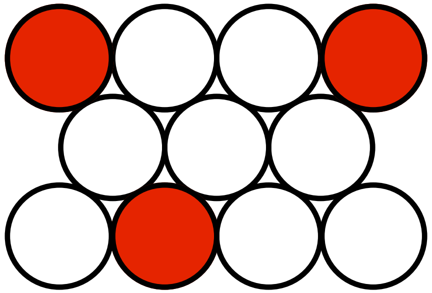
Spin Glas Dynamics
and
Stochastic Optimization Schemes

Karl Heinz Hoffmann
TU Chemnitz

Spin Glasses

spin glass

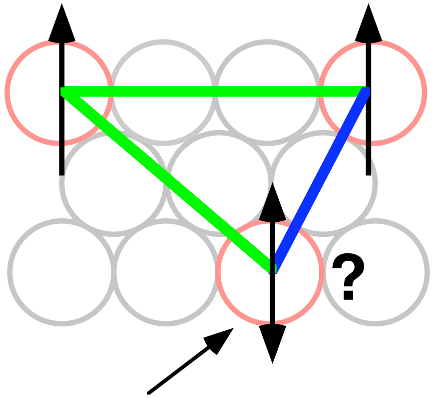


e.g.
AuFe
AuMn
CuMn

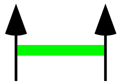
nobel metal (no spin)

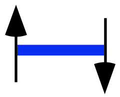
↕

transition metal (spin)   0.1-10 at%



Frustration

 ferromagnetic coupling

 antiferromagnetic coupling

disorder + frustration ⟷ complex system

Spin Glas Models

- Mimic spatial disorder and interaction variation by interaction disorder

- Edwards-Anderson model

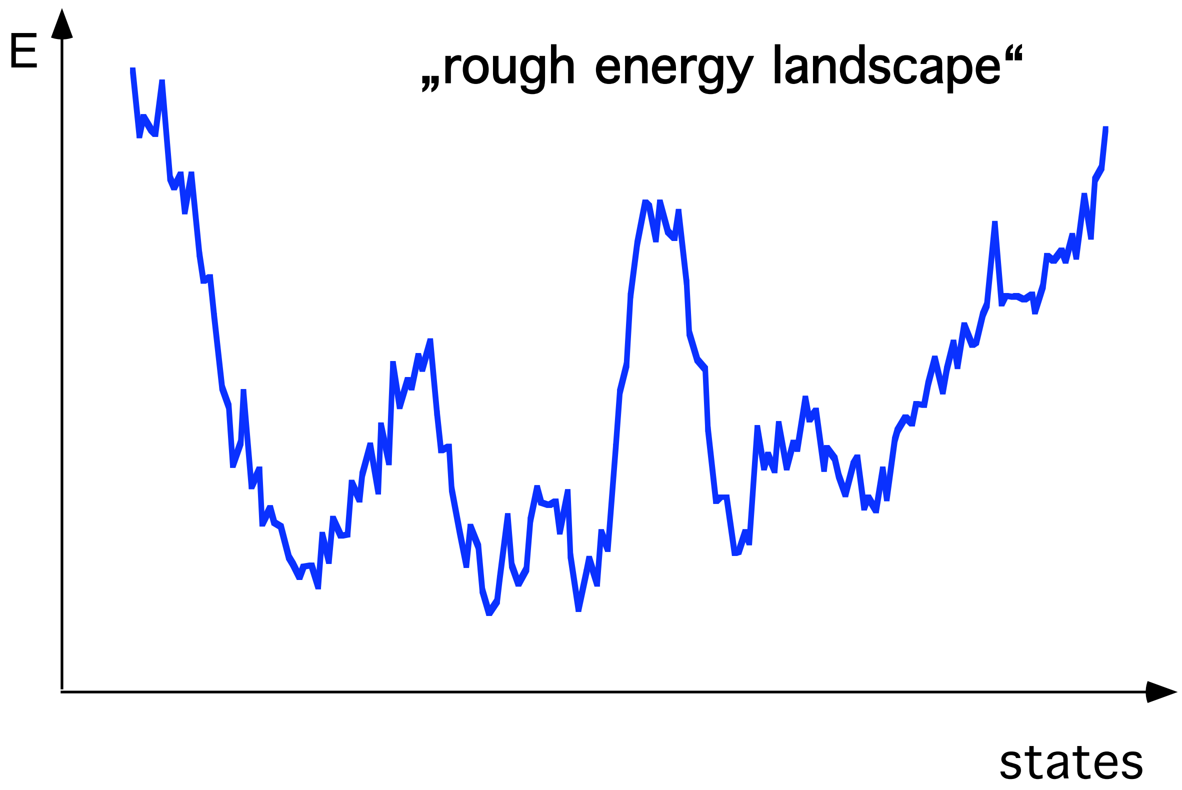
$$H = \sum_{<i,j>} J_{i,j} \, s_i \, s_j + \sum_j h_j \, s_j$$

- Interaction is chosen randomly:

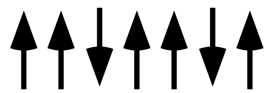
  – Usually symmetric: Gauss, uniform, ±J

Complex Systems

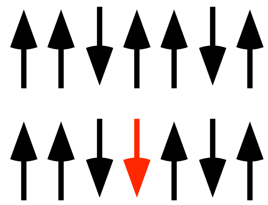
- Complex state space = many, many local minima in the energy function



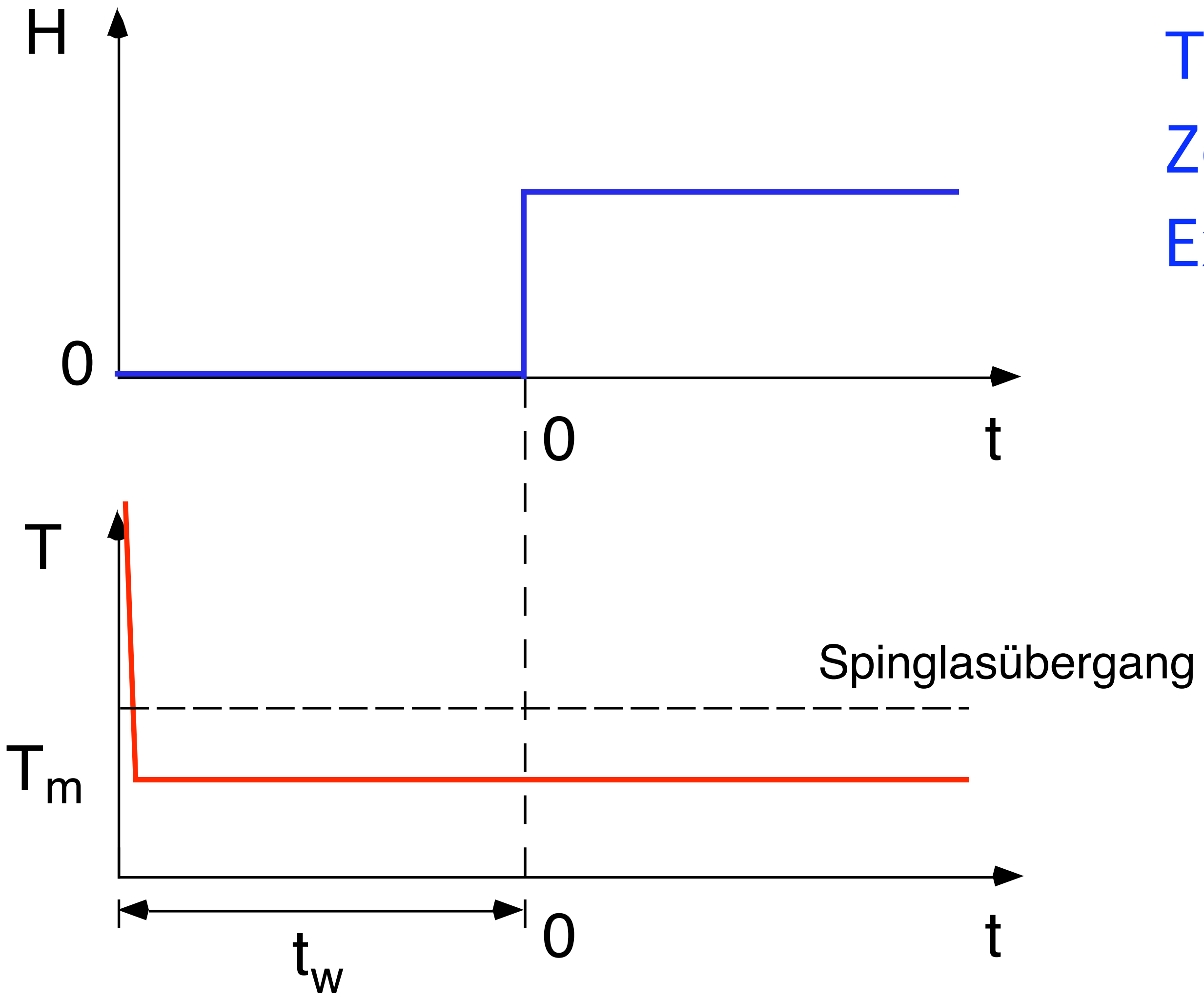
state:
variable to describe degrees of freedom,
e.g. spin configuration:



important:
neighborhood relation

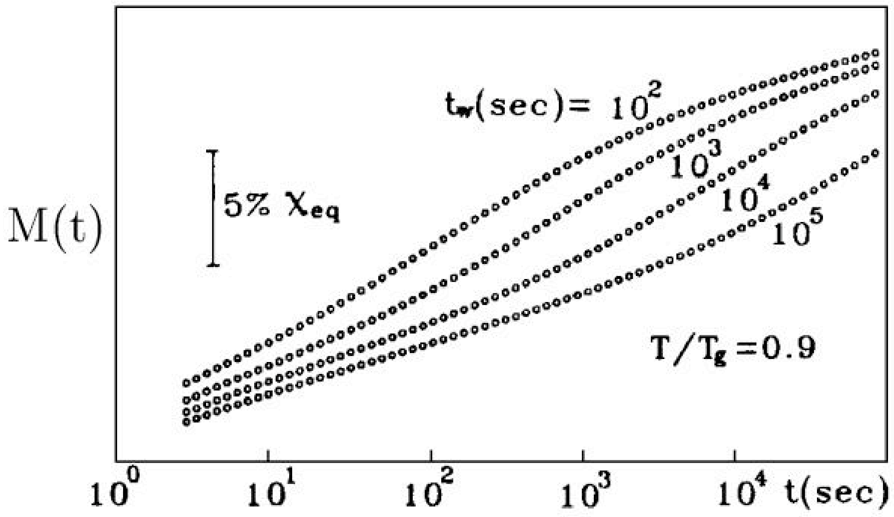


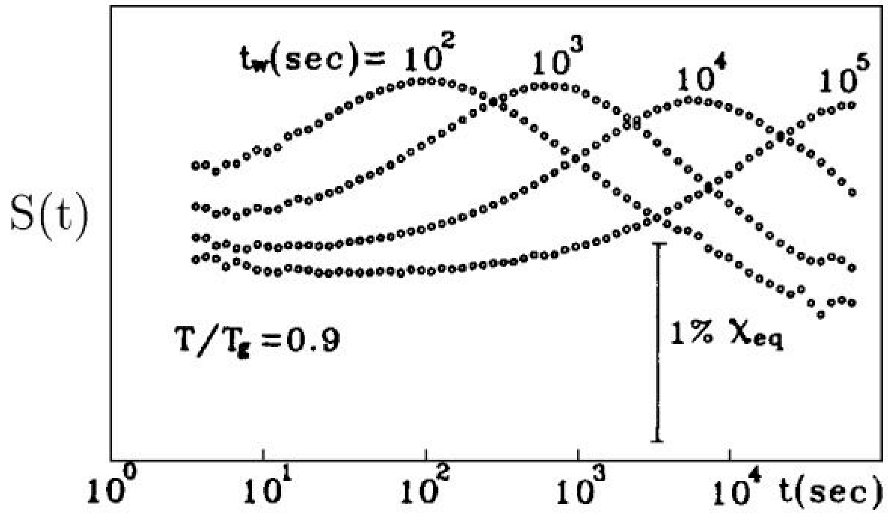
ZFC - Experiment



The Zero-Field-Cooled-Experiment

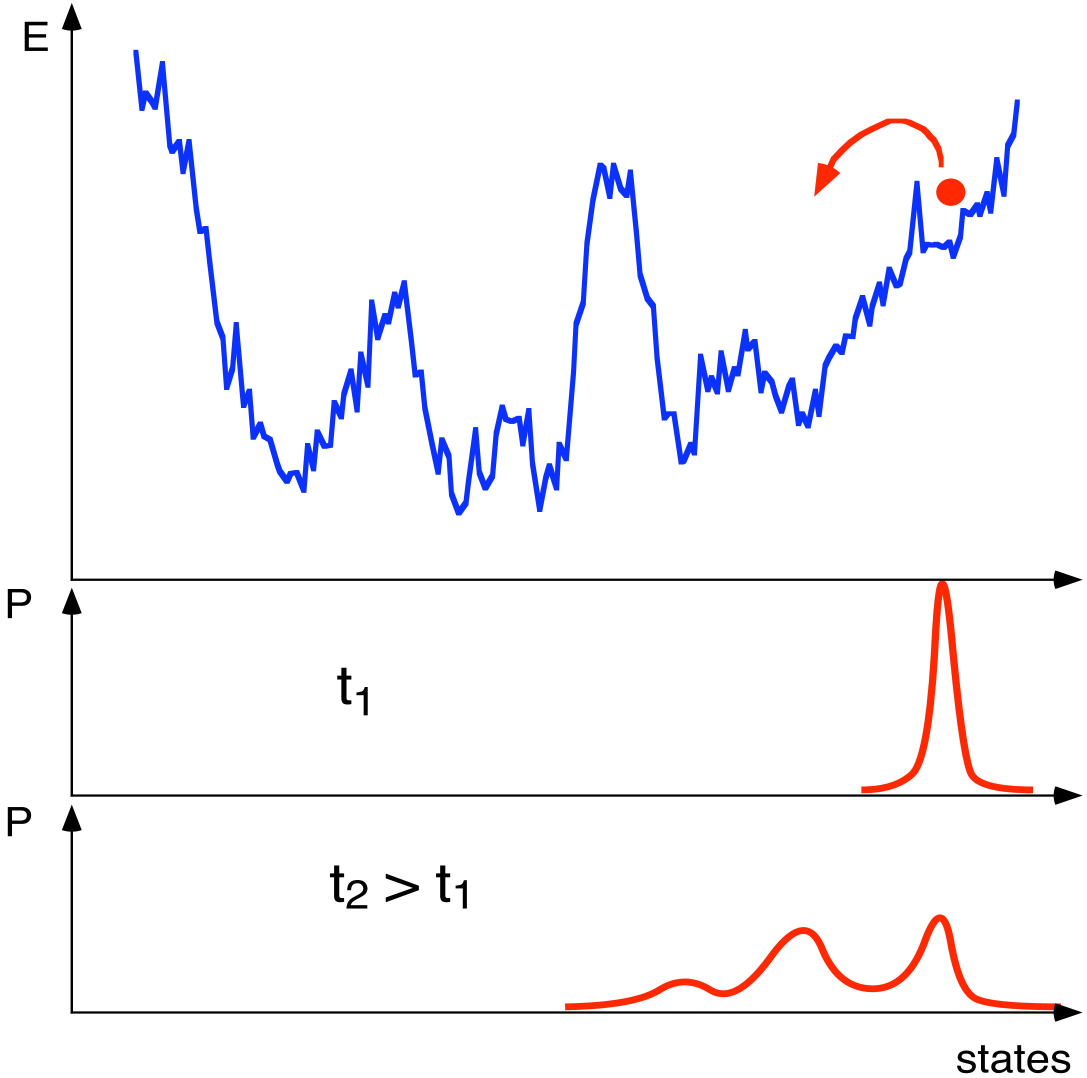
ZFC - Experiment: Aging Phenomena





- $(Fe_{0.15}Ni_{0.85})_{75}P_{16}B_6Al_3$
- measurements depend on waiting time t_w
- $S(t = t_w) = MAX$

Thermally activated Relaxation Dynamics



- „hopping on the energy landscape“
- heat bath = random number generator
- description by a probability density P

Thermally Activated Relaxation Dynamics

- simulation of the thermally activated hopping by the

  Metropolis-Algorithm:

  - choose a neighbor of state j at random, for instance i
  - accept i with

    probabilty

$$P_{metropolis} = \begin{cases} 1 & E_i \leq E_j \\ e^{-(E_i - E_j)/T} & E_i > E_j \end{cases}$$

- random walk on the energy landscape

Master Equation

- modelling the time evolution leads to

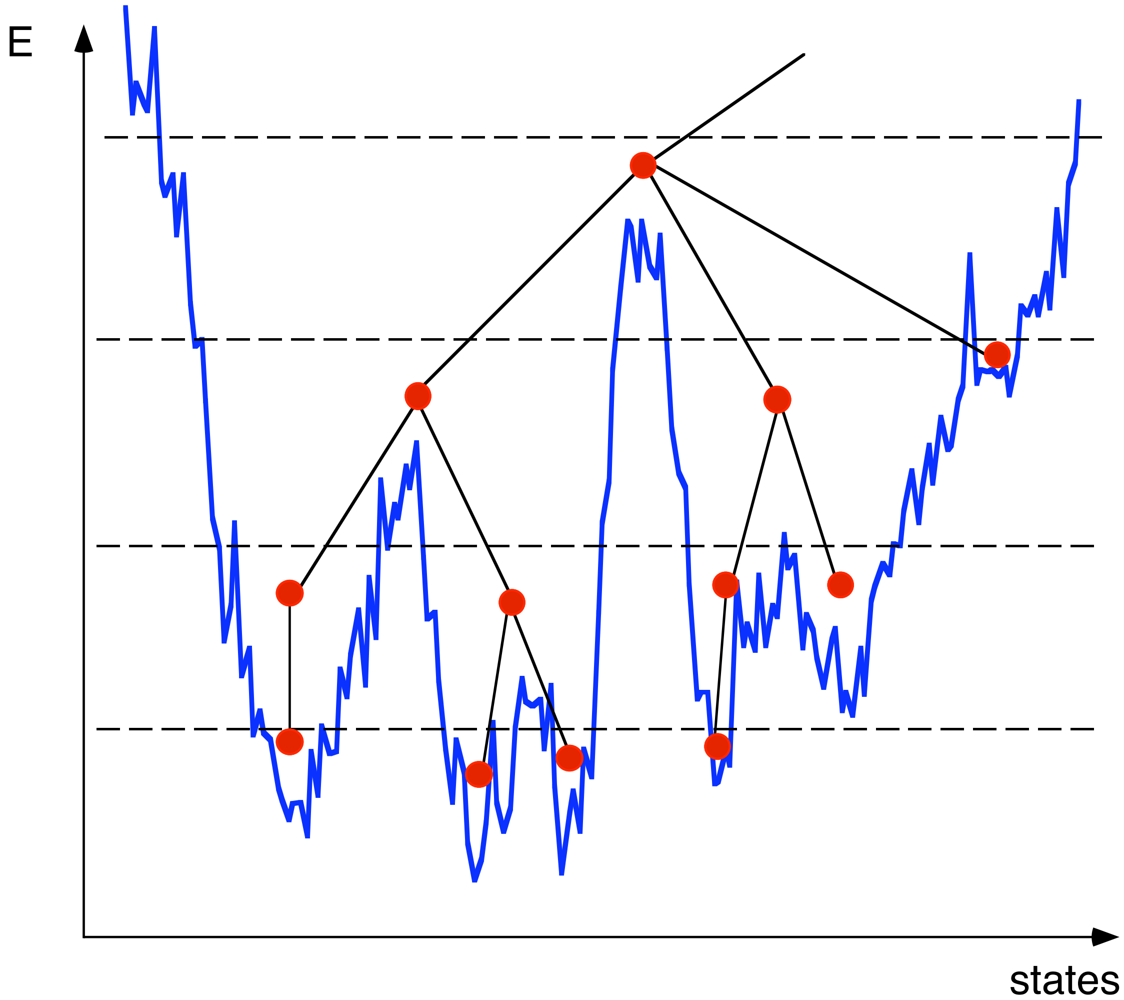
  a stochastic (Marcov) process:

$$P_i(t+\Delta t) = P_i(t) + \sum_j W(i \leftarrow j) P_j(t) - \sum_j W(j \leftarrow i) P_i(t)$$

- transition probabilities $W(i \leftarrow j) \equiv W_{ij} = (P_{ac} N)_{ij}$

- matrix notation $P_i(t+\Delta t) = W_{ij} P_j(t) \qquad P(t+\Delta t) = W P(t)$

