# Static & dynamic properties of spin glasses as seen through the parallel tempering telescope

Helmut G. Katzgraber
http://katzgraber.org

TEXAS A&M
UNIVERSITY

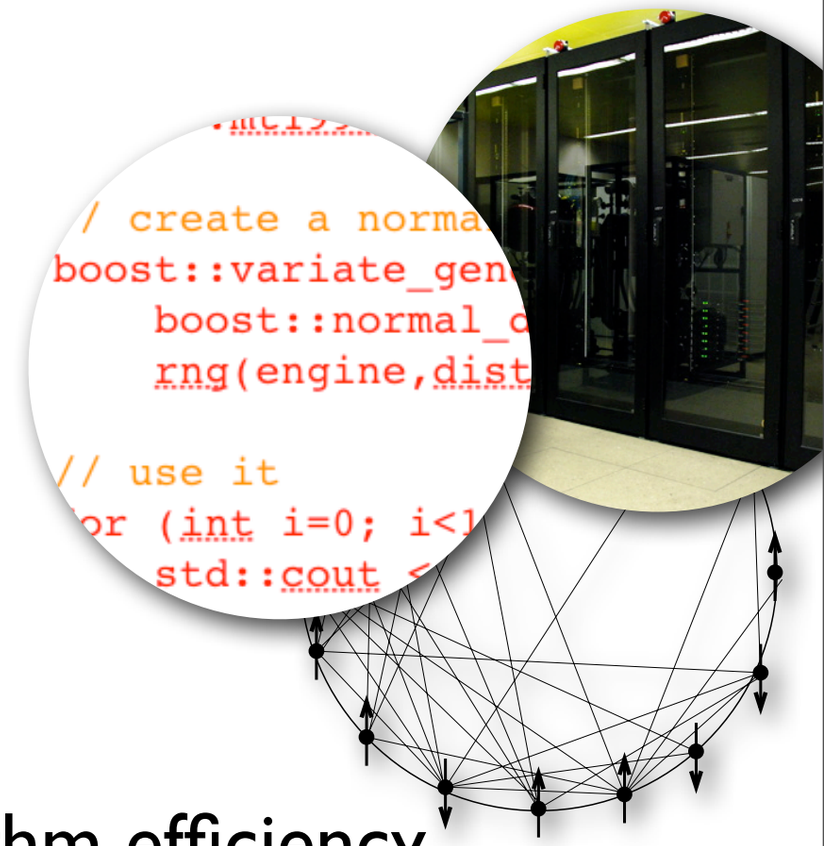
# Outline

- Introduction to spin glasses:
  - In case you forgot…
  - Reminder:  theoretical descriptions.

- Parallel tempering (exchange) Monte Carlo:
  - Why do we need it here?
  - Outline of the algorithm & practical issues.

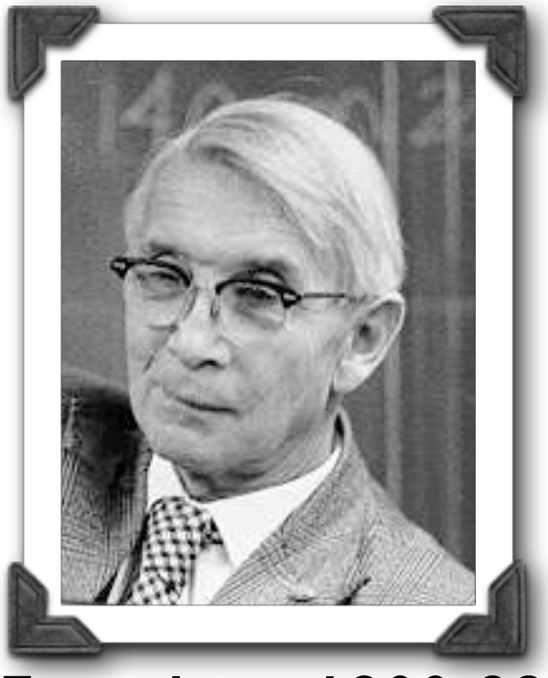
- Applications:
  - Correlations between free-energy and algorithm efficiency.
  - Nature of the spin-glass state.

- Collaborators:  B. Yucesoy, J. Machta & R. Andrist.

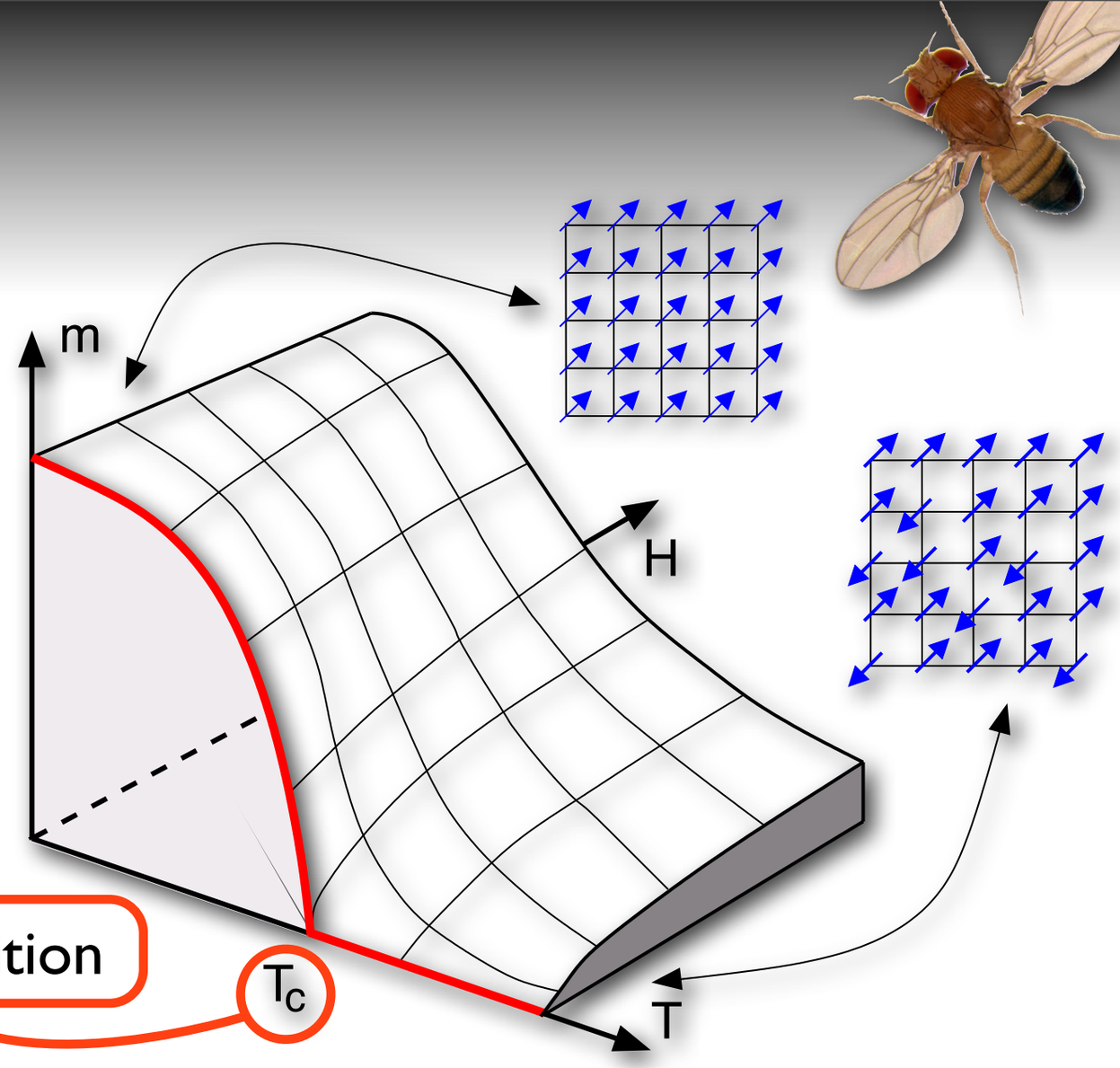


# Rewind: A brief introduction…

# The Ising model



Ernst Ising 1900-98




- Ising Model:

$$\mathcal{H} = -\sum_{\langle ij \rangle} J_{ij} S_i S_j - H \sum_i S_i \qquad S_i \in \{\pm 1\}$$

$$J_{ij} = 1 \;\; \forall i,j \qquad i \neq j$$

- Magnetization: measure of order $m = \frac{1}{N} \sum_i S_i$

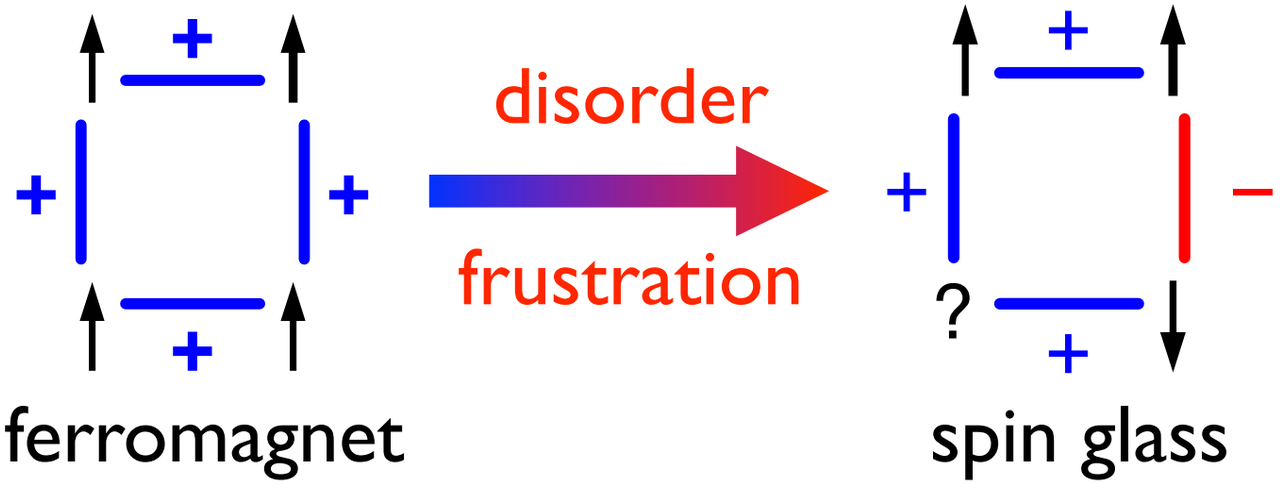
Let's add disorder!

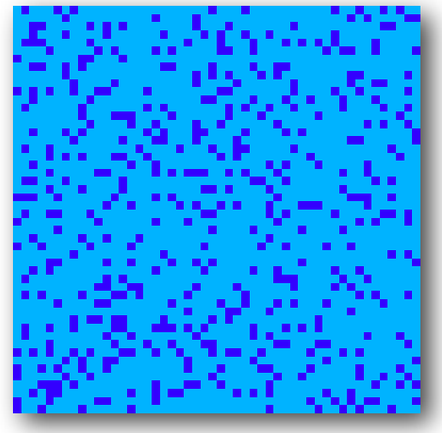
# Cooking up a horrible optimization problem

- Add disorder… Edwards-Anderson spin glass

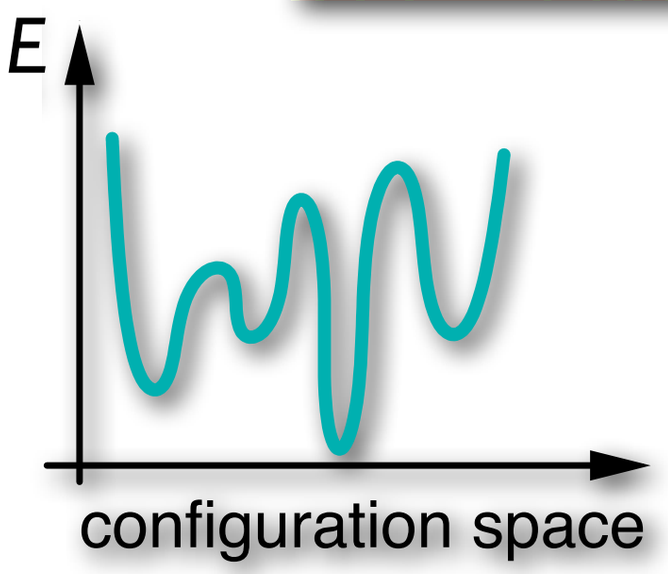
$$\mathcal{H} = -\sum_{ij} J_{ij} S_i S_j - h \sum S_i \qquad J_{ij} \text{ random}$$

- … obtain loads of frustration:

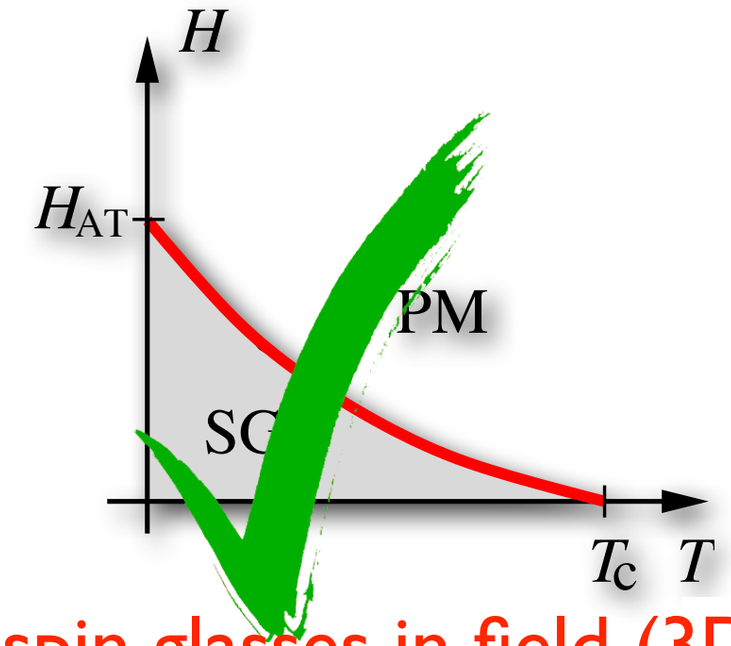


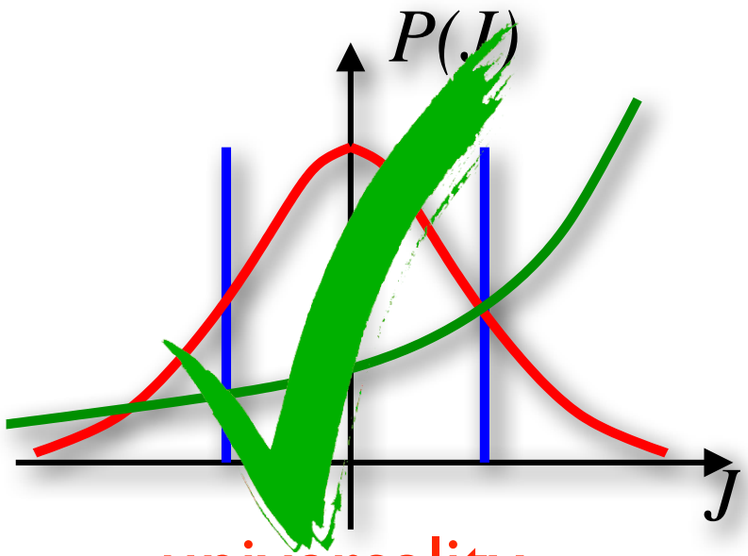
- Many metastable states, slow relaxation
- Rough energy landscape
- Nontrivial memory effects, aging
- Perfect for testing algorithms!



