Lecturenotes 4 MCMC I – Contents

Statistical Physics and Potts Model

MC simulations of systems described by the Gibbs canonical ensemble aim at calculating estimators of physical observables at a temperature T. In the following we choose units so that $\beta = 1/T$ and consider the calculation of the expectation value of an observable $\mathcal{O}$. Mathematically all systems on a computer are discrete, because a finite word length has to be used. Hence,

$$\widehat{\mathcal{O}} = \widehat{\mathcal{O}}(\beta) = \langle \mathcal{O} \rangle \quad = \quad Z^{-1} \sum_{k=1}^{K} \mathcal{O}^{(k)} \, e^{-\beta \, E^{(k)}} \tag{1}$$

$$\text{where} \quad Z \; = \; Z(\beta) \quad = \quad \sum_{k=1}^{K} e^{-\beta \, E^{(k)}} \tag{2}$$

is the partition function. The index $k = 1, \ldots, K$ labels all configurations (or microstates) of the system, and $E^{(k)}$ is the (internal) energy of configuration k. To distinguish the configuration index from other indices, it is put in parenthesis.

We introduce generalized Potts models in an external magnetic field on d-dimensional hypercubic lattices with periodic boundary conditions. Without being overly complicated, these models are general enough to illustrate the essential features we are interested in. In addition, various subcases of these models are by themselves of physical interest. Generalizations of the algorithmic concepts to other models are straightforward, although technical complications may arise.

We define the energy function of the system by

$$-\beta\, E^{(k)} = -\beta\, E_0^{(k)} + H\, M^{(k)} \tag{3}$$

where

$$E_0^{(k)} = -2 \sum_{\langle ij \rangle} J_{ij}(q_i^{(k)}, q_j^{(k)})\ \delta(q_i^{(k)}, q_j^{(k)}) + \frac{2\,d\,N}{q} \tag{4}$$

$$\text{with} \quad \delta(q_i, q_j) = \begin{cases} 1 \text{ for } q_i = q_j \\ 0 \text{ for } q_i \neq q_j \end{cases} \qquad \text{and} \quad M^{(k)} = 2 \sum_{i=1}^{N} \delta(1, q_i^{(k)})\,.$$

The sum $\langle ij \rangle$ is over the nearest neighbor lattice sites and $q_i^{(k)}$ is called the Potts spin or Potts state of configuration k at site i. For the q-state Potts model $q_i^{(k)}$ takes on the values $1, \ldots, q$. The external magnetic field is chosen to interact with the state $q_i = 1$ at each site i, but not with the other states $q_i \neq 1$. The $J_{ij}(q_i, q_j)$, $(q_i = 1, \ldots, q;\ q_j = 1, \ldots, q)$ functions define the exchange coupling constants between the states at site i and site j. The energy function describes a number of physically interesting situations. With

$$J_{ij}(q_i, q_j) \equiv J > 0 \quad (\text{conventionally } J = 1) \tag{5}$$

the original model is recovered and $q = 2$ becomes equivalent to the Ising ferromagnet. The Ising case of Edwards-Anderson spin glasses and quadrupolar Potts glasses are obtained when the exchange constants are quenched random variables. Other choices of the J_{ij} include anti-ferromagnets and the fully frustrated Ising model.

For the energy per spin the notation is: $e_s = E/N$.

The normalization is chosen so that e_s agrees for $q = 2$ with the conventional Ising model definition, $\beta = \beta^{\rm Ising} = \beta^{\rm Potts}/2$.

For the $2d$ Potts models a number of exact results are known in the infinite volume limit, mainly due to work by Baxter. The phase transition temperatures are

$$\frac{1}{2}\,\beta_c^{\rm Potts} \;=\; \beta_c \;=\; \frac{1}{T_c} \;=\; \frac{1}{2}\,\ln(1+\sqrt{q}), \quad q = 2, 3, \ldots \ . \tag{6}$$

