dimensions, both very far from the $N \to \infty$ thermodynamic limit of interest for phase transitions.

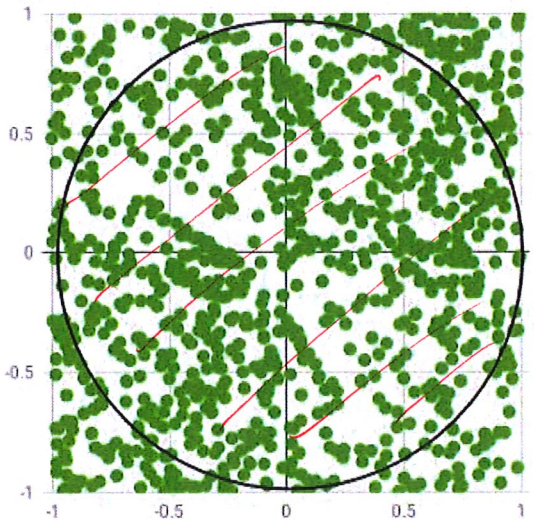
And yet, at the end of Section 9.3 we were able to quote numerical results for the Ising model critical temperature for $3 \leq d \leq 7$, along with a $d = 3$ critical exponent. These results are obtainable because practical numerical computations do not perform a 'brute-force' evaluation of every single micro-state. Instead, they proceed by (pseudo-)randomly **sampling** a very small subset of those micro-states, and using this subset to compute results for the average energy, magnetization, and other thermodynamic quantities. So long as this sampling is done appropriately, the law of large numbers allows us to treat these averages as controlled approximations to the true ensemble expectation values.

As we saw in the computer project, such numerical calculations employ pseudo-random numbers rather than complete randomness, which allows them to be reproducible up to very high precision by different people using different computers. Due to the role of randomness, these numerical approaches have become known as **Monte Carlo** methods, based on a whimsical reference to the famous gambling centre in Monaco. Monte Carlo methods are crucial in statistical physics, and related disciplines, because they are very broadly applicable to interacting systems that no longer benefit from dramatic simplifications through factorization.

We can gain some intuition about how Monte Carlo methods work by using such pseudo-random sampling to numerically evaluate a simple integral. The idea is that the integral can be numerically approximated by evaluating its integrand at randomly sampled points in the integration domain, and normalizing by the number of samples. An amusing example is to compute

$$\pi = \int_{-1}^{1} dx \int_{-1}^{1} dy \, H\big(1 - \{x^2 + y^2\}\big) \qquad H(r) = \begin{cases} 1 & \text{for } r \geq 0 \\ 0 & \text{for } r < 0 \end{cases},$$

where the Heaviside step function $H(r)$ picks out a disk with radius $R = 1$ in a square integration domain with area 4, as shown below. Since the integrand is either 0 or 1 for each randomly sampled point in that domain, simply counting the fraction of the S samples that lie in the disk provides a numerical determination of π, with a *statistical uncertainty* that vanishes $\propto 1/\sqrt{S}$. In just a few minutes, this Python code predicts $\pi = 3.14152 \pm 0.00023$ purely from sampling pseudo-random numbers.



Of course, numerically computing π can be done far more efficiently with other, more specialized, techniques. Monte Carlo integration is most useful when we need to consider very high-dimensional integrals — such as partition functions of interacting statistical systems, interpreted as N-dimensional integrals over the system's N degrees of freedom. To illustrate the scale of computations that can currently be carried out, ongoing theoretical physics research here in Liverpool routinely uses Monte Carlo methods to numerically evaluate roughly billion-dimensional integrals.

At this point, you might be concerned that such sampling can account for only an extremely small fraction of the possible micro-states for the systems under consideration, suggesting a risk of inaccurate results from unrepresentative sampling. This is a new manifestation of the obstacle we encountered when considering brute-force computations above. If the brute-force evaluation of every

single micro-state takes far longer than the age of the universe, then the fraction we could sample in a reasonable amount of time (say, a day) is almost vanishingly small.

As a concrete example, if we generously suppose our computer only needs a few nanoseconds to sample a micro-state of a tiny $N \sim 1000$ Ising system, over the course of a day it would sample roughly ten trillion (10^{13}) spin configurations — only about one part in 10^{287} of the total $2^N \sim 10^{300}$ micro-states. To make the situation even worse, as N increases the number of possible Ising model micro-states grows *exponentially* quickly, $\sim 2^N$, in addition to the more modest growth in the amount of computing required to sample each micro-state. For illustration, 2015 research arXiv:1502.07613 numerically predicting T_c for the Ising model in $d = 5$, 6 and 7 dimensions includes calculations up to $N = 64^5 \approx 10^9$. Out of the roughly $2^{10^9} \sim 10^{323,000,000}$ micro-states for this systems, only $\sim 10^4$ could be sampled in a reasonable amount of time. How much trust should we place in results from such numerical work?

Thinking back to our consideration of the ordered and disordered phases of the Ising model in Section 9.2, we could make a case that everything may work out in the high-temperature disordered phase. In the infinite-temperature limit, all the micro-states become equally probable, and observable expectation values are determined by the degeneracies of the different energy levels. Random sampling is more likely to account for the dominant energy levels with large degeneracies, making it plausible that reasonable results could be obtained by averaging even over such a tiny fraction of the total number of micro-states.

In the low-temperature ordered phase, however, the opposite occurs. As the temperature decreases, the large-scale behaviour of the system in this phase is dominated by a very small number of micro-states. For sufficiently low temperatures, observable expectation values are effectively determined by the two degenerate minimum-energy micro-states with all spins aligned either up or down. Only exponentially suppressed corrections would then be introduced by higher-energy excited states. As there is essentially no chance of randomly sampling either of those two minimum-energy micro-states, the random sampling approach described above seems doomed to fail.

10 May

A key breakthrough that made numerical results truly reliable was the invention of stochastic procedures to sample any micro-state ω_i with a probability proportional to its Boltzmann factor, $p_i \propto e^{-\beta E_i}$. Such automated procedures are known as *algorithms* (a term that evolved from the name of Muhammad ibn Musa al-Khwarizmi), and the overall approach is called **importance sampling**, since it preferentially samples the important micro-states that make the most significant contributions to the partition function and derived quantities. Assuming we have such an algorithm, applying it to the $\beta \to \infty$ low-temperature phase considered above would produce an exponentially enhanced probability of sampling low-energy micro-states, as desired. As $\beta \to 0$ in the high-temperature phase, there would be little change compared to the more straightforward pseudo-random sampling considered above, since all micro-states would become equally probable.