Advanced Monte Carlo Methods in Classical Statistical Physics
A very short bibliography to start:


- *Introduction to phase transitions and critical phenomena.* H.E. Stanley

- *Statistical mechanics of phase transitions* J.M. Yeomans

...and several original papers that will be indicated during the lectures
Plan of the Lectures

• Introduction: Standard Monte Carlo

• Limitations of the Simple Sampling Method: finite size effects, finite temperature steps, critical slowing down, meta-stabilities.

• Corrections to finite temperature steps: Histogram methods

• Corrections to critical slowing down: cluster algorithms

• Corrections to meta-stabilities : extended ensemble methods
I. Introduction

The problem: thermal averages of an observable $A (x_i)$, where $x_i$ : microstate of the system.

$$<A> = \frac{\sum A(\bar{x}_i) e^{-\beta E(\bar{x}_i)}}{\sum e^{-\beta E(\bar{x}_i)}} = \frac{1}{Z} \sum A(\bar{x}_i) e^{-\beta E(\bar{x}_i)}$$

But: to sum over all microstates is not possible!

Toy model: 2D Ising model

$$H = -J \sum_{<i,j>} s_i s_j$$

with $s_i = 1$.

For a very modest lattice, $L=10$, $\#\Omega = 2^{100} \approx (10^3)^{10} = 10^{30}$ !!!

Monte Carlo Techniques
a) **Simple Sampling**: just cut the sums, selecting only $1 << M << 2^N$ states with a uniform distribution.

**But**: most of the chosen states will correspond to high energy region (as DOS grows rapidly with $E$)

bad results! (unless $\beta \to 0$)

b) **Importance Sampling**: choose $1 << M << 2^N$ states $x_l$ among the most probable $\mu$-states at temperature $T$.

With the canonical probability distribution: $P_{eq}(\tilde{x}_l) \propto e^{-\beta E(\tilde{x}_l)}$
\[
\langle A \rangle = \frac{1}{M} \sum_{l=1}^{M} \tilde{A}_l(x_l)
\]

where the \( \tilde{A} \) correspond to the \( \mu \)-states chosen with \( P_{eq}(x_l) \)

How to find those states?

Metropolis dynamics

Build a Markov chain of \( \mu \)-states \( x_{tn} \) with the transition rate verifying *detailed balance condition*

\[
W(x_j \rightarrow x_i) P(x_j) = W(x_i \rightarrow x_j) P(x_i)
\]

This condition is not necessary but enough to guarantee convergence to an equilibrium state.
Master equation:

\[
dP(\vec{x}_j, t) \over dt = - \sum_j W(\vec{x}_j \to \vec{x}_i) P(\vec{x}_j) + \sum_i W(\vec{x}_i \to \vec{x}_j) P(\vec{x}_i)
\]

becomes stationary with detailed balance condition!

This only imposes a ratio between the transition rates different possible dynamics:

\[
W(\vec{x} \to \vec{x}') = \frac{1}{\tau_s} \frac{e^{-\beta \Delta H}}{1 + e^{-\beta \Delta H}}
\]

\(\Delta H = H(\vec{x}') - H(\vec{x})\)

- **Glauber dynamics**
- **Metropolis dynamics**
Metropolis algorithm: single flip dynamics

For each lattice site $i$, propose a local move $x \rightarrow x'$

Evaluate $\Delta H$ between initial and final states

$\Delta H \leq 0$?

Accept the new state $x'$

$W \geq R$?

$R$ uniform RN $\in [0,1]$
General layout of a MC study:

- **Working temperature loop:** for each temperature $T$, start with a given initial configuration of the system.

- **Thermalisation stage:** iterate Metropolis algorithm during $N_{\text{therm}}$ steps and discard the results.

- **Calculation stage:** iterate Metropolis algorithm during $N_{\text{calc}}$ steps and calculate averages of different quantities at the working temperature $T$. 

