Monte Carlo Methods in Classical Statistical Physics

Wolfhard Janke

Universität Leipzig

Contents

Heraeus Summerschool

Computational Many Particle Physics

Greifswald, 18 – 29 September 2006

Motivation

Objectives:

Understanding of phase transitions and critical phenomena in classical and quantum statistical physics.

Tools:

Numerical computer simulations.

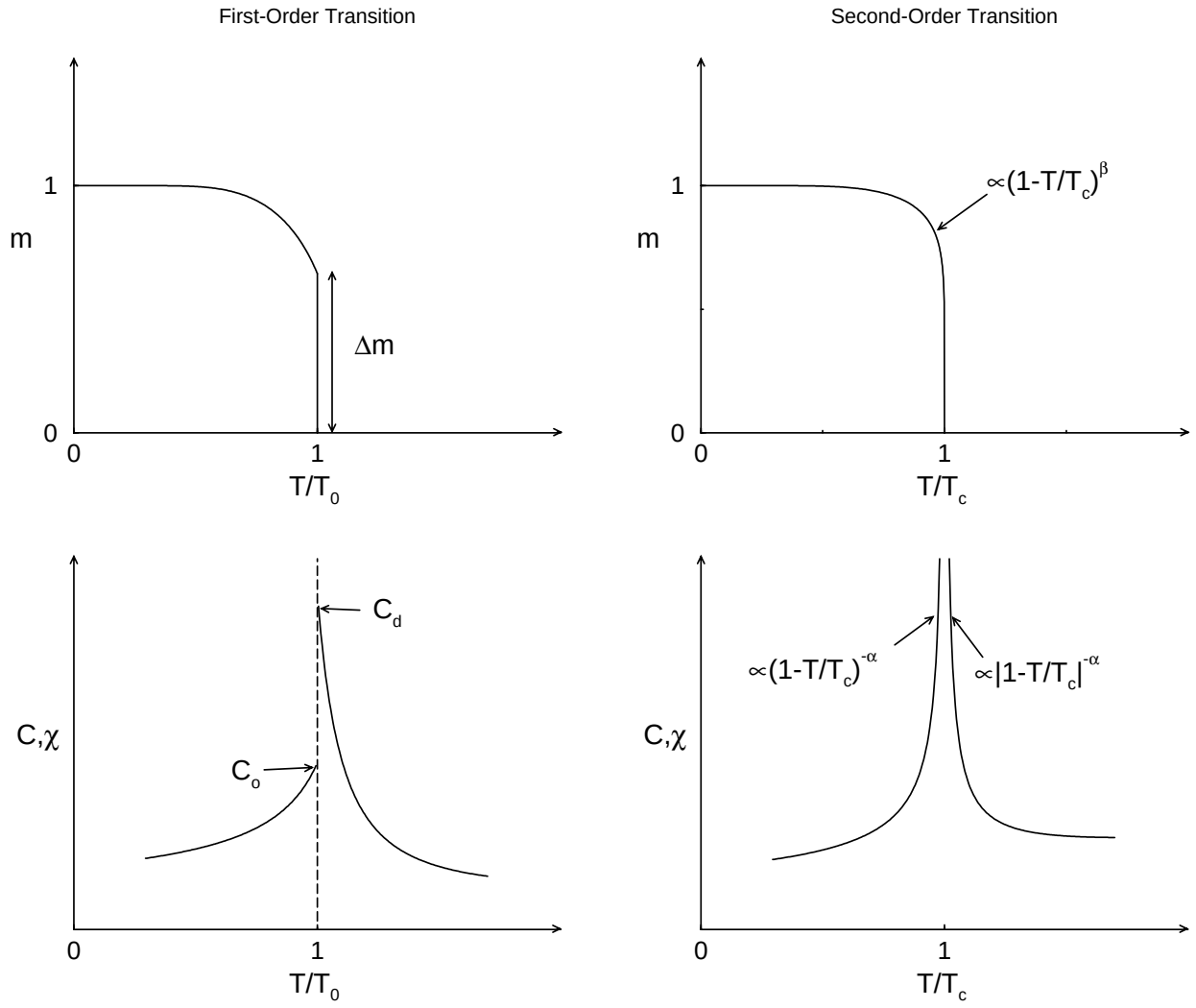
Two major methods:

- Molecular Dynamics (MD)
- Monte Carlo simulations (MC)

Here focus on MC:

- Stochastic method
- Importance sampling
- Temporal (auto-)correlations
- Statistical error estimation of correlated data
- Improved algorithms and analysis tools

Models and Phase Transitions



$$
\begin{aligned}
\xi &\propto t^{-\nu} + \ldots \\
m &\propto t^{+\beta} + \ldots \\
\chi &\propto t^{-\gamma} + \ldots \\
C &= C_{\rm reg} + a t^{-\alpha} + \ldots \\
\\
t &= |1 - T/T_c|
\end{aligned}
$$

Ising Model

$$Z_I(\beta) = \sum_{\{\sigma_i\}} \exp(-\beta H_I) \ , \quad H_I = -\sum_{\langle ij \rangle} \sigma_i \sigma_j \ , \quad \sigma_i = \pm 1$$

Internal energy per site:

$$e = E/V, \quad E = -d \ln Z_I / d\beta \equiv \langle H_I \rangle$$

Specific heat:

$$C/k_B = \frac{\mathrm{d}e}{\mathrm{d}(k_B T)} = \beta^2 \left(\langle H_I^2 \rangle - \langle H_I \rangle^2 \right) / V$$

Magnetization:

$$m = M/V = \langle |\mu| \rangle \ , \qquad \mu = \sum_i \sigma_i / V$$

Susceptibility:

$$\begin{aligned} \chi &= \beta V \left(\langle \mu^2 \rangle - \langle |\mu| \rangle^2 \right) \\ \chi' &= \beta V \langle \mu^2 \rangle \qquad (T \geq T_c) \end{aligned}$$

