

# Lecture notes MCMC II – Contents

1. Statistical Errors of Markov Chain MC Data
2. Autocorrelations
3. Integrated Autocorrelation Time and Binning
4. Illustration: Metropolis generation of normally distributed data
5. Self-consistent versus reasonable error analysis
6. Comparison of Markov chain MC algorithms

## Statistical Errors of Markov Chain MC Data

In large scale MC simulation it may take months, possibly years, of computer time to collect the necessary statistics. For such data a thorough error analysis is a must. A typical MC simulation falls into two parts:

1. **Equilibration:** Initial sweeps are performed to reach the equilibrium distribution. During these sweeps measurements are either not taken at all or have to be discarded when calculating equilibrium expectation values.
2. **Production:** Sweeps with measurements are performed. Equilibrium expectation values are calculated from this statistics.

A rule of thumb is: **Do not spend more than 50% of your CPU time on measurements!** The reason for this rule is that, that one cannot be off by a factor worse than two ( $\sqrt{2}$  in the statistical error).

How many sweeps should be discarded for reaching equilibrium?

In a few situations this question can be rigorously answered with the **Coupling from the Past** method (Propp and Wilson).

The next best thing to do is to measure the integrated autocorrelation time self-consistently and to discard, after reaching a visually satisfactory situation, a number of sweeps which is larger than the integrated autocorrelation time. In practice even this can often not be achieved.

Therefore, it is re-assuring that it is sufficient to pick the number of discarded sweeps approximately right. With an increasing statistics the contribution of the non-equilibrium data dies out like  $1/N$ , where  $N$  is the number of measurements. For large  $N$  the effect is eventually swallowed by the statistical error, which declines only like  $1/\sqrt{N}$ . The point of discarding configurations for reaching equilibrium is that the factor in front of  $1/N$  can be large.

There can be far more involved situations, like that the Markov chain may end up in metastable configurations, which may even stay unnoticed (e.g. complex systems like spin glasses or proteins).

# Autocorrelations

We like to estimate the expectation value  $\hat{f}$  of some physical observable. We assume that the system has reached equilibrium. How many MC sweeps are needed to estimate  $\hat{f}$  with some desired accuracy? To answer this question, one has to understand the autocorrelations within the Markov chain.

Given is a **time series** of  $N$  measurements

$$f_i = f_i(x_i), \quad i = 1, \dots, N$$

from a Markov process, where  $x_i$  are the configurations generated. The label  $i = 1, \dots, N$  runs in the temporal order of the Markov chain and the elapsed time, measured in updates or sweeps, between subsequent measurements  $f_i, f_{i+1}$  is always the same, independently of  $i$ . The estimator of the expectation value  $\hat{f}$  is

$$\bar{f} = \frac{1}{N} \sum f_i .$$

With the notation

$$t = |i - j|$$

the definition of the **autocorrelation function** of the observable  $\hat{f}$  is

$$\hat{C}(t) = \hat{C}_{ij} = \langle (f_i - \langle f_i \rangle) (f_j - \langle f_j \rangle) \rangle = \langle f_i f_j \rangle - \langle f_i \rangle \langle f_j \rangle = \langle f_0 f_t \rangle - \hat{f}^2 \quad (1)$$

where we used that translation invariance in time holds for the equilibrium ensemble. The asymptotic behavior for large  $t$  is

$$\hat{C}(t) \sim \exp\left(-\frac{t}{\tau_{\text{exp}}}\right) \quad \text{for } t \rightarrow \infty,$$

where  $\tau_{\text{exp}}$  is called **(exponential) autocorrelation time** and is related to the second largest eigenvalue  $\lambda_1$  of the transition matrix by

$$\tau_{\text{exp}} = -\ln \lambda_1$$

under the assumption that  $f$  has a non-zero projection on the corresponding eigenstate. Superselection rules are possible so that different autocorrelation times reign for different operators.

