Contents of Lecture III

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 situations, 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. Assume, we count some out off equilibrium configurations as part of our production run. 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 the equilibrium configurations is that the factor in front of $1/N$ may be large, as we have seen.

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

Autocorrelations

We like to estimate the expectation value $\widehat{f}$ of some physical observable. We assume that the system has reached equilibrium. How many MC sweeps are needed to estimate $\widehat{f}$ 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

$$f_i = f_i(x_i), \quad i = 1, \ldots, N \tag{1}$$

from a Markov process, where x_i are the configurations generated. The label $i = 1, \ldots, N$ 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 $\widehat{f}$ is

$$\overline{f} = \frac{1}{N} \sum f_i \ . \tag{2}$$

With the notation

$$t = |i - j|$$

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

$$\widehat{C}(t) = \widehat{C}_{ij} = \langle \left(f_i - \langle f_i \rangle\right)\left(f_j - \langle f_j \rangle\right) \rangle = \langle f_i f_j \rangle - \langle f_i \rangle \langle f_j \rangle = \langle f_0 f_t \rangle - \widehat{f}^2 \quad (3)$$

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

$$\widehat{C}(t) \sim \exp\left(-\frac{t}{\tau_{\exp}}\right) \quad \text{for} \quad t \to \infty, \quad (4)$$

where $\tau_{\exp}$ is called **(exponential) autocorrelation time** and is related to the second largest eigenvalue λ_1 of the transition matrix by

$$\tau_{\exp} = -\ln \lambda_1 \quad (5)$$

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 (3)

$$\widehat{C}(0) = \sigma^2(f) \ . \tag{6}$$

Some algebra shows that the variance of the estimator $\overline{f}$ (2) for the mean and the autocorrelation functions (3) are related by

$$\sigma^2(\overline{f}) \ = \ \frac{\sigma^2(f)}{N} \left[1 + 2 \sum_{t=1}^{N-1} \left(1 - \frac{t}{N} \right) \widehat{c}(t) \right] \quad \text{with} \ \ \widehat{c}(t) = \frac{\widehat{C}(t)}{\widehat{C}(0)} \ . \tag{7}$$

This equation ought to be compared with the corresponding equation for uncorrelated random variables $\sigma^2(\overline{f}) = \sigma^2(f)/N$. The difference is the factor in the bracket of (7) which defines the **integrated autocorrelation time**

$$\tau_{\text{int}} \ = \ \left[1 + 2 \sum_{t=1}^{N-1} \left(1 - \frac{t}{N} \right) \widehat{c}(t) \right] \ . \tag{8}$$

For correlated data the variance of the mean is by the factor τ_{int} larger than the

corresponding **naive variance** for uncorrelated data:

$$\tau_{\mathrm{int}} \;=\; \frac{\sigma^2(\overline{f})}{\sigma^2_{\mathrm{naive}}(\overline{f})} \quad \text{with} \quad \sigma^2_{\mathrm{naive}} = \frac{\sigma^2(f)}{N} \;. \tag{9}$$

In most simulations one is interested in the limit $N \to \infty$ and equation (8) becomes

$$\tau_{\mathrm{int}} \;=\; 1 + 2\sum_{t=1}^{\infty} \widehat{c}(t) \;. \tag{10}$$

The numerical estimation of the integrated autocorrelation time faces difficulties. Namely, the variance of the $N \to \infty$ estimator of τ_{int} diverges:

$$\overline{\tau}_{\mathrm{int}} \;=\; 1 + 2\sum_{t=1}^{\infty} \overline{c}(t) \quad \text{and} \quad \sigma^2(\overline{\tau}_{\mathrm{int}}) \;\to\; \infty \tag{11}$$

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

estimator

$$\overline{\tau}_{\rm int}(t) \;=\; 1 + 2 \sum_{t'=1}^{t} \overline{c}(t') \tag{12}$$

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

$$\widehat{C}(t) \;=\; const \, \exp\left(-\frac{t}{\tau_{\rm exp}}\right) \;. \tag{13}$$

In this case we can carry out the sum (10) for the integrated autocorrelation function and find

$$\tau_{\rm int} = 1 + 2 \sum_{t=1}^{\infty} e^{-t/\tau_{\rm exp}} = 1 + \frac{2\, e^{-1/\tau_{\rm exp}}}{1 - e^{-1/\tau_{\rm exp}}} \;. \tag{14}$$