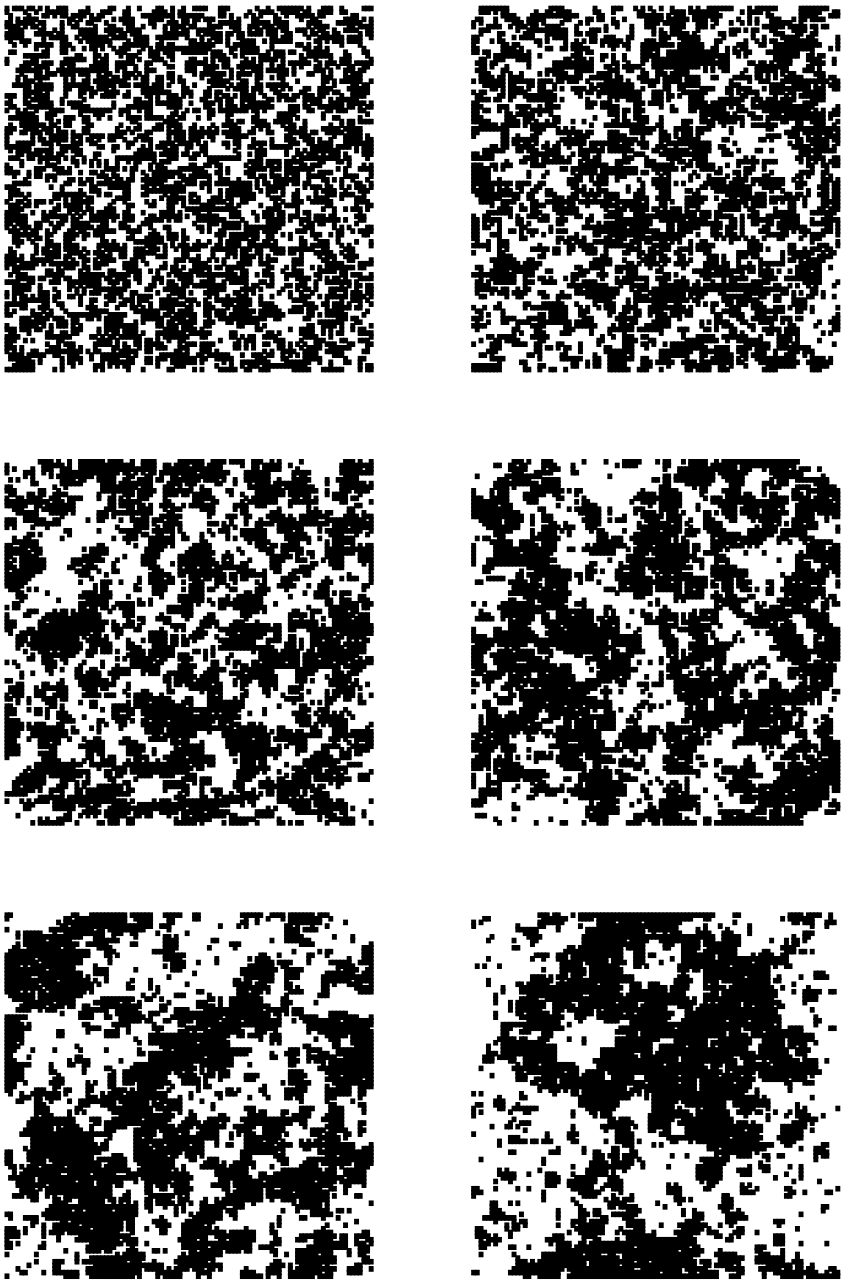
Correlation function ($\vec{x} = \vec{x}_i - \vec{x}_j$):

$$G(\vec{x}_i - \vec{x}_j) = \langle \sigma_i \sigma_j \rangle \sim \exp(-|\vec{x}|/\xi) \qquad \text{for large } |\vec{x}|$$

Spatial correlation length ξ:

$$\xi = -\lim_{|\vec{x}| \to \infty} \left(|\vec{x}| / \ln G(\vec{x}) \right)$$

Spatial Correlations



Ilustration of the growth of spatial correlations when criticality is approached (100×100 Ising model, $\beta/\beta_c = 0.50$, 0.70, 0.85, 0.90, 0.95, and 0.98).

Finite-Size Rounding and Shifts

For systems of finite size L^d, as in any numerical simulation, the correlation length cannot diverge, and the divergencies in all other quantities are rounded and shifted.

Write

$$t \equiv |1 - T/T_c| \propto \xi^{-1/\nu} \longrightarrow L^{-1/\nu} \ ,$$

and use this replacement in the temperature scaling laws to obtain the finite-size scaling (FSS) ansätze

$$\begin{aligned} m &\propto L^{-\beta/\nu} + \ldots \ , \\ \chi &\propto L^{+\gamma/\nu} + \ldots \ , \\ C &= C_{\rm reg} + aL^{+\alpha/\nu} + \ldots \ . \end{aligned}$$

Locations $T_{\rm max}$ of the (finite) maxima of thermodynamic quantities scale with the system size as

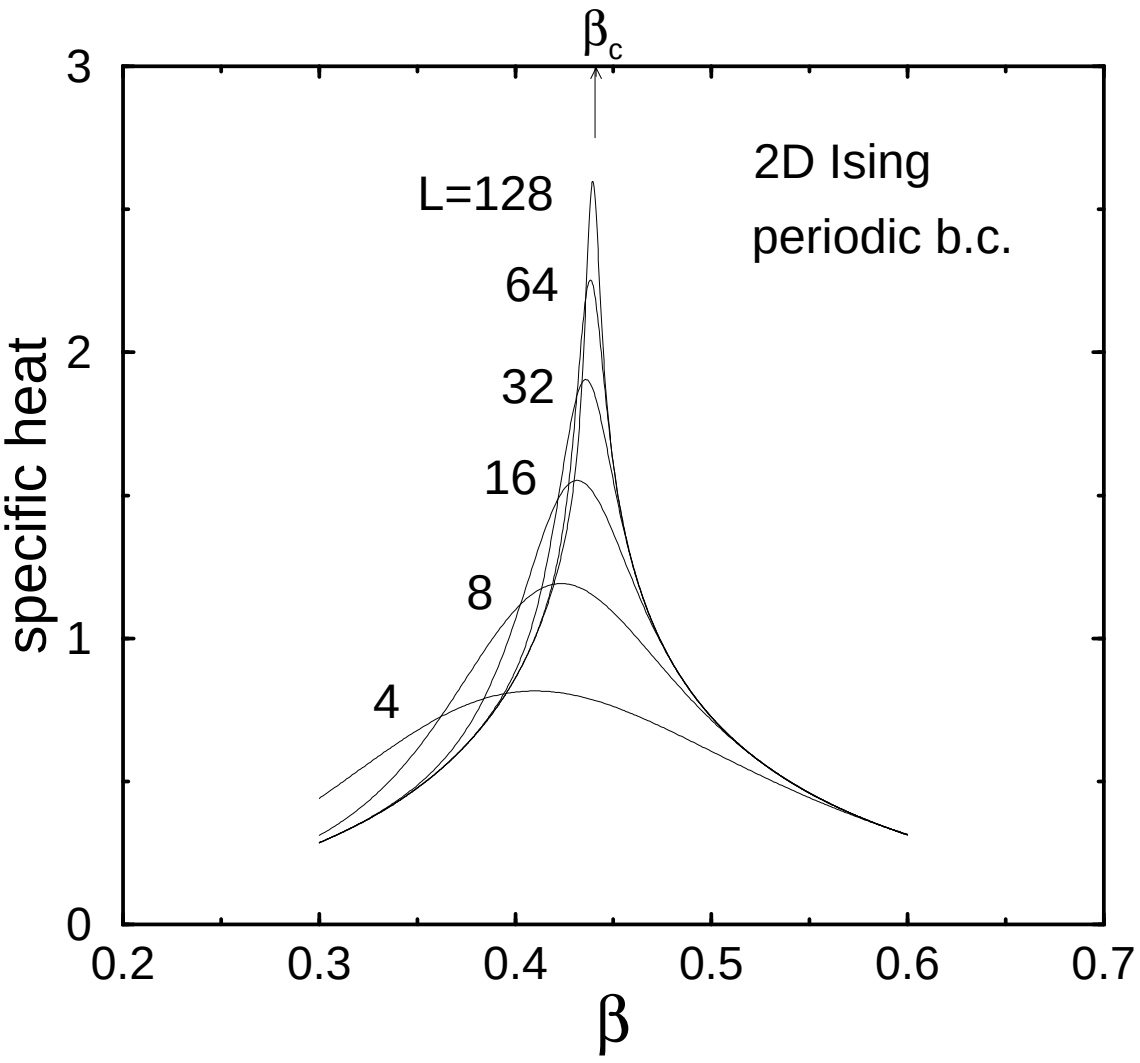
$$T_{\rm max} = T_c + cL^{-1/\nu} + \ldots \ .$$

Finite-scaling functions, e.g., for the susceptibility:

$$\chi(T,L) = L^{\gamma/\nu} f_\chi(x) + \ldots \ ,$$

where $x = (1 - T/T_c)L^{1/\nu}$ is the scaling variable (i.e, $L^{-\gamma/\nu}\chi(T,L)$ does depend only on the single variable x).