The variance of  $f$  is a special case of the autocorrelations (1)

$$\hat{C}(0) = \sigma^2(f) .$$

Some algebra shows that the variance of the estimator  $\bar{f}$  of the mean is now

$$\sigma^2(\bar{f}) = \frac{\sigma^2(f)}{N} \left[ 1 + 2 \sum_{t=1}^{N-1} \left( 1 - \frac{t}{N} \right) \hat{c}(t) \right] \quad \text{with} \quad \hat{c}(t) = \frac{\hat{C}(t)}{\hat{C}(0)} .$$

This equation ought to be compared with the corresponding equation for uncorrelated random variables  $\sigma^2(\bar{f}) = \sigma^2(f)/N$ . The difference is the factor

in the bracket which defines the **integrated autocorrelation time**

$$\tau_{\text{int}} = \left[ 1 + 2 \sum_{t=1}^{N-1} \left( 1 - \frac{t}{N} \right) \hat{c}(t) \right] .$$

For correlated data the variance of the mean is by the factor  $\tau_{\text{int}}$  larger than the corresponding **naive variance** for uncorrelated data:

$$\tau_{\text{int}} = \frac{\sigma^2(\bar{f})}{\sigma_{\text{naive}}^2(\bar{f})} \quad \text{with} \quad \sigma_{\text{naive}}^2 = \frac{\sigma^2(f)}{N} . \quad (2)$$

In most simulations one is interested in the limit  $N \rightarrow \infty$  and  $\tau_{\text{int}}$  becomes

$$\tau_{\text{int}} = 1 + 2 \sum_{t=1}^{\infty} \hat{c}(t) . \quad (3)$$

The numerical estimation of the integrated autocorrelation time faces difficulties.

The variance of the estimator of  $\tau_{\text{int}}$  diverges for  $N \rightarrow \infty$ :

$$\bar{\tau}_{\text{int}} = 1 + 2 \sum_{t=1}^{\infty} \bar{c}(t) \quad \text{and} \quad \sigma^2(\bar{\tau}_{\text{int}}) \rightarrow \infty$$

because for large  $t$  each  $\bar{c}(t)$  adds a constant amount of noise, whereas the signal dies out like  $\exp(-t/\tau_{\text{exp}})$ . To obtain an estimate one considers the  $t$ -dependent estimator

$$\bar{\tau}_{\text{int}}(t) = 1 + 2 \sum_{t'=1}^t \bar{c}(t')$$

of the integrated autocorrelation time and looks out for a **window** in  $t$  for which  $\bar{\tau}_{\text{int}}(t)$  is flat. To give a **simple example**, let us assume that the autocorrelation function is governed by a single exponential autocorrelation time

$$\hat{C}(t) = \text{const} \exp\left(-\frac{t}{\tau_{\text{exp}}}\right).$$



In this case we can carry out the sum (3) for the integrated autocorrelation function and find

$$\tau_{\text{int}} = 1 + 2 \sum_{t=1}^{\infty} e^{-t/\tau_{\text{exp}}} = 1 + \frac{2 e^{-1/\tau_{\text{exp}}}}{1 - e^{-1/\tau_{\text{exp}}}} .$$

In particular, the difference between the asymptotic value and the finite  $t$  definition becomes then

$$\tau_{\text{int}} - \tau_{\text{int}}(t) = 2 e^{-t/\tau_{\text{exp}}} \sum_{t'=1}^{\infty} e^{-t'/\tau_{\text{exp}}} = \frac{2 e^{-(t+1)/\tau_{\text{exp}}}}{1 - e^{-1/\tau_{\text{exp}}}} .$$

For a large exponential autocorrelation time  $\tau_{\text{exp}} \gg 1$  the approximation

$$\tau_{\text{int}} = 1 + \frac{2 e^{-1/\tau_{\text{exp}}}}{1 - e^{-1/\tau_{\text{exp}}}} \cong 1 + \frac{2 - 2/\tau_{\text{exp}}}{1/\tau_{\text{exp}}} = 2 \tau_{\text{exp}} - 1 \cong 2 \tau_{\text{exp}}$$

holds.

## Integrated Autocorrelation Time and Binning

Using binning the integrated autocorrelation time can also be estimated via the variance ratio. We bin the time series into  $N_{bs} \leq N$  bins of

$$N_b = \text{NBIN} = \left[ \frac{N}{N_{bs}} \right] = \left[ \frac{\text{NDAT}}{\text{NBINS}} \right]$$

data each. Here  $[.]$  stands for Fortran integer division, *i.e.*,  $N_b = \text{NBIN}$  is the largest integer  $\leq N/N_{bs}$ , implying  $N_{ba} \cdot N_b \leq N$ . It is convenient to choose the values of  $N$  and  $N_{bs}$  so that  $N$  is a multiple of  $N_{bs}$ . The binned data are the averages

$$f_j^{N_b} = \frac{1}{N_b} \sum_{i=1+(j-1)N_b}^{jN_b} f_i \quad \text{for } j = 1, \dots, N_{bs} .$$

For  $N_b > \tau_{\text{exp}}$  the autocorrelations are essentially reduced to those between neighbor bins and even these approach zero under further increase of the binlength.