General idea: to get an estimate of the $g(E)$ from the canonical probability distribution of the energy at temperature $T$

Let’s consider the Ising Hamiltonian in a magnetic field of intensity $h$:

$$-\beta H = K \sum_{<i,j>} s_i s_j + h \sum_j s_j = KS + hM$$

; $K=J/kT$ and $h=B/kT$

The joint probability distribution of $S,M$ is

$$P_{K,h}(S,M) = \frac{1}{Z(K,h)} N(S,M) \exp(KS + hM)$$
$N(S,M)$ is the number of configurations at the point $(S,M)$ at the phase space (a bi-dimensional density of states) and:

$$Z = \sum_{S,M} N(S, M) \exp(KS + hM)$$

Notice that

$$\Gamma_{K,h}(S,M) \propto P_{K,h}(S,M)$$

$\Gamma(S,M)$: histogram calculated by a on MC simulation at a given point of the parameter space $K,h$.

Then, the probability distribution at another point $K',h'$ may be obtained from the one calculated at $K,h$ provided that $K'-K$ and $h'-h$ are “small”:
\[
P_{(K',h')}(S, M) = \frac{P_{(K,h)}(S, M) \exp[(K' - K)S + (h' - h)M]}{\sum_{S,M} P_{(K,h)}(S, M) \exp[(K' - K)S + (h' - h)M]}
\]

As K and h are continuous variables any average may be calculated as a continuous function of the parameters, using the data issued from a single simulation!!!

Not so quickly....This scheme works only for neighbouring values of the parameters!

Multi-histogram MC!
Convergence problems at second order phase transitions: critical slowing down

As \( T \to T_c \) large correlations appear in the system which are associated to the correlation time of the markovian process at the origin of Metropolis algorithm (Van Hove, Phys. Rev. 93 p1374 1954)

At equilibrium a time correlation function may be defined (for instance for the magnetization):

\[
\Phi_{MM}(t) = \frac{\langle M(0)M(t) \rangle - \langle M \rangle^2}{\langle M^2 \rangle - \langle M \rangle^2}
\]

\( t \to 0; \Phi(t) \to 1 \)

\( t \to \infty; \Phi(t) \to 0 \)
Asymptotic behaviour:

\[ \Phi(t)_{t \to \infty} \to e^{-t/\tau} \]

as \( T \to T_c \)

\[ \tau \propto \xi^z \quad \text{and} \quad \xi \propto \varepsilon^{-\nu} \]

so \( \tau \propto \varepsilon^{-\nu z} \)

with \( z: \textit{dynamical critical exponent} \)

It is usual to define the \textit{integrated correlation time}:

\[ \tau_{\text{int}} = \int_{0}^{\infty} \Phi(t) dt \]

measuring the characteristic time for the decay of correlations

important in error estimation!
Let’s suppose $N$ successive observations of a quantity $A$: $A_\mu$, $\mu=1, N \gg 1$, built by a Markov chain in an importance sampling scheme.

The expectation value of the square of the statistical error of $A$ is:

$$\langle (\delta A)^2 \rangle = \left\langle \left[ \frac{1}{N} \sum_{\mu=1}^{N} (A_\mu - \langle A \rangle) \right]^2 \right\rangle$$
Which can be written:

\[
< (\delta A)^2 > = \frac{1}{N^2} \sum_{\mu=1}^{N} < (A_\mu - < A >)^2 > + \frac{2}{N^2} \sum_{\mu_1=1}^{N} \sum_{\mu_2=\mu_1+1}^{N} (< A_{\mu_1} A_{\mu_2} > - < A >^2)
\]

As Metropolis algorithm guarantees that all the \( \mu \)-states are distributed with a Boltzmann law, the origin of the “time” is indistinguishable from any other time translational invariance, which gives:

\[
< A_{\mu_1} A_{\mu_2} > = < A_0 A_{\mu_1-\mu_2} >
\]

valid as long as the system is at equilibrium at the working temperature \( T \).

This means that thermalization steps are achieved, and the corresponding out of equilibrium \( \mu \)-states had been discarded.
So changing variables: \( \mu = \mu_1 - \mu_2 \)

\[
\langle (\delta A)^2 \rangle = \frac{1}{N} \left[ \langle A^2 \rangle - \langle A \rangle^2 + 2 \sum_{\mu=1}^{N} \left( 1 - \frac{\mu}{N} \right) \langle A_0 A_\mu \rangle - \langle A \rangle^2 \right]
\]

- Let’s introduce a “time” associated to the MC process: \( t = \mu \delta t \) where \( \delta t \) : time interval between 2 successive \( \mu \)-states \( \mu \), et \( \mu + 1 \)

- Let’s \( N \) be the number of degrees of freedom (number of spins in the Ising model) \( \delta t = \tau_s / N \) where \( \tau_s \) is the time constant giving the transition rate in the Metropolis algorithm.
This would mean that there are $\mathcal{N}$ transitions in the unit time → high correlation between subsequent states.