Example: 2D Ising Model



$$C = C_{\rm reg} + aL^{\alpha/\nu} + \ldots$$

$$\beta_{\rm max} = \beta_c + cL^{-1/\nu} + \ldots$$

Importance Sampling Monte Carlo Simulations

Goal: Draw configurations according to their Boltzmann weight:

$$P^{\rm eq}[\{\sigma_i\}] \propto \exp\left(-\beta H[\{\sigma_i\}]\right) \quad .$$

In mathematical terms one sets up a Markov chain:

$$\ldots \xrightarrow{W} \{\sigma_i\} \xrightarrow{W} \{\sigma_i'\} \xrightarrow{W} \{\sigma_i''\} \xrightarrow{W} \ldots \quad ,$$

with a transition operator W satisfying the conditions

(a) $$W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) > 0 \qquad \forall \{\sigma_i\}, \{\sigma_i'\} \ ,$$

(b) $$\sum_{\{\sigma_i'\}} W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = 1 \qquad \forall \{\sigma_i\} \ ,$$

(c) $$\sum_{\{\sigma_i\}} W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) P^{\rm eq}[\{\sigma_i\}] = P^{\rm eq}[\{\sigma_i'\}] \quad \forall \{\sigma_i'\} \ .$$

From (c): $P^{\rm eq}$ is a fixed point of W.
A somewhat simpler sufficient condition is detailed balance:

$$P^{\rm eq}[\{\sigma_i\}] W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = P^{\rm eq}[\{\sigma_i'\}] W(\{\sigma_i'\} \longrightarrow \{\sigma_i\}) \ .$$

After an initial equilibration time, expectation values can be estimated as an arithmetic mean over the Markov chain, e.g.,

$$\langle H \rangle = \sum_{\{\sigma_i\}} H[\{\sigma_i\}] P^{\rm eq}[\{\sigma_i\}] \approx \frac{1}{N} \sum_{j=1}^{N} H[\{\sigma_i\}]_j \ .$$

Metropolis Algorithm

Propose locally a flip of a single spin and accept this update with probability:

$$W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = \begin{cases} \exp\left[-\beta(E'-E)\right] & E' > E \\ 1 & E' \leq E \end{cases},$$

with energies E before and E' after the spin flip.

Note: The practical implementation needs pseudo-random numbers.

If the energy is lowered, the spin flip is always accepted.
If the energy is increased, accept flip with a certain probability $\longrightarrow$ entropic contributions.

It is straightforward to show that the Markov conditions (a) – (c) are satisfied by this update rule.

Selecting spins: At random, random permutation, sequentially, checker-board scheme.

Continuous degrees of freedom: Choose new spin direction at random (but be careful with the proper integration measure!).

This update algorithm is not the most efficient one, but works in virtually all cases!

So, we are basically done, just wait – every 5 years, the computer speed improves by about a factor of 10 (Moore's law; confirmed since about 1950).

Or, if you are impatient, use better update algorithms ...

Heat-Bath Algorithm

Test all spin states of a single spin s_{i_0} in the heat-bath of its (fixed) local neighbors (4 on a square lattice, 6 on a simple-cubic lattice with nearest-neighbour interactions):

$$W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = \frac{e^{-\beta H(\{\sigma_i'\}}}{\sum_{s_{i_0}'} e^{-\beta H(\{\sigma_i'\}}}$$

Detailed balance obviously satisfied:

$$e^{-\beta H(\{\sigma_i\}} \frac{e^{-\beta H(\{\sigma_i'\}}}{\sum_{s_{i_0}'} e^{-\beta H(\{\sigma_i'\}}} = e^{-\beta H(\{\sigma_i'\}} \frac{e^{-\beta H(\{\sigma_i\}}}{\sum_{s_{i_0}} e^{-\beta H(\{\sigma_i\}}}$$

Selecting spins as for Metropolis alg. (random, sequentially, . . .).

But: Only in special cases easy to generalize to continuous degrees of freedom.

Glauber Algorithm

Conceptually similar to Metropolis algorithm:

Propose locally a flip of a single spin and accept this update with probability:

$$W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = \frac{1}{2}\left[1 - \tanh\left(\beta(E' - E)/2\right)\right] \quad ,$$

with energies E before and E' after the spin flip.

Only for this update algorithm the Monte Carlo (pseudo-) dynamics can be calculated analytically – and even only in one dimension.

For two and higher dimensions no exact solutions are known.

Estimators and Autocorrelation Times

Given a time series of N measurements of a quantity $\mathcal{O}$. As estimator for the expectation value $\langle\mathcal{O}\rangle$ one may take the arithmetic mean value

$$\overline{\mathcal{O}} = \frac{1}{N}\sum_{j=1}^{N}\mathcal{O}_j \ .$$

Expectation value $\langle\mathcal{O}\rangle$: Ordinary number

Mean value $\overline{\mathcal{O}}$: fluctuating random number with variance

$$\sigma^2_{\overline{\mathcal{O}}} = \langle[\overline{\mathcal{O}} - \langle\overline{\mathcal{O}}\rangle]^2\rangle = \langle\overline{\mathcal{O}}^2\rangle - \langle\overline{\mathcal{O}}\rangle^2 \ .$$

For uncorrelated measurements $\mathcal{O}_j$:

$$\sigma^2_{\overline{\mathcal{O}}} = \sigma^2_{\mathcal{O}_j}/N$$

$$\sigma^2_{\mathcal{O}_j} = \langle\mathcal{O}_j^2\rangle - \langle\mathcal{O}_j\rangle^2$$

$\sigma^2_{\mathcal{O}_j}$: Variance of individual measurements; for e.g. $\mathcal{O}$ = energy E, $\sigma^2_{\mathcal{O}_j} \propto$ specific heat C.

Correlated measurements $\mathcal{O}_j$:

$$\sigma^2_{\overline{\mathcal{O}}} = \langle \overline{\mathcal{O}}^2 \rangle - \langle \overline{\mathcal{O}} \rangle^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \mathcal{O}_i \mathcal{O}_j \rangle - \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle \ .$$

Collecting diagonal and off-diagonal terms:

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{1}{N^2} \sum_{i=1}^{N} \left(\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle^2 \right) + \frac{1}{N^2} \sum_{i \neq j}^{N} \left(\langle \mathcal{O}_i \mathcal{O}_j \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle \right) \ .$$

Symmetry $i \leftrightarrow j \Rightarrow \sum_{i \neq j}^{N} = 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N}$, time translation invariance:

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{1}{N} \left[\sigma^2_{\mathcal{O}_i} + 2 \sum_{k=1}^{N} \left(\langle \mathcal{O}_1 \mathcal{O}_{1+k} \rangle - \langle \mathcal{O}_1 \rangle \langle \mathcal{O}_{1+k} \rangle \right) \left(1 - \frac{k}{N} \right) \right] \ .$$

Factoring out $\sigma^2_{\mathcal{O}_i}$:

$$\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}} = \frac{\sigma^2_{\mathcal{O}_i}}{N} 2\tau'_{\mathcal{O},\mathrm{int}} \ ,$$

with the (proper) integrated autocorrelation time

$$\tau'_{\mathcal{O},\mathrm{int}} = \frac{1}{2} + \sum_{k=1}^{N} A(k) \left(1 - \frac{k}{N} \right) \ ,$$

$$A(k) = \frac{\langle \mathcal{O}_i \mathcal{O}_{i+k} \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle}{\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle} \ .$$

Asymptotic behavior (governed by exponential autocorrelation time):

$$A(k) \stackrel{k\to\infty}{\longrightarrow} a e^{-k/\tau_{\mathcal{O},\mathrm{exp}}} \ ,$$

Hence $\left(1-\frac{k}{N}\right)$ can be neglected:

$$\tau_{\mathcal{O},\mathrm{int}} = \frac{1}{2} + \sum_{k=1}^{N} A(k) \ .$$

Generically $\tau_{\mathcal{O},\mathrm{int}} \neq \tau_{\mathcal{O},\mathrm{exp}}$.