At β_c the average energy per state is

$$e_{0s}^c = E_0^c/N = \frac{4}{q} - 2 - 2/\sqrt{q} \ . \tag{7}$$

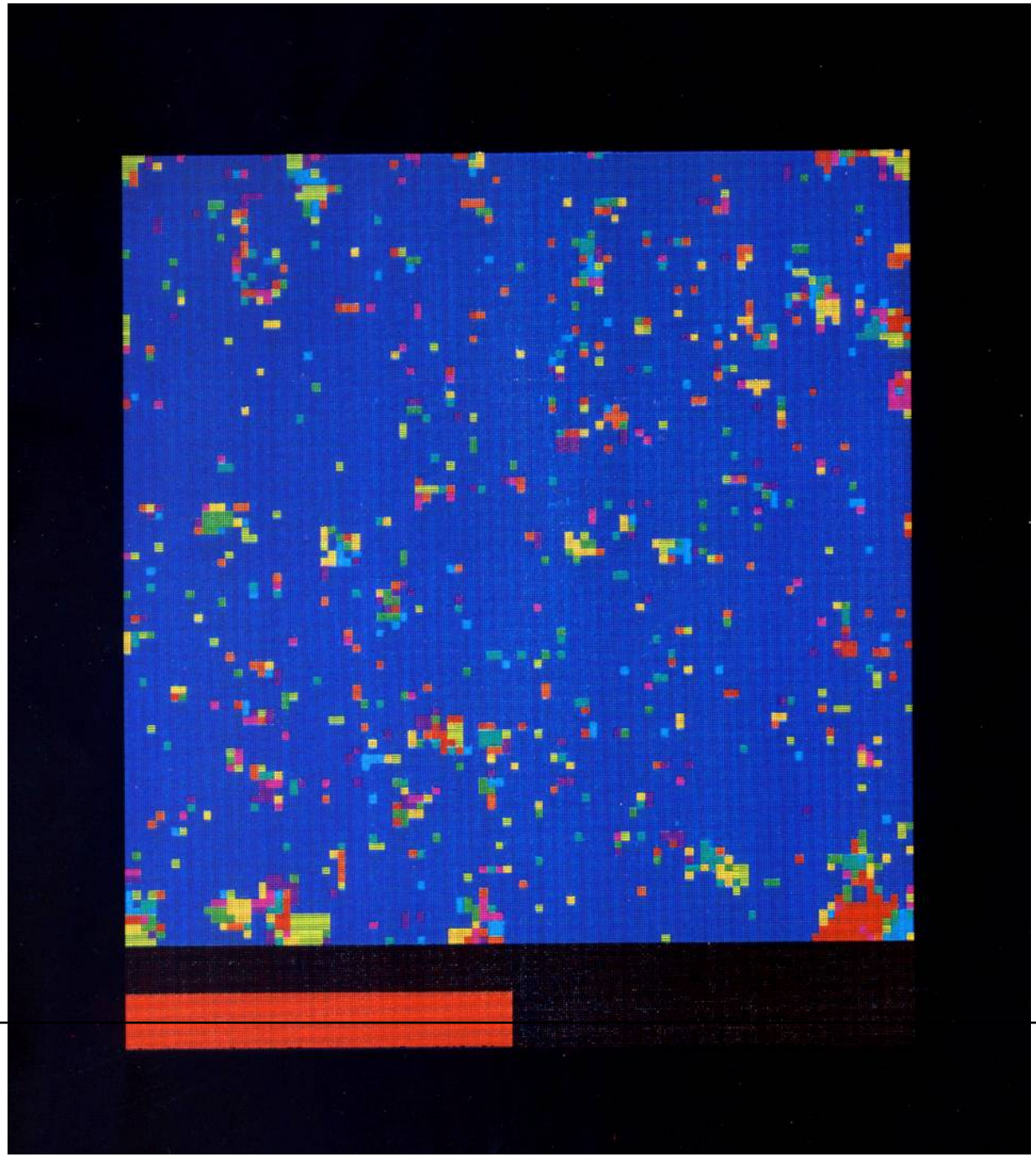
The phase transition is second order for $q \leq 4$ and first order for $q \geq 5$ for which the exact infinite volume latent heats $\triangle e_{0s}$ and entropy jumps $\triangle s$ were also found by Baxter, while the interface tensions f_s were derived later.

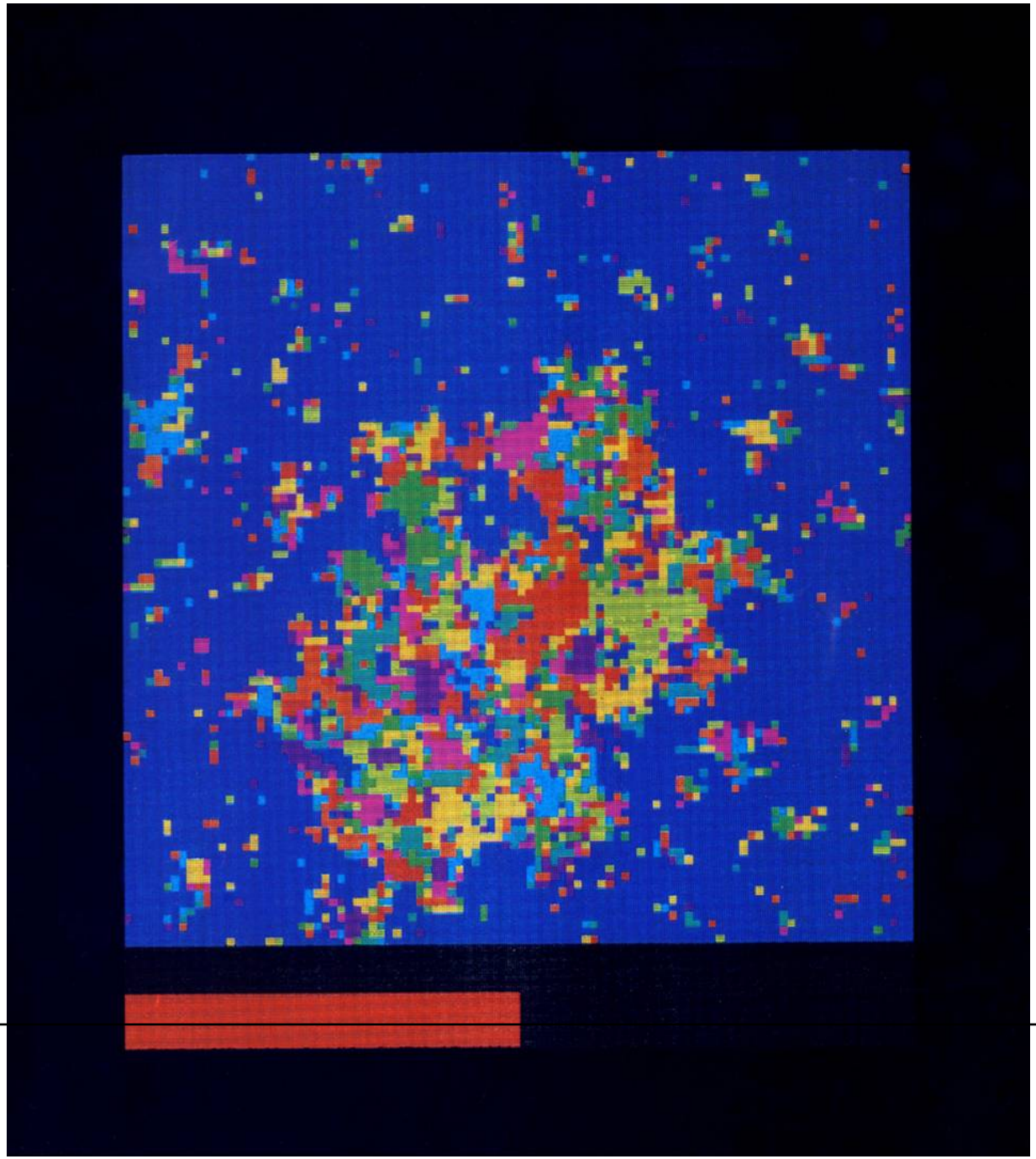
Some Potts Configurations $(2d,\ q = 10)$