For a set of  $N_{bs}$  binned data  $f_j^{N_b}$ , ( $j = 1, \dots, N_{bs}$ ) we may calculate the mean with its **naive error bar**. Assuming for the moment an infinite time series, we find the integrated autocorrelation time (2) from the following ratio of sample variances

$$\tau_{\text{int}} = \lim_{N_b \rightarrow \infty} \tau_{\text{int}}^{N_b} \quad \text{with} \quad \tau_{\text{int}}^{N_b} = \left( \frac{s_{\frac{2}{f}N_b}^2}{s_{\frac{2}{f}}^2} \right). \quad (4)$$

In practice the  $N_b \rightarrow \infty$  limit will be reached for a sufficiently large, finite value of  $N_b$ . The statistical error of the  $\tau_{\text{int}}$  estimate (4) is, in the first approximation, determined by the errors of  $s_{\frac{2}{f}N_b}^2$ . The typical situation is then that, due to the central limit theorem, the binned data are approximately Gaussian, so that the **error of  $s_{\frac{2}{f}N_b}^2$  is analytically known** from the  $\chi^2$  distribution. Finally, the fluctuations of  $s_{\frac{2}{f}}^2$  of the denominator give rise to a small correction which can be worked out.

Numerically most accurate estimates of  $\tau_{\text{int}}$  are obtained for the finite binlength  $N_b$  which is just large enough that the binned data are practically uncorrelated.

For applications it is convenient to choose  $N$  and  $N_b$  to be powers of 2. In the following we assume

$$N = 2^K, \quad K \geq 4 \quad \text{and} \quad N_b = 2^{K_b} \quad \text{with} \quad K_b = 0, 1, \dots, K - 5, K - 4. \quad (5)$$

Choosing the maximum value of  $K_b$  to be  $K - 4$  implies that the smallest number of bins is

$$N_{bs}^{\min} = 2^4 = 16. \quad (6)$$

While the Student distribution shows that the confidence intervals of the error bars from 16 uncorrelated normal data are reasonable approximations to those of the Gaussian standard deviation, about 1000 independent data are needed to provide a decent estimate of the corresponding variance (at the 95% confidence level with an accuracy of slightly better than 10%).

It makes sense to work with error bars from 16 binned data, but the error of the error bar, and hence a reliable estimate of  $\tau_{\text{int}}$ , requires far more data.

## Illustration: Metropolis generation of normally distributed data

We generate normally distributed data according to the Markov process

$$x' = x + 2a x^r - a \quad (7)$$

where  $x$  is the event at hand,  $x^r$  a uniformly distributed random number in the range  $[0, 1)$ , and the real number  $a > 0$  is a parameter which relates to the efficiency of the algorithm. The new event  $x'$  is accepted with the Metropolis probability

$$P_{\text{accept}}(x') = \begin{cases} 1 & \text{for } x'^2 \leq x^2; \\ \exp[-(x'^2 - x^2)/2] & \text{for } x'^2 > x^2. \end{cases} \quad (8)$$

If  $x'$  is rejected, the event  $x$  is counted again. The Metropolis process introduces an autocorrelation time in the generation of normally distributed random data.

We work with  $K = 17$ , *i.e.*,  $N = 2^{17} = 131072$ , data and take  $a = 3$  for the Markov process (7), what gives an acceptance rate of approximately 50%.