**Correction:** it is standard to take $\delta t \gg \tau_s / \mathcal{N}$, commonly $\delta t = \tau_s$ → 2 successive observations entering in the average calculation are separated by a number of Markov chain steps of the order of the degrees of freedom of the system (good for thermodynamical limit) this is the so called “Monte Carlo step per spin=$\text{MCS/s}$”

**ATTN:** it may not be enough to de-correrate the microstates at $T \approx T_c$ !!!
Near $T_c$ correlation time diverges $\tau \gg \delta t$ one can transform the sums into integrals:

$$\sum_{\mu} \ldots \rightarrow \int_0^t \ldots dt'$$

which gives:

$$\langle (\delta A)^2 \rangle = \frac{1}{N} \left[ \sum A^2 - \langle A \rangle^2 \right] + \frac{2}{\delta t} \int_0^t \left( 1 - \frac{t'}{t} \right) \sum A(0)A(t') - \langle A \rangle^2 \right] dt'$$

and using the definition of the time correlation function:

$$\langle (\delta A)^2 \rangle = \frac{1}{N} \left[ \sum A^2 - \langle A \rangle^2 \right] + \frac{2}{\delta t} \int_0^t \left( 1 - \frac{t'}{t} \right) \Phi_A(t') dt'$$
For long enough simulation times $t \gg \tau_A$ the term in $t'/t$ can be neglected in the regime $t'/t \ll 1$ and in the regime $t'/t \leq 1$ because $\Phi_A(t) \to 0$ for $t \sim \tau_A$.

This leads to:

$$
\langle (\delta A)^2 \rangle = \frac{1}{N} \langle A^2 \rangle - \langle A \rangle^2 \left[ 1 + \frac{2}{\delta t} \tau_A \right]
$$

The quadratic mean of the statistical error of the averages of a quantity $A$ calculated by the MC process differs from $V(A)/N$ by a term which is proportional to $\tau_A$ which may become very large for $T \sim T_c$. 
As $\tau_A$ diverges when $T \sim T_c$ → **critical slowing down**

special algorithms to reduce correlations

→ **Cluster flipping algorithms:**

- Swendsen and Wang
- Wolff
Cluster flipping algorithms


FK transformation: ferromagnetic Potts model → percolation problem.

Interesting: percolation is a geometrical transition → No critical slowing down!!

A typical percolation problem:
Let's define $H_{lm}$: the restricted hamiltonian where the sites $l,m$ have been removed and two restricted sums corresponding to the cases where

- $\sigma_l = \sigma_m$ and $\sigma_l \neq \sigma_m$

respectively.

So

One can then write:

$$Z = Z^{\text{same}}_{<l,m>} + e^{-K} Z^{\text{diff}}_{<l,m>}$$

$$Z = \left( 1 - e^{-K} \right) Z^{\text{same}}_{<l,m>} + e^{-K} Z^{\text{ind}}_{<l,m>}$$
the weight factors can be interpreted in terms of probability: $p = 1 - e^{-K}$ of being equal (bond between $l$ and $m$)!!

So the partition function may be written:

$$Z = \sum_{\text{bonds}} p^b (1 - p)^m q^{N_c}$$

Where: $b = \text{number of bonds}$
$m = \text{number of nn not having a bond}$
$N_c = \text{Number of clusters in the system}$

Basis of the cluster methods...

With a starting configuration, choose a site $i$

∀ j nn of i

$\sigma_i = \sigma_j$?

yes

$p = 1 - e^{-K\delta_{\sigma_i\sigma_j}}$

$r \in U[0,1]$

$p > r$?

no

yes

draw a bond linking $i$ and $j$

yes

no

All the bonds considered?