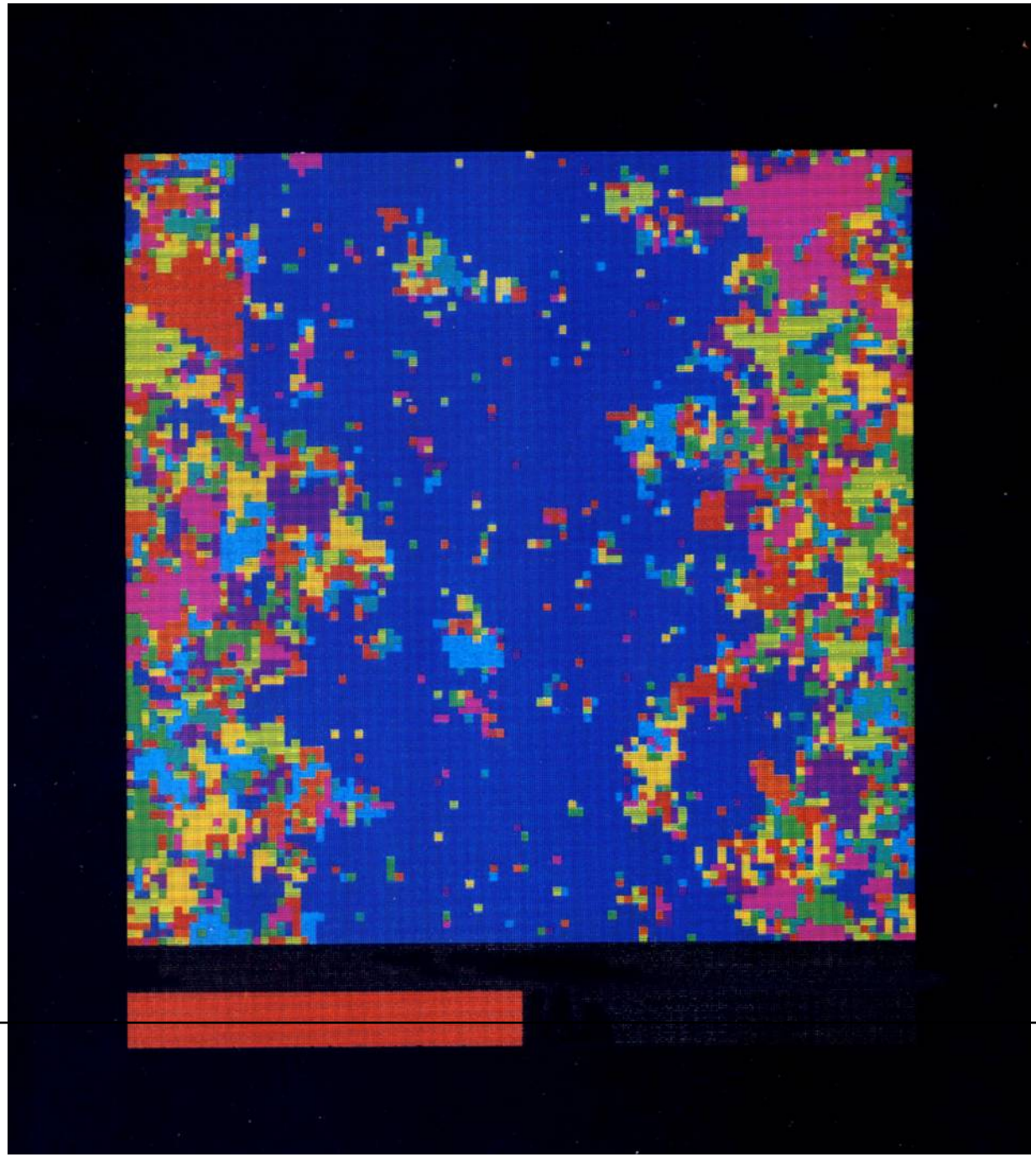
1. Ordered with small fluctuations.

2. Disordered droplet in ordered background.

3. Percolated disordered phase (order-disorder separation)







Sampling and Re-weighting

For the Ising model it is straightforward to sample statistically independent configurations. We simply have to generate N spins, each either up or down with 50% likelihood. This is called random sampling, see the figure.