Coarse Graining the State Space



- Problem:
  too many states
- Solution:
  „coarse graining“
  the state space

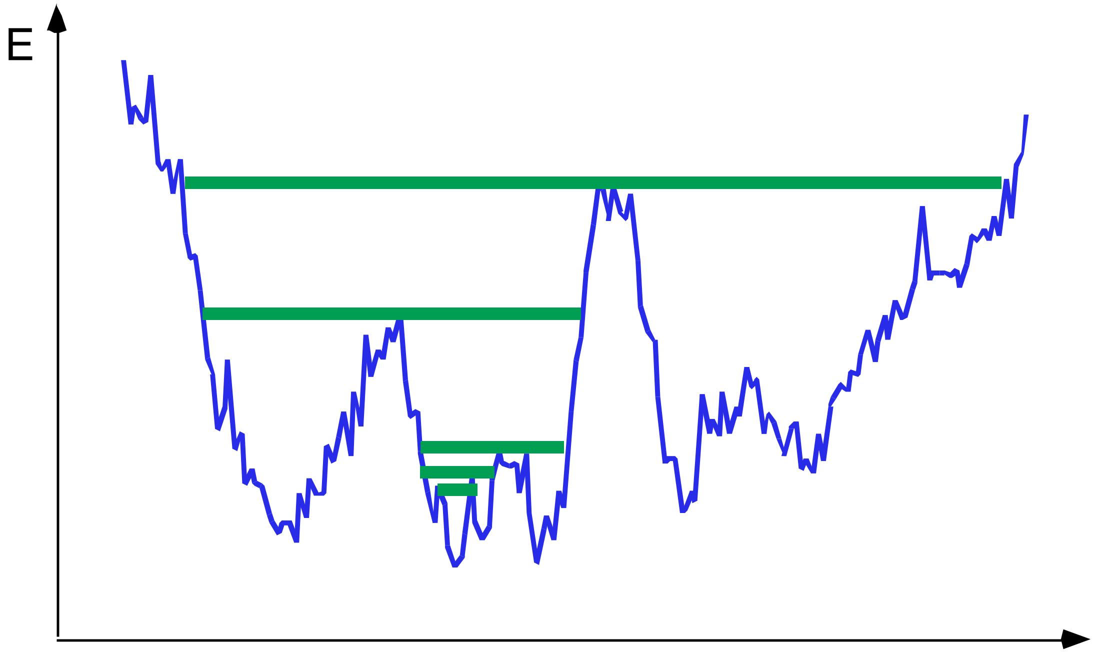
coarse graining

↓

tree structures

Complex Relaxation Dynamics

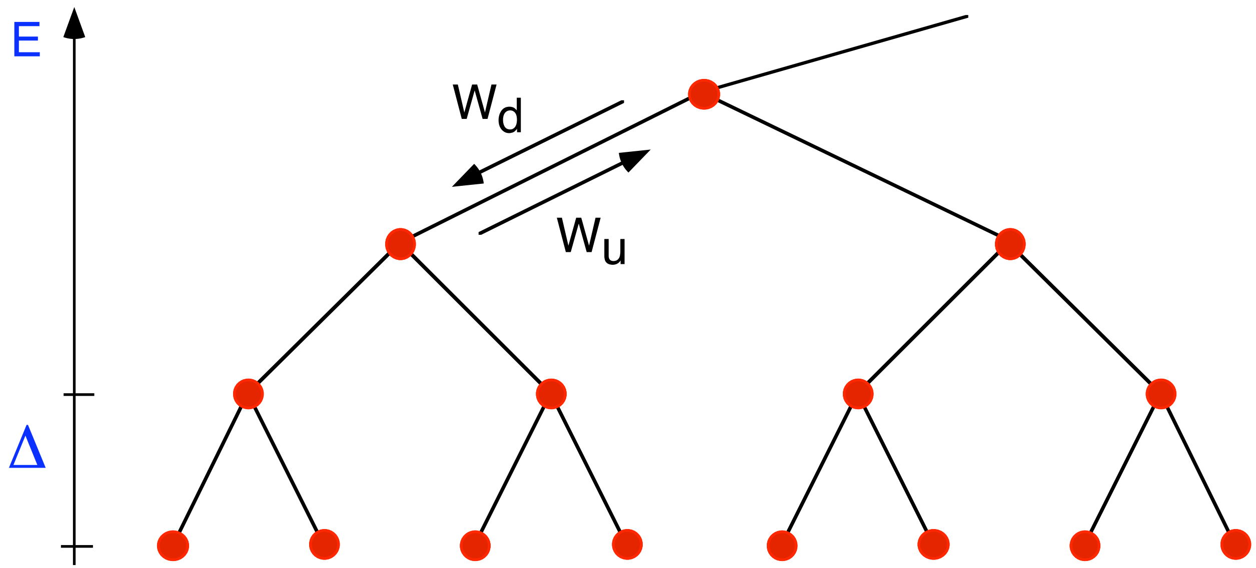
Problem: How does the relaxation proceed ?



Result: the relaxation is a
sequence of partial (or quasi) equilibria
in larger and larger regions of the state space

due to energy barriers on all scales: power laws

Spin Glass Modelling: Simple Tree Models



- Detailed Ballance:

$$\frac{W_u}{W_d} = \kappa\, e^{-\beta\Delta}$$

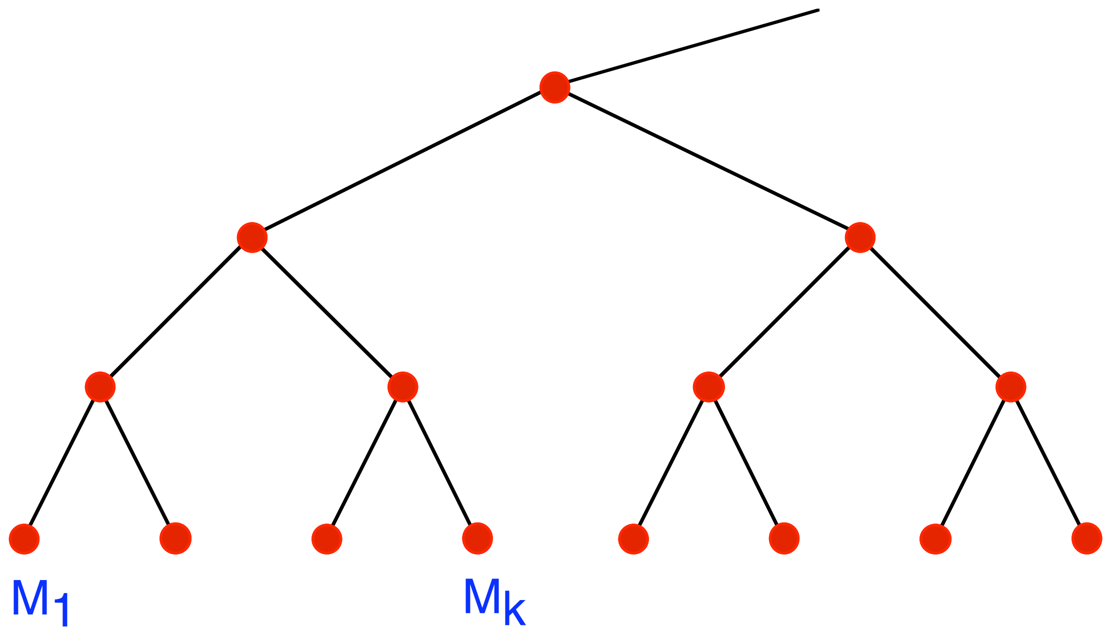
- Relaxation can be treated analytically

- Power Laws: $P(i,t \mid j) \propto t^{\gamma}$

- Result: Relaxation proceeds through a sequence of Quasiequilibria

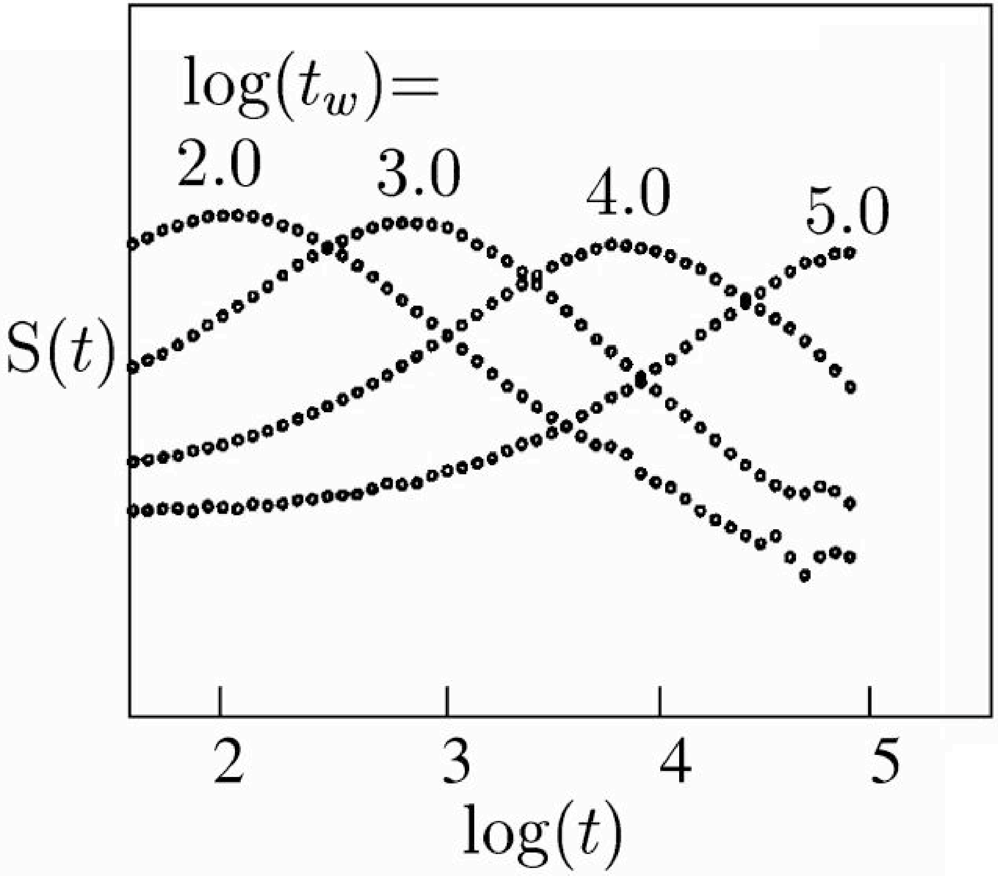
Spin Glass Modelling: Simple Tree Models

- dynamics

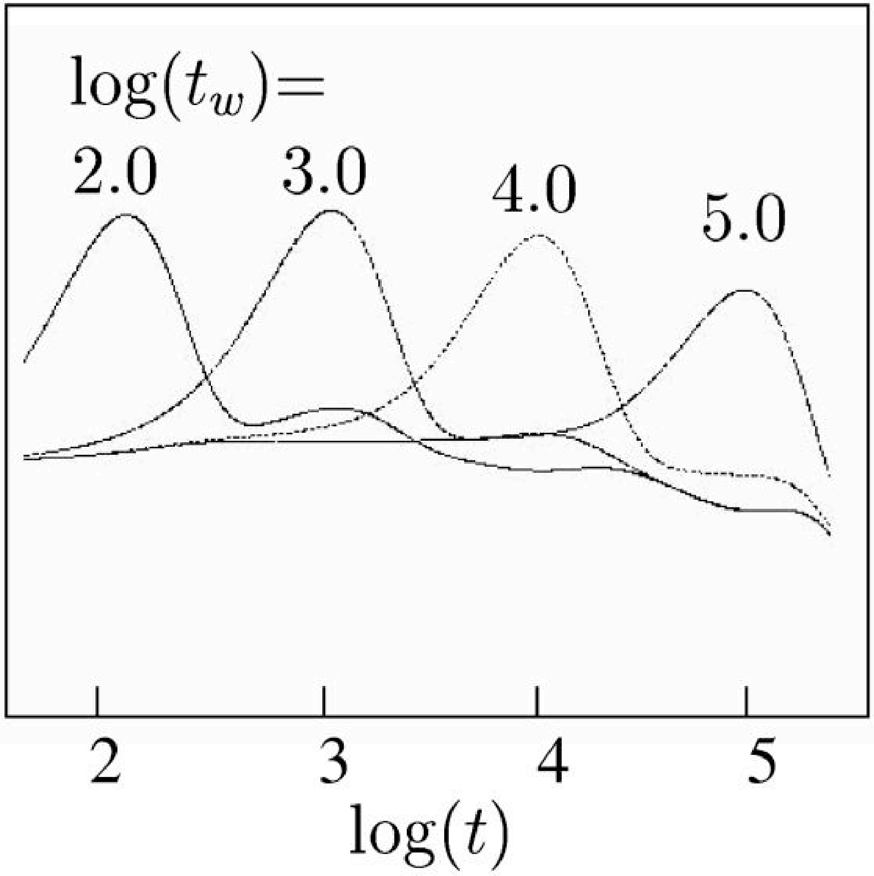


- \+ magnetic properties: magnetic disorder correlation function $\langle M_k M_l \rangle$
- \+ linear response theory: out of equilibrium
- = model results