## The autocorrelation function (assignment a0401\_01):

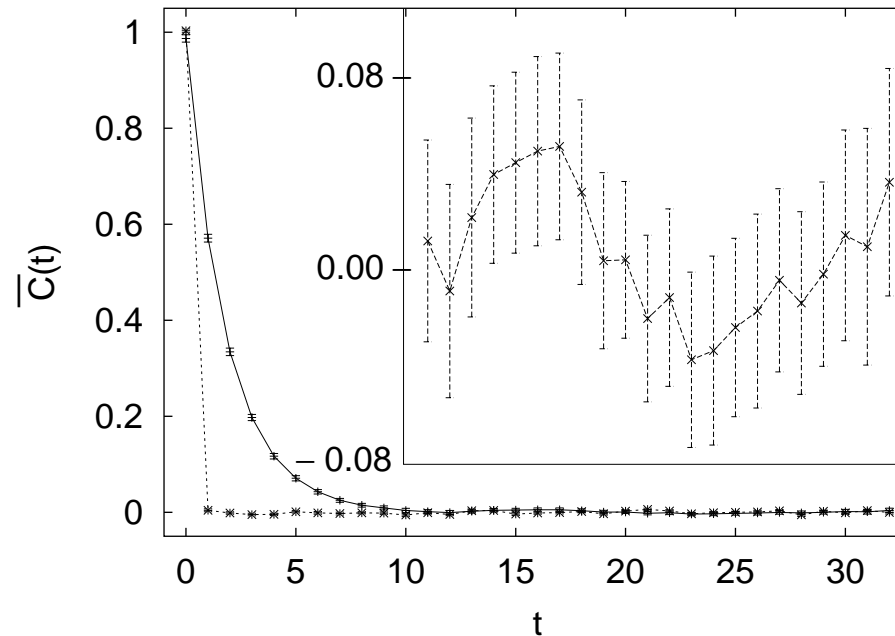


Figure 1: The autocorrelation function. Upper data: Metropolis time series for the normal distribution. Lower data: Gaussian random number generator. For  $t \geq 11$  the inlay shows the autocorrelations on an enlarged ordinate. The straight lines between the data points are just to guide the eyes. The curves start with  $\overline{C}(0) \approx 1$  because the variance of the normal distribution is one.

## Integrated autocorrelation time (assignment a0401\_02):

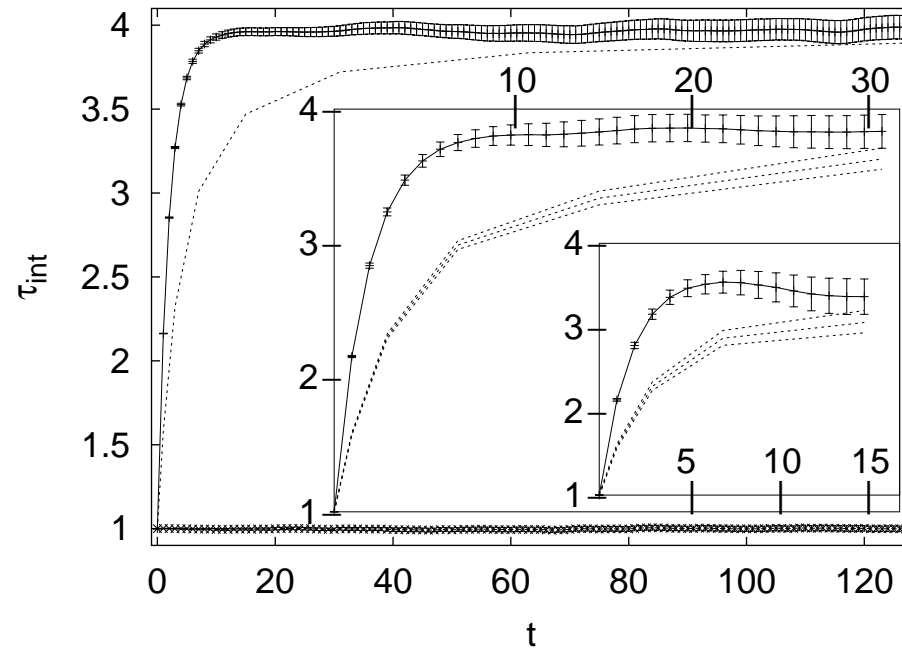


Figure 2: Upper curves in the figure and its inlays:  $\tau_{\text{int}}(t)$ . Lowest curve: Gaussian random number generator. Remaining curves: Binning estimators of the integrated autocorrelation time with one standard deviation bounds. The main figure relies on  $2^{21}$  data, the first inlay on  $2^{17}$  data, and the second inlay relies on  $2^{14}$  data.

We compare the  $\tau_{\text{int}}^{N_b}$  estimators with the direct estimators  $\tau_{\text{int}}(t)$  at

$$t = N_b - 1 . \quad (9)$$

With this relation the estimators agree for binlength  $N_b = 1$  and for larger  $N_b$  the relation gives the range over which we combine data into either one of the estimators. The approach of the binning procedure towards the asymptotic  $\tau_{\text{int}}$  value is slower than that of the direct estimate of  $\tau_{\text{int}}$ .