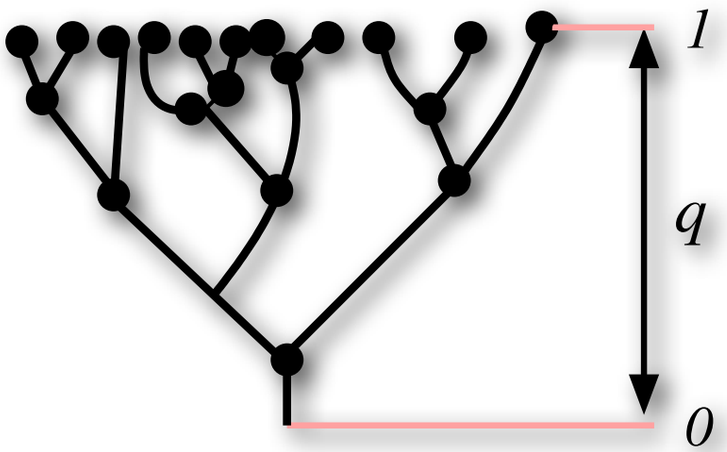
# Still poorly understood: Some challenges…



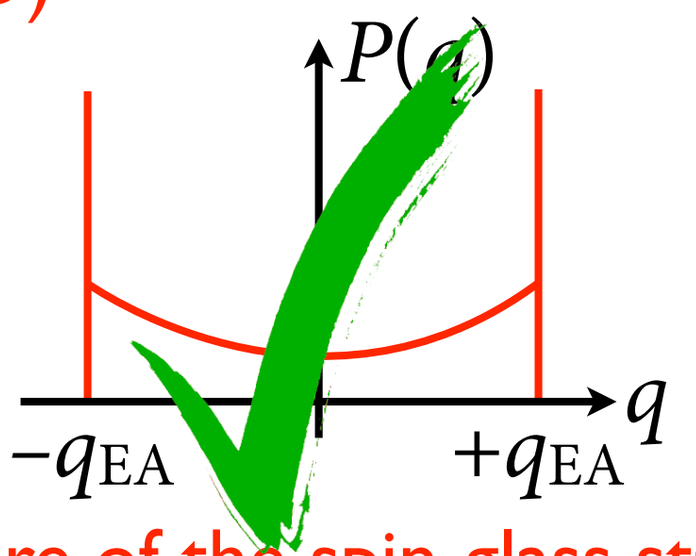

spin glasses in field (3D)



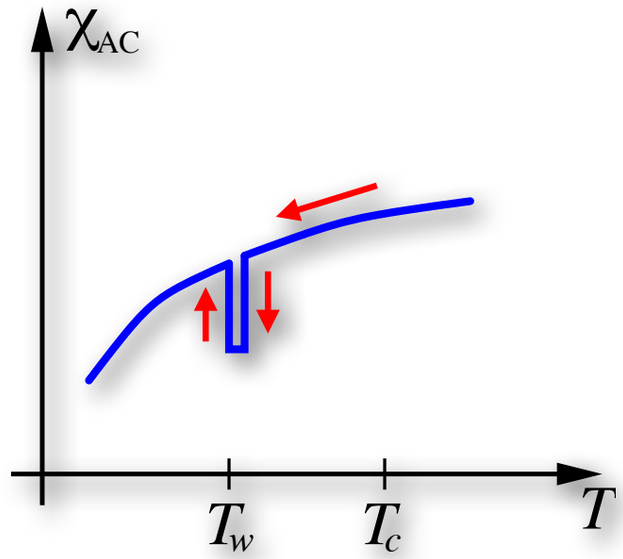

universality




ultrametricity




nature of the spin-glass state

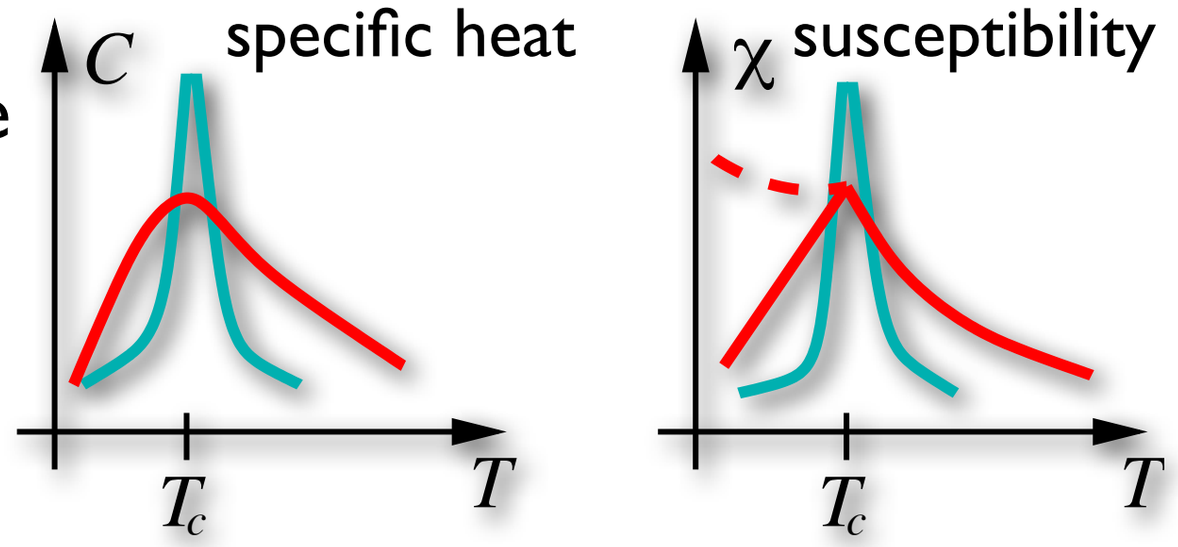



memory effect

# A word on theoretical descriptions…

- Early experimental discovery:
  Canella & Mydosh see a cusp in the susceptibility of a Fe/Au alloy.

$$J_{ij} \sim \frac{\cos(2k_{\mathrm{F}}R_{ij})}{R_{ij}^3}$$

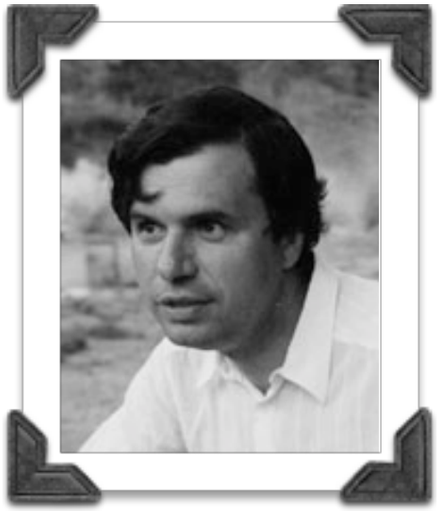



- Brief incomplete history…
  - mid 70's: Edwards-Anderson Ising spin glass model ( $J_{ij}$ random):

$$\mathcal{H} = -\sum_{\langle ij \rangle} J_{ij} S_i S_j \quad \xrightarrow{\text{mean-field approx.}} \quad \sum_{\langle ij \rangle} \rightarrow \sum_{i,j}$$

  - mid 70': Mean-field Sherrington-Kirkpatrick spin glass.
  - 70's: Parisi mean-field solution (replica symmetry breaking - RSB).
  - 80's: Scaling-like droplet picture (DP) for short-range systems.
  - 90's: Chaotic pairs picture (CP) by Newman & Stein.

# A word on theoretical descriptions…



Giorgio Parisi

Daniel Fisher

David Huse

and many more…

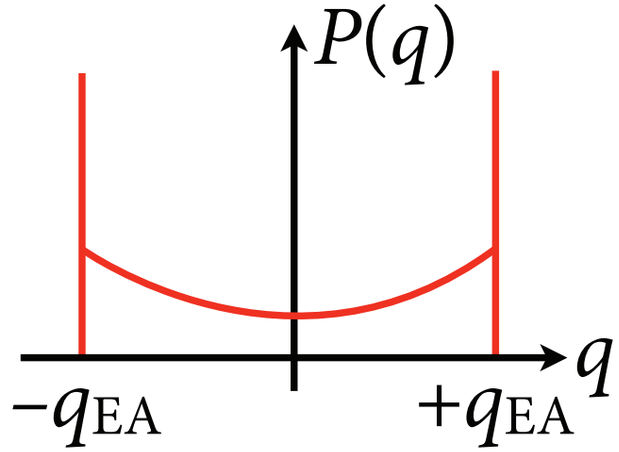
- Brief incomplete history…
  - mid 70's: Edwards-Anderson Ising spin glass model ( $J_{ij}$ random):

$$\mathcal{H} = -\sum_{\langle ij \rangle} J_{ij} S_i S_j \quad \xrightarrow{\text{mean-field approx.}} \quad \sum_{\langle ij \rangle} \rightarrow \sum_{i,j}$$

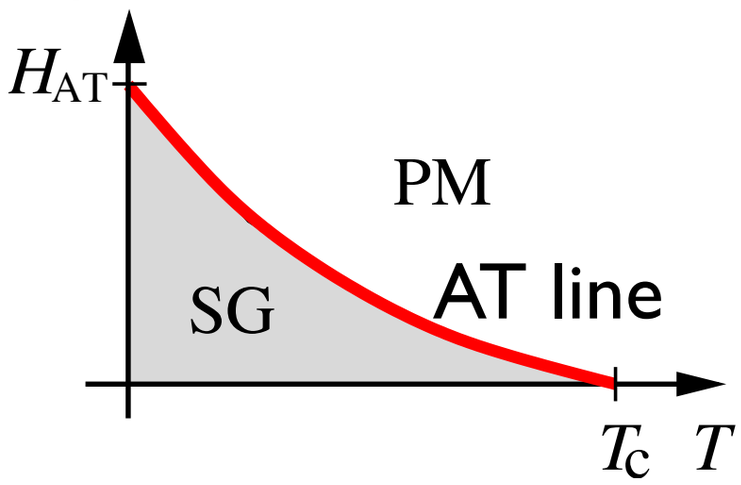
  - mid 70': Mean-field Sherrington-Kirkpatrick spin glass.
  - 70's: Parisi mean-field solution (replica symmetry breaking - RSB).
  - 80's: Scaling-like droplet picture (DP) for short-range systems.
  - 90's: Chaotic pairs picture (CP) by Newman & Stein.

# RSB vs DP

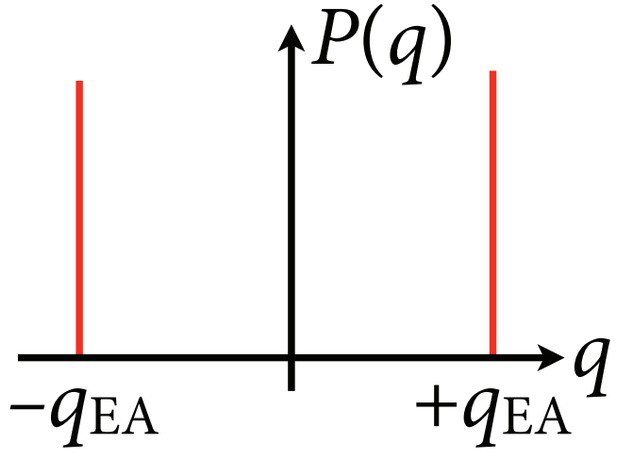
- Countable infinity of pure states in the thermodyn. limit.



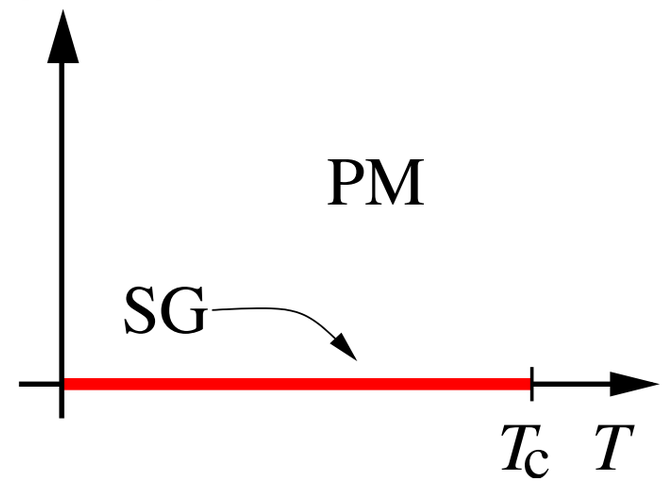

- Nontrivial ground state
- Excitations are space filling with $\Delta E \sim \text{const.}$
- Spin-glass state in a field:




- One pair of pure states in the thermodyn. limit.



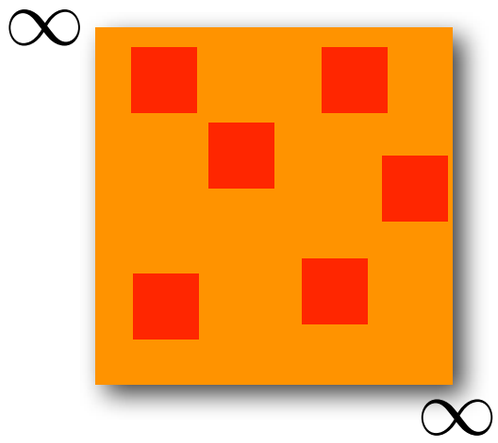

- Trivial ground state
- Excitations are fractal droplets with $\Delta E \sim L^{\theta}$.
- No spin-glass state in a field:



# How can we study these systems?

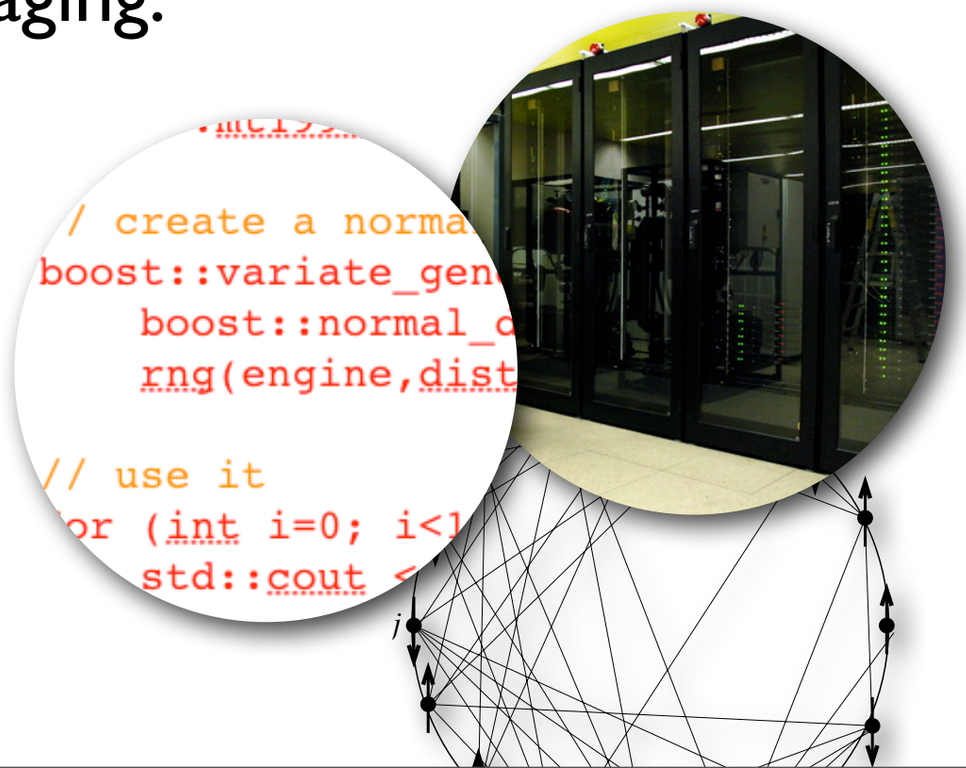
# Computer simulations

- Analytically:
  - Mean field theory: Parisi's RSB
  - Scaling approaches: Droplet-type arguments



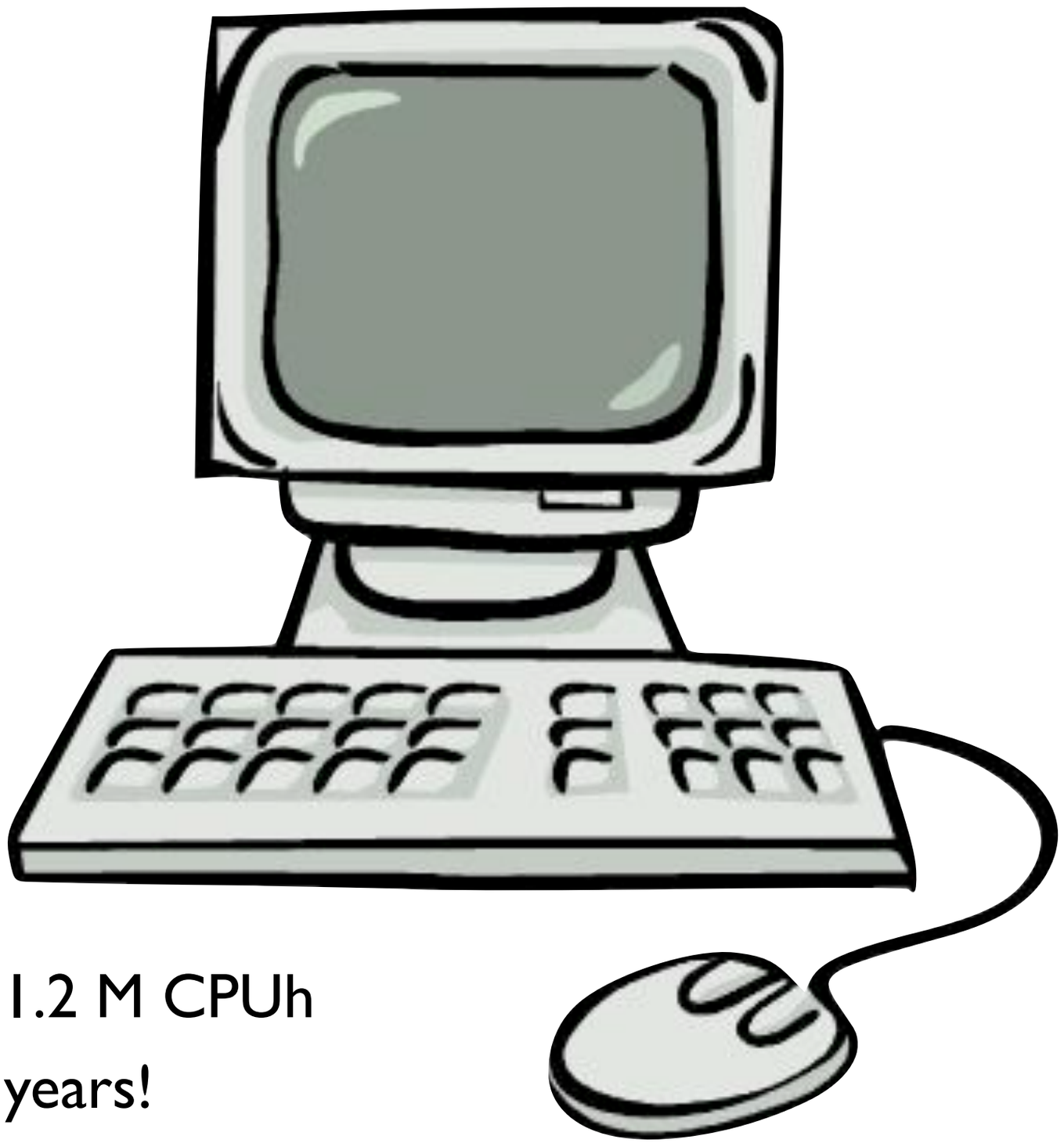

- Numerically many challenges:
  - Exponential number of competing states (usually NP hard).
  - Relaxation times diverge exponentially with the system size.
  - Extra overhead due to disorder averaging.