ZFC - Experiment: Model Results

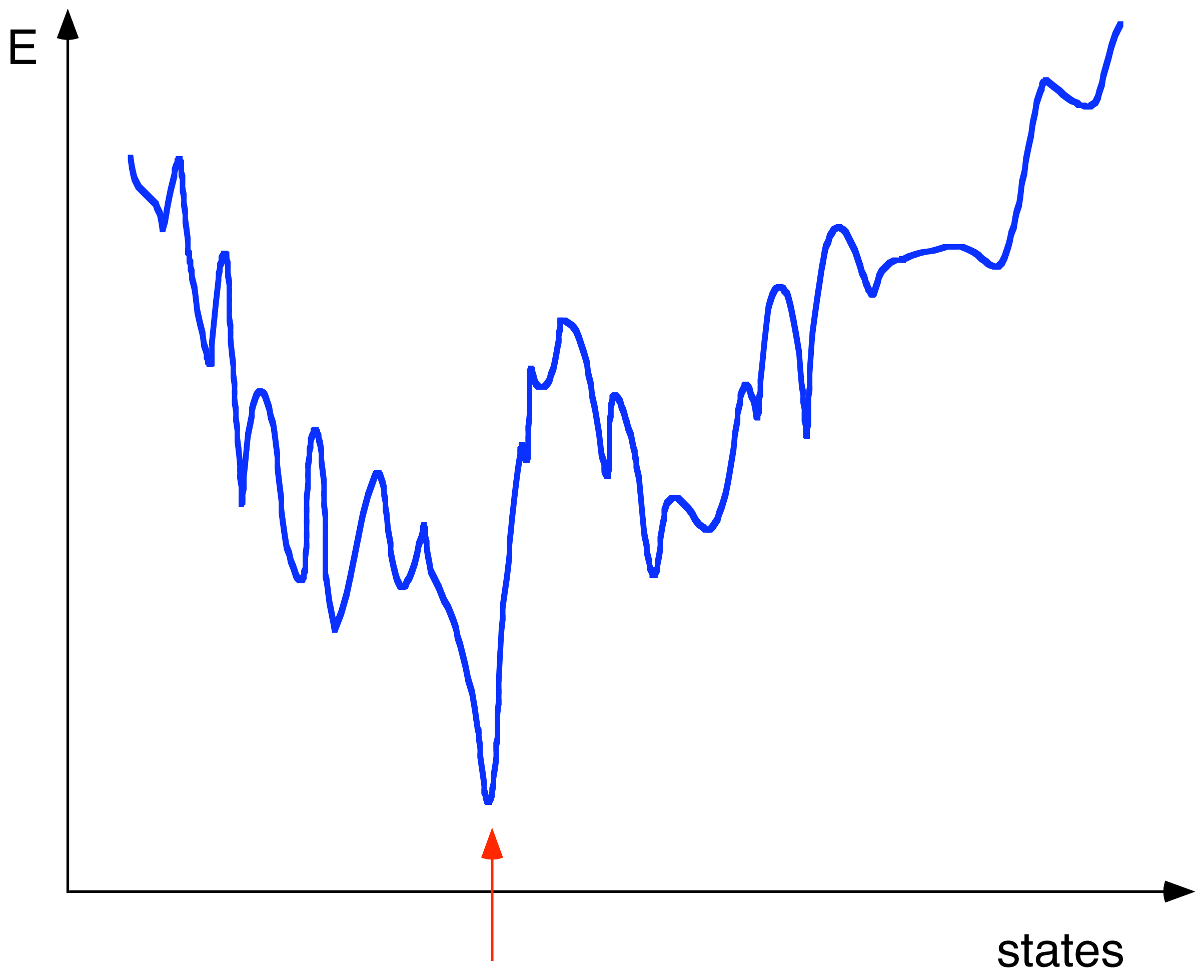


experiment



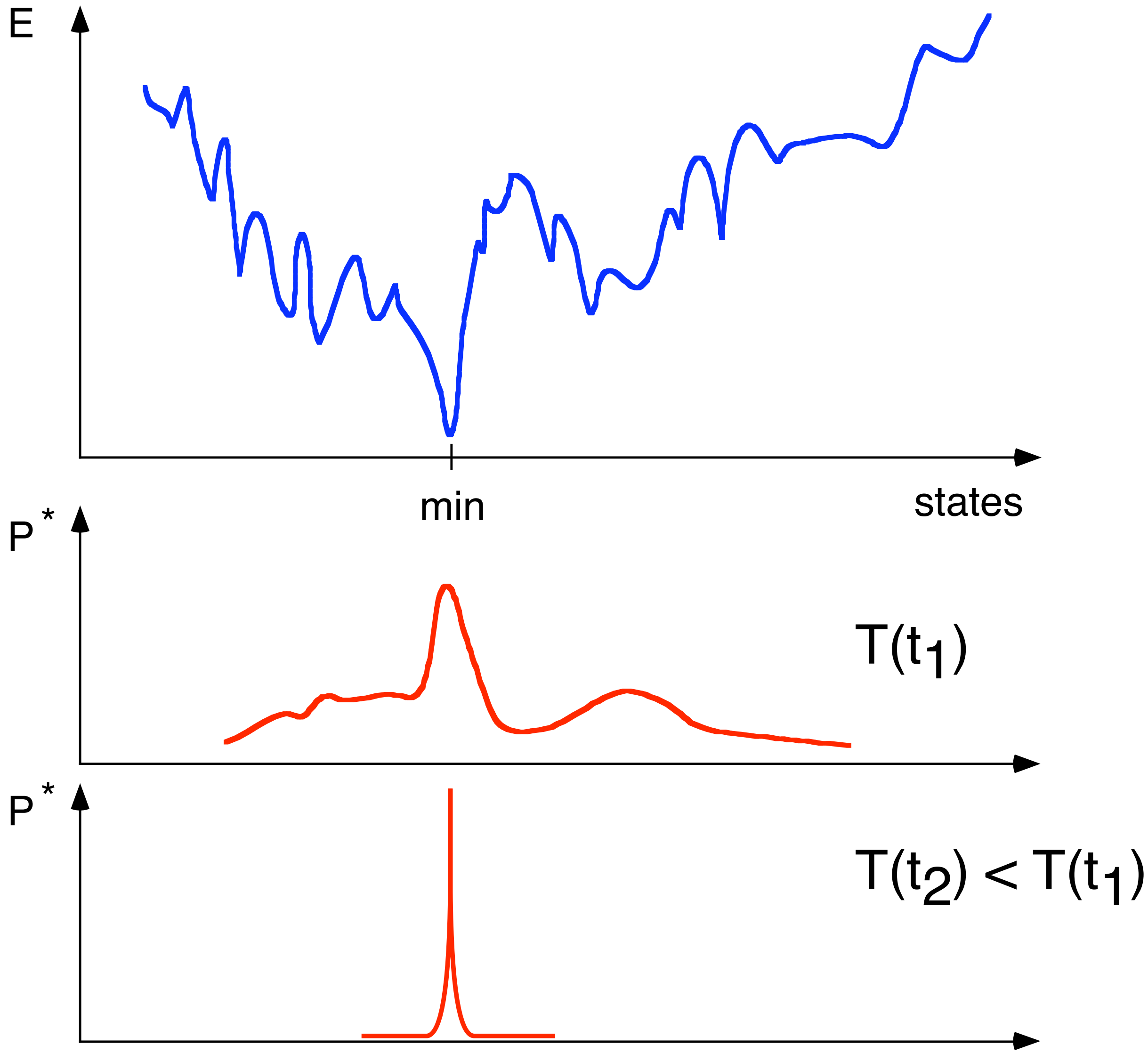
model

Complex Systems: the Ground State Problem



How does one find the ground state in a complex system ?

Stochastic Optimization: Simulated Annealing



simulated annealing:

Metropolis algorithm with

$$T(t) \rightarrow 0$$

$$P^*(i) \propto e^{-E_i/T}$$

$$\rightarrow \delta_{i,MIN}$$

Universal Optimization Scheme

$\Omega = \{i\}$	state	spin configuration
$E : i \rightarrow E_i$	objective function	energy
$\{N_i\}$	neighborhood relation, move class	spin configuration differing by one spin flip
$T(t)$	external parameter	temperature

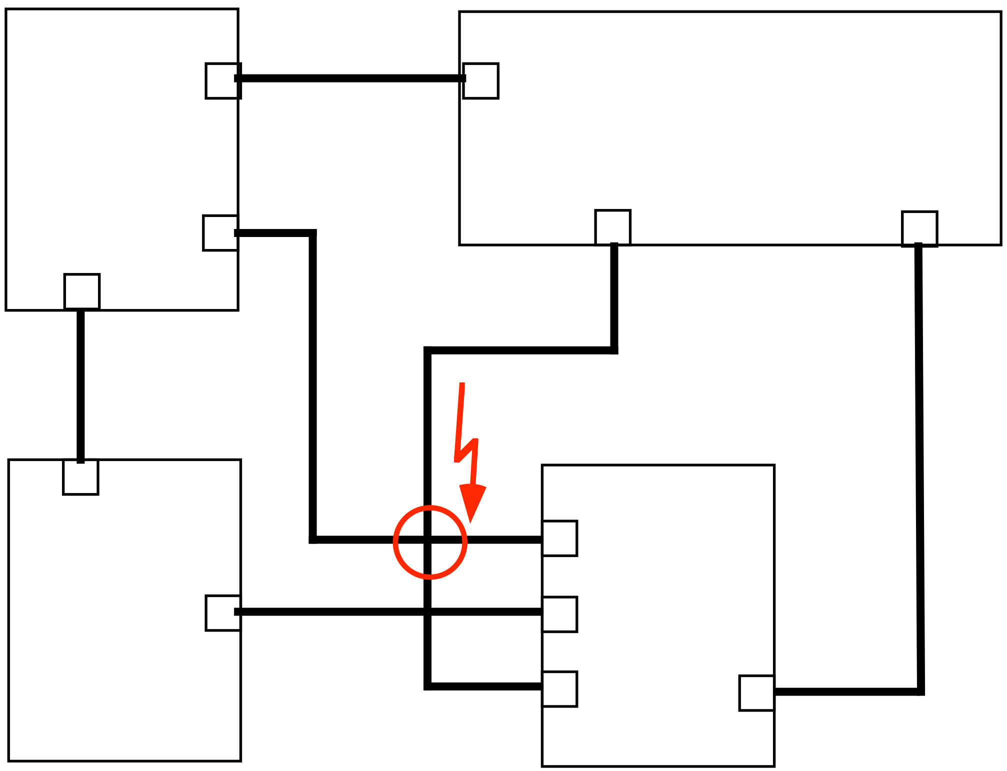
simulated annealing: $E \rightarrow E_{MIN}$

Examples of Complex Optimization Problems

- chip design
- graph partitioning
- neural networks
- seismic deconvolution
- travelling salesman problem
- patern recognition
- ..... NP-hard optimization problems

Chip Design

Chip Design

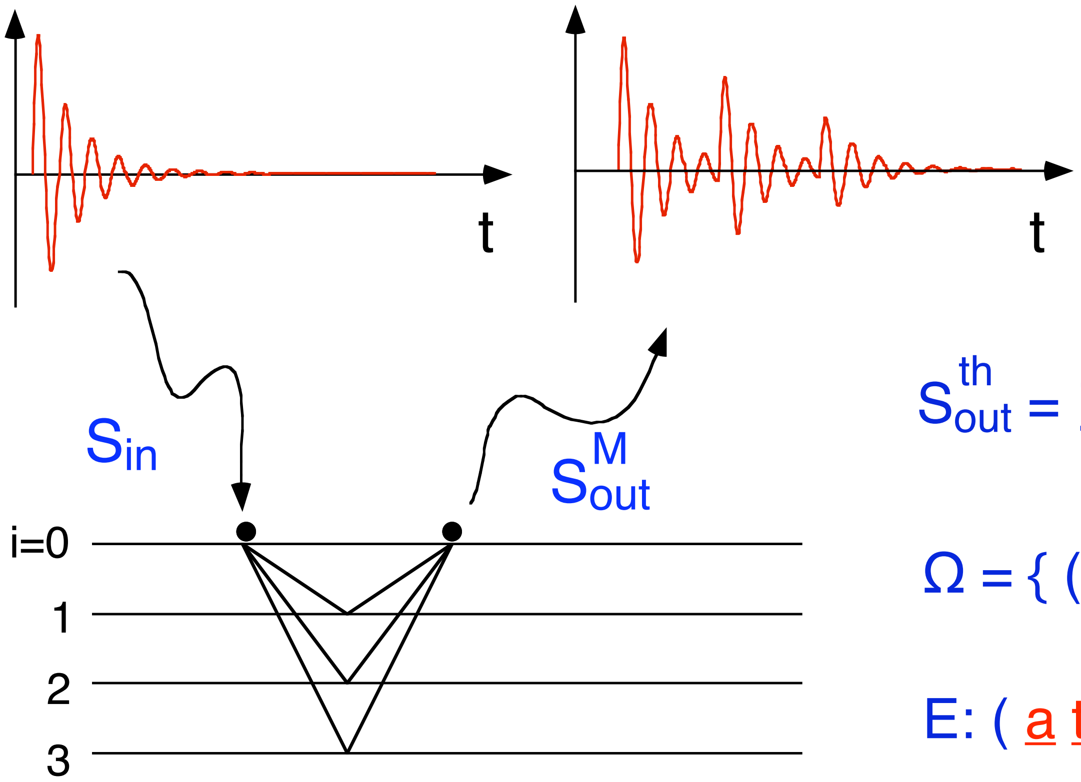


- Find best layout for chip
- typical constraint:

  minimize number of layers needed for connections

Seismic Deconvolution

- typical example for a parameter fit



$$S_{out}^{th} = \sum_i a_i \; S_{in} \; (t - t_i)$$

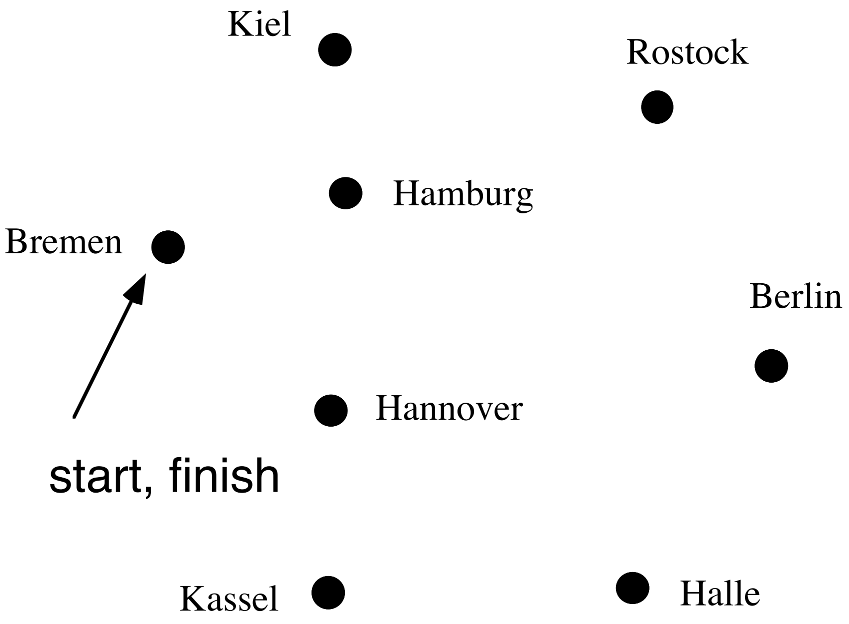
$$\Omega = \{ (\underline{a} \; \underline{t}) \}$$

$$E: (\underline{a} \; \underline{t}) \longrightarrow E = (S_{out}^{M} - S_{out}^{th})^2$$

$$N((\underline{a} \; \underline{t})) = \{ (\underline{a}' \; \underline{t}') \mid \underline{a}' = \underline{a} \pm \Delta\underline{a}, \; \underline{t}' = \underline{t} \pm \Delta\underline{t} \}$$

The Travelling Salesman Problem

- The TSP is a typical NP-hard optimization problem



distance matrix
given

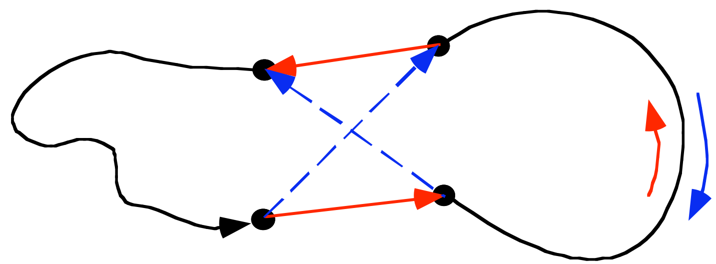
shortest tour?

Ω = { tour = permutation of cities }

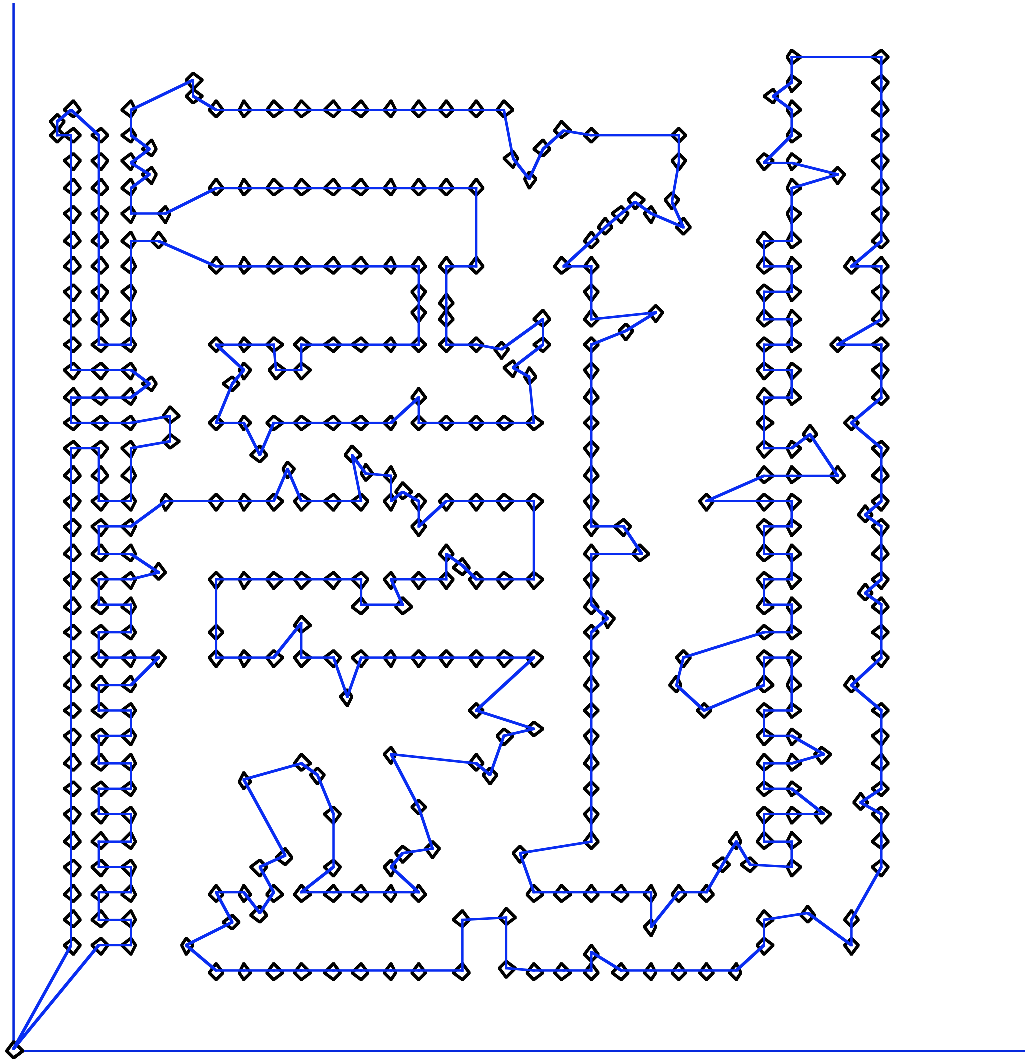
E: tour ⟶ lenght of tour

N(tour) =
{ tour' | two cities in tour exchanged }

or:



The Grötschel Problem: a TSP



- The Grötschel Problem
- an instance of a

  travelling salesman problem (TSP)
- 443 „cities“ to visit

Simulated Annealing: Optimal Annealing Schedule

- guarantee: $E \rightarrow E_{MIN}$ for $t \rightarrow \infty$
- if there is only a finite time τ available: $P(i,\tau) \neq \delta_{i,MIN}$
- then determine $P(i,\tau)$ such that $\bar{E}(\tau) = \sum_i P(i,\tau) E_i \rightarrow MIN$

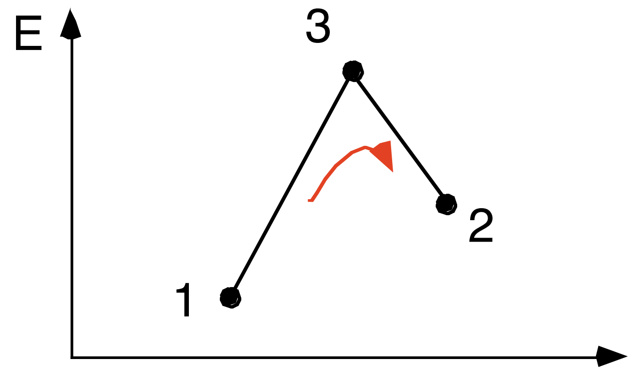
  ↑ dynamics ← $T(t)$

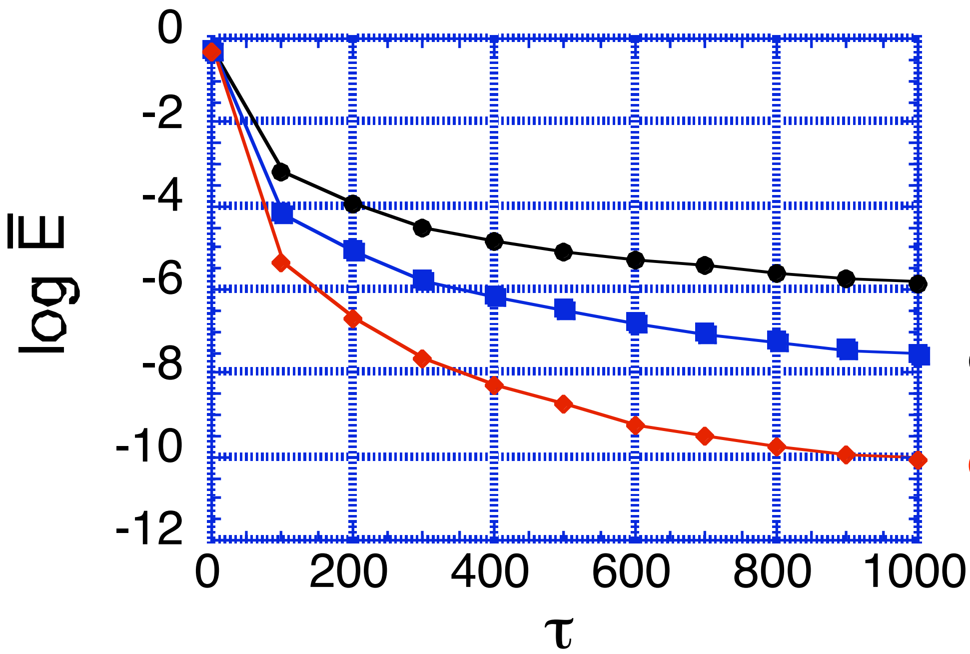
- aim: determine $T(t)$ such that $\bar{E}(\tau) \rightarrow MIN$

⟶ optimal annealing schedule

Simulated Annealing: Optimal Annealing Schedule

- example: a single barrier





linear: $T = T_0 - \varepsilon t$

exponentiell: $T = T_0 \alpha^t$

optimal

Optimality Criteria

- Criteria based on final distribution at time S :
  - Mean final energy (MIN) $$\bar{E}(S) = \sum_i P(i,S) E_i \rightarrow MIN$$
  - Final probability in ground state (MAX)
  - Other features of the final distribution $P(i,S)$
- Criteria based on the Best So Far Energy:

$$E(S) = \mathrm{MIN}_{\ 0 \leq t \leq S}\left[E(S)\right]$$

Features of the BSFE distribution $B^S(E)$

B(est) S(o) F(ar) E(nergy) Distribution

- Modify transition probabilities: make states below E absorbing

$$\Gamma^t_{\alpha\beta;E} = \begin{cases} \delta(\alpha,\beta) & \text{if } E(\beta) \leq E \\ \Gamma^t_{\alpha\beta} & \text{if } E(\beta) > E \end{cases}$$

- Master equation

$$p^t_{\alpha;E} = \sum_{\beta\in\Omega} \Gamma^t_{\alpha\beta;E} p^{t-1}_{\beta;E}$$