Identify clusters (Hoshen-Kopelman)

erase all the bonds

All the clusters done?

yes

no

set $\sigma_i = R \ \forall i$ in the cluster

sort a $\text{rn } R=1,2,...q$

Choose a cluster

no

yes


With a starting configuration, choose a site $i$

∀ j nn of i

$\sigma_i = \sigma_j$?

yes

$p = 1 - e^{-K\delta_{\sigma_i\sigma_j}}$

$r \in U[0,1]$

$p > r$?

no

yes

draw a bond linking $i$ and $j$

yes

no

All the bonds considered?

Identify clusters (Hoshen-Kopelman)

erase all the bonds

All the clusters done?

yes

no

set $\sigma_i = R \ \forall i$ in the cluster

sort a $\text{rn } R=1,2,...q$

Choose a cluster
Notice that:

- the SW algorithm forms “physical clusters” $\neq$ “geometrical clusters.

- low $T \Rightarrow$ big $K \Rightarrow p \rightarrow 1 \Rightarrow$ big clusters
- high $T \Rightarrow$ small $K \Rightarrow p \rightarrow 0 \Rightarrow$ small clusters
- $T \sim T_C$ clusters of all sizes $\Rightarrow$ each $\mu$-state differs substantially from the previous one $\Rightarrow$ reduction of critical slowing down!

- Wang (1990) (Physica A 164 240) calculates $z \rightarrow 0$ in 2D (instead of $z=2$), $z = 0.5$ in 3D.

- Immediate extensions: a) magnetic field
  b) antiferromagnetic Potts
- **Drawback:** a lot of time is lost dealing with small clusters...

**Principle:** build only one cluster and flip it with a given probability

1. Choose a site $i$ at random

2. Add a bond to its neighbours with a probability $P_{add}$ provided $\sigma_i = \sigma_j$

3. For each one of new spins integrated to the cluster do (2)
   
   *Cluster stops growing when all the possible new neighbours have been visited*

4. Flip the resulting cluster with an acceptance probability $A$

   One can wisely calculate $P_{add}$ for a given $A$ (high!)
Performance of Wolff’s algorithm:

Let’s consider 2 \( \mu \)-states \( \mu \) et \( \nu \) differing only in the flipping of a single cluster.

\[ \mu \rightarrow \nu: 7 \text{ bonds to brake} \]

\[ \nu \rightarrow \mu: 9 \text{ bonds to brake} \]
Let’s analyse the move $\mu \rightarrow \nu$ and its reverse one $\nu \rightarrow \mu$:

- Same probability of choosing a given seed to grow the cluster
- Same probability of adding each spin to the growing cluster
- Different acceptance probability because the number of bonds to brake is not the same in direct and reverse moves
- These bonds represent the spins that are oriented in the same direction as those of the cluster that haven’t been chosen for the physical cluster definition of the cluster
- Proba of NOT adding a site: $(1-P_{\text{add}})$, proba of NOT adding $m$ sites: $(1-P_{\text{add}})^m$
Proba of selecting a given cluster: \( g(\mu \rightarrow \nu) \approx (1 - P_{\text{add}})^m \)

Detailed balance condition:

\[
P(\mu) g(\mu \rightarrow \nu) A(\mu \rightarrow \nu) = P(\nu) g(\nu \rightarrow \mu) A(\nu \rightarrow \mu)
\]

with:

\[
P(\mu) \propto e^{-\beta E_{\mu}}
\]

\[
(1 - P_{\text{add}})^{m-n} \frac{A(\mu \rightarrow \nu)}{A(\nu \rightarrow \mu)} = e^{-\beta (E_{\nu} - E_{\mu})}
\]

each one of the \( m \) broken bonds \( \mu \rightarrow \nu : +2J \)

each one of the \( n \) broken bonds \( \mu \rightarrow \nu : +2J \)

\[
\Delta E = E_{\mu} - E_{\nu} = 2J(m-n)
\]
It’s enough to choose \( P_{\text{add}} = 1 - e^{-2\beta J} \) to have an acceptance ratio independent of the move!

Each proposed move is accepted!!!