- What to do?
  - Use large computer clusters.
  - Use better algorithms.
  - Use better models.



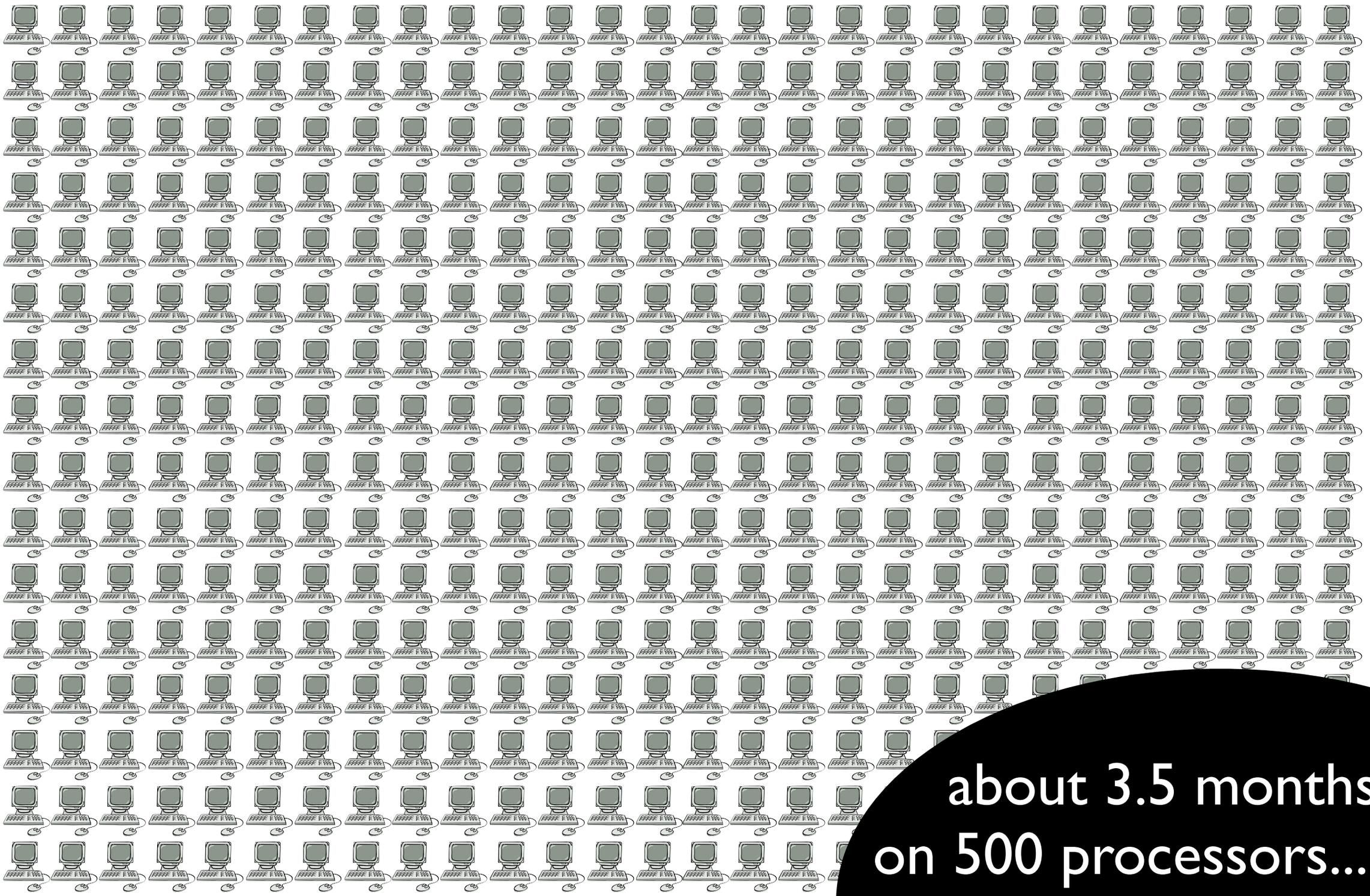
# How CPU intensive are these problems…

# Typical numerical effort



- This project took ~ 1.2 M CPUh
- On one CPU ~ 140 years!

# Typical numerical effort



about 3.5 months
on 500 processors...

# Monte Carlo

# Metropolis Algorithm

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

## Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
*Los Alamos Scientific Laboratory, Los Alamos, New Mexico*

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

### I. INTRODUCTION

THE purpose of this paper is to describe a general method, suitable for fast electronic computing machines, of calculating the properties of any substance which may be considered as composed of interacting individual molecules. Classical statistics is assumed,

### II. THE GENERAL METHOD FOR AN ARBITRARY POTENTIAL BETWEEN THE PARTICLES

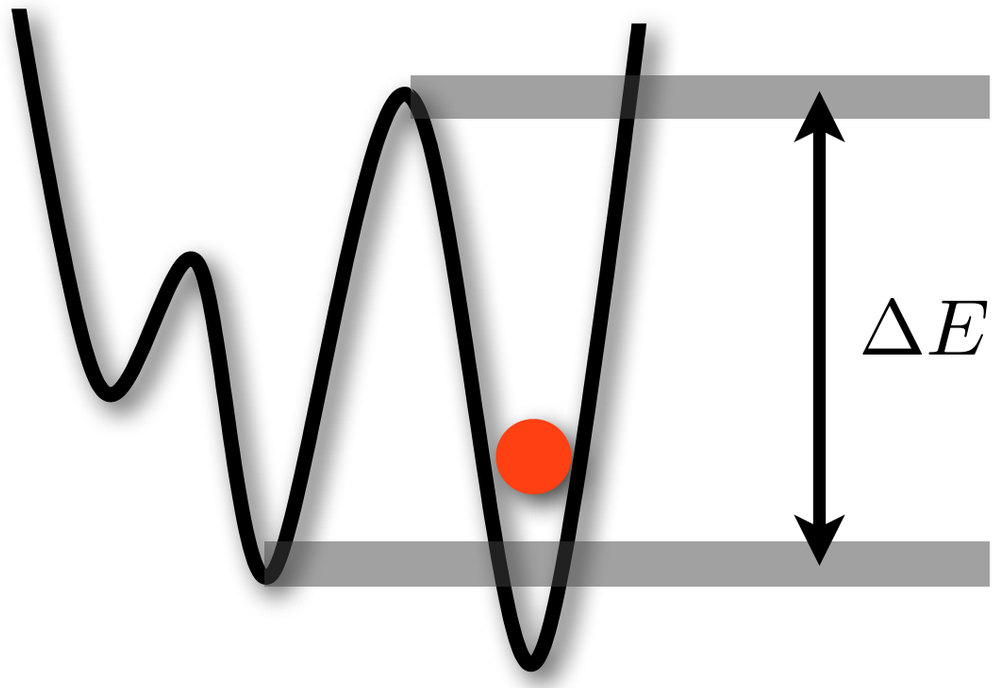
In order to reduce the problem to a feasible size for numerical work, we can, of course, consider only a finite number of particles. This number $N$ may be as high as several hundred. Our system consists of a square† con-

# Efficiently overcoming barriers…

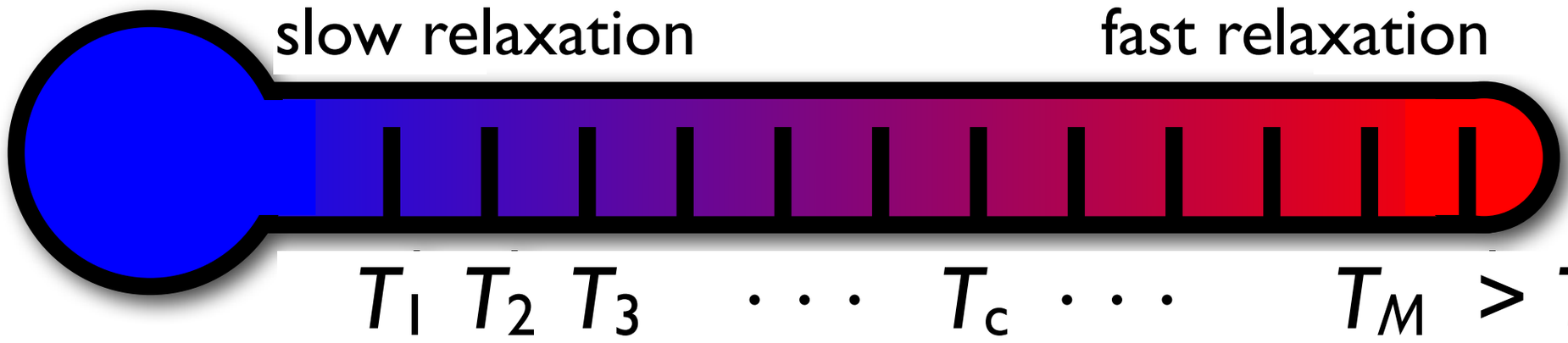
- Simple Monte Carlo:

$$P_{\text{accept}} = \min(1, e^{-\Delta E/T})$$

  - At low *T*, when $\Delta E$ is large, moves are "never" accepted.




- How can we resolve the problem?
  - Parallel tempering (Exchange) Monte Carlo




Hukushima & Nemoto (96)
Geyer (91)

# Algorithm and details

- Brief outline of the algorithm:
  - Perform a Monte Carlo update between *neighboring* replicas:

$$\mathcal{T}[(E_i, T_i) \rightarrow (E_{i+1}, T_{i+1})] = \min\{1, \exp[\Delta E_{i+1,i} \Delta\beta_{i+1,i}]\}$$

$$\Delta E_{i+1,i} = E_{i+1} - E_i$$

[obeys detailed balance]

$$\Delta\beta_{i+1,i} = 1/T_{i+1} - 1/T_i$$

- Practical implementation:

```
algorithm parallel_tempering(*energy,*temp,*spins)

    for(counter = 1 ... (num_temps - 1)) do
        delta = (1/temp[i] - 1/temp[i+1])*(energy[i] - energy[i+1])
        if(rand(0,1) < exp(delta)) then
            swap(spins[i],spins[i+1])
            swap(energy[i],energy[i+1])
        fi
    done
```

# Efficiency example: Ising spin glass in 3D

- Equilibration times:

$$\tau_{\mathrm{eq}}^{\mathrm{PT}} \approx 300\,\mathrm{MCS}$$

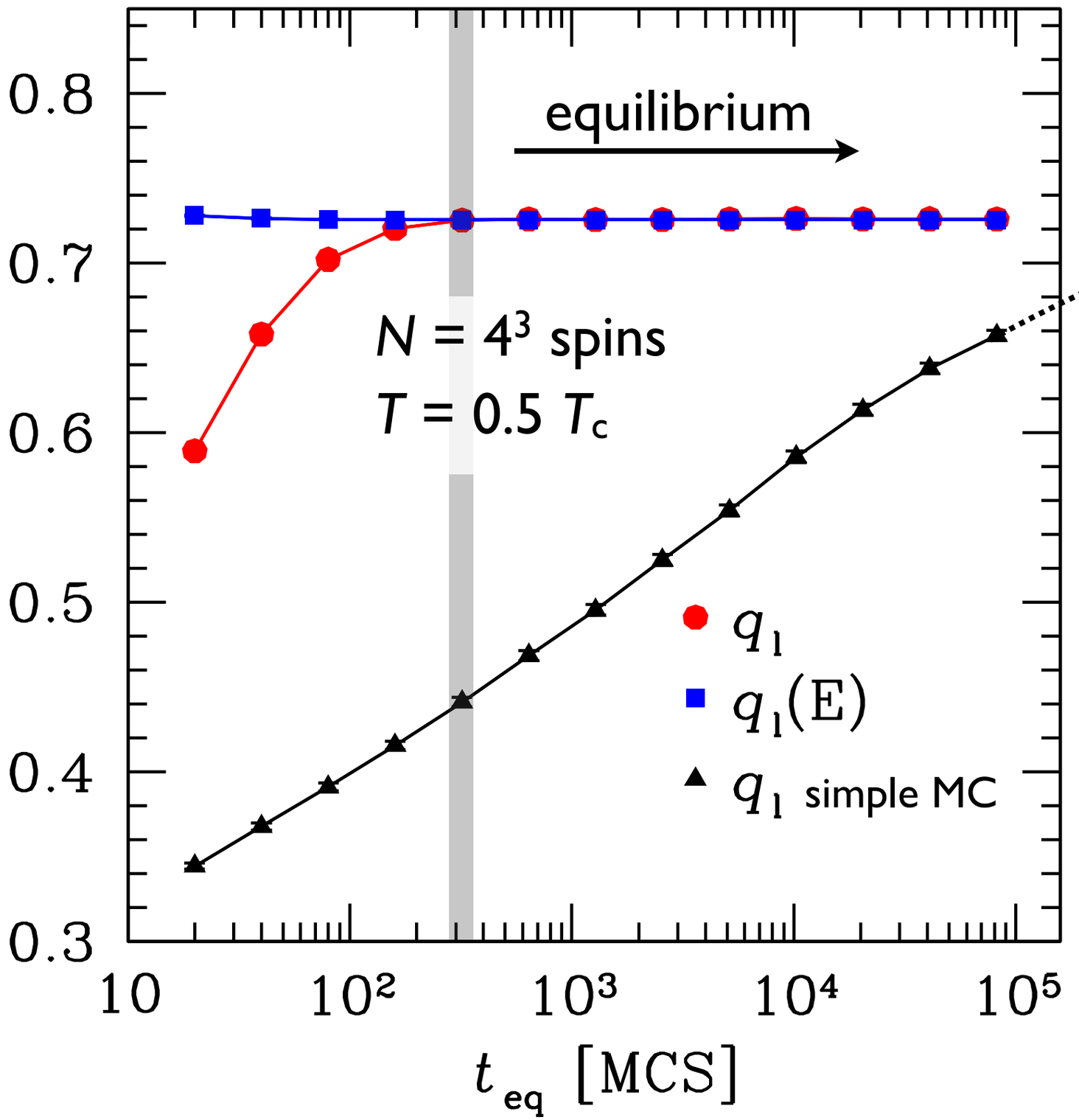
$$\tau_{\mathrm{eq}}^{\mathrm{SM}} \approx 10^{6}\,\mathrm{MCS}$$

- Equilibration test (gaussian disorder):

$$q_l(E) = \frac{2T|E|}{z} + 1$$

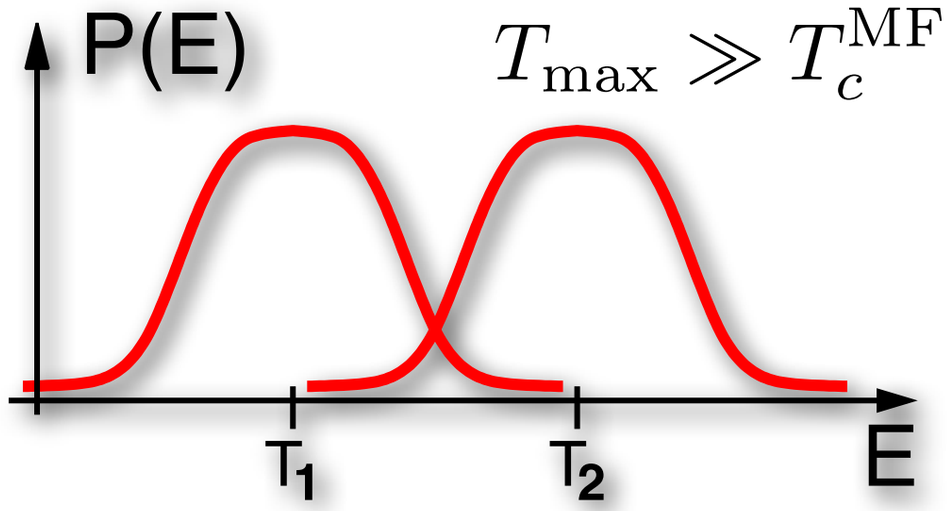
Once both agree, the system is in equilibrium.

Katzgraber *et al.* PRB (01)



# How many temperatures do we need?

- Two possible scenarios:
  - *Temperatures too far apart:*
    parallel simple Monte Carlo chains.
  - *Temperatures too close:*
    massive waste of CPU time.