- Probability to have seen energies below E at time S

$$B^S(E) = \sum_{\alpha:E(\alpha)\leq E} p^S_{\alpha;E}$$

B(est) S(o) F(ar) E(nergy) Distribution

- Finite state space: finite number of energy values
- Sort energies

$$E_1 < E_2 < \ldots < E_R$$

- Probability that the lowest energy visited is E_r

$$b^S(E_r) = B^S(E_r) - B^S(E_{r-1})$$

- Mean BSF energy

$$\langle E_{\mathrm{BSF}}(S) \rangle = \sum_{r=1}^{R} b^S(E_r) E_r$$

B(est) S(o) F(ar) E(nergy) Distribution

- Create enlarged state space

$$\boldsymbol{q}^{t+1} = \begin{pmatrix} \boldsymbol{p}_{E_0}^{t+1} \\ \boldsymbol{p}_{E_1}^{t+1} \\ \vdots \\ \boldsymbol{p}_{E_R}^{t+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Gamma}_{E_0}^{t} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Gamma}_{E_1}^{t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Gamma}_{E_R}^{t} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{p}_{E_0}^{t} \\ \boldsymbol{p}_{E_1}^{t} \\ \vdots \\ \boldsymbol{p}_{E_R}^{t} \end{pmatrix} = \tilde{\boldsymbol{\Gamma}}^{t} \cdot \boldsymbol{q}^{t}$$

- Use new probabilties

$$\langle E_{\mathrm{BSF}}(S)\rangle = \sum_{r=1}^{R} E_r \left(B^S(E_r) - B^S(E_{r-1})\right)$$

$$= \sum_{r=1}^{R} E_r \left(\sum_{\alpha:E(\alpha)\leq E_r} q_{Lr+\alpha}^{S} - \sum_{\alpha:E(\alpha)\leq E_{r-1}} q_{L(r-1)+\alpha}^{S} \right)$$

General Optimality Criterion

- In general

$$F(\boldsymbol{q}^1, \boldsymbol{q}^2, \ldots, \boldsymbol{q}^S) = \sum_{t=1}^{S} (\boldsymbol{F}^t)^{\mathrm{tr}} \cdot \boldsymbol{q}^t = \sum_{t=1}^{S} \sum_{i=1}^{L(R+1)} F_i^t q_i^t \to \min$$

- Control: acceptance probabilities at each point in time

Simulated Annealing: the Ensemble Approach

- problem: the barrier structure is unknown
- solution: adaptive schedules
- method: the ensemble approach to simulated annealing

  take N copies of the system,

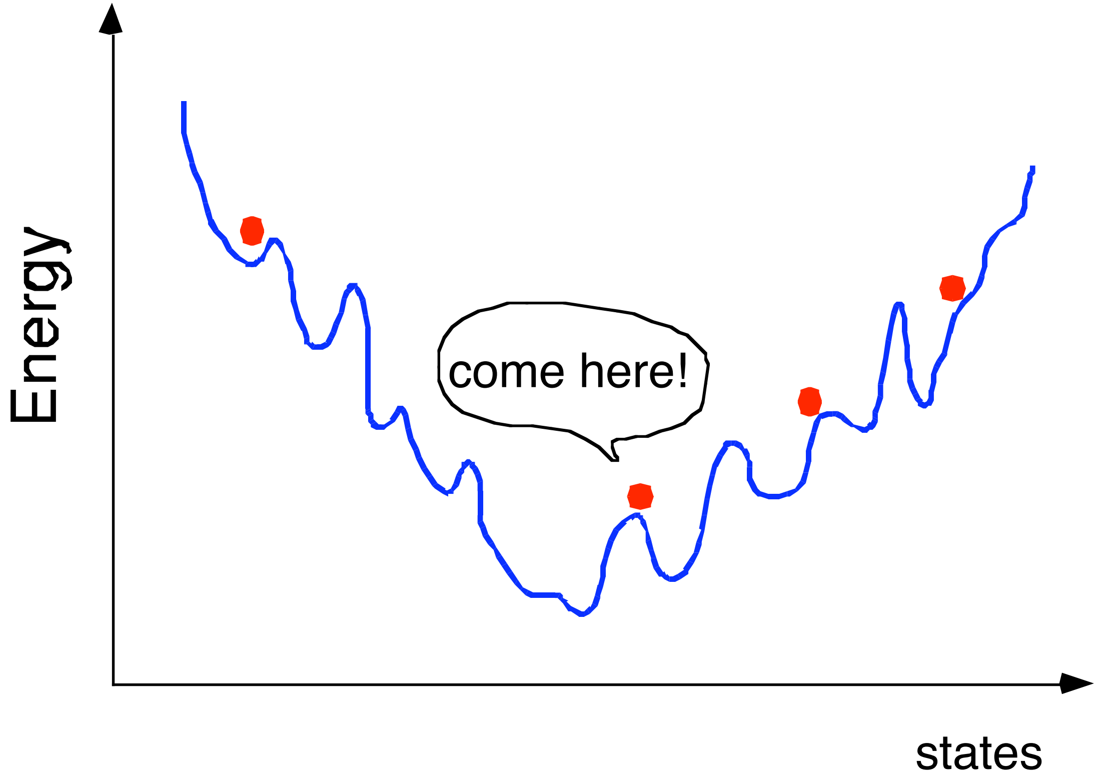
  anneal with the same schedule

  use ensemble properties (mean, variance)

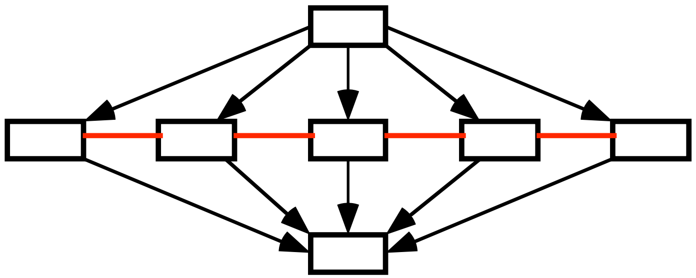
  to control schedule adaptively

The Ensemble Approach to SA

- many at the same time



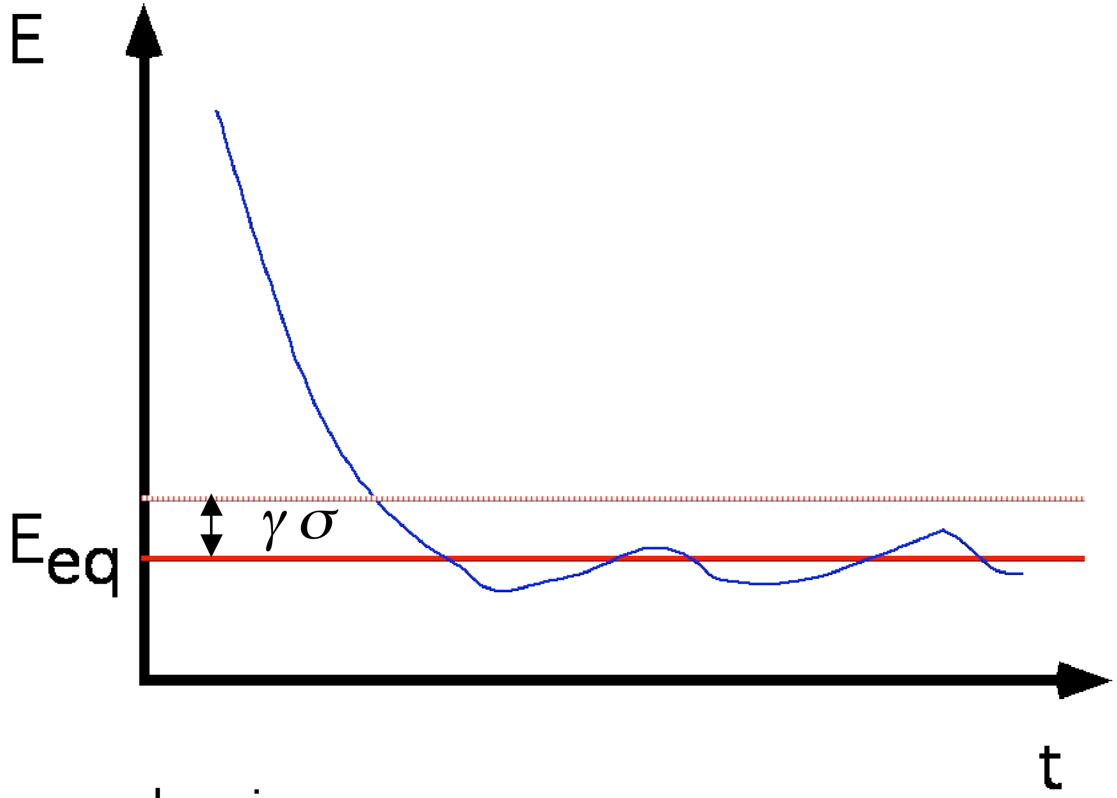
- with communication
- parallel implementation:



- task
- compute farm
- result

Adaptive Schedules for SA

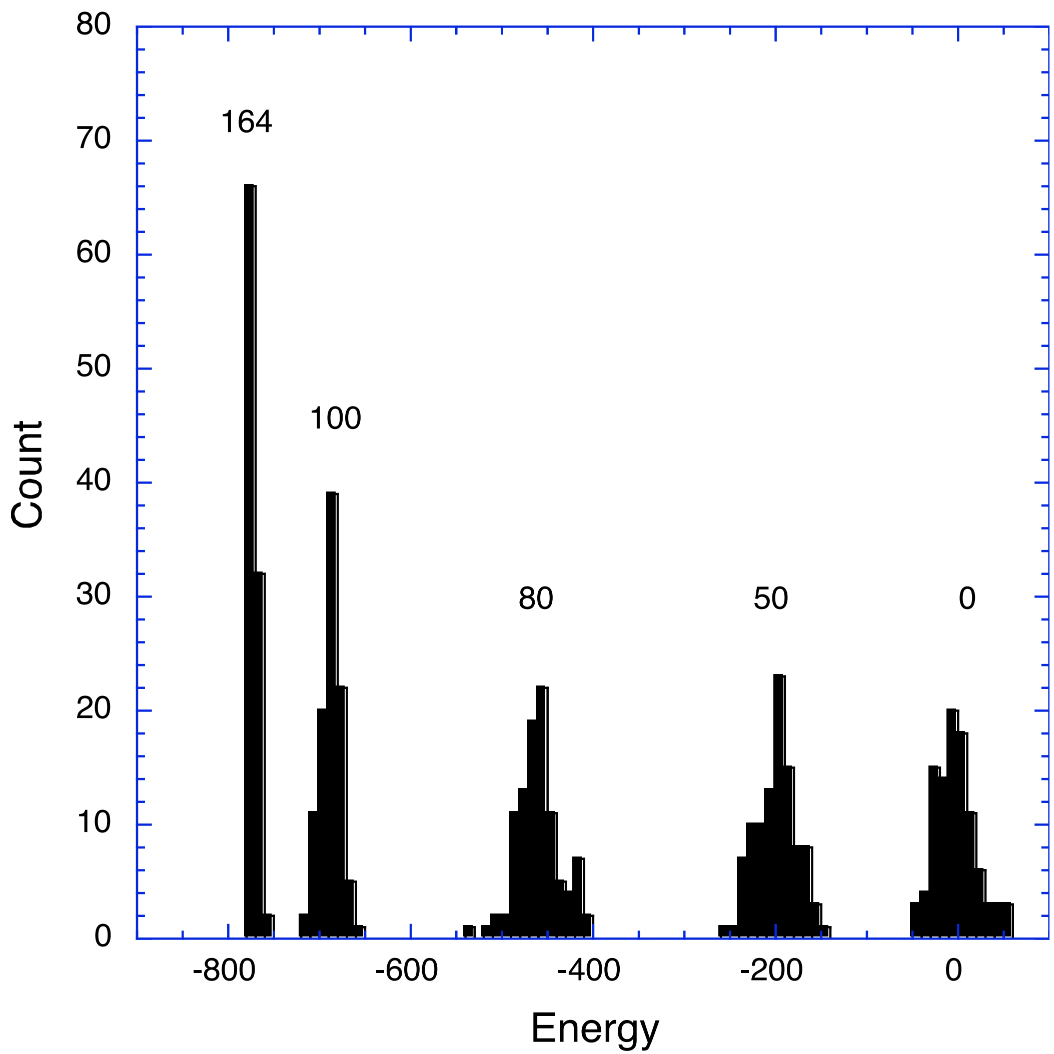
- idea:
  - anneal
    as fast
    as possible
  - but stay
    close to
    equilibrium
    distribution



- the algorithm:
  - determine ensemble mean and variance
  - anneal
  - if ensemble mean starts to fluctuate: lower temperature
  - etc

The Ensemble Approach to SA

- Typical distribution of an ensemble
- Note progress in time towards lower energies



Threshold Accepting

- Simulated Annealing

$$P_{metropolis} = \begin{cases} 1 & E_i - E_j \le 0 \\ e^{-(E_i - E_j)/T} & E_i - E_j > 0 \end{cases}$$

- Threshold Accepting

$$P_{threshold} = \begin{cases} 1 & E_i - E_j \le T \\ 0 & E_i - E_j > T \end{cases}$$

computationally more effective: no exponentials required !