In particular, the difference between the asymptotic value (10) and the finite t

definition (12) becomes then

$$\tau_{\mathrm{int}} - \tau_{\mathrm{int}}(t) = 2\, e^{-t/\tau_{\mathrm{exp}}} \sum_{t'=1}^{\infty} e^{-t'/\tau_{\mathrm{exp}}} = \frac{2\, e^{-(t+1)/\tau_{\mathrm{exp}}}}{1 - e^{-1/\tau_{\mathrm{exp}}}} \; . \tag{15}$$

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

$$\tau_{\mathrm{int}} = 1 + \frac{2\, e^{-1/\tau_{\mathrm{exp}}}}{1 - e^{-1/\tau_{\mathrm{exp}}}} \cong 1 + \frac{2 - 2/\tau_{\mathrm{exp}}}{1/\tau_{\mathrm{exp}}} = 2\,\tau_{\mathrm{exp}} - 1 \;\cong\; 2\,\tau_{\mathrm{exp}} \tag{16}$$

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 (1) into $N_{bs} \leq N$ bins of

$$N_b = \texttt{NBIN} = \left[\frac{N}{N_{bs}}\right] = \left[\frac{\texttt{NDAT}}{\texttt{NBINS}}\right] \tag{17}$$

data each. Here [.] stands for Fortran integer division, *i.e.*, $N_b = \texttt{NBIN}$ 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

$$f_j^{N_b} = \frac{1}{N_b} \sum_{i=1+(j-1)N_b}^{jN_b} f_i \quad \text{for} \quad j = 1, \ldots, N_{bs} \ . \tag{18}$$

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

For a set of N_{bs} binned data $f_j^{N_b}$, $(j = 1, \ldots, N_{bs})$ we may calculate the mean with its **naive error bar**. Assuming for the moment an infinite time series, we find the integrated autocorrelation time (9) from the following ratio of sample variances

$$\tau_{\rm int} \;=\; \lim_{N_b \to \infty} \tau_{\rm int}^{N_b} \quad \text{with} \quad \tau_{\rm int}^{N_b} \;=\; \left(\frac{s^2_{\overline{f}^{N_b}}}{s^2_{\overline{f}}} \right) \quad . \tag{19}$$

In practice the $N_b \to \infty$ limit will be reached for a sufficiently large, finite value of N_b. The statistical error of the $\tau_{\rm int}$ estimate (19) is, in the first approximation, determined by the errors of $s^2_{\overline{f}^{N_b}}$. The typical situation is then that, due to the central limit theorem, the binned data are approximately Gaussian, so that the **error of $s^2_{\overline{f}^{N_b}}$ is analytically known** from the χ^2 distribution. Finally, the fluctuations of $s^2_{\overline{f}}$ of the denominator give rise to a small correction which can be worked out.

Numerically most accurate estimates of $\tau_{\rm int}$ are obtained for the finite binsize N_b which is just large enough that the binned data (18) are practically uncorrelated.

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

$$N = 2^K, \;\; K \geq 4 \;\; \text{and} \;\; N_b = 2^{K_b} \;\; \text{with} \;\; K_b = 0, 1, \ldots, K-5, K-4. \qquad (20)$$

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

$$N_{bs}^{\min} \;=\; 2^4 \;=\; 16 \; . \qquad (21)$$

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 $\tau_{\rm int}$, requires far more data.

Illustration: Metropolis generation of normally distributed data

We generate normally distributed data according to the Markov process

$$x' = x + 2\,a\,x^r - a \tag{22}$$

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

$$P_{\rm accept}(x') = \begin{cases} 1 \quad \text{for} \quad x'^{\,2} \leq x^2; \\ \exp[-(x'^{\,2} - x^2)/2] \ \text{for} \ x'^{\,2} > x^2. \end{cases} \tag{23}$$

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 (22), what gives an acceptance rate of approximately 50%.