Effective statistics:

$$\sigma^2_{\overline{\mathcal{O}}} = \sigma^2_{\mathcal{O}_j}/N_{\mathrm{eff}} \ ,$$

$$N_{\mathrm{eff}} = N/2\tau_{\mathcal{O},\mathrm{int}} \leq N \ .$$

Scaling behavior of autocorrelation time at criticality:

$$\tau_{\mathcal{O},\mathrm{int}} \propto \xi^z \propto t^{-\nu z} \ .$$

In finite systems:

$$\tau_{\mathcal{O},\mathrm{int}} \propto L^z \ .$$

z: Dynamical critical exponent ($z \approx 2$ for local update algorithms)

Bias

Physical example: Specific heat $C = \beta^2 V \left(\langle e^2 \rangle - \langle e \rangle^2 \right) = \beta^2 V \sigma_{e_i}^2$.

Estimator for variance:

$$\hat{\sigma}_{e_i}^2 = \overline{e^2} - \overline{e}^2 = \overline{(e - \overline{e})^2} = \frac{1}{N} \sum_{i=1}^{N} (e_i - \overline{e})^2 \quad .$$

What is the *expected* value of $\hat{\sigma}_{e_i}^2$?

$$\langle \hat{\sigma}_{e_i}^2 \rangle = \langle \overline{e^2} - \overline{e}^2 \rangle = \langle \overline{e^2} \rangle - \langle \overline{e} \rangle^2 - \left(\langle \overline{e}^2 \rangle - \langle \overline{e} \rangle^2 \right) \quad ,$$

This gives

$$\langle \hat{\sigma}_{e_i}^2 \rangle = \sigma_{e_i}^2 \left(1 - \frac{2\tau_{e,\mathrm{int}}}{N} \right) = \sigma_{e_i}^2 \left(1 - \frac{1}{N_{\mathrm{eff}}} \right) \neq \sigma_{e_i}^2 \quad ,$$

showing the weak bias of oder $1/N_{\mathrm{eff}} = 2\tau_{\mathcal{O},\mathrm{int}}/N$.

Uncorrelated measurements: Bias $= 1/N$ can be corrected by using

$$\hat{\sigma}_{e_i,\mathrm{corr}}^2 = \frac{N}{N-1} \hat{\sigma}_{e_i}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (e_i - \overline{e})^2 \quad ,$$

$$\langle \hat{\sigma}_{e_i,\mathrm{corr}}^2 \rangle = \sigma_{e_i}^2 \quad .$$

Binning and Jackknife Analysis

N_B non-overlapping blocks of length k , $N = N_B k$. Block average

$$\mathcal{O}_{B,n} \equiv \frac{1}{k} \sum_{i=1}^{k} \mathcal{O}_{(n-1)k+i} \ , \qquad n = 1, \ldots, N_B \ .$$

Error estimate on the mean value:

$$\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}} = \sigma^2_B / N_B = \frac{1}{N_B(N_B-1)} \sum_{n=1}^{N_B} (\mathcal{O}_{B,n} - \overline{\mathcal{O}}_B)^2 \ .$$

Since $\sigma^2_B / N_B = 2\tau_{\mathcal{O},\mathrm{int}} \sigma^2_{\mathcal{O}_i} / N$:

$$2\tau_{\mathcal{O},\mathrm{int}} = k\sigma^2_B / \sigma^2_{\mathcal{O}_i}$$

Better: N_B large Jackknife blocks $\mathcal{O}_{J,n}$ containing all data but one of the binning blocks:

$$\mathcal{O}_{J,n} = \frac{N\overline{\mathcal{O}} - k\mathcal{O}_{B,n}}{N-k} \ , \qquad n = 1, \ldots, N_B \ .$$

Re-using of data leads to an extra factor $(N_B - 1)^2$:

$$\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}} = \frac{N_B - 1}{N_B} \sum_{n=1}^{N_B} (\mathcal{O}_{J,n} - \overline{\mathcal{O}}_J)^2 \ .$$

A Simplified Model

Bivariate Gaussian variables ($0 \le \rho < 1$):

$$\begin{aligned} e_0 &= e_0' \ , \\ e_i &= \rho e_{i-1} + \sqrt{1-\rho^2}\, e_i' \ , \qquad i \ge 1 \ , \end{aligned}$$

e_i': *independent* Gaussian random variables satisfying $\langle e_i' \rangle = 0$ and $\langle e_i' e_j' \rangle = \delta_{ij}$. Thus

$$e_k = \rho e_{k-1} + \sqrt{1-\rho^2}\, e_k' = \rho^k e_0 + \sqrt{1-\rho^2} \sum_{l=1}^{k} \rho^{k-l} e_l' \ ,$$

and consequently

$$A(k) = \langle e_0 e_k \rangle = \rho^k \equiv e^{-k/\tau_{\rm exp}} \ ,$$

$$\tau_{\rm exp} = -1/\ln \rho \ ,$$

$$\tau_{\rm int} = \frac{1}{2} + \sum_{k=1}^{\infty} A(k) = \frac{1}{2}\frac{1+\rho}{1-\rho} = \frac{1}{2}{\rm cth}(1/2\tau_{\rm exp}) \ .$$

Practical implementation:

$$\begin{aligned}\tau_{\rm int}(k_{\rm max}) &\equiv \frac{1}{2}+\sum_{k=1}^{k_{\rm max}} A(k) \\ &= \frac{1}{2}{\rm cth}(1/2\tau_{\rm exp})\left[1-\frac{2e^{-(k_{\rm max}+1)/\tau_{\rm exp}}}{1+e^{-1/\tau_{\rm exp}}}\right] .\end{aligned}$$

For $\tau_{\rm exp} \gg 1$:

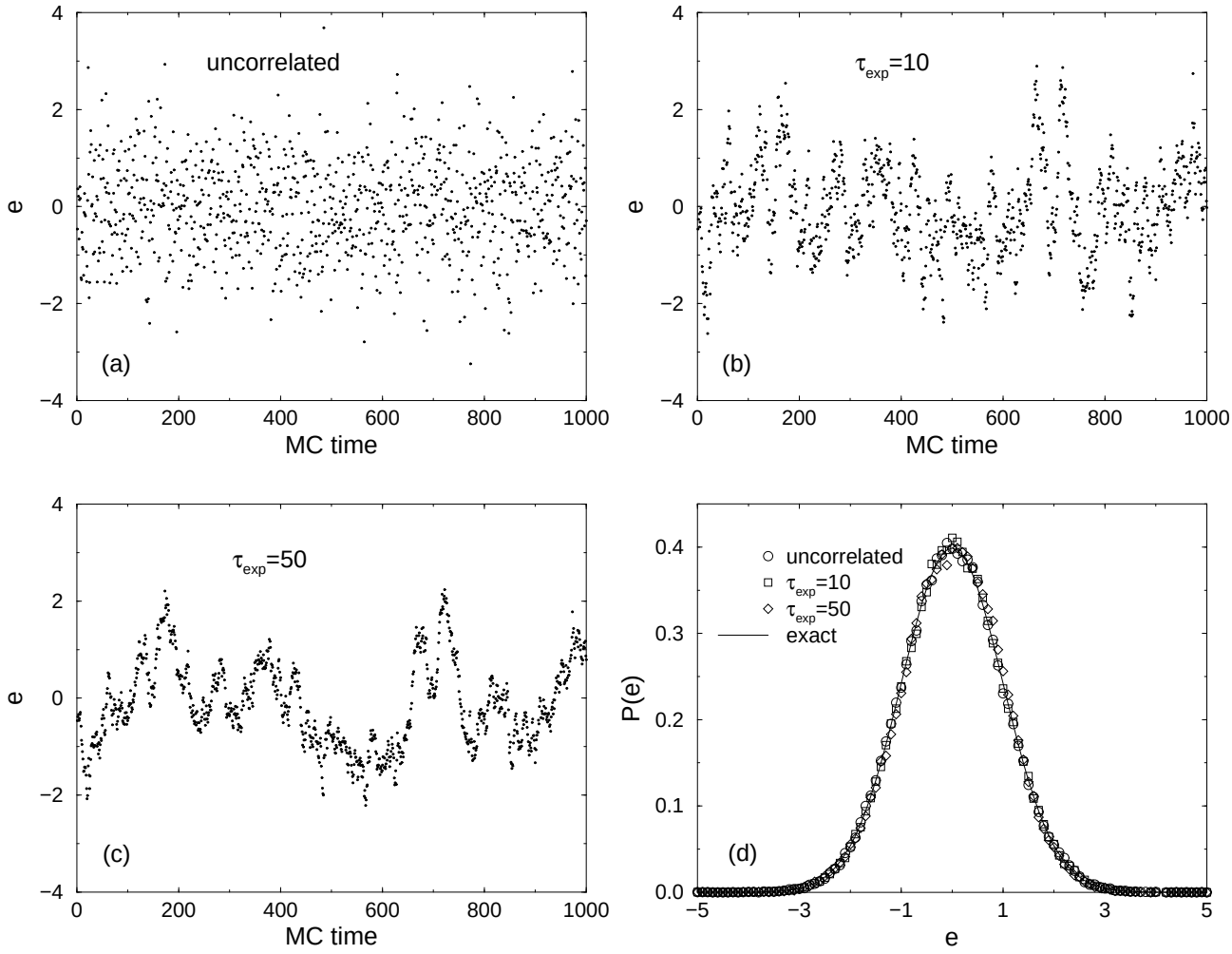
$$\tau_{\rm int}(k_{\rm max}) = \tau_{\rm int}\left[1-\frac{2\tau_{\rm exp}}{2\tau_{\rm exp}+1}e^{-k_{\rm max}/\tau_{\rm exp}}\right] .$$

Often employed: Self-consistent cutoff at $k_{\rm max} = 6\tau_{\rm int}$.

Binning analysis:

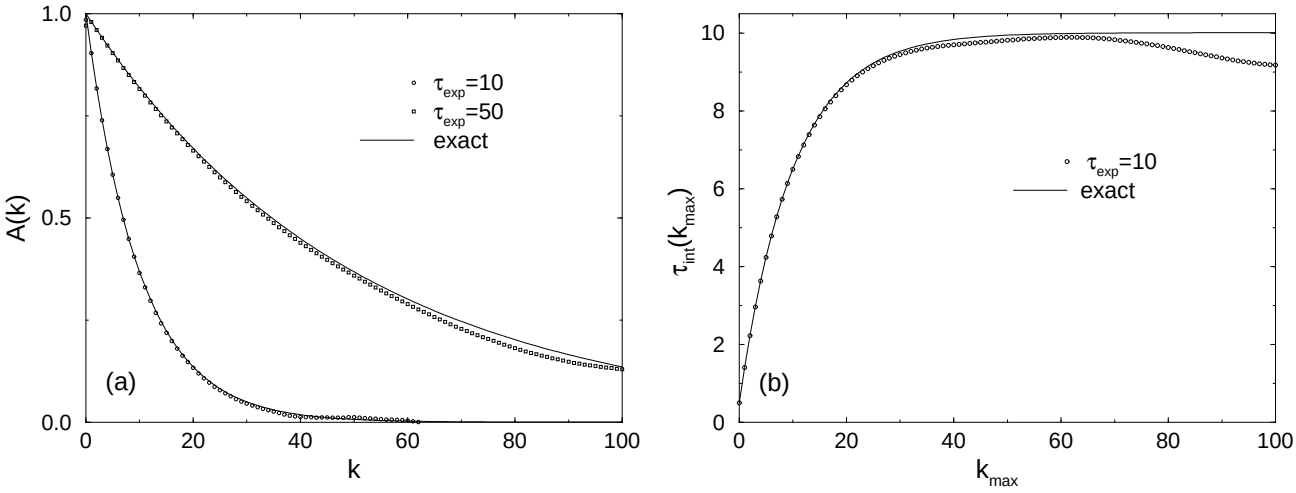
$$\begin{aligned}\sigma_B^2 = \langle e_{B,n}^2\rangle &= \frac{1}{k^2}\sum_{i,j=1}^{k}\rho^{|i-j|} \\ &= \frac{1}{k^2}\left[k+2\sum_{i=1}^{k}\sum_{j=1}^{i-1}\rho^{i-j}\right] \\ &= \frac{1}{k}\left[1+\frac{2\rho}{1-\rho}-\frac{2\rho}{k}\frac{1-\rho^k}{(1-\rho)^2}\right] \\ &\approx 2\tau_{\rm exp}\left[1-\frac{\tau_{\rm exp}}{k}\left(1-e^{-k/\tau_{\rm exp}}\right)\right] .\end{aligned}$$

Time Evolutions

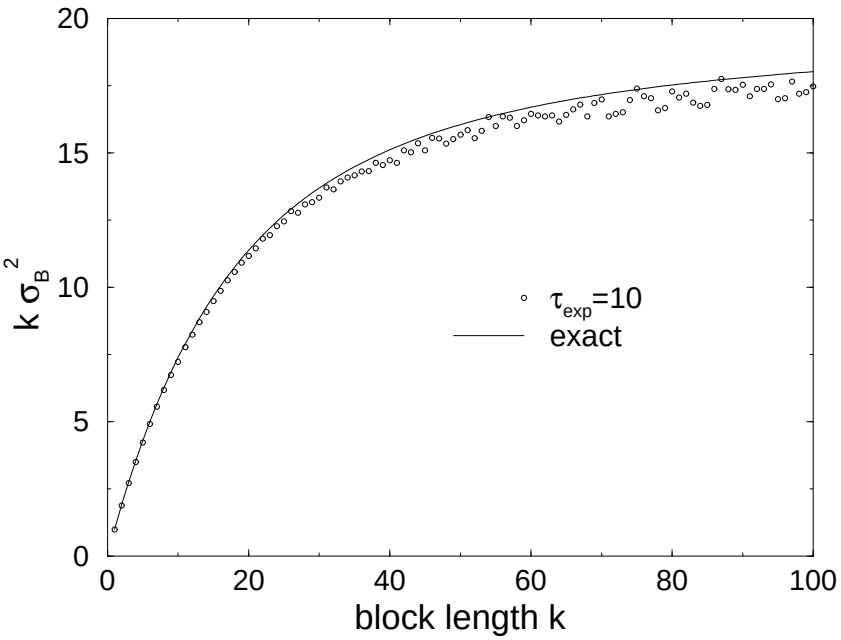


One percent of the "MC time" evolution of the bivariate Gaussian process with a total of 100 000 "measurements".