By a single passage through the lattice

cluster identification
determination of the cluster sizes

Let’s suppose a crystal composed of 2 types of sites A and B

- Sweep the lattice in order:
  - label: \( m_j^\alpha \) if molecule A
  - label: 0 if molecule B

Notice:
\( \alpha \) stands for the name of the cluster, a given cluster may have different labels \( \{ m_1^\alpha m_2^\alpha \ldots m_p^\alpha m_s^\alpha \ldots \} \), \( m_s^\alpha \) is the smallest: *proper cluster label*
In order to assign the label, look for the previously assigned nn of each site (sweep the lattice in an ordered way):

- one nn found → give to the new site the same label of that neighbour
- more than one nn found → update “label of the labels” $N(m_j^\alpha)$

$N(m_j^\alpha)<0$ → $m_j^\alpha$ is not the smallest label of the cluster $\alpha$

$N(m_s^\alpha)>0$ → $m_s^\alpha$ is the smallest label of cluster $\alpha$ and gives the number of sites in the cluster
\[ m^1_1 = 1, \quad N(1) = 1 \]
\[ m^1_2 = 1, \quad N(1) = 2 \]
\[ m^2_1 = 2, \quad N(2) = 1 \]
\[ m^3_1 = 3, \quad N(3) = 1 \]
\[ m^3_2 = 3, \quad N(3) = 2 \]
\[ m^4_1 = 4, \quad N(4) = 1 \]
\[ m^4_2 = 4, \quad N(4) = 2 \]
\[ m^5_1 = 5, \quad N(5) = 1 \]
\[ m^5_2 = 5, \quad N(5) = 2 \]
\[ m^6_1 = 6, \quad N(6) = 1 \]
\[ m^6_2 = 6, \quad N(6) = 2 \]
\[ m^6_3 = 6, \quad N(6) = 3 \]
\[ m^7_1 = 7, \quad N(7) = 1 \]
\[ m^7_2 = 7, \quad N(7) = 2 \]
\[ m^7_3 = 2, \quad N(7) = 3 \]

\[ N(m^2_1) = N(2) = 1 \]
\[ N(2) \leftarrow N(2) + N(7) + 1 = 4 \]
\[ N(7) = -2 \]

\[ N(m^4_3) = N(4) = 3 \]
\[ N(2) \leftarrow N(2) + N(3) + 1 = 9 \]
\[ N(3) = -2 \]

\[ N(m^7_6) = 2 \]
\[ N(2) \leftarrow N(2) + N(3) + 1 = 9 \]
\[ N(3) = -2 \]

\[ m^7_7 = 2, \quad N(2) = 10 \]
nothing to do with cluster 3 as \( N(3) < 0 \)
Lattice size : 29x29

“A” concentration c = 0.57

Largest cluster : label 5
N(5) = 209 (“Z”)

Other clusters (“I”)
ex: N(2) = 45
III. Extended Ensemble Methods
(recent review: Yukito Iba cond-mat/0012323v2)

**Aim:** to overcome regions of high potential barriers, typically meta-stable states in first order transitions

**Ancestor:** Histogram Method (Ferrenberg & Swendsen 1988)
PRL 61 p2635

Loosely speaking, two kinds of algorithms:

A) Just enlarge the ensemble by including T as a variable and *recovering the canonical averages as marginal laws*

Examples: Simulated and Parallel Tempering

B) Methods based on the calculation of the density of states (DOS), generally working in an artificial ensemble *recovering the canonical averages through a reweighting procedure.*
Examples:

• Multicanonical; Berg&Neuhaus, (1991) PRL B267 p249

• Entropic Sampling; Lee (1993) PRL 71 p211


• Flat Histogram MC; Wang (2000) Physica A 281 p147

• Transition Matrix Monte Carlo; Wang&Swendsen (2001) cond-mat/0104148v1

• Wang- Landau Monte Carlo (2001) PRL 86 p2050
B. Methods based in the calculation of DOS

1st order transitions + local dynamics → bad convergence as energy barriers grow with the size of the system

General idea:

\[
\int dE g(E) = N
\]

N total number of configurations, for Ising

\[ N = 2^{N_s} \]

where \( N_s \) is the total number of spins

g(E) is a big number!!! → ln[g(E)]

Lower bound: ground state,
Upper bound: depends on the system
Detailed balance:

\[ W(\vec{x}_j \rightarrow \vec{x}_i) P_{eq}(\vec{x}_j) = W(\vec{x}_i \rightarrow \vec{x}_j) P_{eq}(\vec{x}_i) \]