The autocorrelation function:

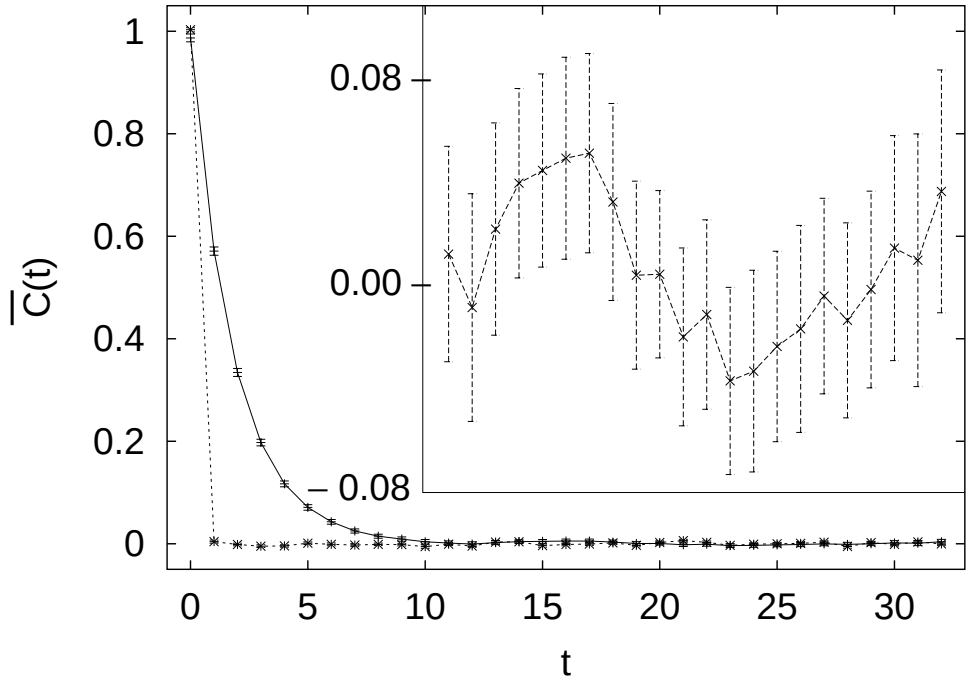

Figure 1: The autocorrelation function (3) of a Metropolis time series for the normal distribution (upper data) in comparison with those of our Gaussian random number generator (lower data). For $t \geq 11$ the inlay shows the autocorrelations on an enlarged ordinate (assignment a0401_01). The straight 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):

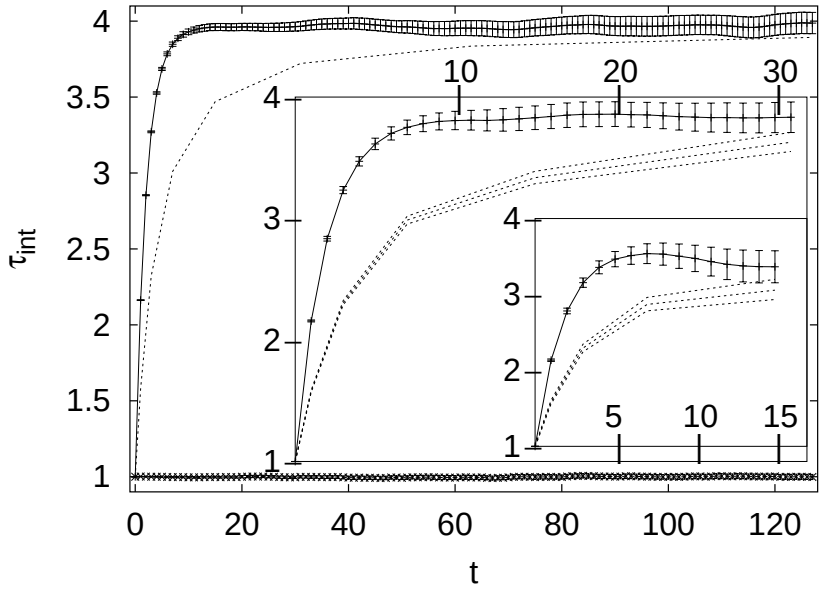

Figure 2: The upper curves in the figure and its inlays display the estimators obtained by direct calculation. The lowest curve is for the Gaussian random number generator. The remaining curves are binning procedure estimators of the integrated autocorrelation time with one standard deviation bounds. The main figure relies on 2^{21} data and depicts estimators up to $t = 127$. The first inlay relies on 2^{17} data and depicts estimators up to $t = 31$. The second inlay relies on 2^{14} data and depicts estimators up to $t = 15$.