Rényi Entropy

- Another information measure:

$$R_q = \frac{1}{1-q} \ln \sum_i P_i^q$$

- for $q \to 1$: $R_{q \to 1} = \sum_i P_i \ln P_i$ Shannon

- interesting property:
  - non-extensive
  - minimum "Rényi information": No Boltzmann distribution !!

Tsallis Entropy

- A new, non-extensive information measure:

$$S_q = \frac{1}{1-q} \sum_i P_i^q = \sum_i P_i \frac{(1 - P_i^{q-1})}{1-q}$$

- for $q \to 1$:

$$S_{q \to 1} = \sum_i P_i \ln P_i$$

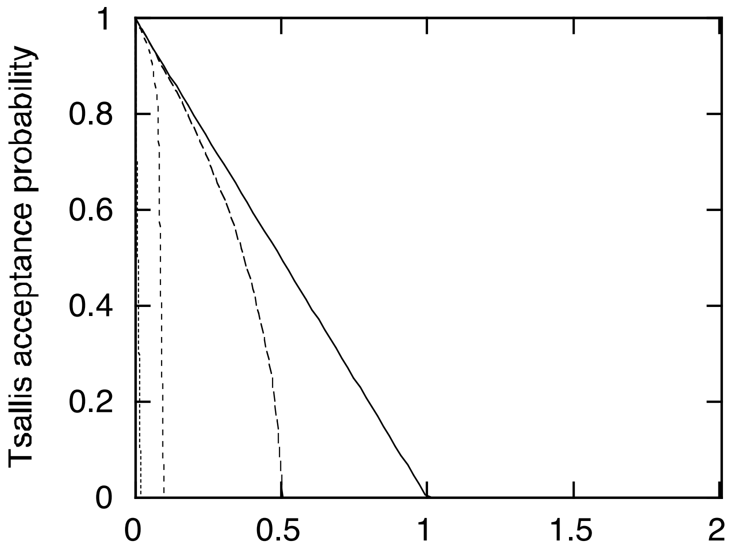
Shannon

- interesting properties:
  - non-extensive
  - minimum "Tsallis information": No Boltzmann distribution !!

Tsallis Annealing

- Tsallis Annealing acceptance probability

$$P_{tsallis} = \begin{cases} 1 & \Delta E \leq 0 \\ \left(1-(1-q)\dfrac{\Delta E}{T}\right)^{\frac{1}{1-q}} & \Delta E > 0 \quad \textit{and} \quad (1-q)\dfrac{\Delta E}{T} \leq 1 \\ 0 & \Delta E > 0 \quad \textit{and} \quad (1-q)\dfrac{\Delta E}{T} > 1 \end{cases}$$



$q \rightarrow -\infty$ Quench

Modified Tsallis Annealing

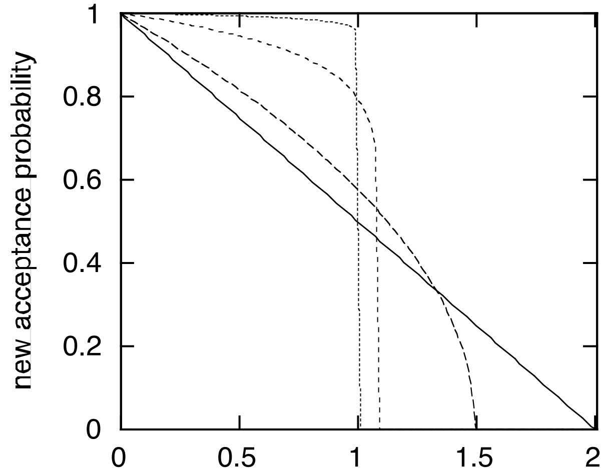
- Modified Tsallis Annealing (Astrid Franz & KHH)

$$P_{MT} = \begin{cases} 1 & \Delta E \leq 0 \\ \left(1 - \frac{1-q}{2-q}\frac{\Delta E}{T}\right)^{\frac{1}{1-q}} & \Delta E > 0 \quad \textit{and} \quad \frac{1-q}{2-q}\frac{\Delta E}{T} \leq 1 \\ 0 & \Delta E > 0 \quad \textit{and} \quad \frac{1-q}{2-q}\frac{\Delta E}{T} > 1 \end{cases}$$

- important property:

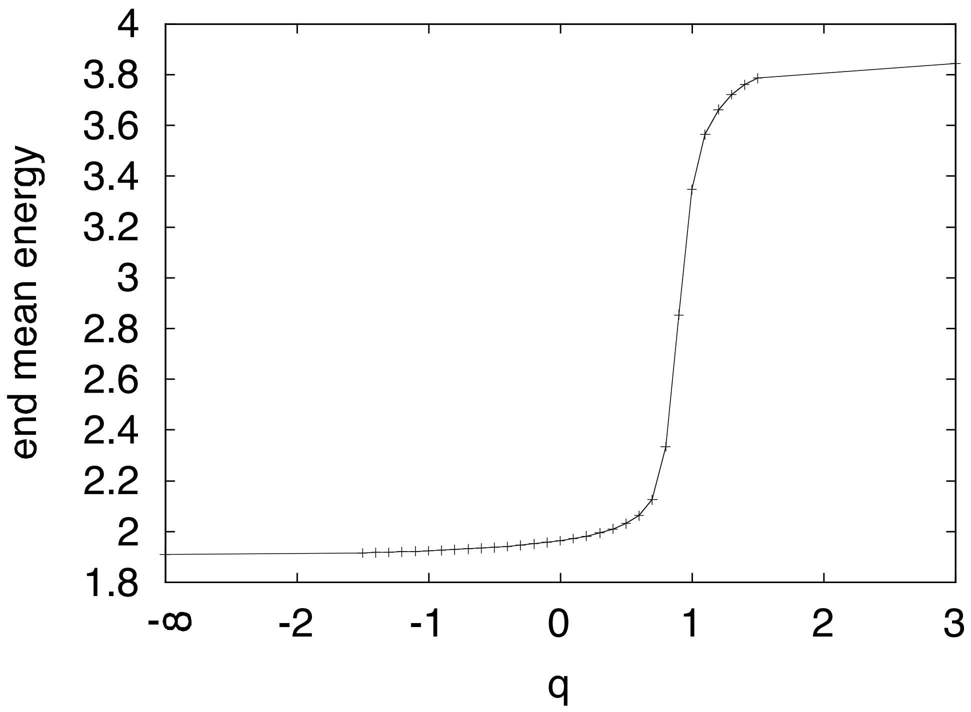
$$q \to 1 \qquad P_{MT} = P_{metropolis}$$

$$q \to -\infty \qquad P_{MT} = P_{threshold}$$



Optimal MT Annealing

- threshold accepting is best !!



- test many different systems

threshold accepting is always best !!

Optimal Strategies for Finding Ground States

- What is the best acceptance probability ??

- Theorem:
  on finite state spaces within the class of acceptance probabilities with these features:

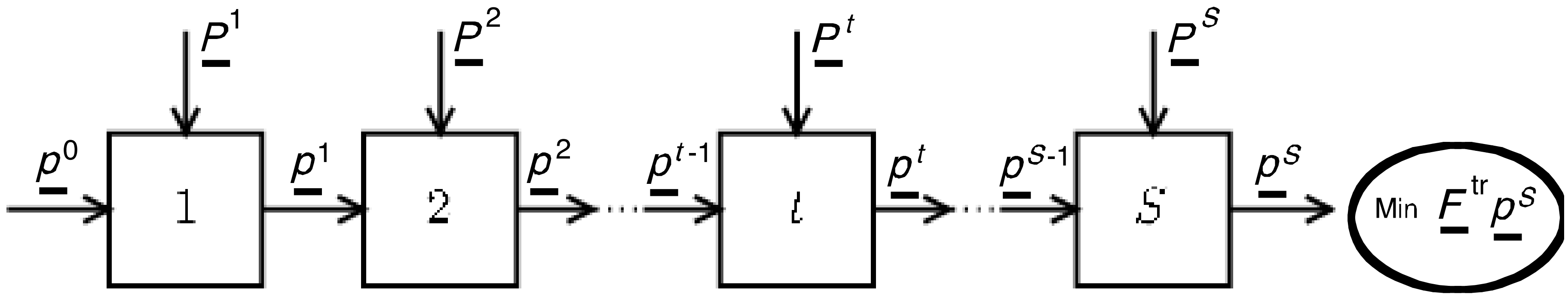
  - $P_{\text{acceptance}} = P_{\text{acceptance}}(\Delta E)$
  - $P_{\text{acceptance}}(\Delta E)$ is monotone decreasing as a function of ΔE

  Threshold Accepting is the best possible strategy (for linear criteria)

- Note that Metropolis, TA, Tsalis, MT have these features

Optimizing Acceptance Probabilities

- At each step, an acceptance probability controls the further time evolution of the distribution in the state space



- Linear criterion:

$$\sum F_i P_i(t_{final}) = \sum F_i \overset{(P_{ac} N)_{ij}}{W_{ij}} W_{jk} W_{kl} \ldots W_{mn} P_n(t_{initial})$$

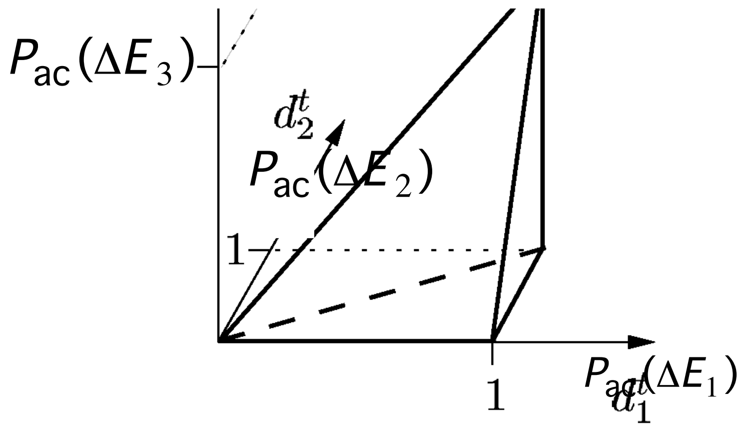
Optimal Strategies for Finding Ground States

- finite state space $\longrightarrow$ finite number of ΔE_α
- $0 < P_{\text{ac}}(\Delta E_\alpha) < 1$
- $P_{\text{acceptance}}(\Delta E)$ is montone decreasing as a function of ΔE

$$P_{\text{ac}}(\Delta E_1) \geq P_{\text{ac}}(\Delta E_2) \geq P_{\text{ac}}(\Delta E_3) \geq \ldots \quad \text{for} \quad \Delta E_1 < \Delta E_2 < \Delta E_3 < \ldots$$