Autocorrelation and Binning Analysis



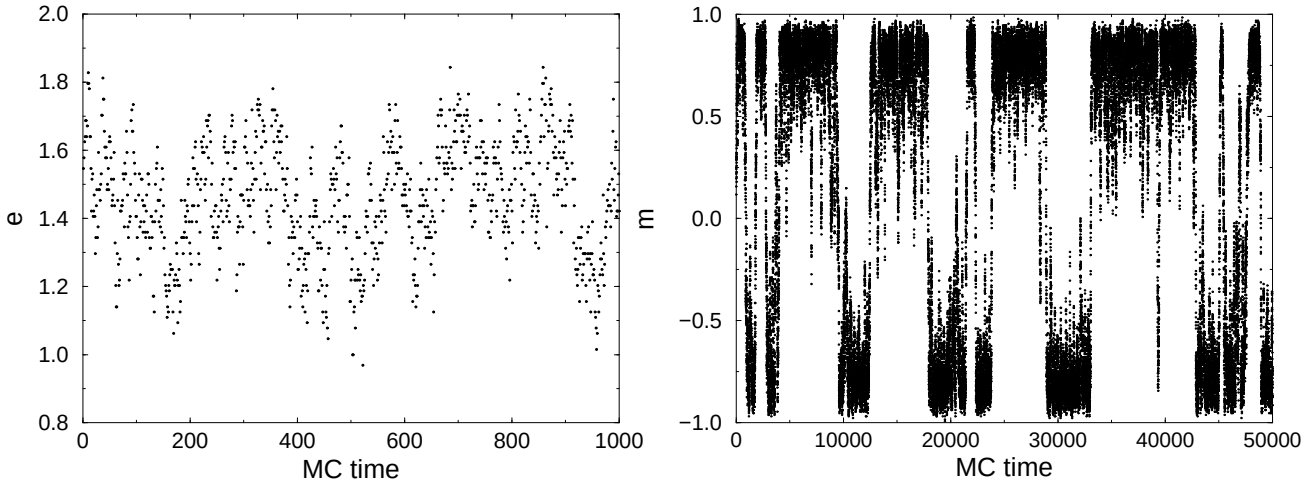
(a) Autocorrelation functions and (b) integrated autocorrelation time for $\tau_{\rm exp} = 10$. The solid lines show the exact results.



Binning analysis for $\tau_{\rm exp} = 10$. The solid line shows the exact result.

A Realistic Example: 2D Ising Model

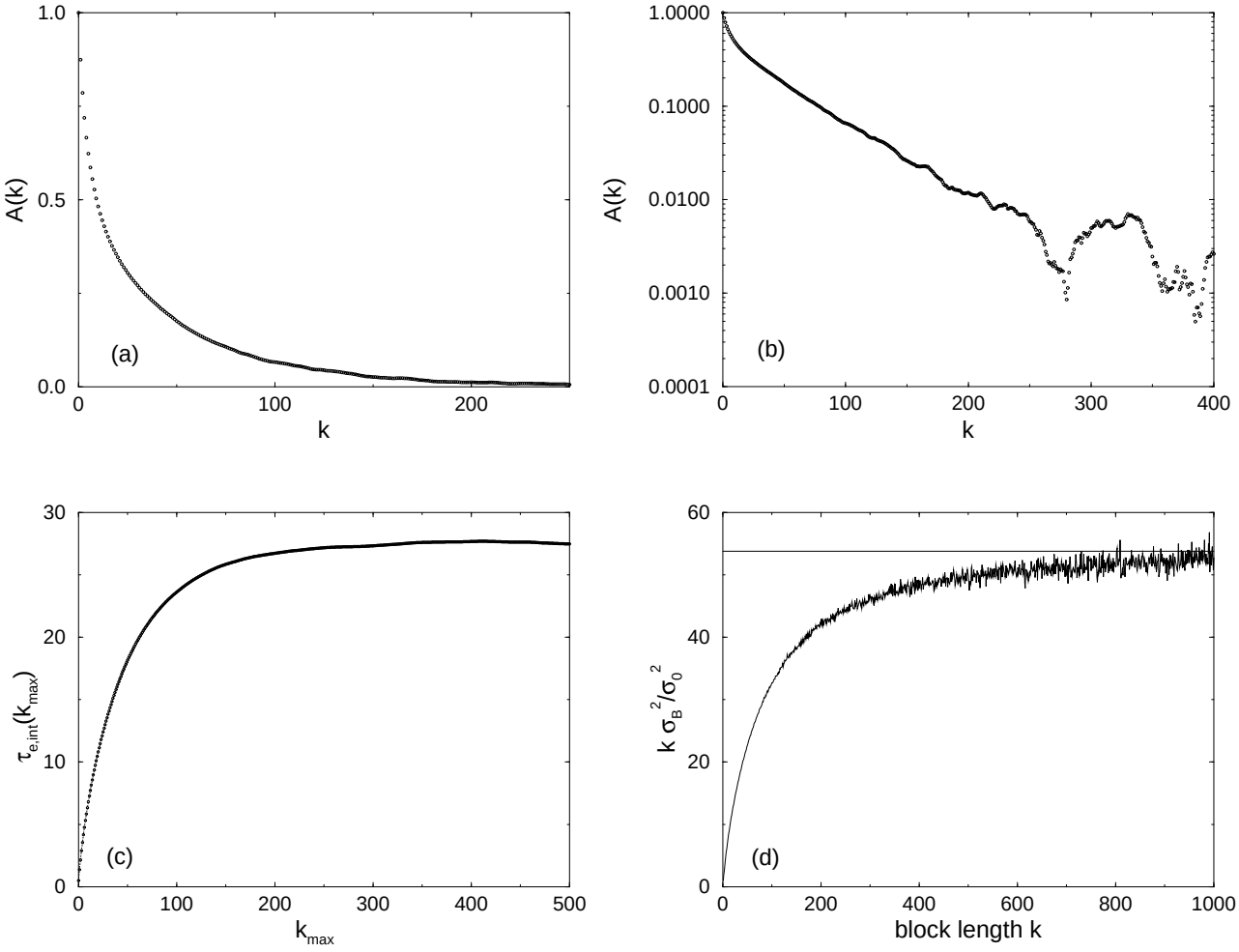
Metropolis simulations, 16×16 lattice at β_c:



Part of the time evolution of the energy e and magnetization m.

Total length of time series: 1 000 000 measurements.

Autocorrelation and Binning Analysis



Autocorrelation function of the energy on (a) linear and (b) log scale. (c) Integrated autocorrelation time and (d) binning analysis.

16×16 lattice, $\beta = \beta_c$.

Horizontal line in (d): $2\tau_{e,\mathrm{int}} \approx 54$.

Local Update Algorithms

. . . such as Metropolis, heat-bath, Glauber, . . .

Dynamical critical exponent ($\tau \propto L^z$):

$$z \approx 2$$

Critical slowing down

Example: 100×100 square lattice ($L = 100$)

$\Rightarrow \tau \simeq 10\,000$

$\Rightarrow$ Only every $\approx 10\,000$th sweep through the lattice, a statistically independent measurement can be taken ($L = 1000 \Rightarrow 1\,000\,000$)

Questions:

Is there a proper way of flipping (non-locally) whole clusters of spins?

Does that reduce the dynamical critical exponent z?