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

$$t = N_b - 1 \ . \tag{24}$$

With this relation the estimators agree for binsize $N_b = 1$ 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 $\tau_{\rm int}$ value is slower than that of the direct estimate of $\tau_{\rm int}$.

For our large NDAT $= 2^{21}$ data set $\tau_{int}(t)$ 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:

$$\tau_{\rm int} = \tau_{\rm int}(20) = 3.962 \pm 0.024 \ \ \text{from} \ \ 2^{21} = 2,097,152 \ \text{data}. \tag{25}$$

The binning procedure, on the other hand, shows an increase of $\tau_{\rm int}^{N_b}$ all the way to $N_b = 2^7 = 128$, where the estimate with the one confidence level error bounds is

$$3.85 \leq \tau_{\rm int}^{128} \leq 3.94 \ \ \text{from} \ \ 2^{14} = 16,384 \ \text{bins} \ \ \text{from} \ \ 2^{21} \ \text{data}. \tag{26}$$

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

For the statistics of $\texttt{NDAT} = 2^{17}$ 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

$$\tau_{\rm int} = \tau_{\rm int}(20) = 3.86 \pm 0.11 \quad \text{from} \quad 2^{17} \text{ data.} \tag{27}$$

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

$$3.55 \leq \tau_{\rm int}^{32} \leq 3.71 \quad \text{from} \quad 2^{12} = 4,096 \text{ bins} \quad \text{from} \quad 2^{17} = 131,072 \text{ data.} \tag{28}$$

Our best value (25) is no longer covered by the two standard deviation zone.

For the second inlay the statistics is reduced to $\texttt{NDAT} = 2^{14}$. With the integrated autocorrelation time rounded to 4, this is 4096 times $\tau_{\rm int}$. For binsize $N_b = 2^4 = 16$ 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 (24), our $\tau_{\rm int}(t)$ plot to the range $t \leq 15$. Still, we find a quite nice window of nearly

constant $\tau_{\rm int}(t)$, namely all the way from $t = 4$ to $t = 15$. By a statistical fluctuation (assignment a0401_03) $\tau_{\rm int}(t)$ takes its maximum value at $t = 7$ and this makes $\tau_{\rm int}(7) = 3.54 \pm 0.13$ a natural candidate. However, this value is inconsistent with our best estimate (25). The true $\tau_{\rm int}(t)$ 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

$$\tau_{\rm int} = 3.54 \pm 0.21 \quad \text{for} \quad 2^{14} = 16,384 \ \text{data}, \tag{29}$$

which achieves consistency with (25) in the two error bar range. For binsize $N_b = 16$ the binning estimate is

$$2.93 \leq \tau_{\rm int}^{16} \leq 3.20 \quad \text{from} \quad 2^{10} = 1,024 \ \text{bins} \quad \text{from} \quad 2^{14} \ \text{data}. \tag{30}$$

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

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 still choose

$$\texttt{nequi} \gg \tau_{\rm int} \ , \tag{31}$$

to allow the system to settle. That is a first reason, why it appears necessary to control the integrated autocorrelation time of a MC simulations. A second reason is that we have to control the error bars of the equilibrium part of our simulation. Ideally the error bars are calculated as

$$\triangle \overline{f} \ = \ \sqrt{\sigma^2(\overline{f})} \quad \text{with} \quad \sigma^2(\overline{f}) \ = \ \tau_{\rm int} \, \frac{\sigma^2(f)}{N} \ . \tag{32}$$

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

However, the calculation of the integrated autocorrelation time may be out of reach. Many more than the about twenty independent data are needed, which

according to the Student distribution are sufficient to estimate mean values with reasonably reliable error bars.

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

$$\texttt{NBINS} \geq 16 \tag{33}$$

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 $t_{\max} > \texttt{NBINS}\,\tau_{\rm int}$ for the largest t of our time series, our small number of binned data become effectively independent and our error analysis is justified.

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 $\tau_{\rm int}$, 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?