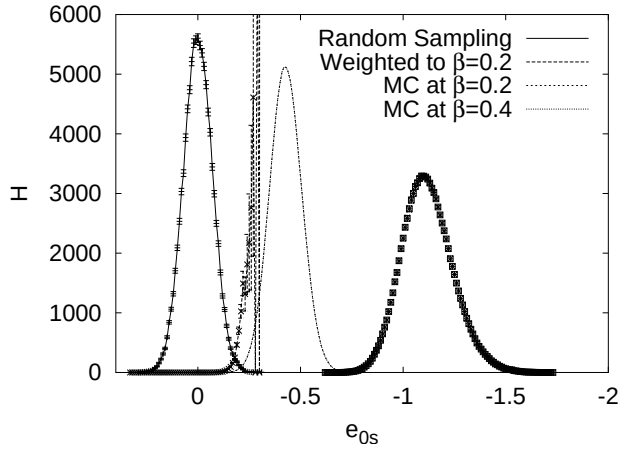


Figure 1: Energy histograms of 100 000 entries each for the Ising model on an 20×20 lattice: Random Sampling gives statistically independent configurations at $\beta = 0$. Histograms at $\beta = 0.2$ and $\beta = 0.4$ are generated by Markov chain MC. Re-weighting of the $\beta = 0$ random configurations to $\beta = 0.2$ is shown to fail (assignments a0301_02 and a0303_02).

Is is very important to distinguish the energy measurements on single configurations from the expectation value. The expectation value $\widehat{e}_s$ is a single number, while e_s fluctuates. From the measurement of many e_s values one finds estimators of its moments. The mean is is denoted by $\overline{e}_s$ and fluctuates.

The histogram entries at $\beta = 0$ can be re-weighted, so that they correspond to other β values. We simply have to multiply the entry corresponding to energy E by $c_\beta \exp(-\beta E)$. Similarly histograms corresponding to the Gibbs ensemble at some value β_0 can be re-weighted to other β values. (Care has to be taken to ensure that the involved arguments of the exponential function do not become too large! This can be done by logarithmic coding.)

Re-weighting has a long history. For Finite Size Scaling (FSS) investigations of second order phase transitions its usefulness of the re-weighting has been stressed by Ferrenberg and Swendsen (accurate determinations of peaks of the specific heat or of susceptibilities).

In the figure re-weighting is done from $\beta_0 = 0$ to $\beta = 0.2$. But, by comparison to the histogram from a Metropolis MC calculation at $\beta = 0.2$, the result is seen to be disastrous. The reason is easily identified: In the range where the $\beta = 0.2$ histogram takes on its maximum, the $\beta = 0$ histogram has not a single entry, $i.e.$, our naive sampling procedure misses the important configurations at $\beta = 0.2$. Re-weighting to new β values works only in a range $\beta_0 \pm \triangle\beta$, where $\triangle\beta \to 0$ in the infinite volume limit.

Important Configurations

Let us determine the important contributions to the partition function. The partition function can be re-written as a sum over energies

$$Z = Z(\beta) = \sum_E n(E)\, e^{-\beta\, E} \tag{8}$$

where the unnormalized spectral density $n(E)$ is defined as the number of microstates k with energy E (remember, on the computer energy values are always discrete).

For a fixed value of β the energy probability density

$$P_\beta(E) = c_\beta \, n(E) \, e^{-\beta E} \tag{9}$$

is peaked around the average value $\widehat{E}(\beta)$, where c_β is a normalization constant so that the $\sum_E P_\beta(E) = 1$ holds.

Away from first and second order phase transitions, the width of the energy distribution is $\triangle E \sim \sqrt{V}$ follows from the fact that, away from phase transition points, the fluctuations of the $N \sim V$ lattice spins are essentially uncorrelated, so that the magnitude of a typical fluctuations is $\sim \sqrt{N}$. From this we find that the re-weighting range is $\triangle\beta \sim 1/\sqrt{V}$, as the energy is an extensive quantity $\sim V$ so that $\triangle\beta E \sim \sqrt{V}$ can stay within the fluctuation of the system.

Interestingly, the re-weighting range increases at second order phase transitions point, because critical fluctuations are larger than non-critical fluctuations. Namely, one has $\triangle E \sim V^x$ with $1/2 < x < 1$ and the requirement $\triangle\beta E \sim V^x$ yields $\triangle\beta \sim V^{x-1}$.

For first order phase transitions one has a latent heat $\triangle V \sim V$, but this does not mean that the re-weighting range becomes of order one. In essence, the fluctuations collapse, because the two phases become separated by an interfacial tension. One is back to fluctuations within either of the two phases, *i.e.* $\triangle\beta \sim 1/\sqrt{V}$.

The important configurations at temperature $T = 1/\beta$ are at the energy values for which the probability density $P_\beta(E)$ is large. To sample them efficiently, one needs a procedure which generates the configurations with their Boltzmann weights

$$w_B^{(k)} = e^{-\beta E^{(k)}} \quad . \tag{10}$$

The number of configurations $n(E)$ and the weights combine then so that the probability to generate a configuration at energy E becomes precisely $P_\beta(E)$ as given by equation (9).