Answers:

YES – YES: Cluster update algorithms

Cluster Algorithms

Employ Fortuin-Kasteleyn representation:

$$\begin{aligned} Z &= \sum_{\{\sigma_i\}} \exp\left(\beta \sum_{\langle ij\rangle} \sigma_i\sigma_j\right) \\ &= \sum_{\{\sigma_i\}} \prod_{\langle ij\rangle} e^{\beta}\left[(1-p) + p\delta_{\sigma_i\sigma_j}\right] \\ &= \sum_{\{\sigma_i\}} \sum_{\{n_{ij}\}} \prod_{\langle ij\rangle} e^{\beta}\left[(1-p)\delta_{n_{ij},0} + p\delta_{\sigma_i\sigma_j}\delta_{n_{ij},1}\right] \quad , \end{aligned}$$

with

$$p = 1 - e^{-2\beta} \quad .$$

Here use was made of the fact that $\sigma_i = \pm 1$ so that

$$\exp(\beta\sigma_i\sigma_j) = x + y\delta_{\sigma_i\sigma_j} \quad ,$$

and of the "deep" identity

$$a + b = \sum_{n=0}^{1} (a\delta_{n,0} + b\delta_{n,1}) \quad .$$

n_{ij} are bond variables with the interpretation:

$n_{ij} = 0$: "deleted" bond
$n_{ij} = 1$: "active" bond

Swendsen-Wang Cluster

According to the last line of the Fortuin-Kasteleyn representation, a cluster update sweep then consists of alternating updates of the bond variables n_{ij} for given spins with updates of the spins σ_i for a given bond configuration:

1. Set $n_{ij} = 0$ if $\sigma_i \neq \sigma_j$, or assign values $n_{ij} = 1$ and 0 with probability p and $1 - p$, respectively, if $\sigma_i = \sigma_j$, cp. Fig.
2. Identify clusters of spins that are connected by "active" bonds ($n_{ij} = 1$).
3. Draw a random value ± 1 independently for each cluster (including one-site clusters), which is then assigned to all spins in a cluster.

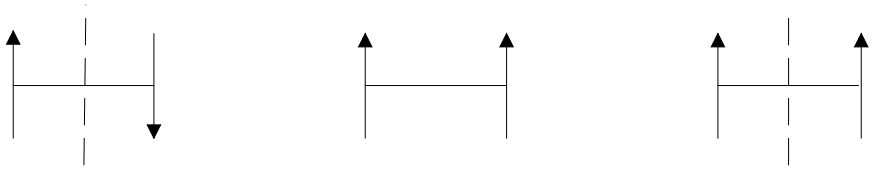


Illustration of the bond variable update. The bond between unlike spins is always "deleted" as indicated by the dashed line. A bond between like spins is only "active" ($n_{ij} = 1$) with probability

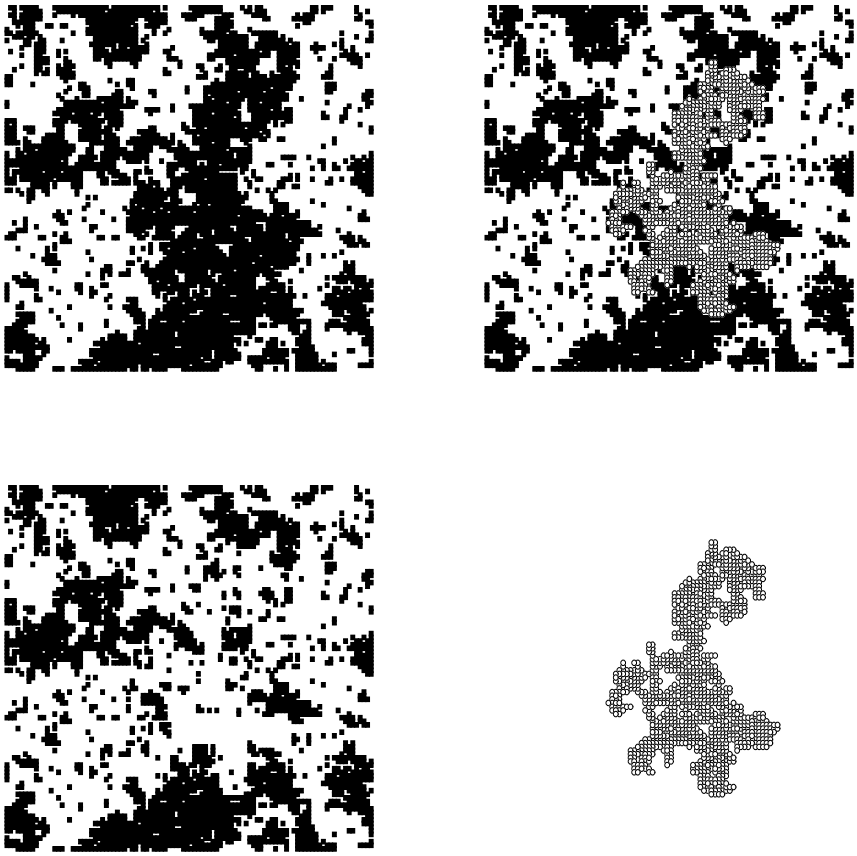
$$p = 1 - \exp(-2\beta) \ .$$

Only at zero temperature ($\beta \longrightarrow \infty$) stochastic and geometric clusters coincide.

Wolff Single Cluster

Here only a single cluster is flipped at a time:

1. Choose lattice site at random.
2. Construct the cluster around this site.
3. Flip the spins in this cluster always.



The spin configuration is from an actual simulation of the 2D Ising model at $0.97 \times \beta_c$ on a 100×100 lattice.

Adjusting work (= time) scale: A sweep consists of $V/\langle C \rangle$ single cluster steps, where $\langle C \rangle$ denotes the average cluster size, which is a very convenient so-called "improved estimator" for the magnetic susceptibility, i.e., scales with system size L as $L^{\gamma/\nu}$.

Single-Histogram Technique

Rewrite partition function in terms of density of states:

$$Z = \sum_{\{s\}} e^{-\beta H} = \sum_E \Omega(E) e^{-\beta E} \sim \sum_E P_\beta(E)$$

At simulation point β_0:

$$P_{\beta_0}(E) \sim \Omega(E) e^{-\beta_0 E}$$

Then:

$$\begin{aligned} P_\beta(E) &\sim \Omega(E) e^{-\beta E} \\ &= \Omega(E) e^{-\beta_0 E} e^{-(\beta-\beta_0)E} \\ &\sim P_{\beta_0}(E) e^{-(\beta-\beta_0)E} \ , \end{aligned}$$

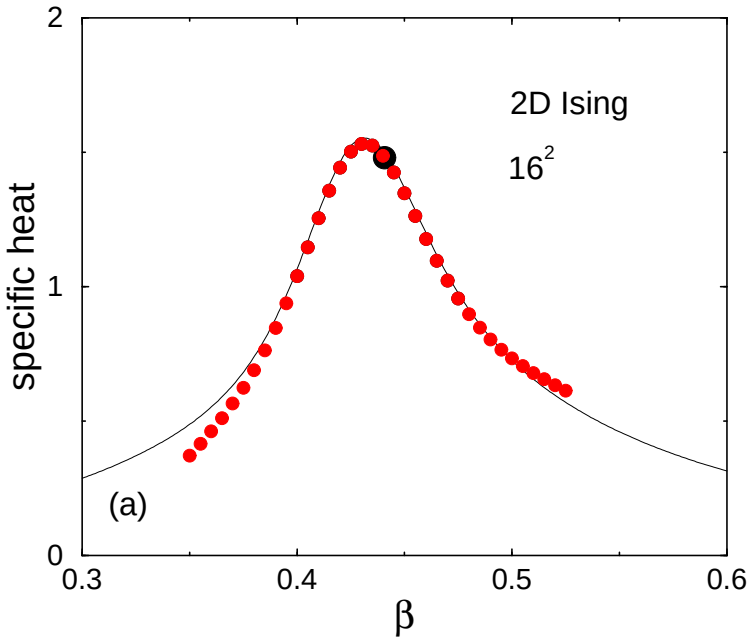
i.e., up to normalization the histogram at any point β can be (trivially) computed from the MC sampled histogram at the simulation point β_0.

Normalization unimportant (cancels) for expectation values.