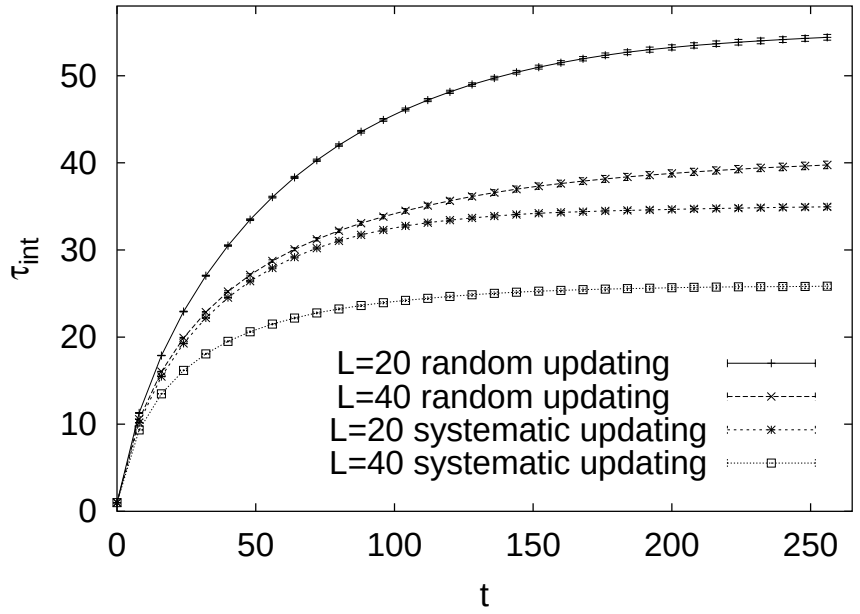

Figure 3: Comparison of the integrated autocorrelation time of the Metropolis process with random updating versus sequential updating for the $d = 2$ Ising model at $\beta = 0.4$ (assignment a0402_01 B) The ordering of the curves is identical with the ordering of the labels in the figure.

The $d = 2$ Ising model off and on the critical point

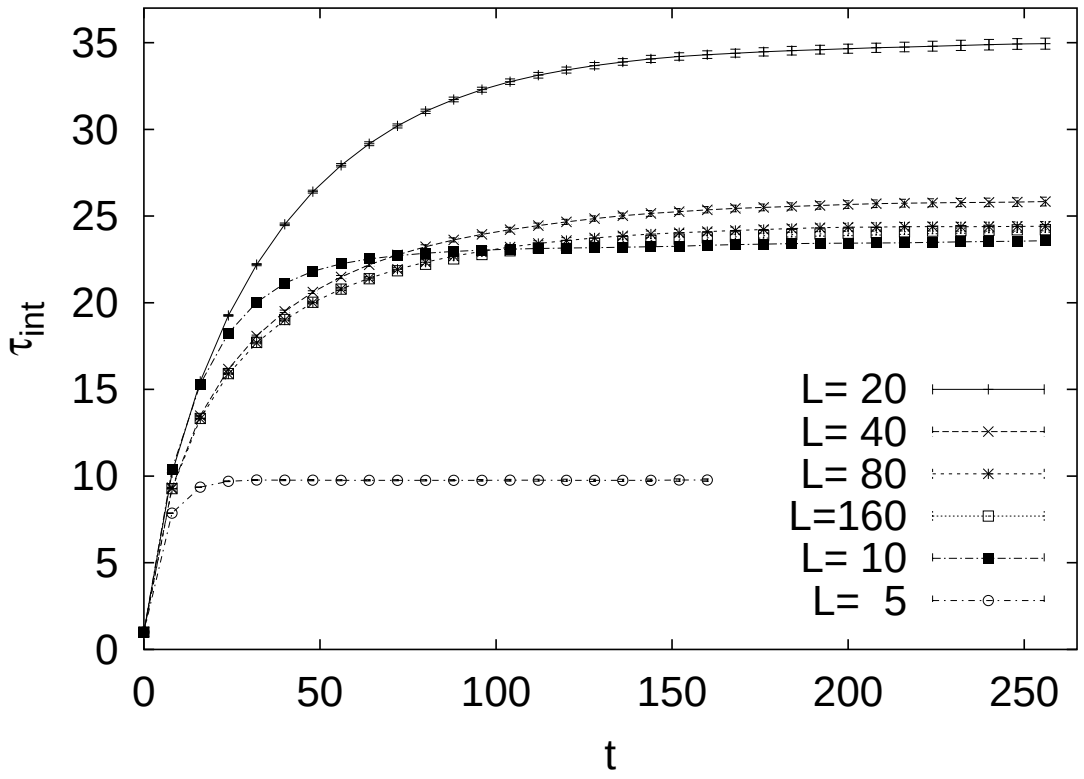

Figure 4: One-hit Metropolis algorithm with sequential updating: Lattice size dependence of the integrated autocorrelation time for the $d = 2$ Ising model at $\beta = 0.4$ (assignment a0402_01 A). The ordering of the curves is identical with the ordering of the labels in the figure.