For our large  $\text{NDAT} = 2^{21}$  data set  $\tau_{\text{int}}(t)$  reaches its plateau before  $t = 20$ . All the error bars within the plateau are strongly correlated. Therefore, it is not recommendable to make an attempt to combine them. Instead, it is save to pick an appropriate single value and its error bar as the final estimate:

$$\tau_{\text{int}} = \tau_{\text{int}}(20) = 3.962 \pm 0.024 \quad \text{from } 2^{21} = 2,097,152 \text{ data.} \quad (10)$$

The binning procedure, on the other hand, shows an increase of  $\tau_{\text{int}}^{N_b}$  all the way to



$N_b = 2^7 = 128$ , where the estimate with the one confidence level error bounds is

$$3.85 \leq \tau_{\text{int}}^{128} \leq 3.94 \quad \text{from } 2^{14} = 16,384 \text{ bins from } 2^{21} \text{ data .}$$

How many data are needed to allow for a meaningful estimate of the integrated autocorrelation time?

For the statistics of  $\text{NDAT} = 2^{17}$  the autocorrelation signal disappears for  $t \geq 11$  into the statistical noise. Still, there is clear evidence of the hoped for window of almost constant estimates. A conservative choice is to take  $t = 20$  again, which now gives

$$\tau_{\text{int}} = \tau_{\text{int}}(20) = 3.86 \pm 0.11 \quad \text{from } 2^{17} \text{ data .}$$

Worse is the binning estimate, which for the  $2^{17}$  data is

$$3.55 \leq \tau_{\text{int}}^{32} \leq 3.71 \quad \text{from } 2^{12} = 4,096 \text{ bins from } 2^{17} = 131,072 \text{ data .}$$

Our best value (10) is no longer covered by the two standard deviation zone.

For the second inlay the statistics is reduced to  $\text{NDAT} = 2^{14}$ . With the integrated autocorrelation time rounded to 4, this is 4096 times  $\tau_{\text{int}}$ . For binlength  $N_b = 2^4 = 16$  we are then down to  $N_{bs} = 1024$  bins, which are needed for accurate error bars of the error. To work with this number we limit, in accordance with equation (9), our  $\tau_{\text{int}}(t)$  plot to the range  $t \leq 15$ . Still, we find a quite nice window of nearly constant  $\tau_{\text{int}}(t)$ , namely all the way from  $t = 4$  to  $t = 15$ . By a statistical fluctuation (assignment a0401\_03)  $\tau_{\text{int}}(t)$  takes its maximum value at  $t = 7$  and this makes  $\tau_{\text{int}}(7) = 3.54 \pm 0.13$  a natural candidate. However, this value is inconsistent with our best estimate (10). The true  $\tau_{\text{int}}(t)$  increases monotonically as function of  $t$ , so we know that the estimators have become bad for  $t > 7$ . The error bar at  $t = 7$  is obviously too small to take care of our difficulties. One may combine the  $t = 15$  error bar. In this way the result is

$$\tau_{\text{int}} = 3.54 \pm 0.21 \quad \text{for } 2^{14} = 16,384 \text{ data,} \quad (11)$$

which achieves consistency in the two error bar range. For binlength  $N_b = 16$  the

binning estimate is

$$2.93 \leq \tau_{\text{int}}^{16} \leq 3.20 \quad \text{from } 2^{10} = 1,024 \text{ bins from } 2^{14} \text{ data.} \quad (12)$$

Clearly, the binlength  $N_b = 16$  is too small for an estimate of the integrated autocorrelation time. We learn from this investigation that one needs a binlength of at least ten times the integrated autocorrelation time  $\tau_{\text{int}}$ , whereas for the direct estimate it is sufficient to have  $t$  about four times larger than  $\tau_{\text{int}}$ .

## Self-consistent versus reasonable error analysis

By visual inspection of the time series, one may get an impression about the length of the out-of-equilibrium part of the simulation. On top of this one should allow  $> \tau_{\text{int}}$  sweeps for the system to settle. A second reason why it appears necessary to control the the integrated autocorrelation times are the statistical errors of our measurements. Ideally the error bars are calculated as

$$\Delta \bar{f} = \sqrt{\sigma^2(\bar{f})} \quad \text{with} \quad \sigma^2(\bar{f}) = \tau_{\text{int}} \frac{\sigma^2(f)}{N} .$$

This constitutes a **self-consistent error analysis** of a MC simulation.