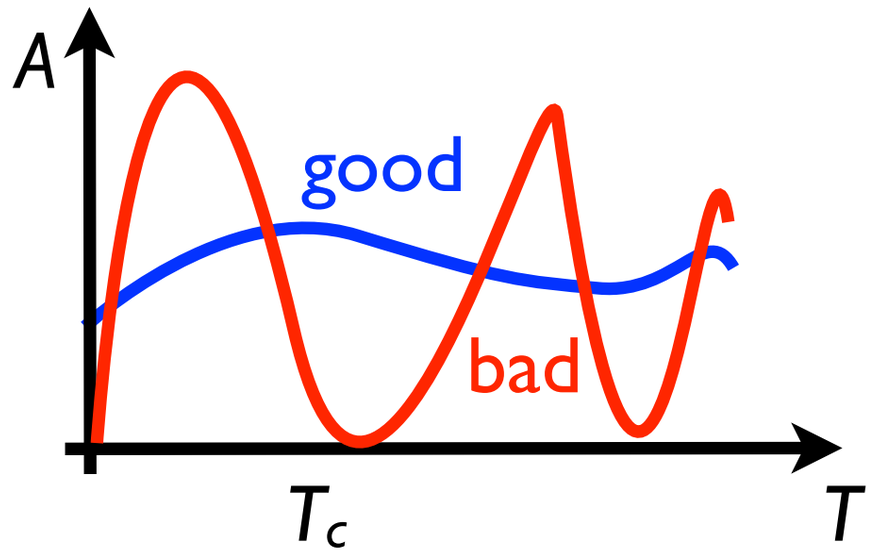



- What determines the number *M* of temperatures?
  - The energy distributions of the system at $T_1$ and $T_2$ have to overlap.
  - Because $\Delta E \sim C_{\mathrm{V}} \longrightarrow M \sim \sqrt{N}$
  - In principle, we need as many temperatures such that the method works. Practical measure? *Acceptance rates*.

# Acceptance rates

- Definition:

$$A = \frac{N_{\mathrm{accept}}}{N_{\mathrm{trial}}}$$




Rathore *et al.*, J. Chem. Phys. (05)

- Traditional wisdom: Tune the temperature set such that...
  - ... $0.2 \leq A \leq 0.9$ ($A \sim 0.3$ optimal)
  - ... $A$ is approximately independent of temperature.

- How do we select the temperatures?
  - *By hand*: A quick run produces stable rates.
  - *Incomplete beta function law*: Detailed implementation which gives flat acceptance rates: [uses $A = f(C_{\mathrm{V}})$]. Predescu *et al.*, JSTAT (03)
  - *Ensemble optimization*: Tune $\{T_i\}$ at a fixed number of $T$'s. Katzgraber *et al.*, JSTAT (06)
  - *Tune* $t_{\mathrm{auto}}$: Perform more PT moves where $t_{\mathrm{auto}}$ is large. Bittner *et al.* (08)

# Ruben method when *Cv* is "well behaved"

Andrist (13)

- Generic frustrated systems:
  - Soft 2nd order transition
  - *Cv* rather smooth
  - "Fancy" methods to select temperatures are a waste of time.

$$\mathcal{T} = \min\{1, \exp[\Delta E_{i+1,i} \Delta\beta_{i+1,i}]\}$$

- Efficient approach:
  - Perform a quick run (no need to equilibrate) and estimate $E(T_i)$ .
  - Spline interpolate the discrete set $\{E(T_i)\}$ and estimate $E(T)$.
  - Compute *T*-set via:
    - Choose target acceptance rate $\tilde{A}$.
    - Compute $\beta_2 \ldots \beta_M$ via $\tilde{A} = \exp[\Delta E_{i+1,i} \Delta\beta_{i+1,i}]$.
    - Adjust $\tilde{A}$ until the range $[\beta_1, \beta_M]$ is covered.

$$\beta = 1/T$$

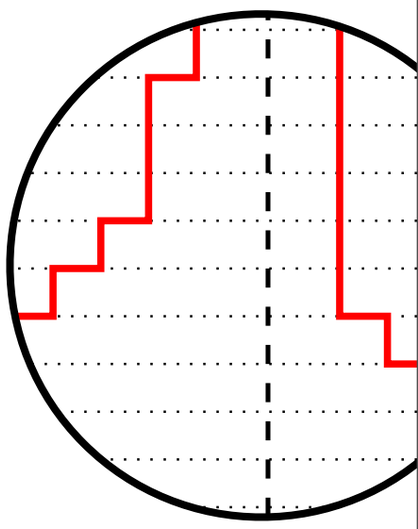
Nonequilibrium:

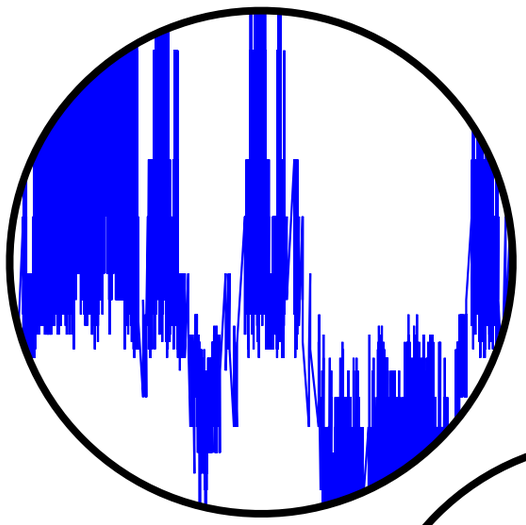
Effect of the energy landscape on the algorithm efficiency

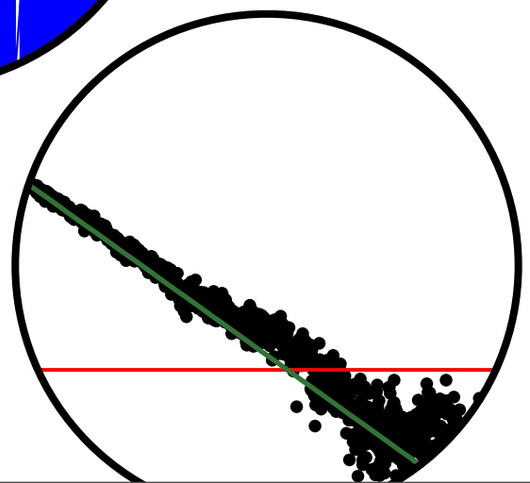
# Motivation

- Questions to be answered:
  - Which model properties correlate with simulation timescales?
  - Which timescale describes thermalization best?
  - What is the sample-to-sample distribution of time scales?
  - How do different time scales correlate?
  - Is it enough to look at the *average* equilibration time?

- Time scales studied:
  - Round-trip time $\tau_{\mathrm{RT}}$
  - Integrated autocorrelation time $\tau_{\mathrm{int}}$
  - Equilibration time $\tau_{\mathrm{eq}}$

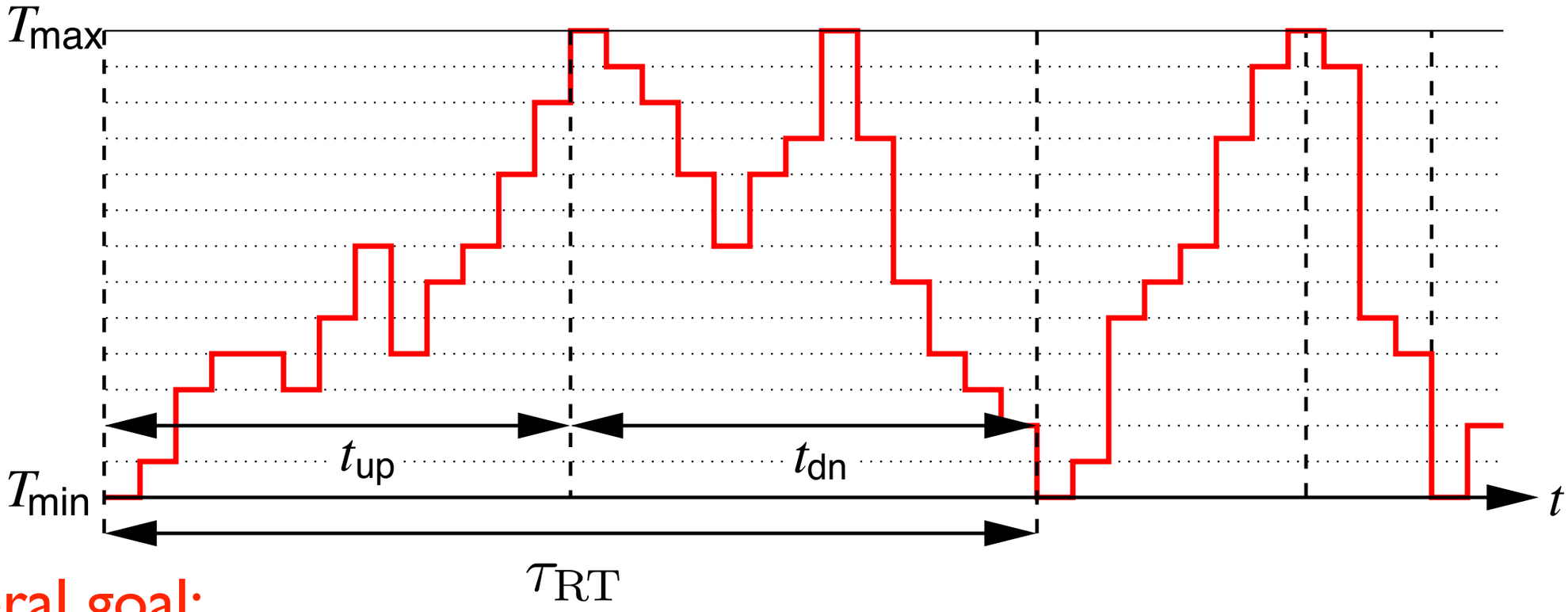






# Round-trip time

- Definition:




- General goal:
  - Minimize the round-trip time and ensure that $t_{\text{up}} \sim t_{\text{dn}}$.
  - This shall ensure an efficient sampling of T-space.
- How?
  - Tune the ensemble $\{T_i\}$ at a fixed number of temperatures $M$.
  - Perform more PT moves where $t_{\text{auto}}$ is large.

Katzgraber *et al.*, JSTAT (06)

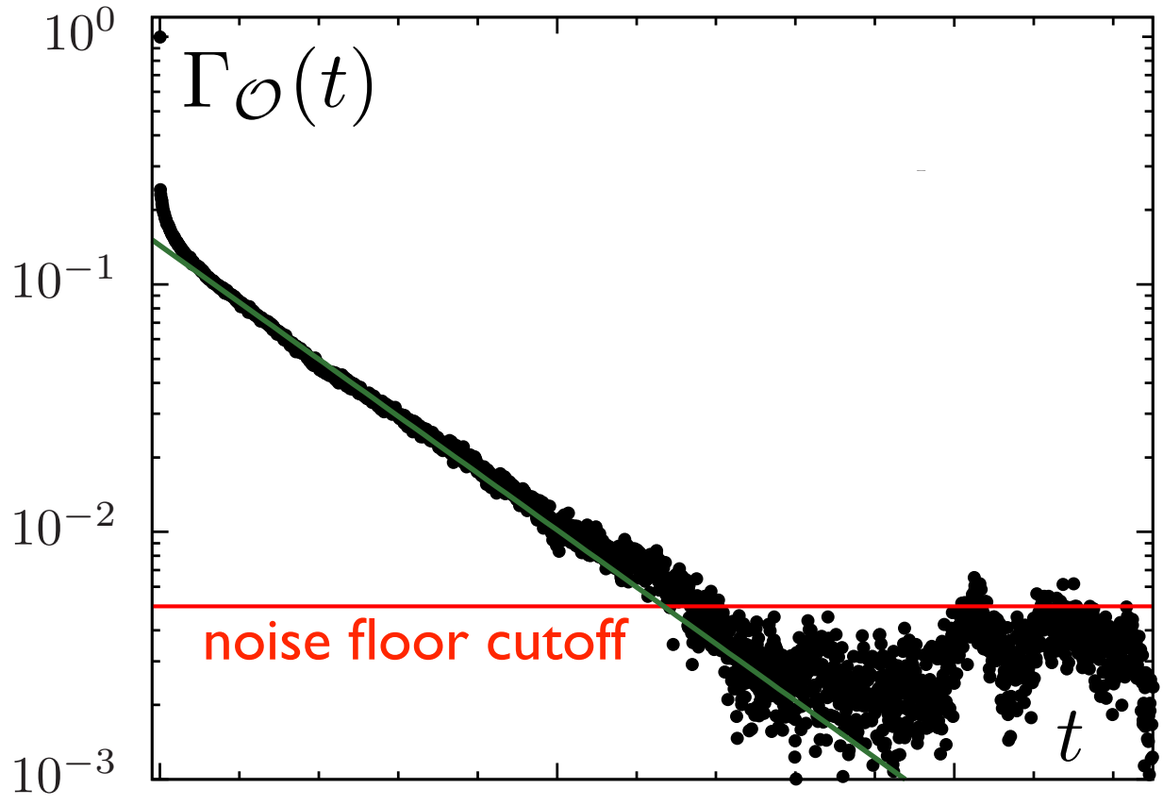
Bittner *et al.* (08)

# Integrated autocorrelation time

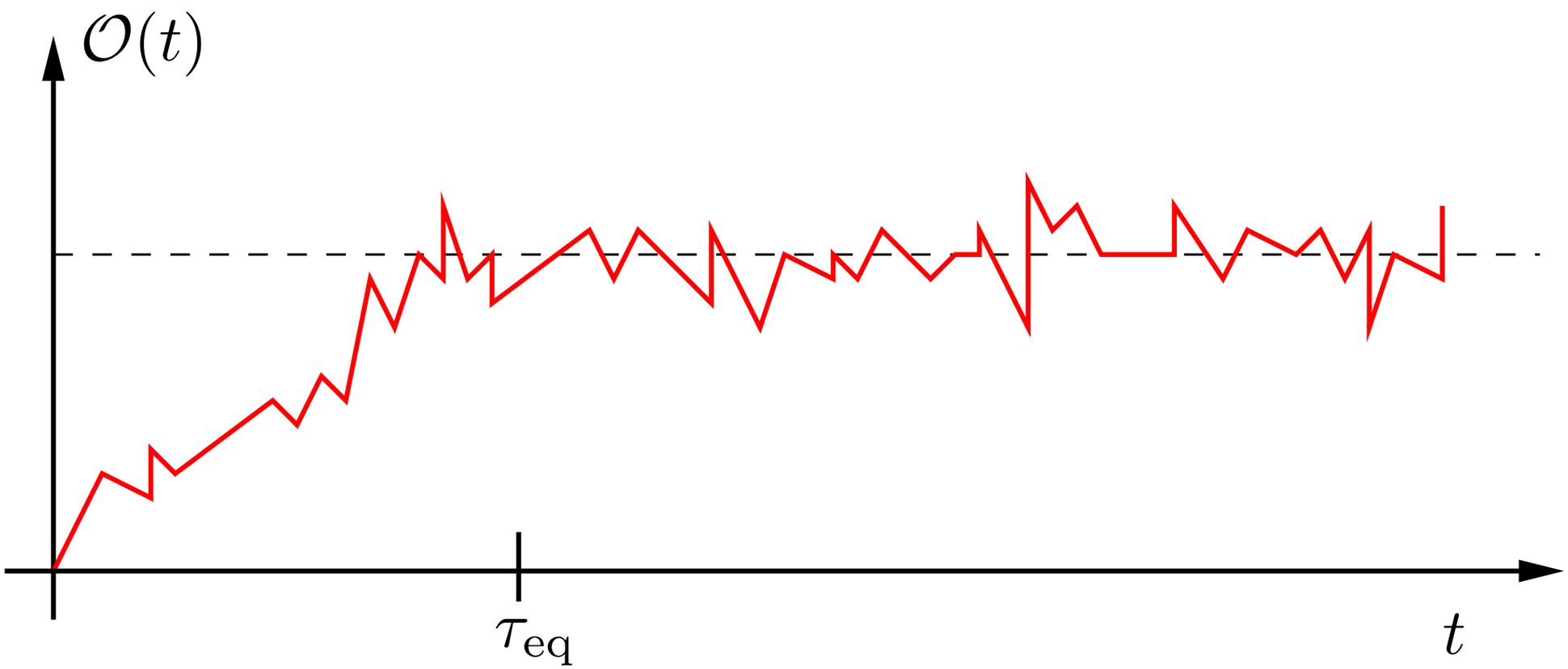
- Definition:
  - Time needed for two subsequent measurements of $\mathcal{O}$ to be decorrelated.
  - Autocorrelation function: $\Gamma_{\mathcal{O}}(t) = \dfrac{\langle \mathcal{O}(0)\mathcal{O}(t)\rangle - \langle \mathcal{O}\rangle^2}{\langle \mathcal{O}^2\rangle - \langle \mathcal{O}\rangle^2} \sim e^{-t/\tau_{\text{exp}}}$
  - Compute the integral:

$$\tau_{\text{int}}^{\mathcal{O}} = \frac{1}{2} + \sum_{t=1}^{\infty} \Gamma_{\mathcal{O}}(t)$$

- Note:
  - Lower bound for equilibration time.
  - We truncate the sum at the noise floor (small error).