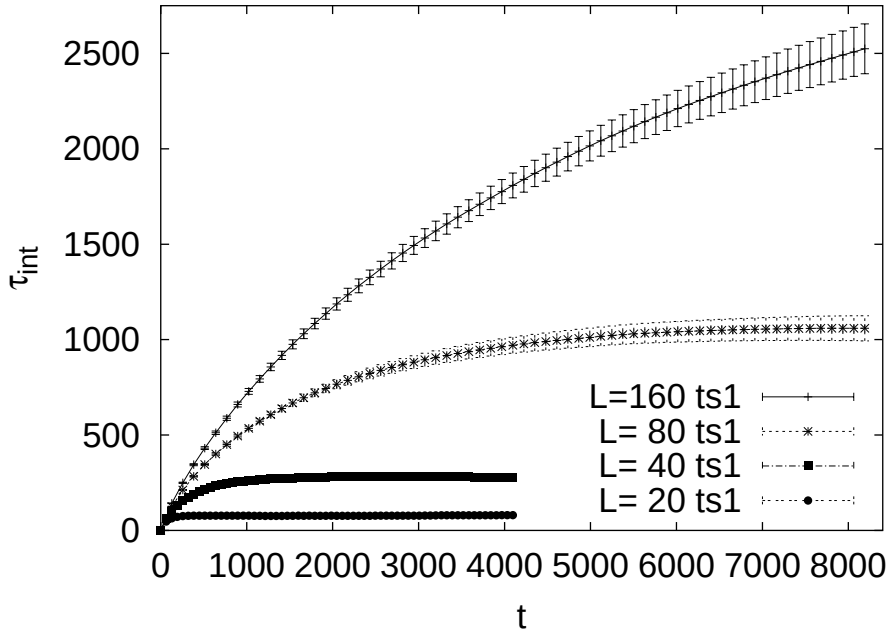

Figure 5: One-hit Metropolis algorithm with sequential updating: Lattice size dependence of the integrated autocorrelation time for the $d = 2$ Ising model at its critical temperature (assignment a0402_02 D). The ordering of the curves is identical with the ordering of the labels in the figure.

The increase of the autocorrelation at the critical point is called **critical slowing down**, $\tau_{\mathrm{int}} \sim L^z$ with $z \approx 2.17$ is the **dynamical critical exponent**. Estimates of z are compiled in the book by Landau and Binder.