However, the calculation of the integrated autocorrelation time may be out of reach. According to the Student distribution about **twenty** independent data are sufficient to estimate mean values with reasonably reliable error bars, while **one thousand** are needed for a 10% error on the error bar at the two sigma confidence level.

In practice, one has to be content with what can be done. Often this means to rely on the binning method. We simply calculate error bars of our ever increasing statistics with respect to a fixed number of NBINS bins, where

$$\text{NBINS} \geq 16$$

are large enough values. In addition, we may put 10% of the initially planned simulation time apart for reaching equilibrium. *A-posteriori*, this can always be increased. Once the statistics is large enough, this means at least  $t_{\max} > \text{NBINS} \tau_{\text{int}}$  for the largest  $t$  of our time series, our small number of binned data become effectively independent.

How do we know that the statistics has become large enough? In practical applications there can be indirect arguments, which tell or suggest us that the integrated autocorrelation time is in fact (much) smaller than the achieved bin length. This is no longer self-consistent, as we perform no explicit measurement of  $\tau_{\text{int}}$ , but it is a **reasonable error analysis**.

## Comparison of Markov chain MC algorithms

Is the 1-hit Metropolis algorithm more efficient with sequential updating or with random updating (assignment a0402\_01 B)?

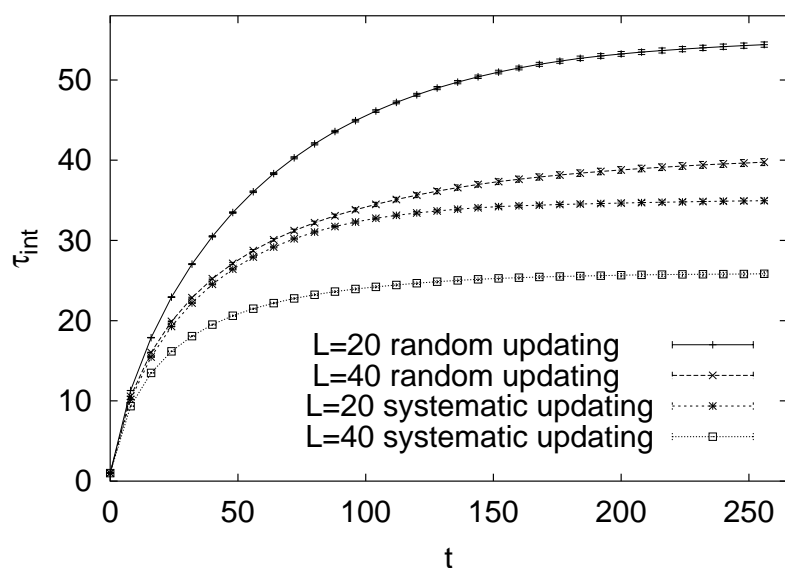


Figure 3: Integrated autocorrelation times of the Metropolis process with random updating versus sequential updating for the  $d = 2$  Ising model at  $\beta = 0.4$ . The ordering of the curves is identical with that of the labels in the figure.

The  $d = 2$  Ising model off and on the critical point (assignment a0402\_01 A).