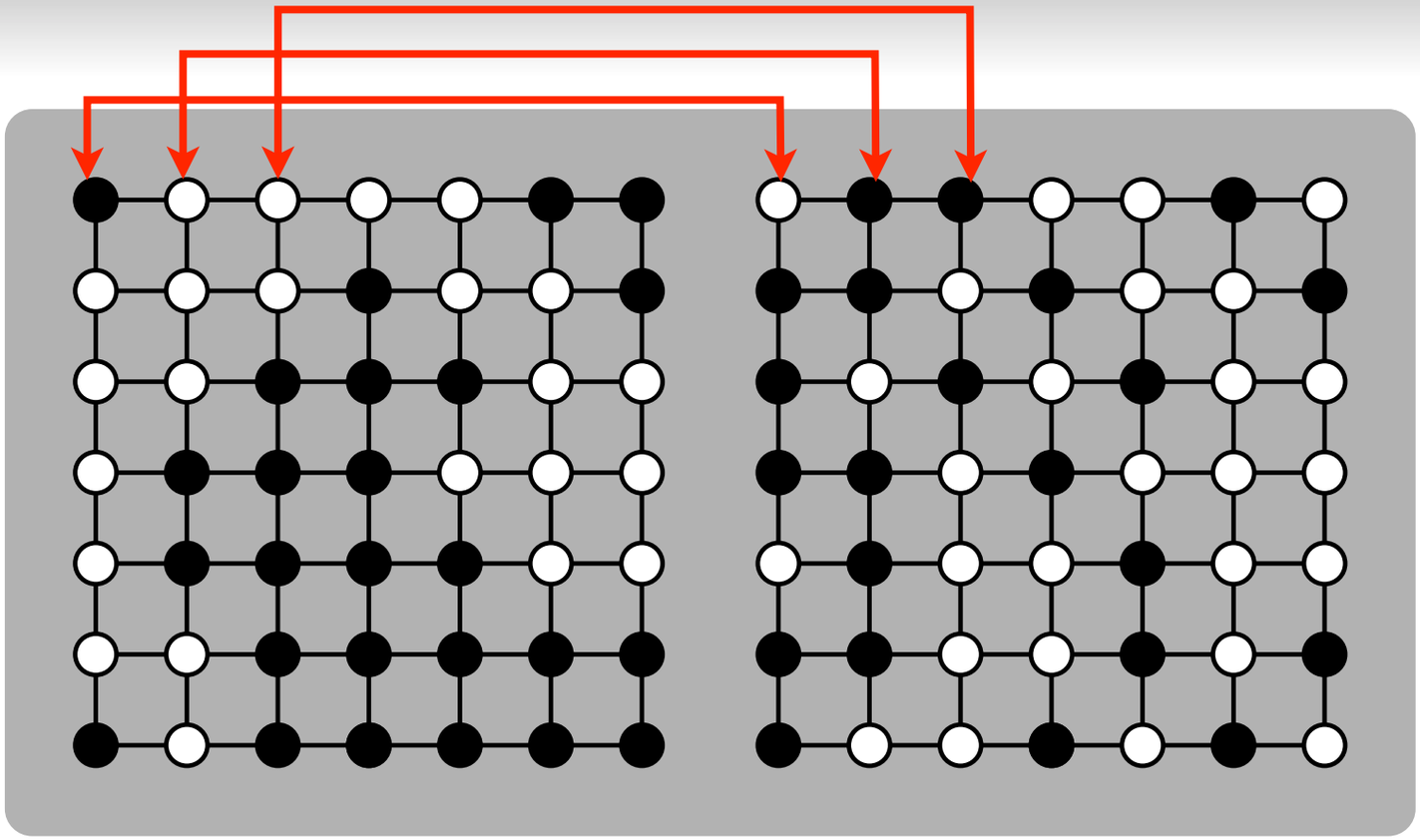
# Equilibration time




- Definition:
  - Time it takes for $\mathcal{O}$ to be independent of MC time *on average*.
  - Note: the energy is not a representative quantity!

# Reminder: How we measure "order"?

- Facts:
  - The ground state has *no spatial order* ($m = 0$).
  - Above $T_c$ spins fluctuate.
  - Below $T_c$ spins freeze.

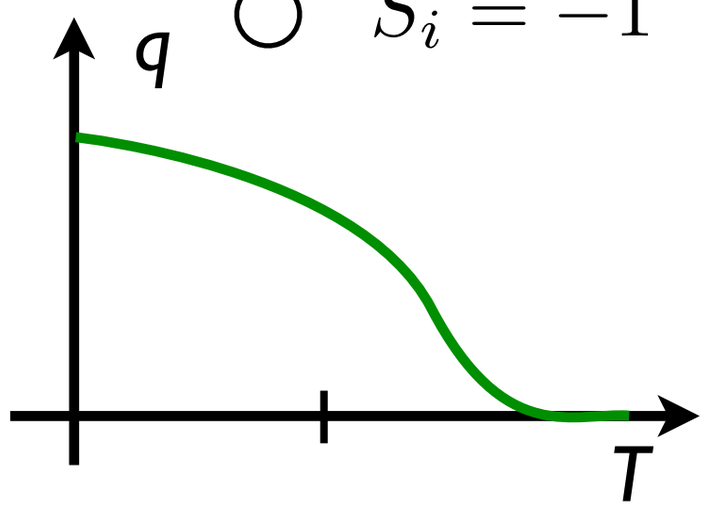



- Solution:
  - Compare spins at time $t_0$ with spins at time $t + t_0$.
  - Not practical in simulations. Better:

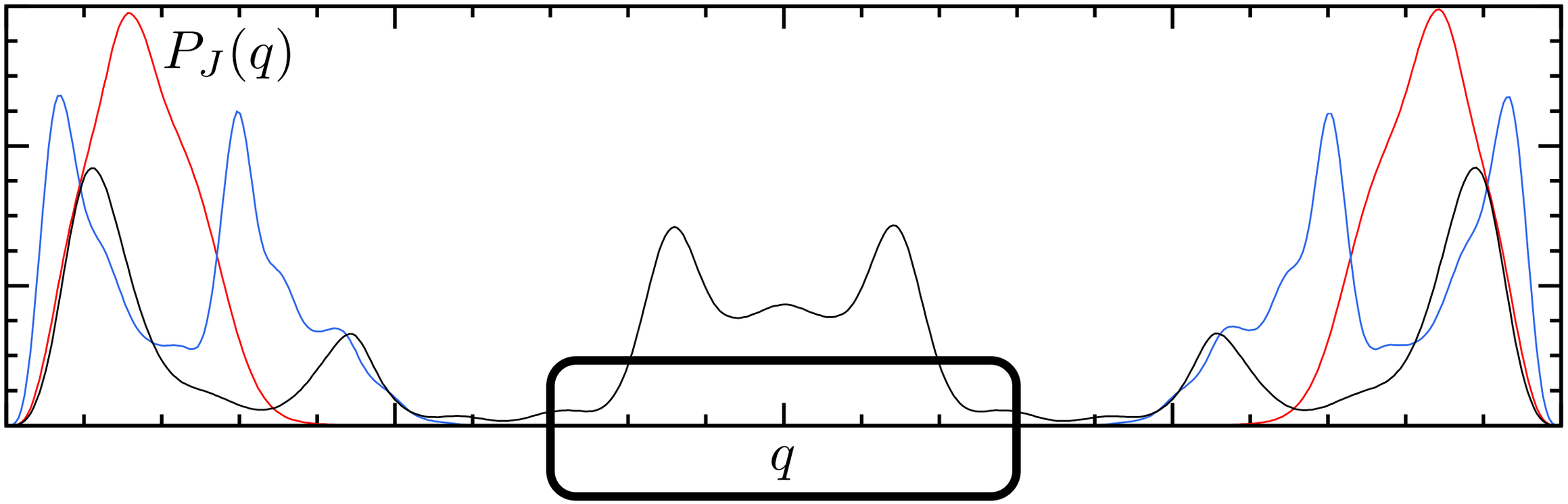
$$q = \frac{1}{N} \sum_{i=1}^{N} S_i^{\alpha} S_i^{\beta}$$ spin overlap

$S_i = +1$

$S_i = -1$



# Proxy towards complexity

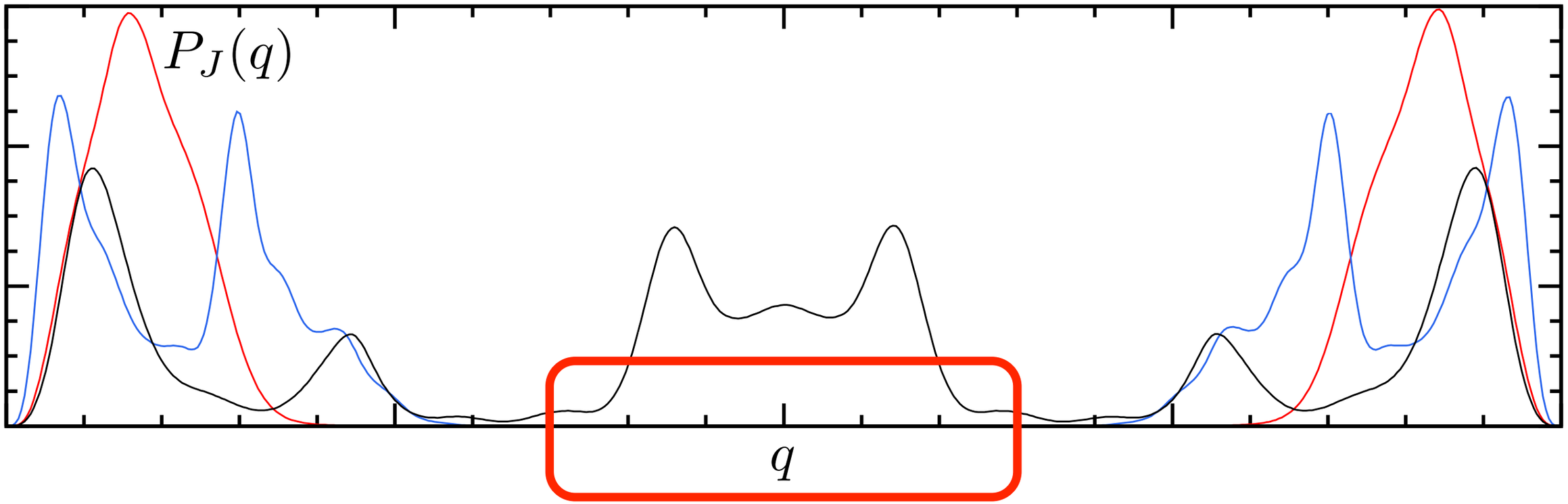



$$P(|q| \leq q_0) > 0$$

- Number of basins in the free-energy landscape correlate with $P_J(q)$.
- Complexity of $P_J(q)$ correlates with:

$$I_J(q_0) = \int_{-q_0}^{q_0} P_J(q)$$

# Proxy towards complexity

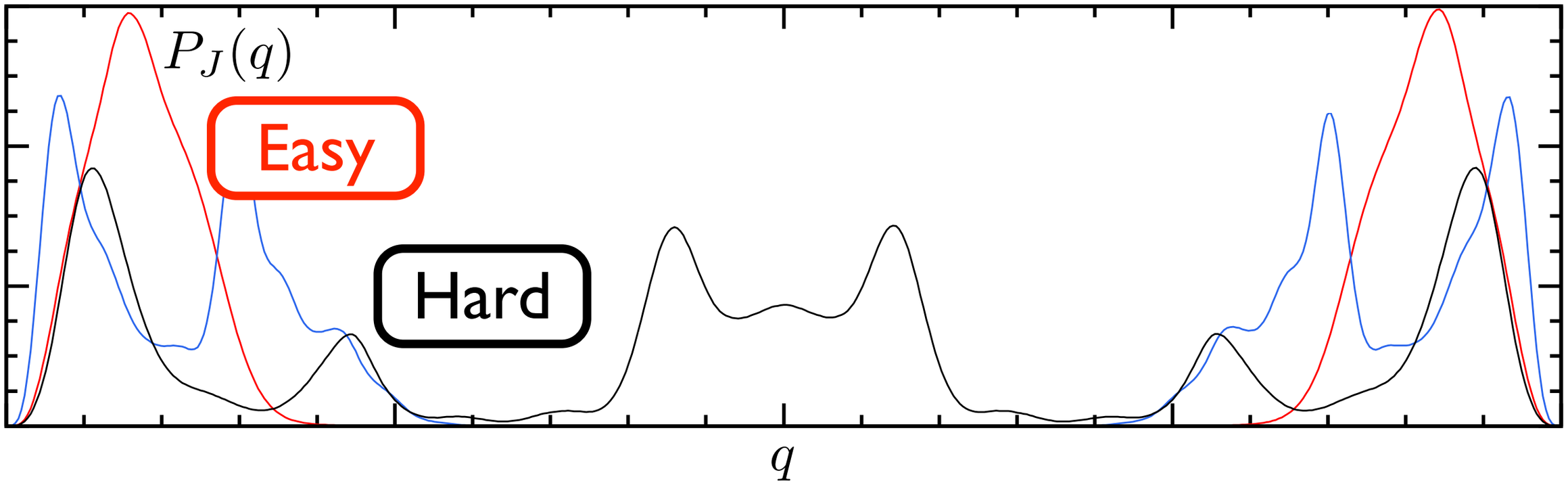



$$P(|q| \leq q_0) \approx 0$$

- Number of basins in the free-energy landscape correlate with $P_J(q)$.
- Complexity of $P_J(q)$ correlates with:

$$I_J(q_0) = \int_{-q_0}^{q_0} P_J(q)$$

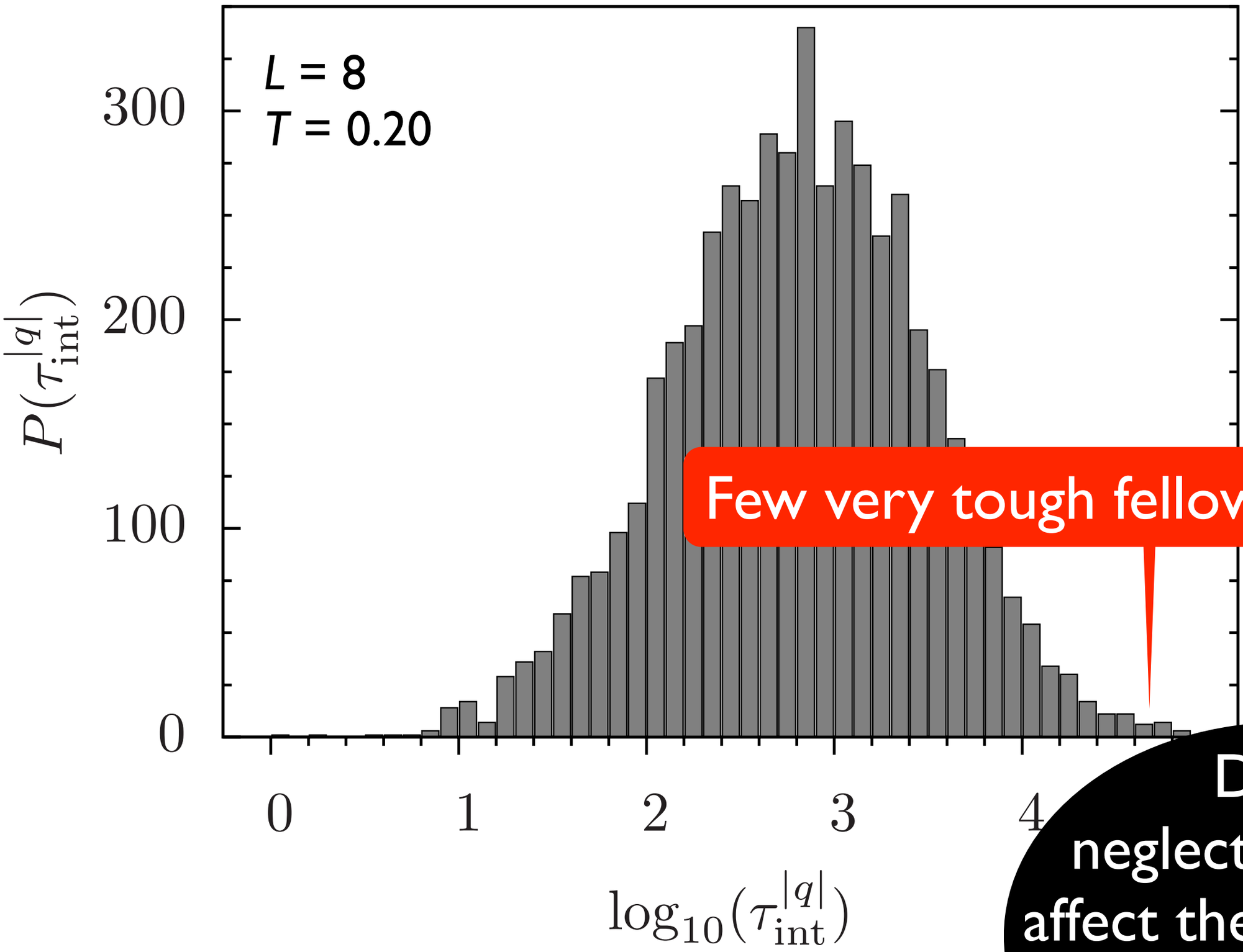
# Proxy towards complexity




- Number of basins in the free-energy landscape correlate with $P_J(q)$.
- Complexity of $P_J(q)$ correlates with:

$$I_J(q_0) = \int_{-q_0}^{q_0} P_J(q)$$

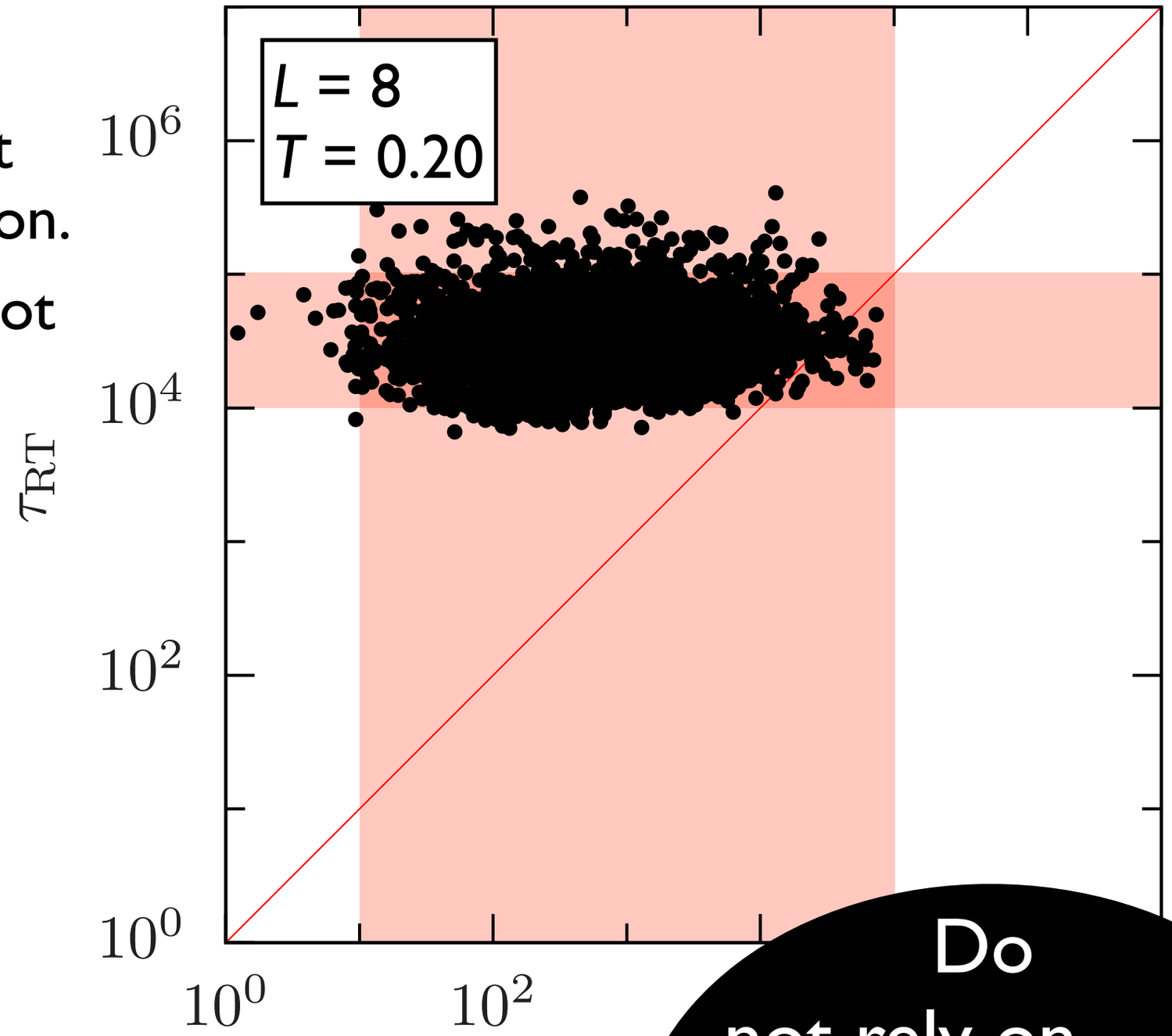
# Distribution of |q| int. autocorrelation times




Does neglecting these affect the physics?

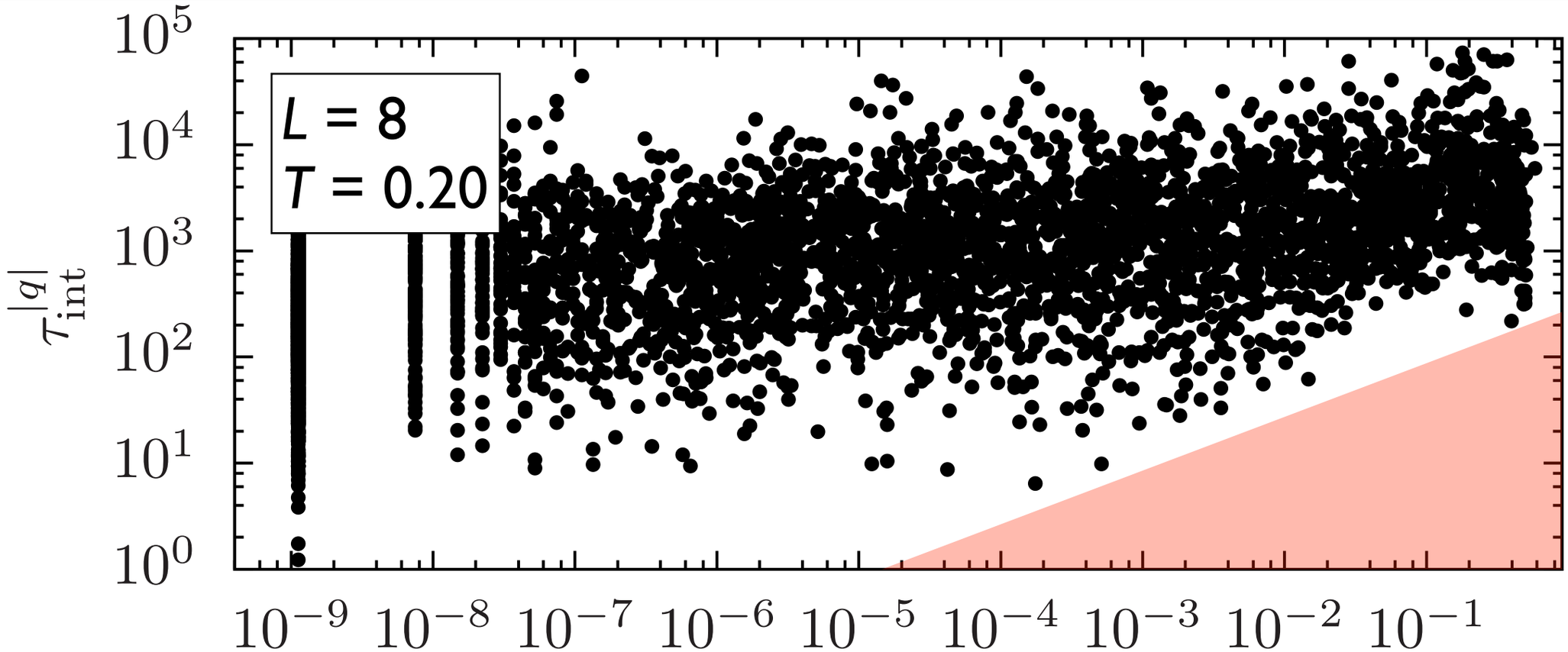
# Round-trip times vs Autocorrelation times

- Broad distribution!
- Full round trips are not needed for decorrelation.
- Round-trip times are not affected by landscape.




Do not rely on round-trip times!

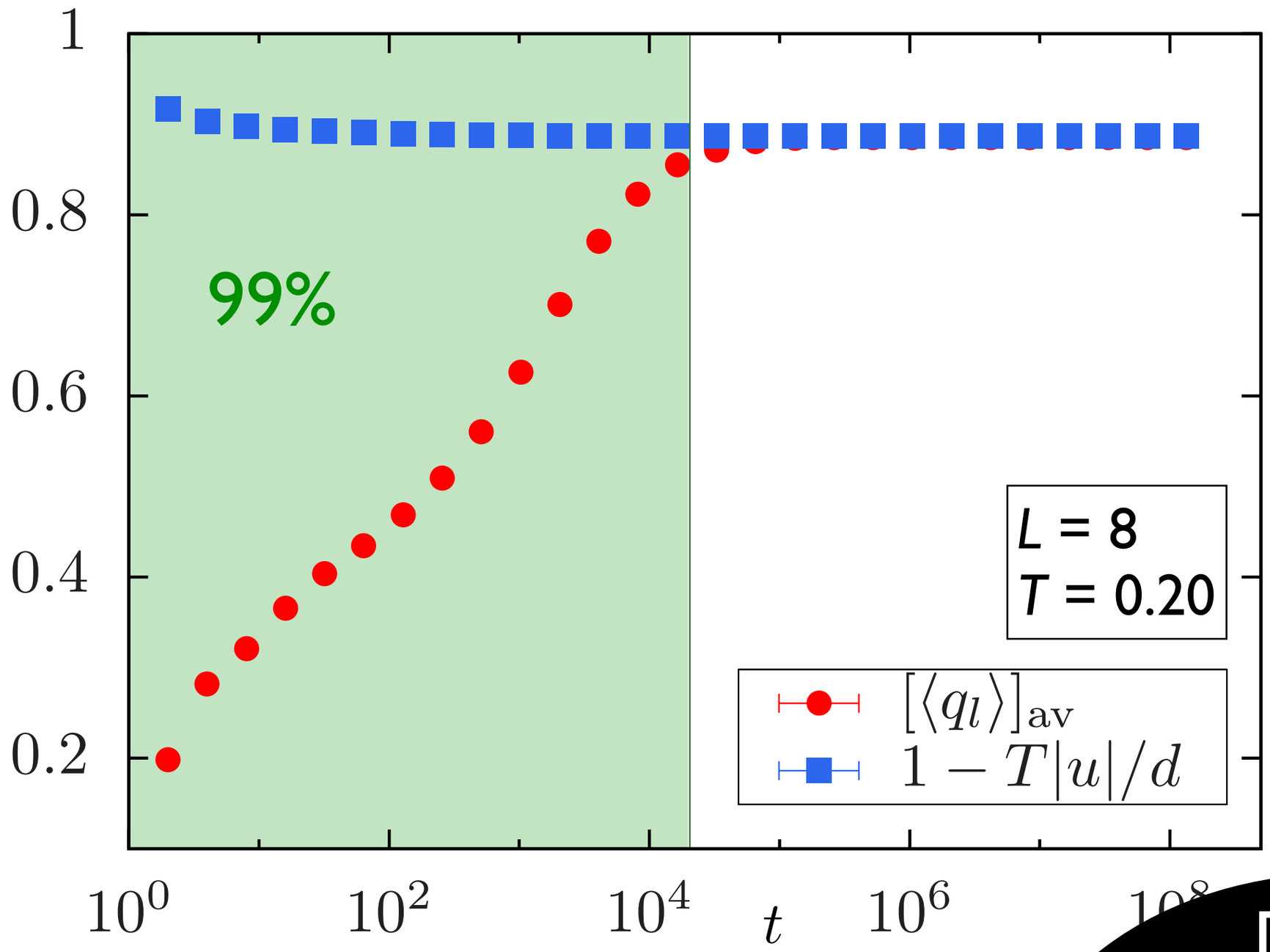
# Correlations with space complexity




$$I_J(q_0) = \int_{-q_0}^{q_0} P_J(q)$$

- Strong correlations between time scales and the overlap complexity.
- Can we (ab)use this to optimize the method?

# Equilibration times



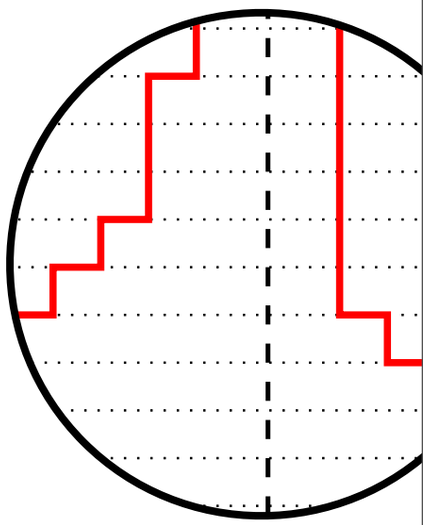

- 99% of $\tau_{\rm int}$ are smaller than – . $\tau_{\rm eq}$ upper bound
- Less than 1% samples not equilibrated after $2^{30}$ M

Does neglecting these affect the physics?

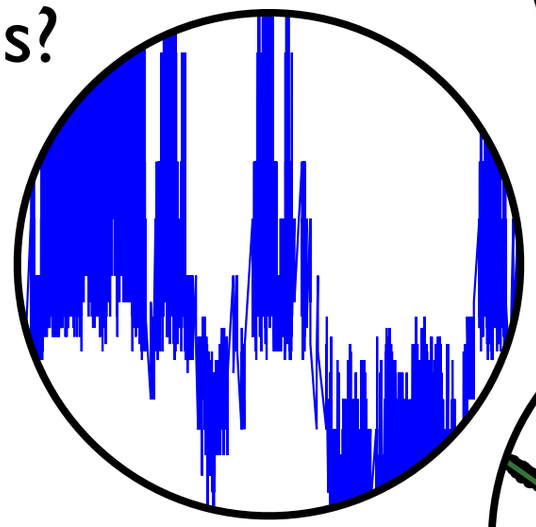
# Back to the questions…

- Which model properties correlate with simulation timescales?
  - Free-energy landscape, except for round-trip times.

- Which timescale describes thermalization best?
  - Autocorrelation and equilibration time.

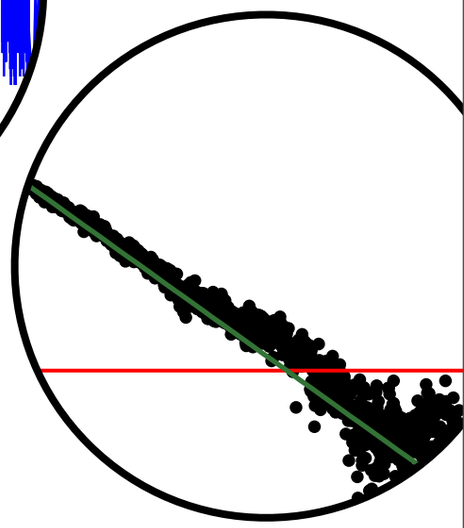


- What is the sample-to-sample distribution of time scales?
  - Very broad! Can we neglect hard samples?



- How do different time scales correlate?
  - Forget about the round-trip times!

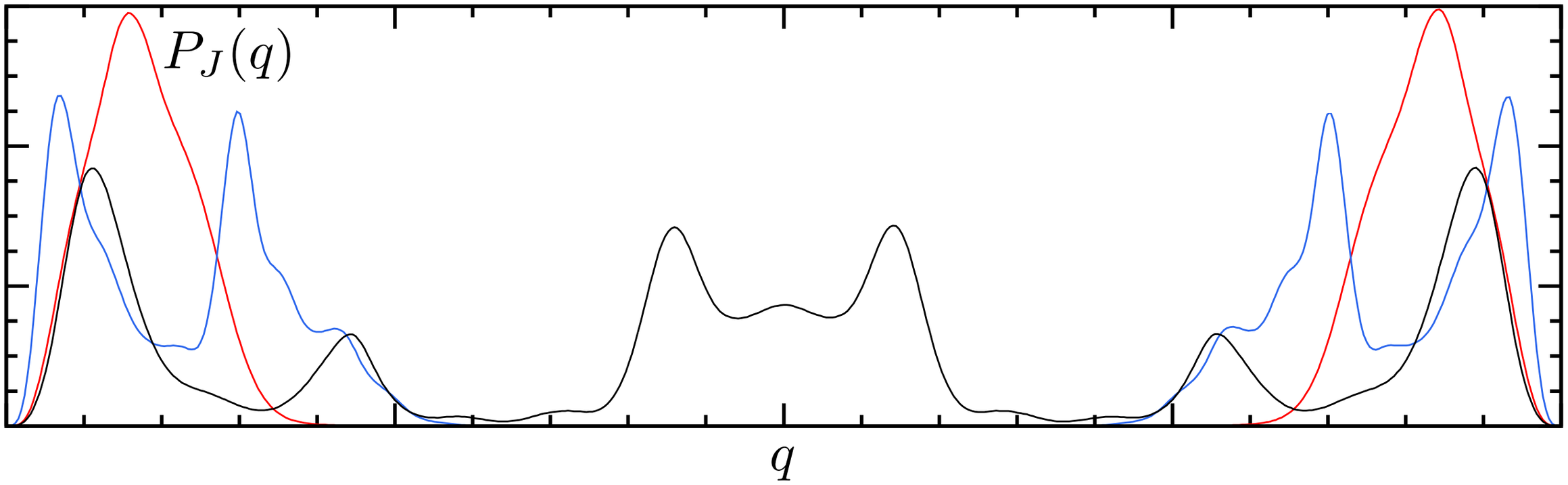
- Is it enough to look at the average equilibration time?
  - Yes, if done very conservatively.