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

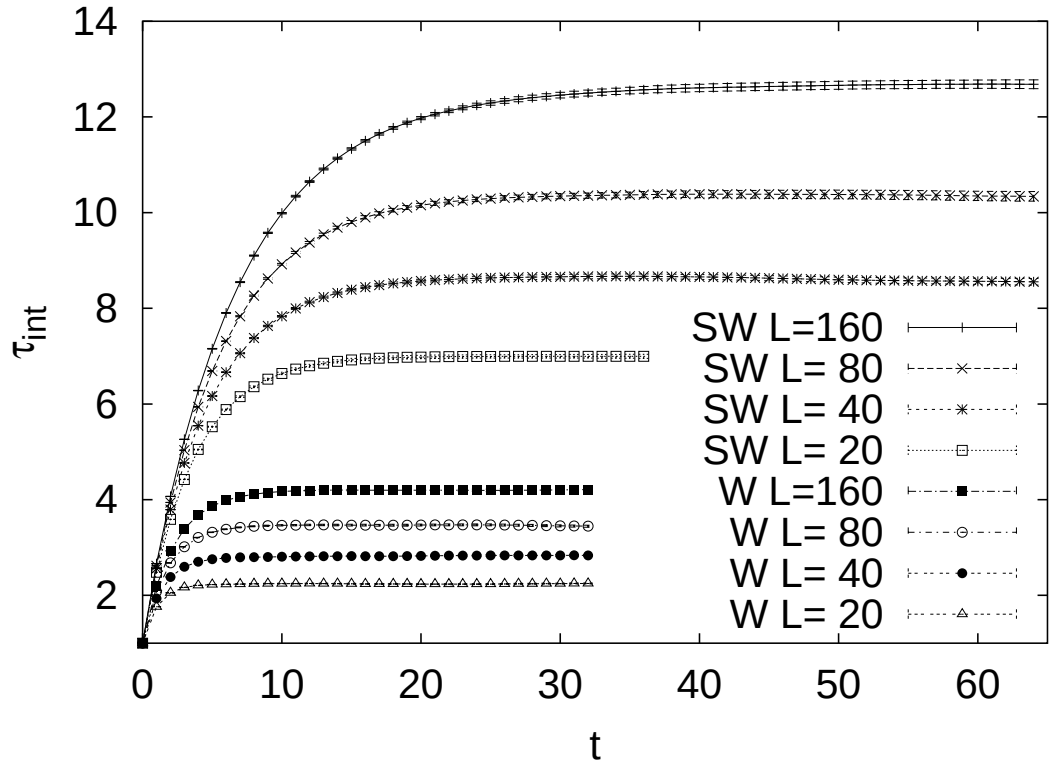

Figure 6: Estimates of integrated autocorrelation times from simulations of the $d = 2$ Ising model at the critical temperature $\beta_c = 0.44068679351$ (assignment a0503_05).

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

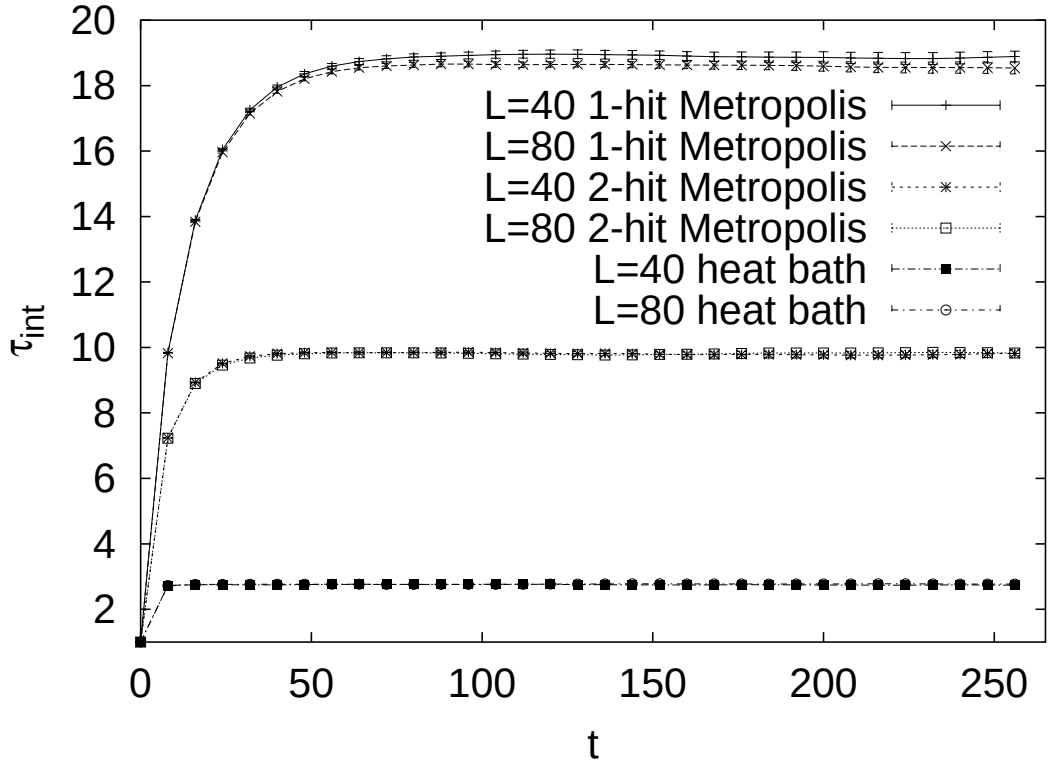

Figure 7: Systematic updating: Comparison of the integrated autocorrelation times of the 1-hit and 2-hit Metropolis algorithms and the heat bath algorithm for the 10-state Potts model on $L \times L$ lattices at $\beta = 0.62$ (assignment a0402_06). The $L = 40$ and $L = 80$ curves lie almost on top of one another.