Importance Sampling and Markov Chain Monte Carlo

For the canonical ensemble Importance sampling generates configurations k with probability

$$P_B^{(k)} = c_B \, w_B^{(k)} = c_B \, e^{-\beta E^{(k)}} , \qquad c_B \;\; \text{constant} . \tag{11}$$

The state vector $(P_B^{(k)})$, for which the configurations are the vector indices, is called Boltzmann state. The expectation value becomes the arithmetic average:

$$\widehat{\mathcal{O}} = \widehat{\mathcal{O}}(\beta) = \langle \mathcal{O} \rangle = \lim_{N_K \to \infty} \frac{1}{N_K} \sum_{n=1}^{N_K} \mathcal{O}^{(k_n)} . \tag{12}$$

When the sum is truncated we obtain an estimator of the expectation value:

$$\overline{\mathcal{O}} = \frac{1}{N_K} \sum_{k=1}^{N_K} \mathcal{O}^{(k_n)} . \tag{13}$$

Normally, we cannot generate configurations k directly with probability (11). But they may be found as members of the equilibrium distribution of a dynamic process. In practice Markov chains are used. A Markov process is a particularly simple dynamic process, which generates configuration k_{n+1} stochastically from configuration k_n, so that no information about previous configurations $k_{n-1}, k_{n-2}, \ldots$ is needed. The elements of the Markov chain time series are the configurations. Assume that the configuration k is given. Let the transition probability to create the configuration l in one step from k be given by $W^{(l)(k)} = W[k \to l]$. In essence, the matrix

$$W = \left(W^{(l)(k)} \right) \tag{14}$$

defines the Markov process. Note, that this transition matrix is a very big (never stored in the computer!), because its labels are the configurations. To achieve our goal to generate configurations with the desired probabilities, the matrix W is required to satisfy the following properties:

(i) **Ergodicity (Irreducible in Math Literature):**

$$e^{-\beta E^{(k)}} > 0 \quad \text{and} \quad e^{-\beta E^{(l)}} > 0 \quad \text{imply :} \tag{15}$$

an integer number $n > 0$ exists so that $(W^n)^{(l)(k)} > 0$ holds.

(ii) **Normalization:**

$$\sum_l W^{(l)(k)} = 1 \ . \tag{16}$$

(iii) **Balance (Stationarity):**

$$\sum_k W^{(l)(k)} \, e^{-\beta E^{(k)}} \ = \ e^{-\beta E^{(l)}} \ . \tag{17}$$

Balance means: The Boltzmann state (11) is an eigenvector with eigenvalue 1 of the transition matrix $W = (W^{(l)(k)})$.

In statistical physics the ensemble notation is frequently used. By definition, an ensemble is a collection of configurations so that to each configuration k a probability $P^{(k)}$ is assigned, $\sum_k P^{(k)} = 1$. The Gibbs or Boltzmann ensemble E_B is defined to be the ensemble with probability distribution (11).

An equilibrium ensemble E_{eq} of the Markov process is defined by its probability distribution P_{eq} satisfying

$$W\,P_{eq} = P_{eq}\,, \quad \text{in components} \quad P_{eq}^{(l)} = \sum_k W^{(l)(k)} P_{eq}^{(k)}\ . \tag{18}$$

Statement: Under the conditions (i), (ii) and (iii) the Boltzmann ensemble is the only equilibrium ensemble of the Markov process and an attractive fixed-point.

Proof: Let us first define a distance between ensembles. Suppose we have two ensembles E and E', each of which is a collection of many configurations. Denote the probability for configuration k in E by $P^{(k)}$ and in E' by $P'^{(k)}$. We define the distance between E and E' to be

$$||E - E'|| \;=\; \sum_k |P^{(k)} - P'^{(k)}| \;, \tag{19}$$

where the sum goes over all configurations. Suppose that E' resulted from the application of the transition matrix W to the ensemble E. We can compare the distance of E' from the Boltzmann ensemble with the distance of E from the Boltzmann ensemble:

$$||E' - E_B|| \;=\; \sum_l \left| \sum_k W^{(l)(k)} \left(P^{(k)} - P_B^{(k)}\right) \right| \qquad \text{(using balance)}$$

$$\leq \; \sum_l \sum_k \left| W^{(l)(k)} \left(P^{(k)} - P_B^{(k)}\right) \right| \qquad \text{(using the triangle inequality)}$$

$$= \sum_k \left| P^{(k)} - P_B^{(k)} \right| \qquad \text{(using } W^{(l)(k)} \geq 0 \text{ and normalization)}$$

$$= ||E - E_B|| \; . \tag{20}$$

The last line is obtained by making use of the condition $\sum_l W^{(l)(k)} = 1$ and of $W^{(l)(k)} \geq 0$. This shows that the Markov process can reduce the distance between an ensemble E and the Boltzmann ensemble E_B and will never increase this distance. Ergodicity is needed to proof that the Boltzmann ensemble is the only equilibrium ensemble and to rule out the equality in the $\leq$ sign in equation (20). We now study the approach to the fixed point and derive and explicit equation for the convergence.

The matrix W by itself is normally not ergodic. Instead, due to the finiteness of the system, a number n exists, so that the matrix $\mathcal{W} = W^n$ is ergodic. This means, all matrix elements of $\mathcal{W}$ are larger than zero. In particular, excluding states with infinite energy from our considerations,

$$1 > w_{\min} = \min_{k,l} \left(\mathcal{W}^{(k)(l)} \right) > 0 \tag{21}$$

holds. Let us assume that the state vectors P and P' are related by

$$P' \;=\; \mathcal{W}\, P\,. \tag{22}$$

As shown in the following, this implies the inequality

$$||E' - E_B|| \;\leq\; (1 - w_{\min})\, ||E - E_B|| \;. \tag{23}$$

Let $\epsilon = ||E - E_B||$. We can decompose the contributions to $||E - E_B||$ in $P^{(k)} - P_B^{(k)} \geq 0$ and in $P^{(k)} - P_B^{(k)} < 0$ terms,

$$||E - E_B|| \;=\; \sum_{k \in K^+} \left(P^{(k)} - P_B^{(k)} \right) + \sum_{k \in K^-} \left(P_B^{(k)} - P^{(k)} \right) \;. \tag{24}$$