Equilibrium:

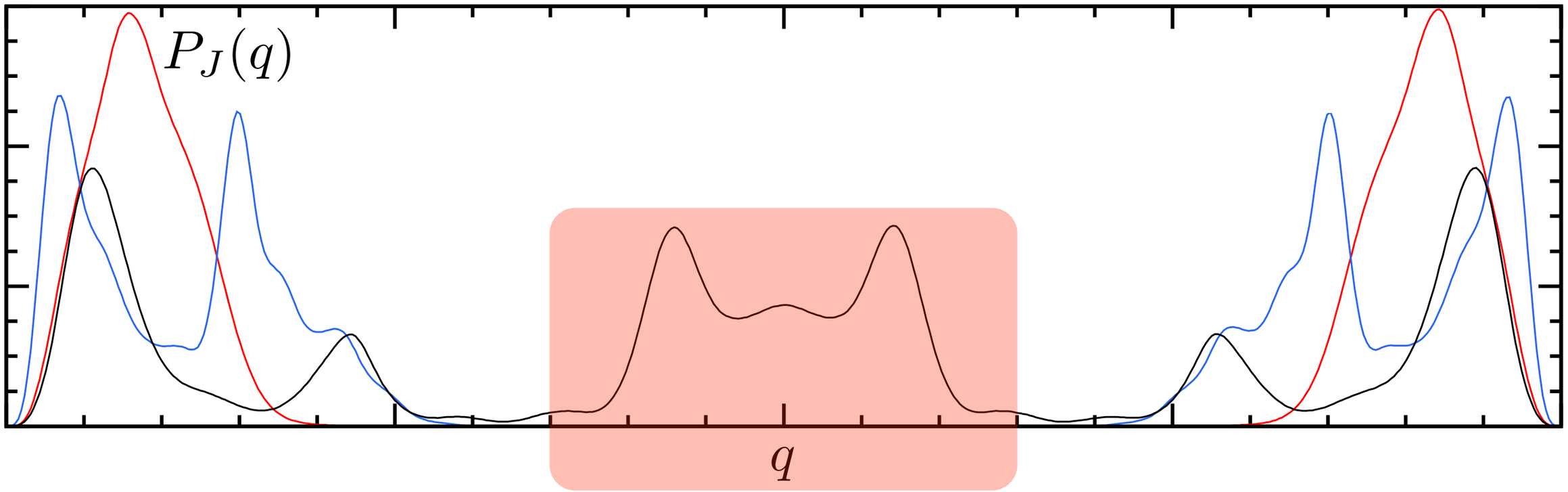
Nature of the spin-glass state

# Overlap distribution $P_J(q)$ for a *given* sample

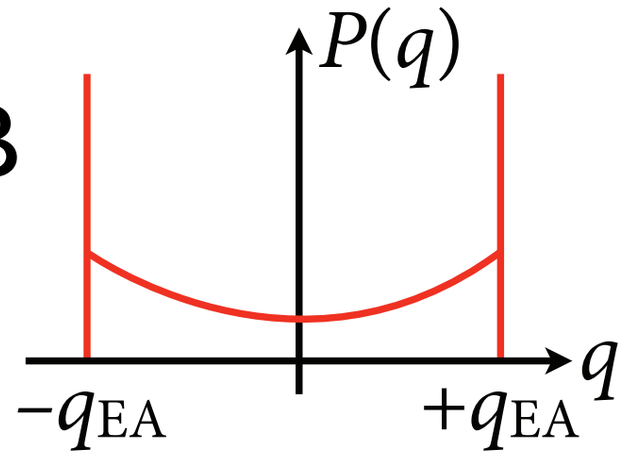



- Remember:
  - $P_J(q)$ is strongly sample dependent. Yucesoy *et al.* (12)
  - The complexity of $P_J(q)$ correlates with algorithmic time scales.

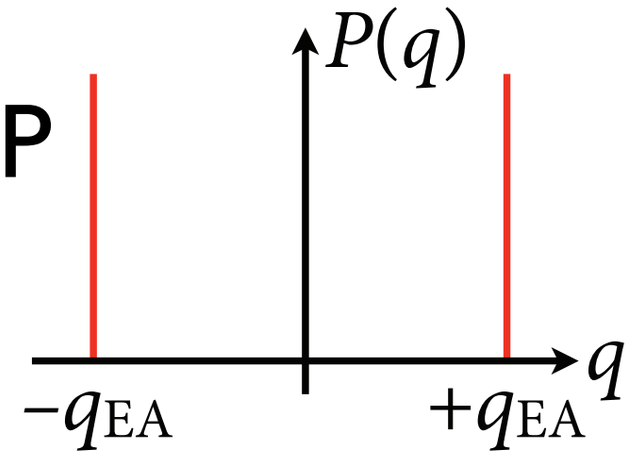
# Overlap distribution $P_J(q)$ for a *given* sample




RSB



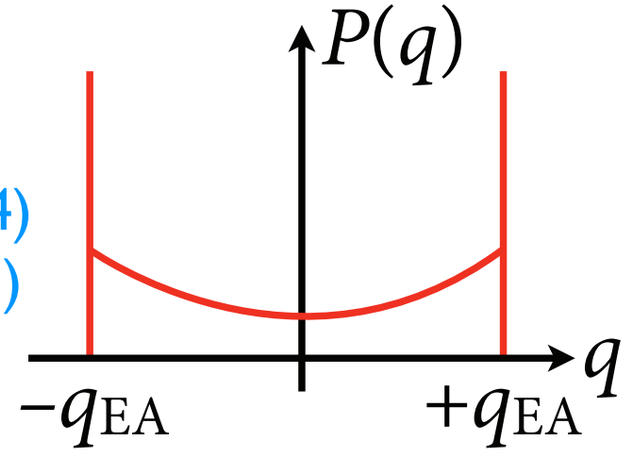

DP or CP




- Question:
  - Does the weight around $q = 0$ vanish after a disorder average?

# What has been done?

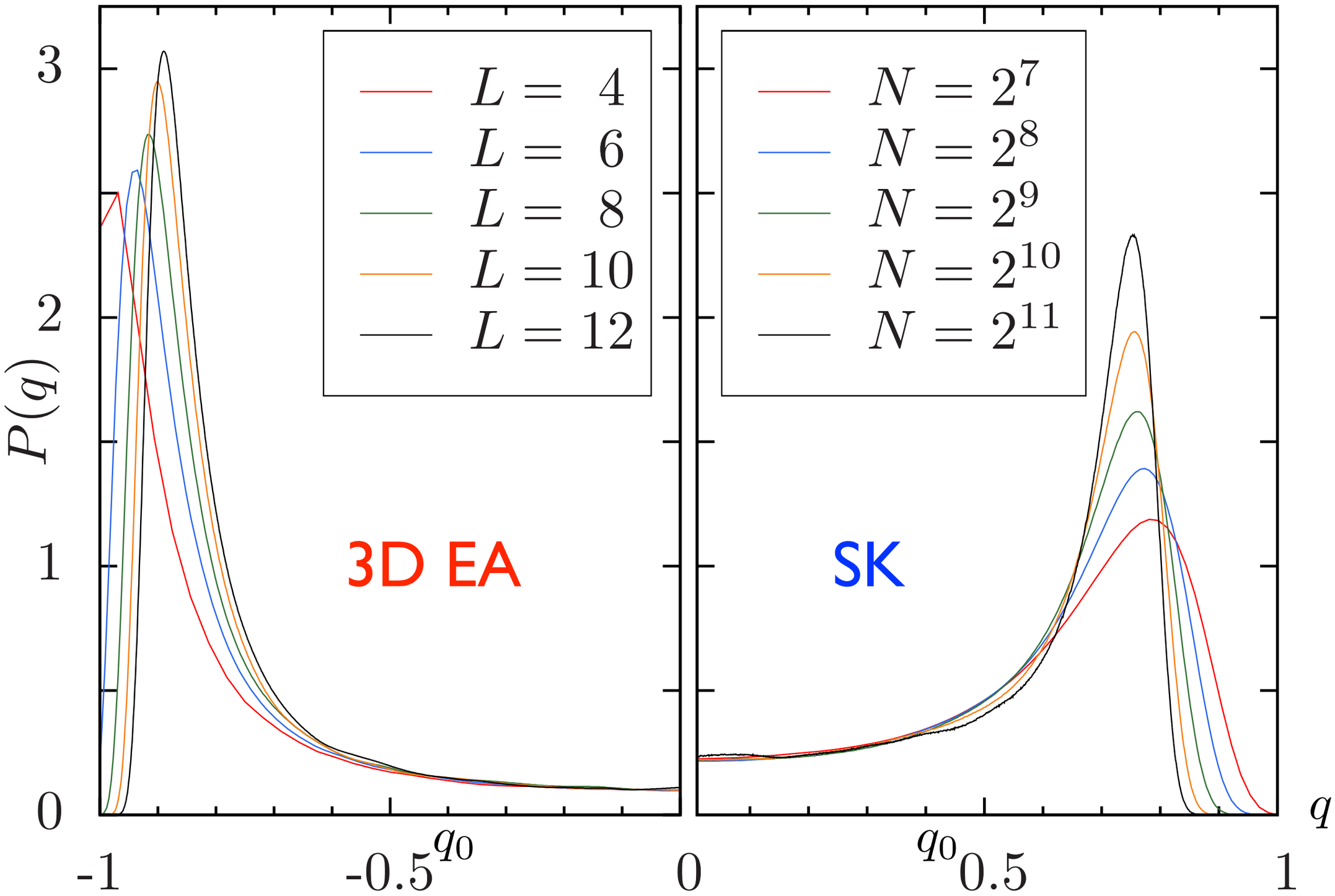
- Apparent agreement with RSB in 3D:
  - Study of the order parameter distribution $P(q)$ suggests many pure states. Reger (90), Marinari (94) Katzgraber *et al.* (01, 03)



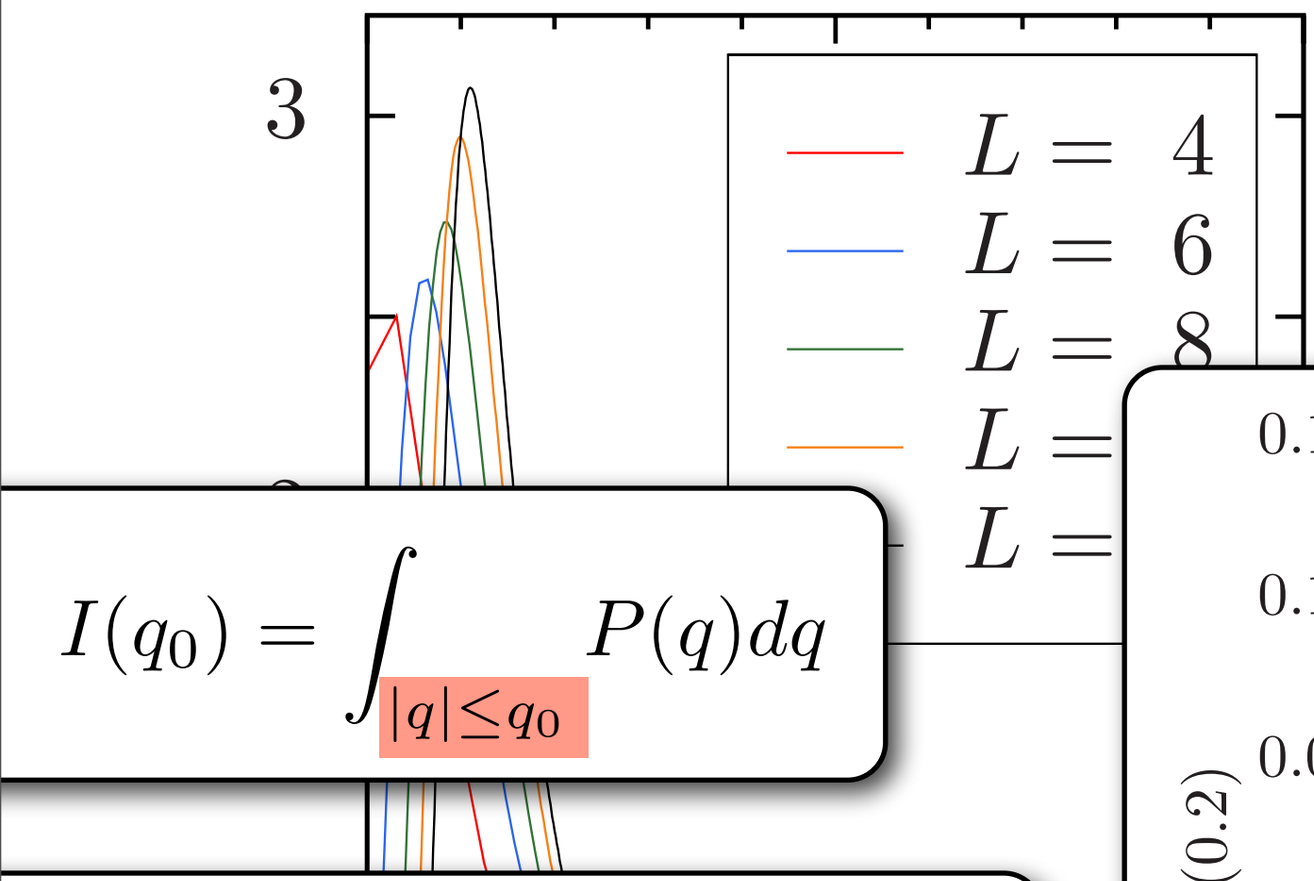

- Indications for intermediate (TNT) scenario in 3D:
  - Study of the spin & link overlap. Palassini & Young (00) Krzakala & Martin (00) Katzgraber *et al.* (01, 03)
  - Many pure states, however…
  - … excitations are fractal (agreement with DP).

- Problems:
  - Is $P(q)$ a sensitive measure?
  - Are the systems large enough? Even in 1D simulations? Katzgraber *et al.* (03)

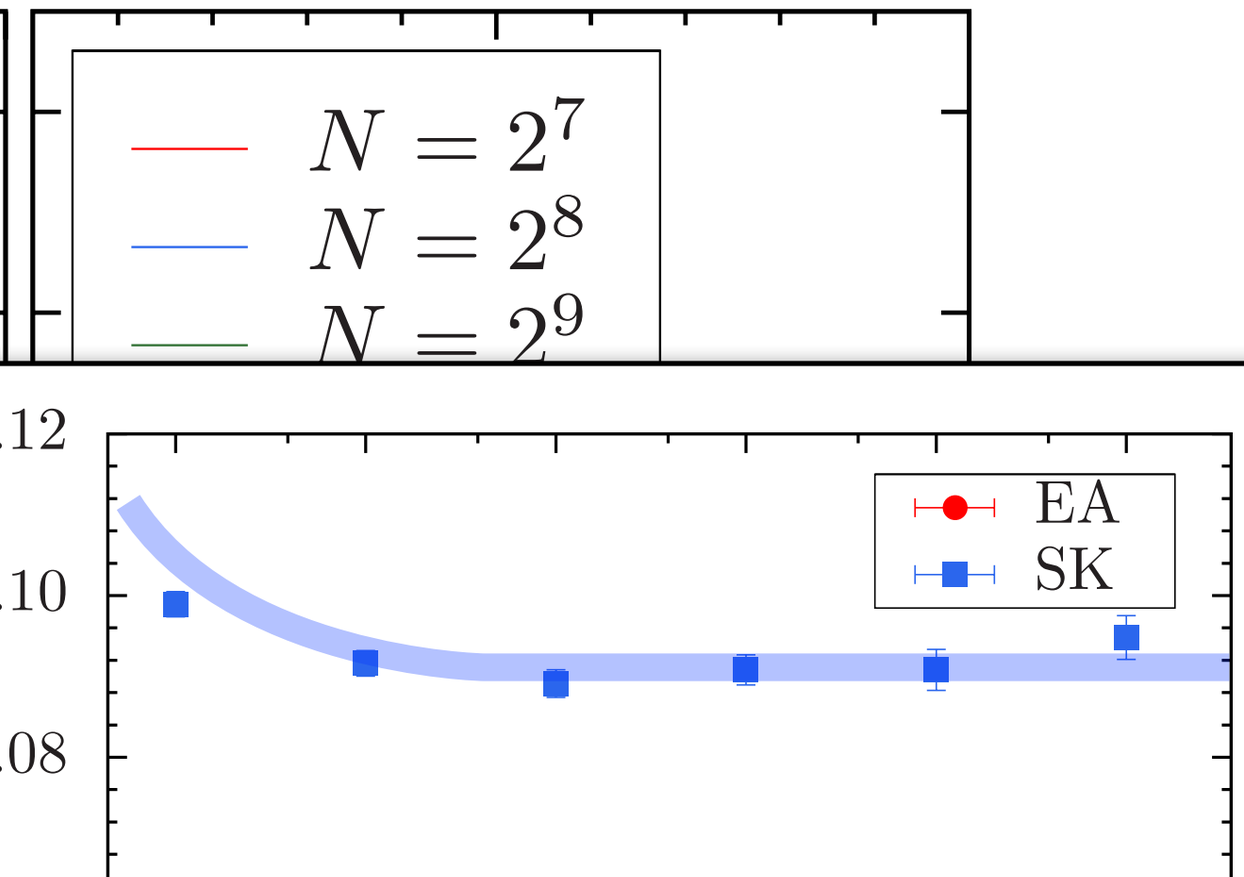
# Traditional approach: Study $P(q \sim 0)$