Can be expressed in terms of energy:

\[ \Pi(E \rightarrow E') \exp(-\beta E) = \Pi(E' \rightarrow E) \exp(-\beta E') \]

If one records the histogram of the energy \( E \) at temperature \( \beta \):

\[ H_\beta(E) \propto g(E) \exp(-\beta E) \]
If one considers a different equilibrium distribution in the detailed balance:

\[
\Pi(E \to E') \exp(-\beta E + w(E)) = \Pi(E' \to E) \exp(-\beta E' + w(E'))
\]

Then the energy histogram becomes:

\[
H_\beta(E) \propto g(E) \exp(-\beta E + w(E))
\]

If one chooses

\[
w(E) \approx \beta E - \ln \left[ \frac{1}{E} \right]
\]

The histogram becomes flat independent of T!!!

Random walk in the energy space!
No more energy barriers!!!
But...we don’t know $g(E)$ beforehand!

Iterative learning procedure:
The Wang-Landau algorithm
General scheme:

1. Start with any configuration \( x \) of the system and set \( g(E) = \text{cst} \)

2. Propose a local modification: \( x \rightarrow x' \) evaluate \( E, E' \)

3. Accept the new state \( x' \) with probability

\[
\Pi(E \rightarrow E') = \min \left( 1, \frac{g(E)}{g(E')} \right)
\]

4. If \( x' \) is rejected count the old one again \( x' \leftarrow x \)

5. Update the density of states:

\[
\ln \left[ h(E') \right] \leftarrow \ln \left[ h(E') \right] + \ln f
\]

and the histogram:

\[
h(E') \leftarrow h(E) + 1
\]

6. Go to (2) till the histogram is flat

7. Update the WL factor, reset \( h(E) \) **but NOT** \( g(E) \)!!! Go to (2)
Remarks:

• \( f \) : modification factor or WL factor. It decreases at each \( f \)-iteration. Typically: \( f \leftarrow f^{1/2} \), with \( f_0 = e, \ln(f) = 10^{-8} \)

• One needs a flatness criterion to stop the simulation at each \( f \). Typically: histogram is flat if points are within 5% or 10% of the average uniform distribution (even less restrictive for large lattices).

• Notice that only in the last WL iterations detailed balance is verified as the transition probability between two given states remains constant.

• Convergence problems if \( (E_{\text{min}}, E_{\text{max}}) \) large (all the possible energies of the system)
Recovering thermodynamics: reweighting procedure

Once we have \( g(E) \):

\[
\langle E(T) \rangle = \frac{\sum_i E_i \exp(-\beta E_i) + \ln g(E_i)}{\sum_i \exp(-\beta E_i) + \ln g(E_i)}
\]

**Warning:** \( g(E) \) very large numbers

need of calculation tricks...

Also by fluctuation-dissipation theorem:

\[
C_v(T) = k_B \beta^2 \left< E^2 \right> - \left< E \right>^2
\]
Averages of any other physical quantity $A$:

$$\langle A(T) \rangle = \sum_i \frac{A(E_i) \exp(-\beta E_i) + \ln \mathcal{Z}(E_i)}{\sum_i \exp(-\beta E_i) + \ln \mathcal{Z}(E_i)}$$

where $A(E_i)$ is the micronanonical average at energy $E_i$.

The probability distribution of the energy:

$$P_T(E) = \frac{g(E) e^{-\beta E}}{Z(T)}$$

*In particular one can calculate $S(T)$ and $F(T)$ which is difficult with canonical MC!*
Non trivial problem: dealing with multi-range simulations
⇒ joining parts of \( \ln[g(E)] \) obtained from independent simulations in different energy intervals
   a lot of care to avoid (human) errors.

Solutions: i) Critical Minimum Energy Subspace (CrMES)

   ii) Adaptive windows WL method

Idea: in the thermal average sums at a given $T$, only some terms give an important contribution: those whose $E$ is in the neighbourhood of $\bar{E}$ where the max of $P_T(E)$

Iterative procedure to determine the minimum interval $\Delta \bar{E}$, (MES) where the WL simulation needs to be run.