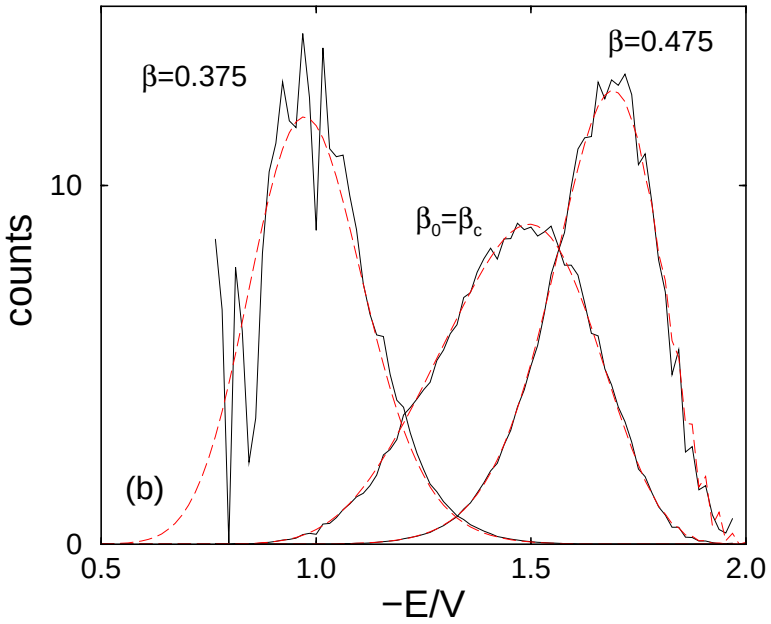
Disclaimer: Be careful with "any" in practice – for finite statistics, β must be "close enough" to β_0.

Ferrenberg & Swendsen (1988)

Example: 2D Ising Model



Specific heat computed by reweighting (●) from a single MC simulation at $\beta_0 = \beta_c$ (●). Continuous line is the exact solution.



Energy histograms at the simulation point $\beta_0 = \beta_c$, and reweighted to $\beta = 0.375$ and $\beta = 0.475$. Red dashed lines are exact.

Reweighting Range

Width of energy histogram:

$$\Delta e^2 \equiv \langle e^2 \rangle - \langle e \rangle^2 = C(\beta_0)/\beta_0^2 V$$

Taylor expansion of energy around T_0 ($C(T_0) = (\partial e/\partial T)|_{T_0}$):

$$e(T) = e(T_0) + C(T_0)(T - T_0) + \ldots$$

Require:

$$e(T) - e(T_0) = \Delta e$$

Hence:

$$C(T_0)(T - T_0) = T_0\sqrt{C(T_0)/V}$$

or

$$\Delta t \equiv \frac{T - T_0}{T_0} = \frac{1}{\sqrt{V}}\frac{1}{\sqrt{C(T_0)}}$$

Three cases:

Off-critical:	$C(T_0) \approx \text{const.}$	$\Delta t \sim L^{-d/2}$
Critical:	$C(T_0) \sim L^{\alpha/\nu}$	$\Delta t \sim L^{-1/\nu}$
1st order:	$C(T_0) \sim L^d = V$	$\Delta t \sim L^{-d} = 1/V$

(Hyperscaling employed: $\alpha = 2 - d\nu$)

Multiple-Histogram Technique

- Perform m MC simulations at $\beta_1, \beta_2, \ldots, \beta_m$
- Reweight all data to one reference point β_0
- Take care of (reweighted) error bars and compute error weighted average at β_0
- Reweight "combined" (averaged) histogram to any other β (in the reliable reweighting range)

By conveniently chosing $\beta_0 = 0$, this yields:

$$\Omega(E) = \frac{\sum_{i=1}^{m} P_{\beta_i}(E)}{\sum_{i=1}^{m} N_i Z(\beta_i)^{-1} e^{-\beta_i E}}$$

Self-consistency determines

$$Z(\beta_i) = \sum_E \Omega(E) e^{-\beta_i E} = \sum_E e^{-\beta_i E} \frac{\sum_{k=1}^{m} P_{\beta_k}(E)}{\sum_{k=1}^{m} N_k Z(\beta_k)^{-1} e^{-\beta_k E}} \quad ,$$

up to an overall, unimportant constant.

Ferrenberg & Swendsen (1989)

Tempering Methods

Loosely speaking, tempering methods may be characterized as "dynamical multi-histogramming".

Consider m simulation points $\beta_1 < \beta_2 < \ldots < \beta_m$ and treat them as dynamical variables to be updated, similar to the spin degrees of freedom.

Two different variants:

- Simulated tempering (expanded ensemble)
- Parallel tempering (exchange Monte Carlo, multiple Markov chain Monte Carlo)

Both are very useful for complex, disordered systems (e.g., spin glasses).

Simulated Tempering

Joint partition function (expanded ensemble):

$$Z_{\rm ST} = \sum_{i=1}^{m} e^{g_i} \sum_{\{s\}} e^{-\beta_i H(\{s\})} \ ,$$

where $g_i = \beta_i f(\beta_i)$ and the inverse temperature β is treated as an additional dynamical degree of freedom that can take the values $\beta_1, \ldots, \beta_m$.

Propose move from $\beta = \beta_i$ to β_j and accept a la Metropolis with probability

$$\min\left[1, \exp[-(\beta_j - \beta_i) H(\{s\})] + g_j - g_i\right] .$$

Similar to multi-histogram reweighting (and also to multicanonical simulations discussed below), the free-energy parameters g_i are a priori unknown and have to be adjusted iteratively.

To assure a reasonable acceptance rate for the β-update moves (usually between neighbouring β_i-values), the histograms at β_i and β_{i+1}, $i = 1, \ldots, m-1$, must overlap. An estimate for a suitable spacing $\delta\beta = \beta_{i+1} - \beta_i$ is hence immediately given by our previous results for the reweighting range:

$$\delta\beta \propto \begin{cases} L^{-d/2} & \text{off-critical} \ , \\ L^{-1/\nu} & \text{critical} \ , \\ L^{-d} & \text{first-order} \ . \end{cases} \qquad (1)$$

Parallel Tempering

While the "expanded" ensemble of the simulated tempering method is built on a sum of partition functions, here the starting point is the product of partition functions ("extended" ensemble)

$$Z_{\rm PT} = \prod_{i=1}^{m} Z(\beta_i) = \prod_{i=1}^{m} \sum_{\{s\}_i} e^{-\beta_i H(\{s\}_i)} \ ,$$

All m systems at different simulation points $\beta_1 < \beta_2 < \ldots < \beta_m$ are simulated in parallel, using any legitimate update algorithm (Metropolis, cluster,. . .).

After a certain number of sweeps, exchanges of the current configurations $\{s\}_i$ and $\{s\}_j$ are attempted (equivalently, the β_i may be exchanged, as is done in most implementations for efficiency reasons).

Accept proposed exchange a la Metropolis with probability

$$W = \min(1, e^{\Delta}) \ , \quad \Delta = (\beta_j - \beta_i)[E(\{s\}_j) - E(\{s\}_i)] \ .$$

To assure a reasonable acceptance rate, usually only "nearest-neighbour" exchanges ($j = i \pm 1$) are attempted, with spacing $\delta\beta = \beta_{i+1} - \beta_i$ choosen as for simulated tempering.

Notice that in parallel tempering no free-energy parameters must be adjusted. The method is thus very flexible and moreover can be almost trivially parallelized.

Summary

- Great improvements of MC simulation methodology $\Rightarrow$ high precision results $\Rightarrow$ need for careful and reliable statistical error analyses.
- Autocorrelations require rather involved analyses.
- However, this extra effort is worth spending when compared with the months or even years of computer time needed for the generation of the raw data.