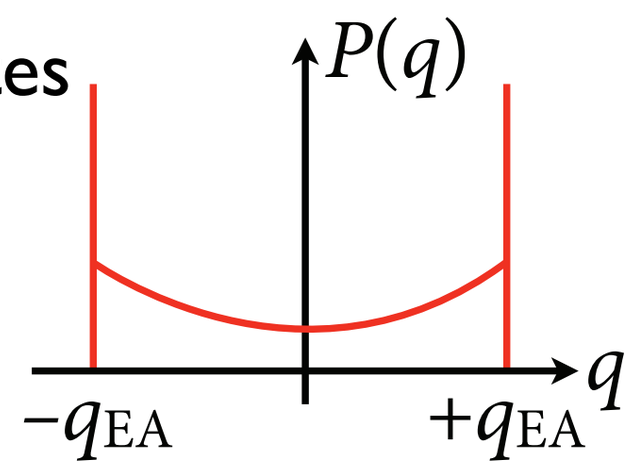
# Traditional approach: Study $P(q \sim 0)$




$$I(q_0) = \int_{|q| \le q_0} P(q)dq$$



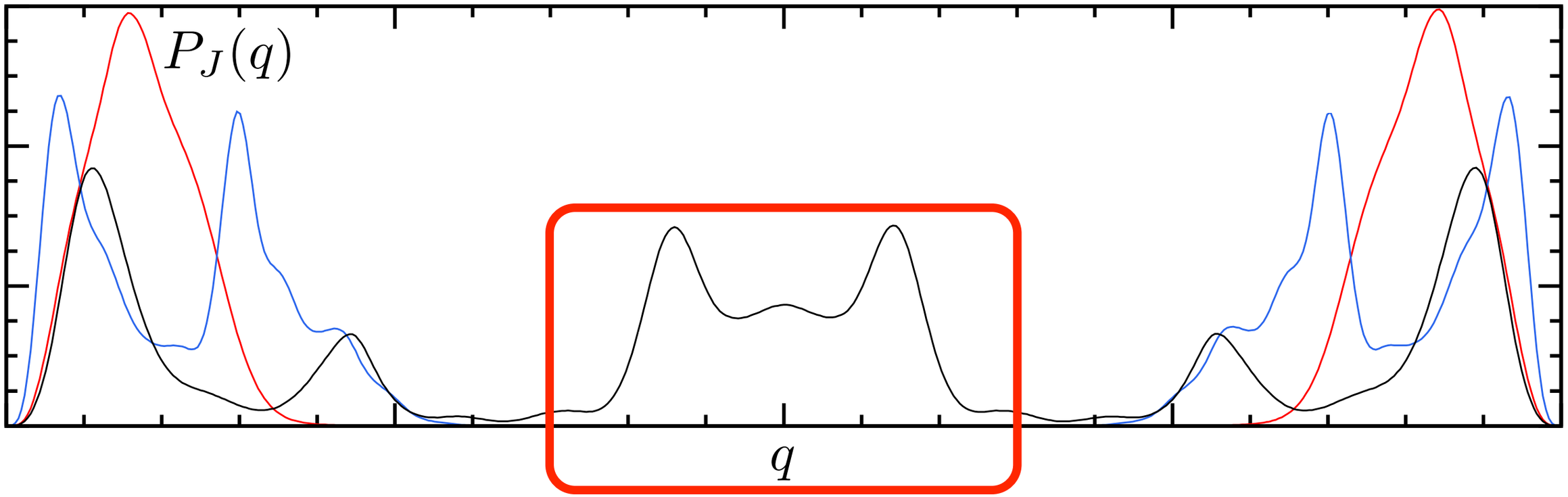

Many states picture?




Large finite-size effects!

# Our approach: Study *individual* distributions

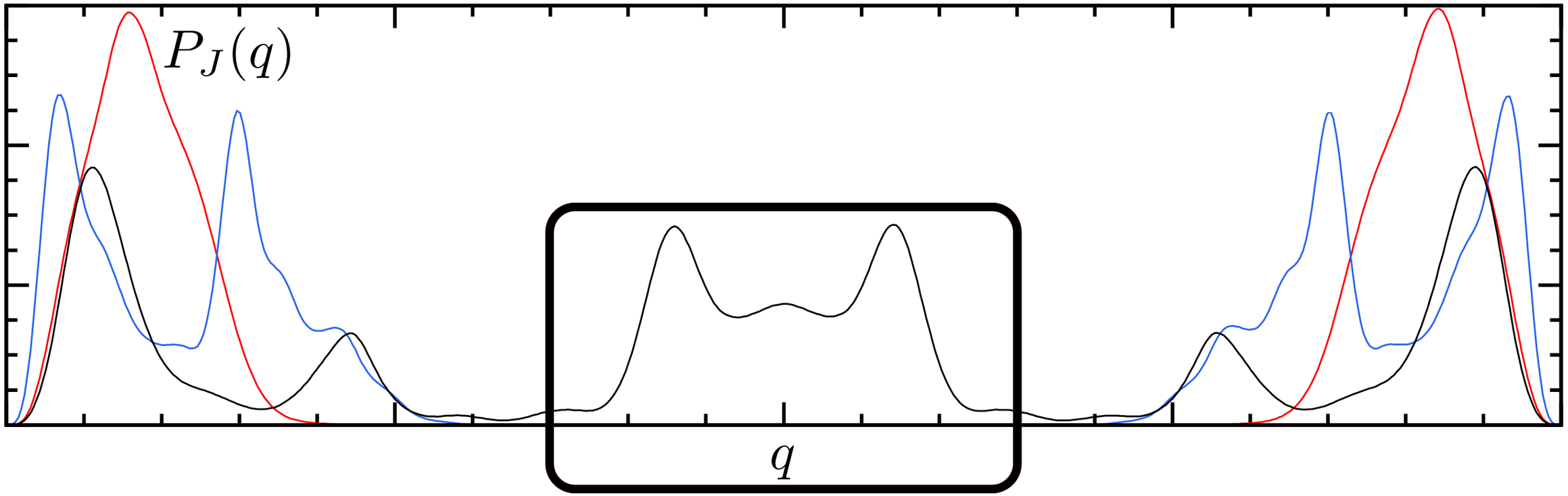
Yucesoy *et al.* (12)




$$P(|q| \leq q_0) \approx 0$$

# Our approach: Study *individual* distributions

Yucesoy *et al.* (12)

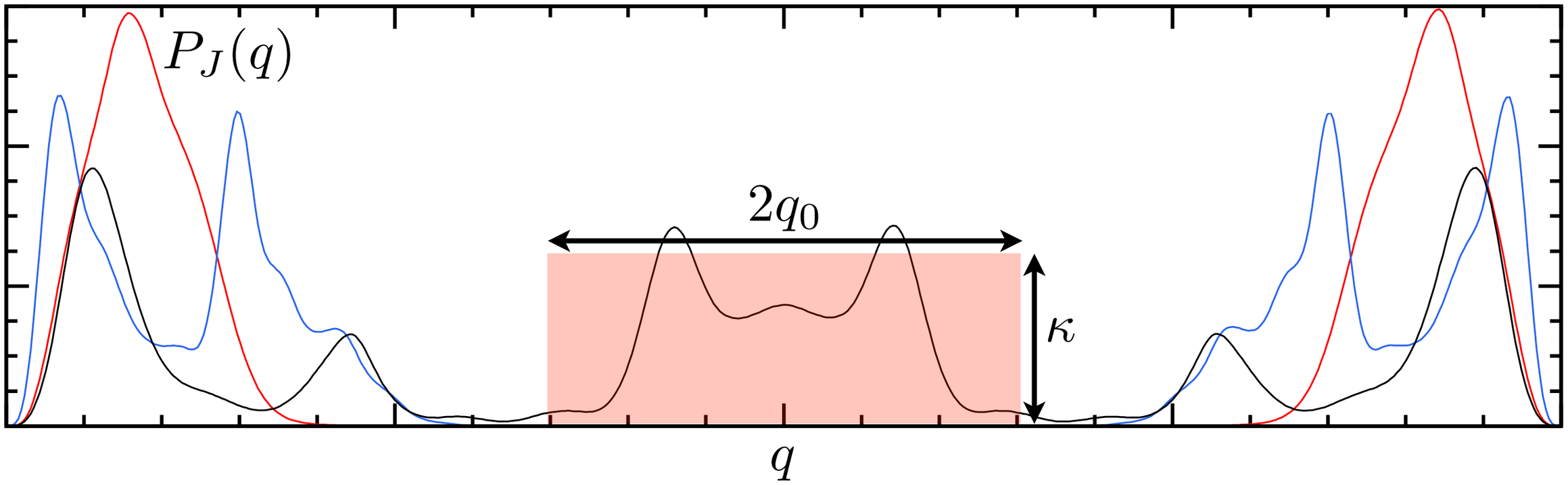



$P(|q| \leq q_0) \approx 0$

$P(|q| \leq q_0) > 0$

# Our approach: Study *individual* distributions

Yucesoy *et al.* (12)

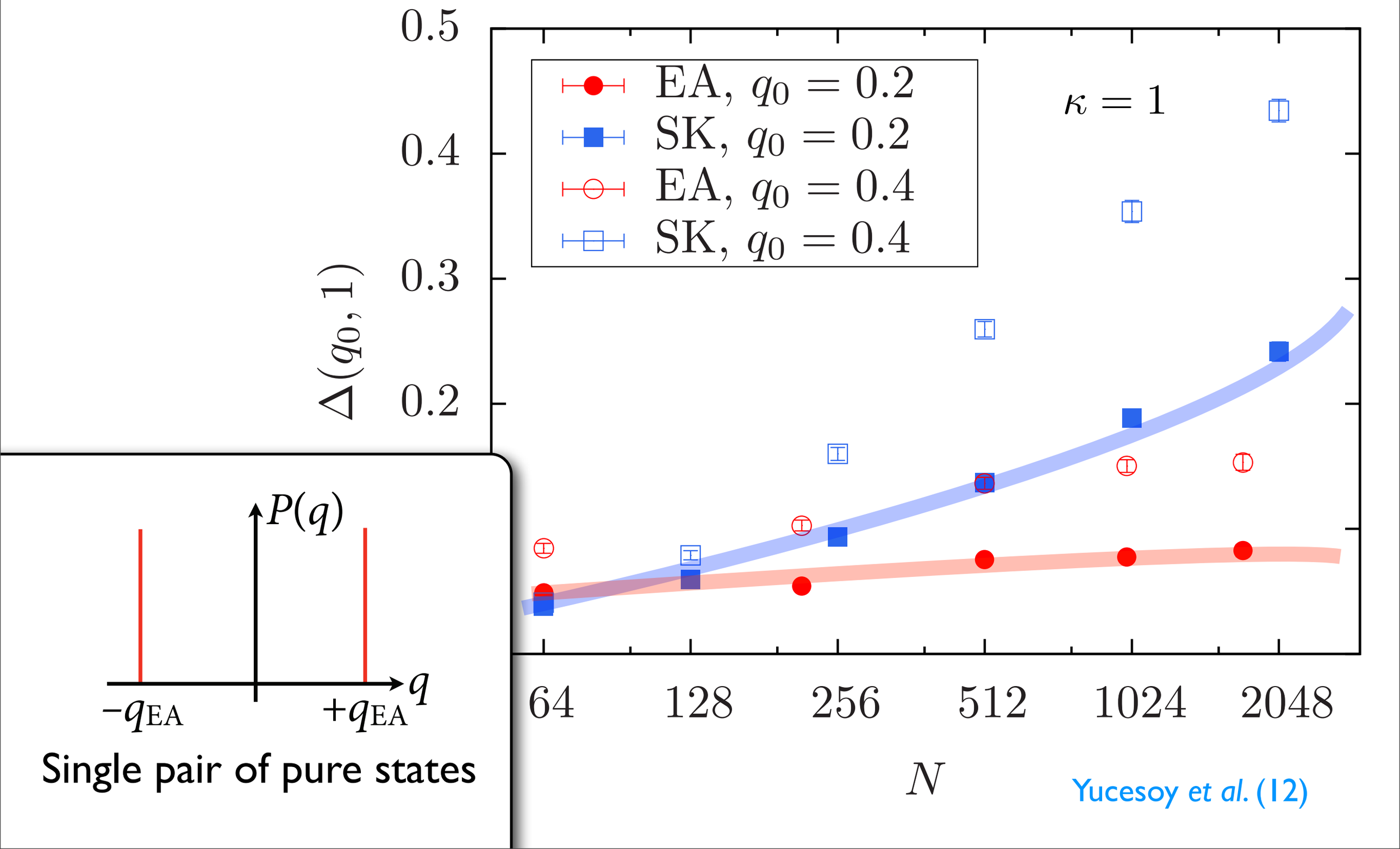



- Define a new "order parameter"

$$\Delta = \frac{\text{number of samples with } P(|q| < q_0) > \kappa}{\text{total number of samples}}$$

  - If $\Delta \to 1$ for $N \to \infty$ Many pairs of pure states (RSB)
  - If $\Delta \to 0$ for $N \to \infty$ Single pair of pure states (DP or CP)

# 3D systems do not behave mean-field like




Single pair of pure states

Yucesoy *et al.* (12)

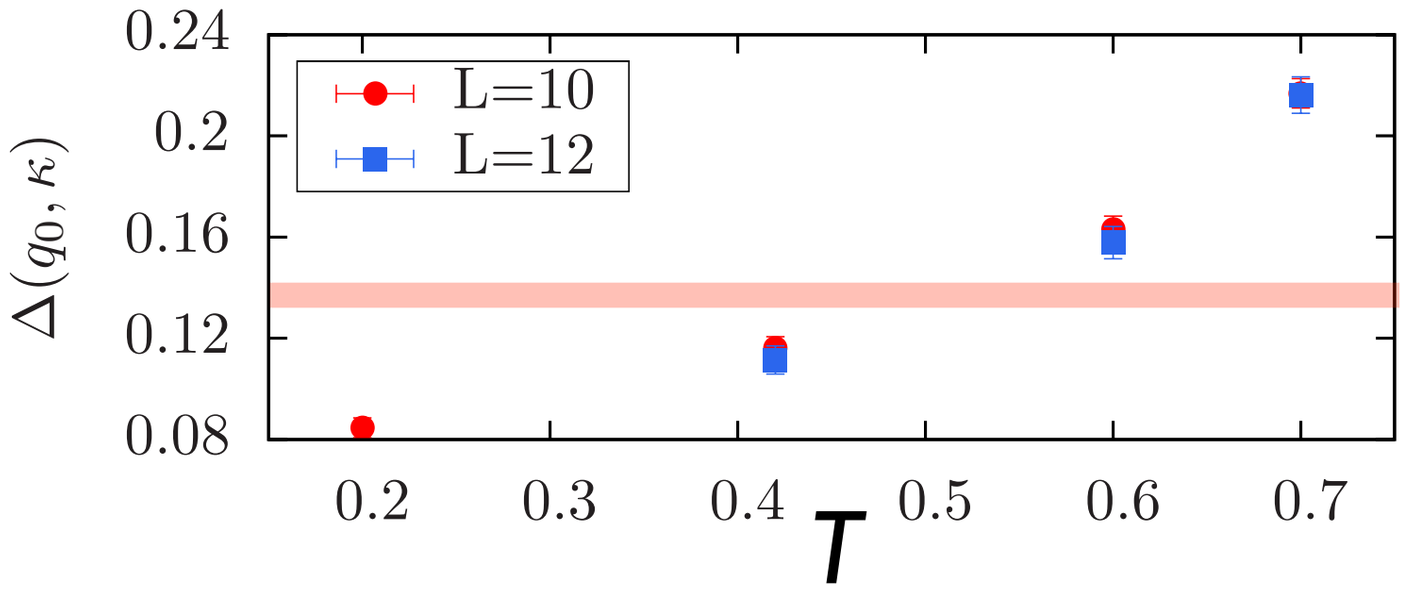
# Comment by Billoire *et al.* (2013)

- Comment: Billoire *et al.* (13)
  - Our system sizes are too small.
  - One should compare EA and SK for similar $P(q)$ [not $T/T_c$].
  - Find "no quantitative difference between SK and EA" ⟶ RSB.

- Reply: Yucesoy *et al.* (13)
  - Our $P(q)$ at lower $T$ are more "peaked" ⟶ closer to infinite vol.
  - Janus data for largest system sizes seems anomalous. Equilibration?
  - Proposed scaling does not work: For fixed $P(q_{\mathrm{EA}})/\kappa$ and $I(q_0)$:

$$\Delta(q_0, \kappa) = \text{const.}$$

Not the case!



# Static & dynamic properties of spin glasses as seen through the parallel tempering telescope

*Nonequilibrium:*

- *Do not rely on round-trip times.*
- *Standard equilibration test good, if used conservatively.*
- *Time scales clearly correlate with the energy landscape.*

*Equilibrium:*

- *Evidence for a two-state scenario.*

*Take-home message: study distributions, not averages.*