Then the normalization $\sum_k P^{(k)} = \sum_k P_B^{(k)} = 1$ implies

$$\sum_{k\in K^+} \left(P^{(k)} - P_B^{(k)}\right) = \sum_{k\in K^-} \left(P_B^{(k)} - P^{(k)}\right) = \epsilon/2 \tag{25}$$

as is seen by adding (gives $2\sum_{k\in K^+}$)

$$0 = \sum_{k\in K^+} \left(P^{(k)} - P_B^{(k)}\right) + \sum_{k\in K^+} \left(P_B^{(k)} - P^{(k)}\right)$$

to equation (24) and, similarly, this equation with K^+ replaced by K^-. Using $\mathcal{W}P_B = P_B$, we have (back in original order to apply $\mathcal{W}$)

$$||E' - E_B|| \; = \tag{26}$$

$$\sum_l \left| \sum_{k\in K^+} \mathcal{W}^{(l)(k)} \left(P^{(k)} - P_B^{(k)}\right) - \sum_{k\in K^-} \mathcal{W}^{(l)(k)} \left(P_B^{(k)} - P^{(k)}\right) \right|$$

$$= \sum_{l\in L^+}\left(\sum_{k\in K^+}\mathcal{W}^{(l)(k)}\left(P^{(k)}-P_B^{(k)}\right)-\sum_{k\in K^-}\mathcal{W}^{(l)(k)}\left(P_B^{(k)}-P^{(k)}\right)\right)$$

$$+ \sum_{l\in L^-}\left(\sum_{k\in K^-}\mathcal{W}^{(l)(k)}\left(P_B^{(k)}-P^{(k)}\right)-\sum_{k\in K^+}\mathcal{W}^{(l)(k)}\left(P^{(k)}-P_B^{(k)}\right)\right)$$

Here L^+ is defined as the set of configurations l which fulfills the inequality

$$\sum_{k\in K^+}\mathcal{W}^{(l)(k)}\left(P^{(k)}-P_B^{(k)}\right)\geq\sum_{k\in K^-}\mathcal{W}^{(l)(k)}\left(P_B^{(k)}-P^{(k)}\right)\ ,$$

while L^- is the set of configurations l with

$$\sum_{k\in K^-}\mathcal{W}^{(l)(k)}\left(P_B^{(k)}-P^{(k)}\right)>\sum_{k\in K^+}\mathcal{W}^{(l)(k)}\left(P^{(k)}-P_B^{(k)}\right)\ .$$

Normalization (16) and that the smallest matrix element is larger than zero (21) imply

$$1 \;\geq\; \sum_{l \in L^+} \mathcal{W}^{(l)(k)} \;\geq\; w_{\min}$$

and the same equation with L^+ replaced by L^-. Inserting this into equation (26), picking the extremes 1 and $w_{\min}$ appropriately, yields

$$||E' - E_B|| \leq (1 - w_{\min})\,\epsilon/2 \;+\; (1 - w_{\min})\,\epsilon/2 \;=\; (1 - w_{\min})\,\epsilon\;. \qquad (27)$$

Under repeated application of the matrix we obtain a state

$$P^t \;=\; \mathcal{W}^t\,P, \quad \text{with} \;\; t = 1, 2, \ldots \qquad (28)$$

and find for the approach to the equilibrium ensemble

$$||E^t - E_B|| \;\leq\; \exp(-\lambda\,t)\,||E - E_B||\,, \quad \lambda = -\ln(1 - w_{\min}) > 0\,. \qquad (29)$$

Hence, the equilibrium ensemble is an attractive fixed-point.

There are many ways to construct a Markov process satisfying (i), (ii) and (iii). A stronger condition than balance (17) is

(iii') Detailed balance:

$$W^{(l)(k)}\, e^{-\beta E^{(k)}} \;=\; W^{(k)(l)} e^{-\beta E^{(l)}} \;. \tag{30}$$

Using the normalization $\sum_k W^{(k)(l)} = 1$ detailed balance implies balance (iii).

At this point we have replaced the canonical ensemble average by a time average over an artificial dynamics. Calculating then averages over large times, like one does in real experiments, is equivalent to calculating ensemble averages. One distinguishes dynamical universality classes. The Metropolis and heat bath algorithms discussed in the following fall into the class of model A or Glauber dynamics, which imitates the thermal fluctuations of nature to some extent. Cluster algorithms discussed constitute another universality class. Some recent attention has focused on dynamical universality classes of non-equilibrium systems.

The Metropolis Algorithm

Detailed balance still does not uniquely fix the transition probabilities $W^{(l)(k)}$. The Metropolis algorithm can be used whenever one knows how to calculate the energy of a configuration. Given a configuration k, the Metropolis algorithm proposes a configuration l with probability

$$f(l,k) \quad \text{normalized to} \quad \sum_l f(l,k) = 1 \ . \tag{31}$$

For $f(l,k)$ we derive a symmetry condition which ensures detailed balance. The new configuration l is accepted with probability

$$w^{(l)(k)} = \min\left[1, \frac{P_B^{(l)}}{P_B^{(k)}}\right] = \begin{cases} 1 & \text{for} \quad E^{(l)} < E^{(k)} \\ e^{-\beta(E^{(l)}-E^{(k)})} & \text{for} \quad E^{(l)} > E^{(k)}. \end{cases} \tag{32}$$

If the new configuration is rejected, the old configuration has to be counted again.