The method is based in the central limit theorem: $P_T(E)$ approaches a Gaussian.

Moreover: one assumes it to remain a Gaussian even at $T_c(L)$

Then one expects that the standard deviation of the energy:
Then one expects that the standard deviation of the energy:

\[ \Delta \hat{E} = \sqrt{NT^2C} \]

\[ N = L^d \]

The CrMES is built, starting with the central value \( \hat{E} \) and extending its borders on both sides till the difference between \( C'(T) \) calculated in this way with respect to \( C(T) \) calculated using the whole energy interval (or in any other known way) is less than a given error.

Imposing the same level of error to all the sizes, there’s still a dependence on \( L \) affecting \( \Delta \hat{E} \). Near the critical temperature one can use the finite size scaling law for \( C_{T=T_c}(L) \sim L^{\alpha/2\nu} \).
The MES scales with $L$ as:

$$\frac{\Delta \hat{E}}{L^{d/2}} \approx L^{\alpha/2v}$$

Performing a WL simulation in the whole range for a small system, allows to estimate the restricted interval for larger systems.

This diminishes the number of junctions for $g(E)$!

Problems:

- not adapted to first order transitions and disordered systems:
- one should know critical exp of the model
ii) **Adaptive windows WL**: Cunha-Netto et. al. (2008)

1. For a given \( f \): perform a given WL in the whole energy range \( (E_{\text{min}}, E_{\text{max}}) \). But check whether the histogram is flat only in the range \( E_{\text{max}} - W \). Stop when flat

2. Try to recover energy levels for \( E < E_{\text{max}} - W \), determine \( E_1 \)

3. Perform a new simulation within \( (E_{\text{min}}, E_1 + \Delta E) \), stop if \( h(E) \) is flat in window \( W \) at the highest energy levels.

4. Cover all the energy range
5. Keep the records of the junction energies $E_j$

Start an new f-iteration with the joint $g_f(E)$ as an input.

*Notice*: eventual errors done at the junction of $g_f(E)$ are corrected at the next f-iteration!
A. Replica Exchange MC (Parallel Tempering): E. Marinari (1996) and many others....

**Main Idea:** Try to generalize the concept of simulated annealing to multivariate distributions: map an “easy” problem (the region of high entropy) into a difficult one (low entropy ground state).

**Notice that:** simulated annealing, good to escape from shallow potential wells but doesn’t accelerate jumps between deep meta-stable states need of an “up and down” strategy.

But: changing the Temperature in any way during the simulation, would destroy detailed balance!
Let’s consider a set of distributions \( \{ p_k(x) \} \) with different parameters \( \lambda_k \), \( k=1,K \), and let’s suppose that they are ordered:
\[ \lambda_1 > \lambda_2 > \ldots > \lambda_K \]

For the Gibbs distribution: \( \lambda_k = \beta_k \)

\[
p_k(\bar{x}) = \frac{\exp(-\beta_k E(\bar{x}))}{Z(\beta_k)}
\]

If we denote the variables of k-th system (k-th replica) as \( x_k \), the simultaneous distribution of \( \{ x_k \} \) is written as:

\[
\tilde{p} \left\{ x_k \right\} = \prod_k p_k(\bar{x}_k)
\]
Two types of updates:

- Conventional update in replica $k$ (detailed balance is verified)

- Replica exchange between replicas $k+1$ and $k$, with acceptance ratio $A = \min\{1, r\}$

For the canonical distribution:

$$r = \frac{p_k(\tilde{x}_k) p_{k+1}(\tilde{x}_{k+1})}{p_k(x_k) p_{k+1}(x_{k+1})} = \frac{p_k(x_{k+1}) p_{k+1}(x_k)}{p_k(x_k) p_{k+1}(x_{k+1})}$$

$$r = \exp \left[ \beta_k - \beta_{k+1} \left( E(\tilde{x}_k) - E(\tilde{x}_{k+1}) \right) \right]$$
So the replica exchange part of the transition probability:

\[ W(X, \beta_m \rightarrow X', \beta_n) = \begin{cases} 
1 \text{ for } \Delta < 0 \\
\exp(-\Delta) \text{ for } \Delta > 0 
\end{cases} \]

Fast mixing of independent Markov chains (coming from non-interacting replicas)