- optimization over a simplex

$$(P_{\text{ac}}(\Delta E_1), P_{\text{ac}}(\Delta E_2), P_{\text{ac}}(\Delta E_3), \ldots)$$

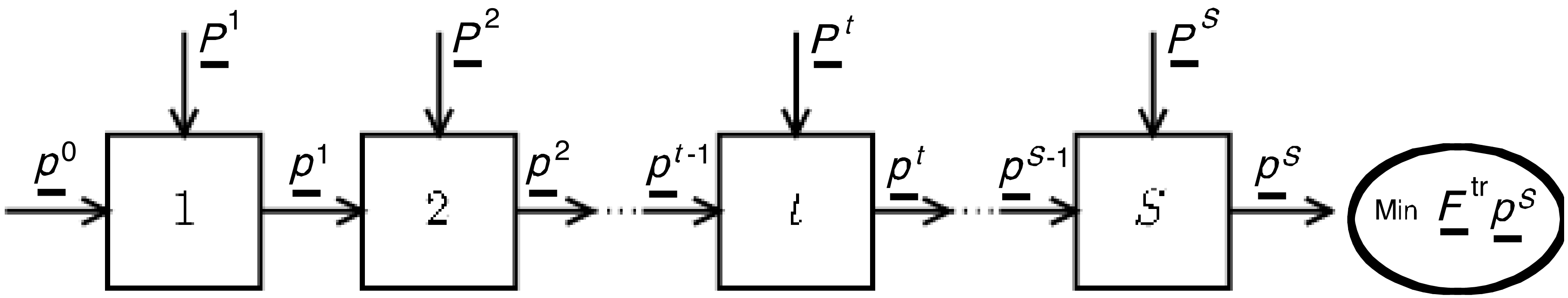


Optimizing Acceptance Probabilities

- a typical vertex of the simplex $(1, 1, 1, \ldots, 1, 0, 0, \ldots, 0)$

- linear criterion:
  optimum lies on vertex of simplex, i.e. threshold accepting

- iterate along path:



Extremal Optimization

- invented by Böttcher and Percus
- a state has internal structure

$$i = \left\{ i_1, i_2, i_3, \cdots, i_j, \cdots, i_{n-1}, i_n \right\}$$

- each DoF may have several values

$$i_j \in i^{(j)} = \left\{ i^{(j)}{}_1, i^{(j)}{}_2, i^{(j)}{}_3, \cdots, i^{(j)}{}_{m-1}, i^{(j)}{}_m \right\}$$

- example: spin configuration
  - each spin is a DoF
  - Ising spins: only two values $s_j \in s^{(j)} = \left\{ -1, 1 \right\}$

Extremal Optimization: the Algorithm

- In a state $i = \{i_1, i_2, i_3, \cdots, i_{n-1}, i_n,\}$

  determine a fitness for each DoF: $\lambda_j(i_j)$

  – example: for spins take the local field

$$H = \sum_{<i,j>} J_{i,j}\, s_i\, s_j + \sum_j h_j\, s_j = \sum_j \lambda_j\, s_j$$

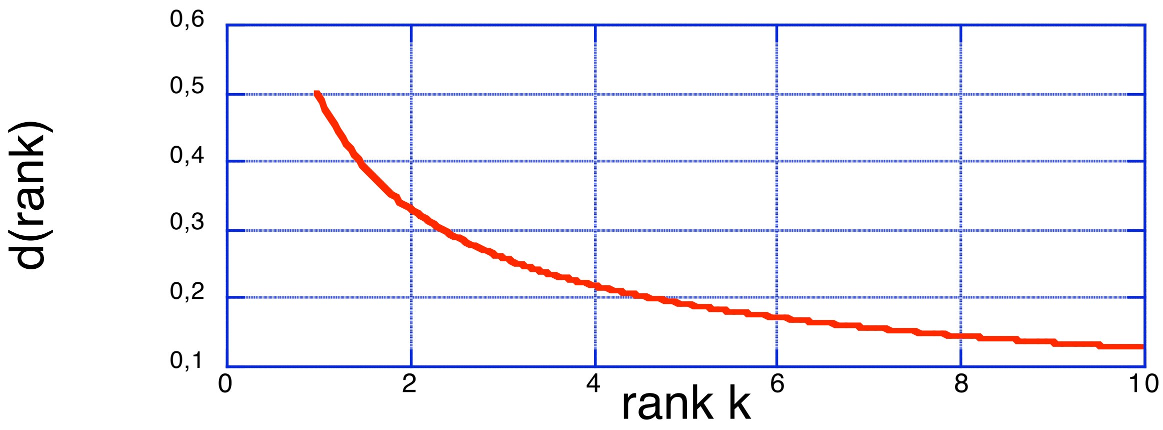
- rank the DoFs with respect to their fitness
- change the worst one to one of its other values randomly

Extremal Optimization: the Algorithm

- EO creates Marcov-process
- EO walks fast through state space:

  there are no rejected moves

- tau-EO: choose DoF according to a rank selection probabilty



$$d^t(k) \propto k^{-\tau}$$

Optimal Rank Selection Probability for EO

- What is the best rank selection probability for Extremal Optimization ??

- Is a power law distribution (with no scale) on the ranks the best ??

- Constraints:

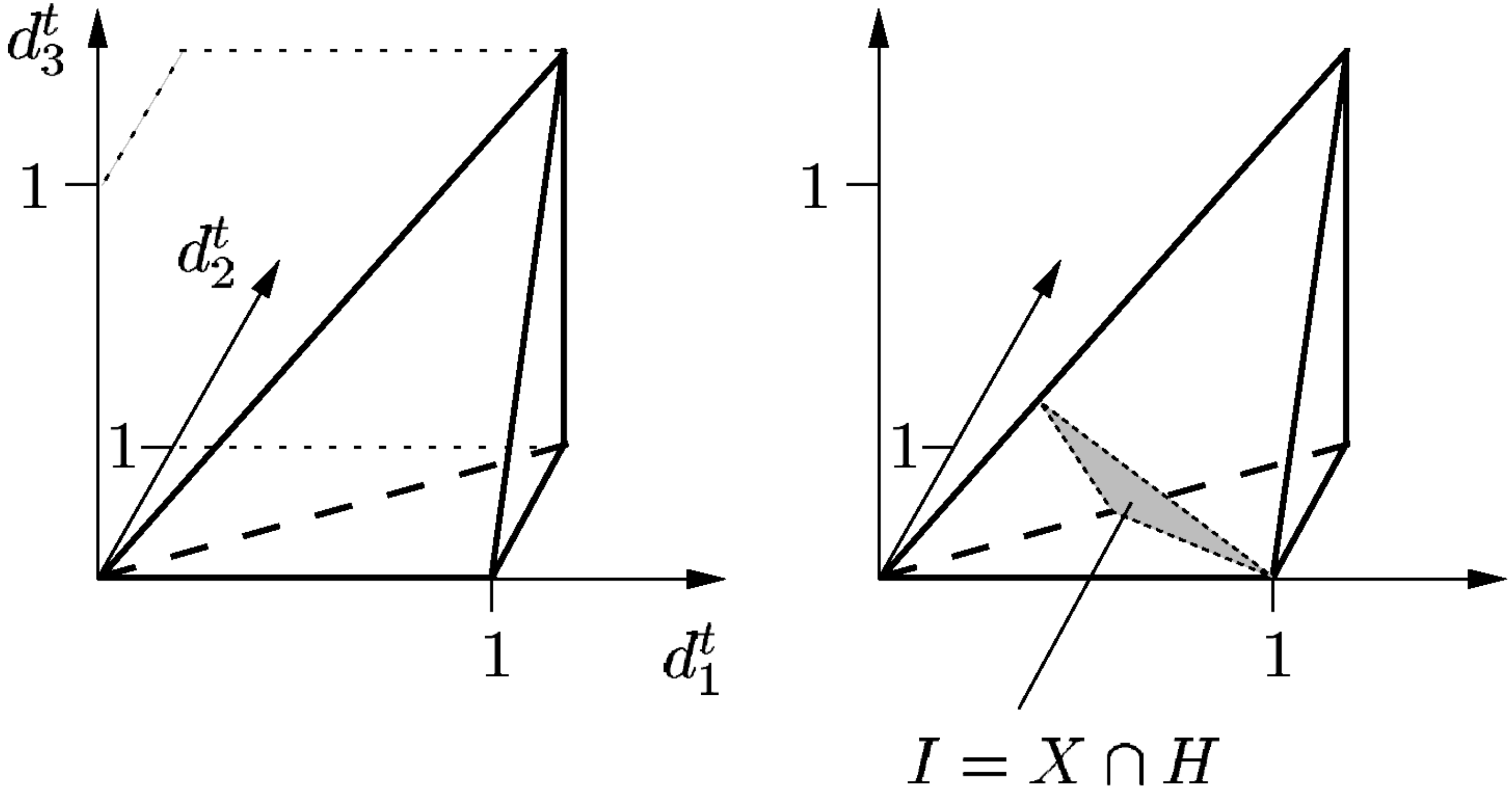
  - all time steps independent

  - $1 \geq d^t(1) \geq d^t(2) \geq \dots$

  - normalization $\sum_i d^t(k_i) = 1$

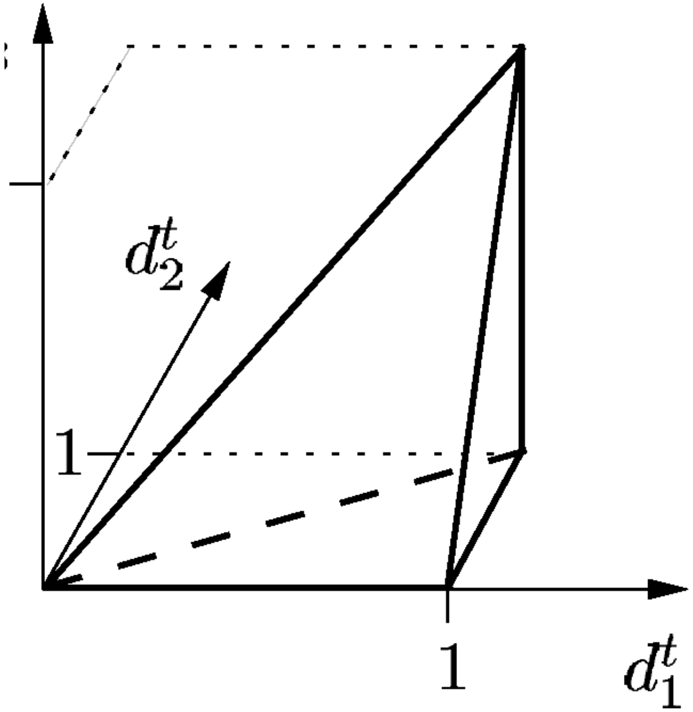
Optimal Rank Selection Probability for EO

- Problem: optimize linear criterion over simplex



Optimal Rank Selection Probability for EO

- result: for linear criteria a step disribution is the best !!



1

$(1, 0, \ldots, 0)^{tr}$

$(1/2, 1/2, 0, \ldots, 0)^{tr}$

$(1/3, 1/3, 1/3, 0, \ldots, 0)^{tr}$

$\ldots$

$(1/n, 1/n, \ldots, 1/n)^{tr}$

Fitness Threshold Accepting

Outlook and Open Problems

- Optimal thresholds
  - Threshold Accepting
  - Fitness Threshold Accepting

- Controlled dynamics on energy landscapes

- Protein folding
  - Functional relevant minima

Summary

- Spin glas dynamics
  - aging experiments
- Complex state space dynamics
- Optimal schedules for SA
- Best strategy
  - Threshold Accepting
  - Fitness Threshold Accepting