The Metropolis procedure gives rise to the transition probabilities

$$W^{(l)(k)} = f(l,k)\, w^{(l)(k)} \quad \text{for} \quad l \neq k \tag{33}$$

$$\text{and} \quad W^{(k)(k)} = f(k,k) + \sum_{l \neq k} f(l,k)\,(1 - w^{(l)(k)})\ . \tag{34}$$

Therefore, the ratio $\left(W^{(l)(k)}/W^{(k)(l)}\right)$ satisfies detailed balance (30) if

$$f(l,k) = f(k,l) \quad \text{holds}\,. \tag{35}$$

Otherwise the probability density $f(l,k)$ is unconstrained. So there is an amazing flexibility in the choice of the transition probabilities $W^{(l)(k)}$. One can even use acceptance probabilities distinct from those of equation (32) and the proposal probabilities are then not necessarily symmetric anymore (Hastings). Also, the algorithm generalizes immediately to arbitrary weights.

The acceptance rate is defined as the ratio of accepted changes over proposed moves (moves proposing the at hand configuration are not counted as accepted).

Potts Model Heatbath algorithm

The heatbath algorithm chooses a state q_i directly with the local Boltzmann distribution defined by its nearest neighbors. The state q_i can take on one of the values $1, \ldots, q$ and, with all other states set, determines a value of the energy function. We denote this energy by $E(q_i)$ and the Boltzmann probabilities are

$$P_B(q_i) \;=\; \text{const}\; e^{-\beta\, E(q_i)} \tag{36}$$

where the constant is determined by the normalization condition

$$\sum_{q_i=1}^{q} P_B(q_i) \;=\; 1\;. \tag{37}$$

In equation (36) we can define $E(q_i)$ to be just the contribution of the interaction of q_i with its nearest neighbors to the total energy and absorb the other contribution into the overall constant. The $E(q_i)$ values depend only on how the nearest

neighbors of the spin q_i partition into the values $1, \ldots, q$. For low values of q and the dimension d the most efficient implementation of the heatbath algorithm is to tabulate all possibilities. However, here we prefer to give a generic code which works for arbitrary values of q and d.

For this we calculate the cumulative distribution function of the heat bath probabilities

$$P_{HB}(q_i) = \sum_{q_i'=1}^{\mathtt{q_i}} P_B(q_i') \ . \tag{38}$$

The normalization condition (37) implies $P_{HB}(q) = 1$. Comparison of these cumulative probabilities with a uniform random number x^r yields the heat bath update $q_i \to q_i'$. Note that in the heat bath procedure the original value $q_i^{\rm in}$ does not influence the selection of $q_i^{\rm new}$.

Start and Equilibration

Initially we have to start with a microstate which may be far off the Boltzmann distribution. Far off means, that the Boltzmann probability (at temperature T) for a typical state of the initially generated distribution can be very, very small. Suppression factors like 10^{-10000} are well possible. Although the weight of states decreases with $1/n$ where n is the steps of the Markov process, one should exclude the initial states from the equilibrium statistics. In practice this means we should allow for a certain number of sweeps `nequi` to equilibrate the system.

Many ways to generate start configurations exist. Two natural and easy to implement choices are:

1. Generate a random configuration corresponding to $\beta = 0$. This defines a random or disordered start of a MC simulation.

2. Generate a configuration for which all Potts spins take on the same q-value. This is called an ordered start of a MC simulation.

Examples of initial time series:

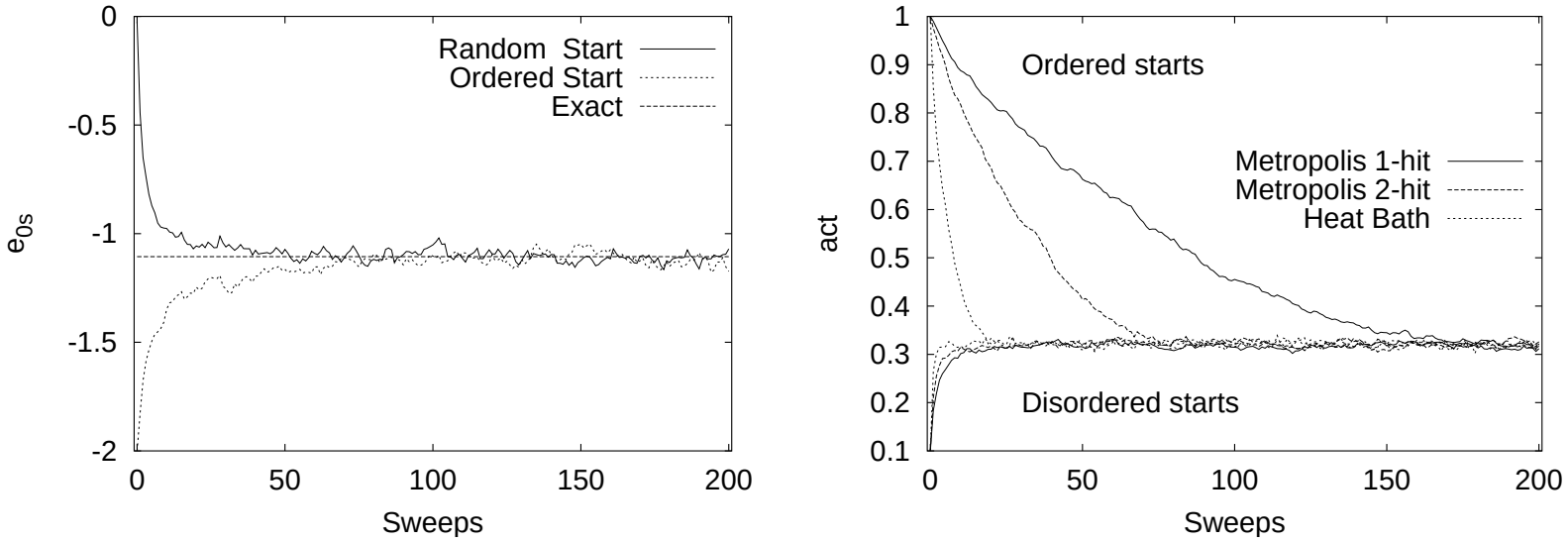


Figure 2: Left: Two Metropolis time series of 200 sweeps each for a $2d$ Ising model on a 80×80 lattice at $\beta = 0.4$ are shown. Ordered and disordered starts are used. The exact mean value $\widehat{e}_{0s} = -1.10608$ is also indicated (assignment a0302_01). Right: $q = 10$ Potts model time series of 200 sweeps on an 80×80 lattice at $\beta = 0.62$. Measurements of the action variable after every sweep are plotted for ordered and disordered starts (assignment a03003_05).

Energy Checks

For the $2d$ Ising model we can test against the exact finite lattice results of Ferdinand and Fisher. We simulate an 20^2 lattice at $\beta = 0.4$.

Now we use a statistics of 10 000 sweeps for reaching equilibrium and assume that this is an overkill. A more careful analysis is the subject of the next lecture.

The statistics for measurement is chosen 32 times larger: 64 bins of 5 000 sweeps each. The number 64 is taken, because according to the student distribution the approximation to the Gaussian approximation is then already excellent, while the binsize of 5 000 ($\gg 200$) is argued to be large enough to neglect correlations between the bins. With this statistics we find (assignment a0303_06)

$$\overline{e}_{0s} = -1.1172\ (14)\ \ (\text{Metropolis}) \quad \text{versus} \quad \widehat{e}_s = -1.117834\ \ (\text{exact})\ . \tag{39}$$

Performing the Gaussian difference test gives a perfectly admissible value $Q = 0.66$.

Specific Heat

With $\widehat{E} = \langle E \rangle$ the **specific heat** is defined by

$$C \;=\; \frac{d\,\widehat{E}}{d\,T} \;=\; \beta^2 \left(\langle E^2 \rangle - \langle E \rangle^2 \right) \;.$$

The last equal sign follows by working out the temperature derivative and is known as **fluctuation-dissipation theorem**. It is often used to estimate the specific heat from equilibrium simulations without relying on numerical derivatives. An equivalent formulation is

$$C \;=\; \beta^2 \left\langle \left(E - \widehat{E} \right)^2 \right\rangle$$

as is easily shown by expanding $\left(E - \widehat{E} \right)^2$ and working out the expectation values.

Defining

$$\texttt{act2lm} = \frac{\overline{\texttt{iactm}^2}}{\texttt{mlink}}$$

we have a notation close to the computer code:

$$C \;=\; \frac{\beta^2\, N\, d^2}{n} \sum_{i=1}^{n} \left(\texttt{act2l}_i - \texttt{actlm}^2\right) \,,$$

or

$$C \;=\; \frac{\beta^2\, N\, d^2}{n} \sum_{i=1}^{n} (\texttt{actl}_i - \texttt{actlm})^2 \,,$$

where the sums are over all measurements in the times series. When energy histograms are available, this is calculated by equations in `potts_mu2.f` of `ForLib`.

In the limit of an infinite statistics specific heat estimates from all these equation agree. But with a finite statistics a number of problems emerge. In the binning approach, where `nrpt` of the production run defines the number of blocks, one may

want to use for $\widehat{E}$ estimators $\overline{E}_i$ which are constructed from the histograms of the blocks. With N_b the number of data in each bin and H_i the energy histogram of block i:

$$\overline{E}_i = \frac{1}{N_b} \sum_{j \in \text{block}(i)} E_j \; = \frac{\sum_E E\, H_i(E)}{\sum_E H_i(E)} \; .$$

Estimates of $\overline{C}_i$ from either equation agree then and Gaussian error bars are expected due to the binning. However, for a not so good statistics as bias towards too small $\overline{C}_i$ values is occurs, because E_j and $\overline{E}_i$ come from the same block in the estimate

$$\overline{C}_i = \frac{\beta^2}{N_b} \sum_{j \in \text{block}(i)} \left(E_j^2 - \overline{E}_i^{\,2} \right) = \frac{\beta^2 \sum_E \left(E^2 - \overline{E}_i^{\,2} \right) H_i(E)}{\sum_E H_i(E)}$$

$$= \frac{\beta^2}{N_b} \sum_{j \in \text{block}(i)} \left(E_j - \overline{E}_i \right)^2 = \frac{\beta^2 \sum_E (E - \overline{E}_i)^2 \, H_i(E)}{\sum_E H_i(E)} \; .$$

The $\overline{E}_i$ estimators are certainly inferior to the estimate $\overline{E}$, which relies on the entire statistics. Therefore, one may consider to use $\overline{E}$ instead of $\overline{E}_i$. However, then one does not know anymore how to calculate the error bar of $\overline{C}$ as the $\overline{C}_i$ estimates would rely on overlapping data. The situation gets even worse when we include reweighting, as this non-linear procedure implies that the two estimators will in general differ. These difficulties are overcome by the jackknife method.

When histograms are used the fast way to create jackknife bins is to sum first the entire statistics:

$$H(E) = \sum_{i=1}^{N_b} H_i(E) \ .$$

Subsequently jackknife histograms (superscript J) are defined by

$$H_i^J(E) = H(E) - H_i(E)$$

and jackknife estimates $\overline{C}_i^J$ are obtained by using $H_i^J(E)$ instead of $H_i(E)$.