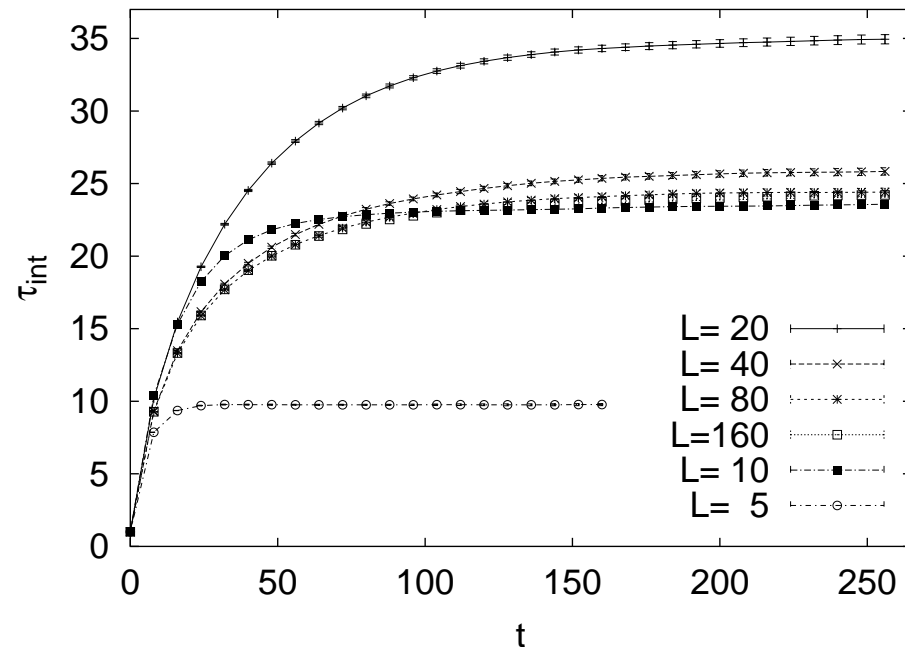


Figure 4: One-hit Metropolis algorithm with sequential updating: Lattice size dependence of the integrated autocorrelation time for the  $d = 2$  Ising model at  $\beta = 0.4$ . The ordering of the curves is identical with that of the labels in the figure.

**Critical slowing down:** Increase of the  $\tau_{\text{int}} \sim L^z$  at the critical point, where  $z \approx 2$  is the **dynamical critical exponent** (assignment a0402\_02 D). Estimates of  $z$  are compiled in the book by Landau and Binder.

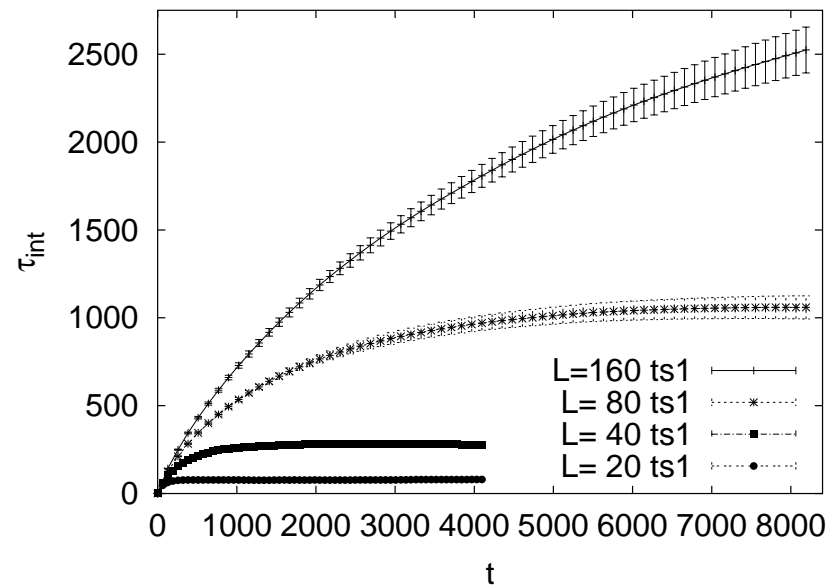


Figure 5: One-hit Metropolis algorithm with sequential updating:  $L$  dependence of the integrated autocorrelation time for the  $d = 2$  Ising model at its critical temperature. The ordering of the curves is identical with that of the labels.



Another MC dynamics, Swendsen-Wang (SW) and Wolff (W) cluster algorithm:

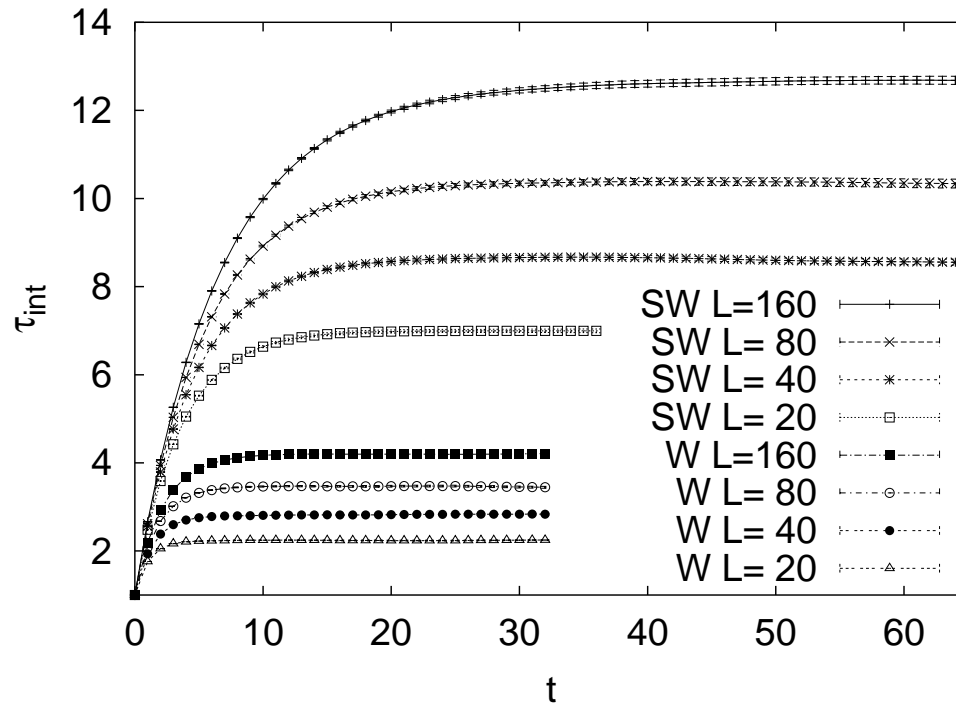


Figure 6: Estimates of integrated autocorrelation times from simulations of the  $d = 2$  Ising model at the critical temperature  $\beta_c = 0.44068679351$  (assignment a0503\_05).

## Metropolis versus heat bath updating for the 10-state $d = 2$ Potts model:

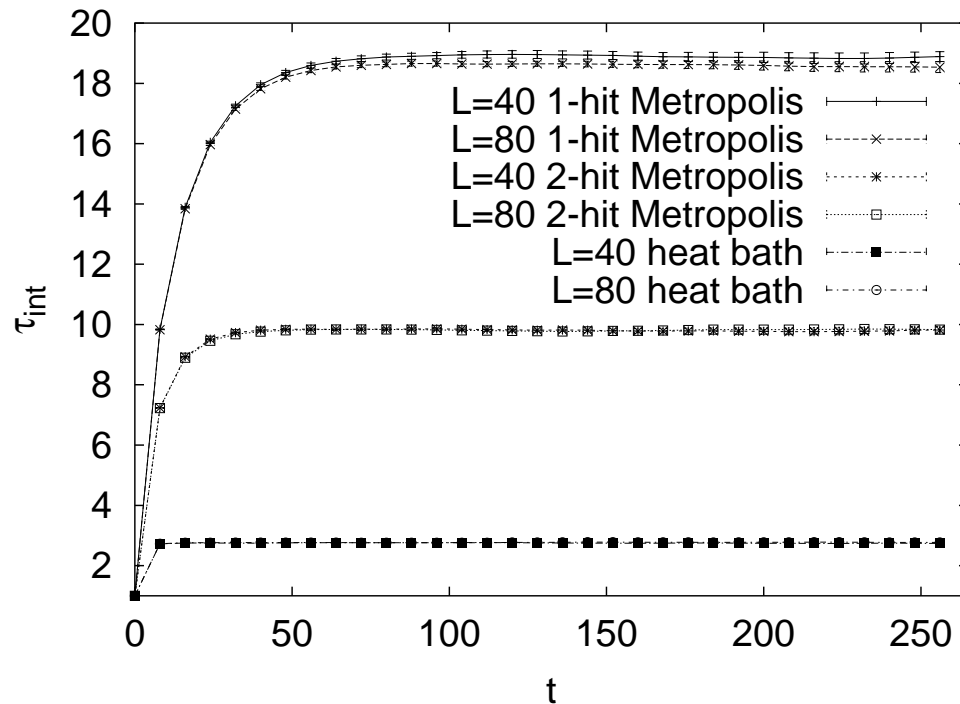


Figure 7: Systematic updating: Comparison of the integrated autocorrelation times of the 1-hit and 2-hit Metropolis algorithms and the heat bath algorithm for the 10-state Potts model on  $L \times L$  lattices at  $\beta = 0.62$  (assignment a0402\_06). The  $L = 40$  and  $L = 80$  curves lie almost on top of one another.