Meccanica statistica dei problemi di ottimizzazione

Federico Ricci-Tersenghi
Dip. Fisica, Università Sapienza di Roma

collaborating with (over the years)
R. Zecchina, M. Weigt, S. Franz, M. Leone, A. Braunstein,
A. Montanari, W. Barthel, A. Hartmann, M. Mézard,
G. Parisi, G. Semerjian, F. Krzakala, L. Zdeborova

Problem definition

Optimization problem

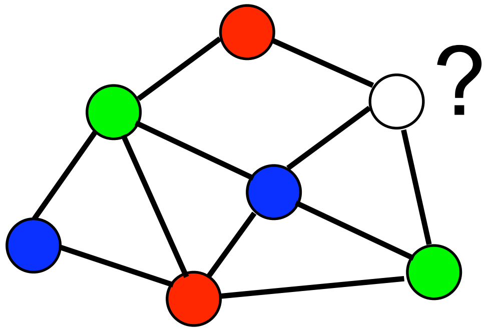
Find a configuration minimizing a cost function
$H(\vec{\sigma})$ = number of violated constraints

With $H_{\min} = 0$

Constraint Satisfaction Problem

Find a configuration of
N variables satisfying *M* constraints

q-colorability (q-COL) of a graph



N q-states Potts variables $\sigma_i \in \{1, 2, \ldots, q\}$

M pairwise interactions avoiding monochromatic edges

$$H(\vec{\sigma}) = \sum_{\langle ij \rangle} \delta_{\sigma_i, \sigma_j} \longleftarrow$$

counts the number of edges connecting vertices of the same color

K-satisfiability (K-SAT)

N binary variables $\sigma_i \in \{-1, 1\}$

M constraints involving *K* variables each

each constraint (clause) prohibits 1 among the 2^K configurations of the *K* variables it contains, e.g.

$(\sigma_7 \vee \bar{\sigma}_4 \vee \sigma_{13})$ forbids $\sigma_7 = \mathrm{F},\ \sigma_4 = \mathrm{T},\ \sigma_{13} = \mathrm{F}$

$$H(\vec{\sigma}) = \sum_{a=1}^{M} \left| \frac{\sigma_{i_a(1)} - J_{a,1}}{2} \frac{\sigma_{i_a(2)} - J_{a,2}}{2} \dots \frac{\sigma_{i_a(K)} - J_{a,K}}{2} \right|$$

Looking for hard instances...

- Benchmarks for solving algorithms
- What makes an instance hard to solve?

- Worst vs. typical case analysis
- These problems are NP-complete
  - hard instances do exist
  - need to find an ensemble concentrated on these

...in the case of SAT

- K-SAT with *K*>2 is NP-complete (Cook ’71) but...
- ...formulas from the fluctuating *K* ensemble are typically easy to be solved.
- Hint: the hardness of a constraint is $\simeq 2^{-K}$

  satisfy first constraints with the smallest *K*
- This ensemble undergoes a phase transition, but when a solution exists it is typically easy to find it

Random K-SAT

- *M* clauses (constraints) of fixed length *K*
- For each clause:
  - choose randomly *K* indices $i_a(1), \ldots, i_a(K)$
  - choose randomly $J_{a,i} = \pm 1$ with prob. 1/2

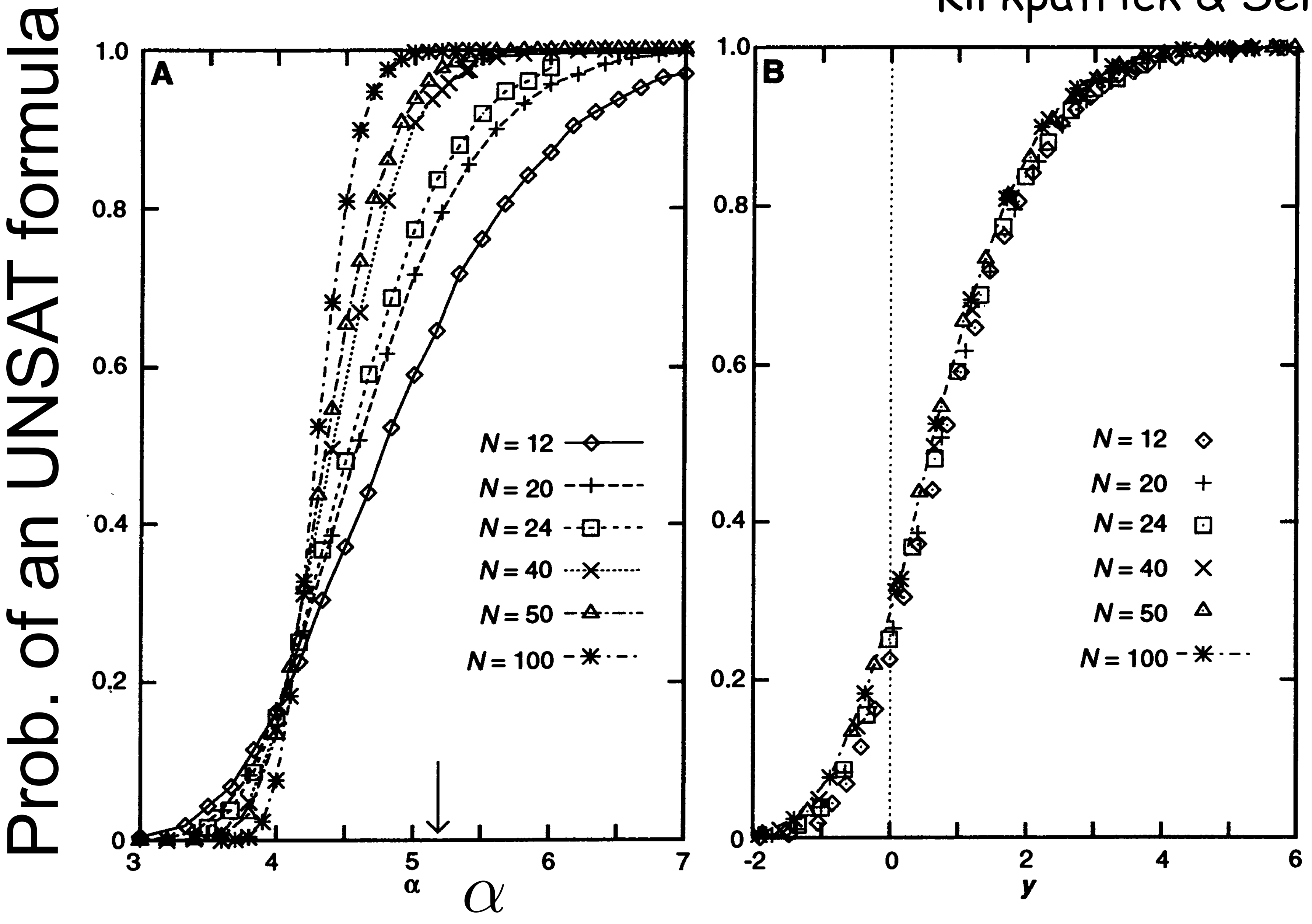
$$(\sigma_7 \vee \bar{\sigma}_4 \vee \sigma_{13}) \wedge (\sigma_{10} \vee \bar{\sigma}_{13} \vee \bar{\sigma}_2) \wedge \ldots$$

Random CSP

- *random q-col*
  - q-coloring a random graph with *M* links
- *random K-SAT*
  - *M* randomly generated clauses (constraints) of fixed length *K*

$$\boxed{\alpha = M/N}$$

SAT/UNSAT phase transition



Kirkpatrick & Selman, Science '94

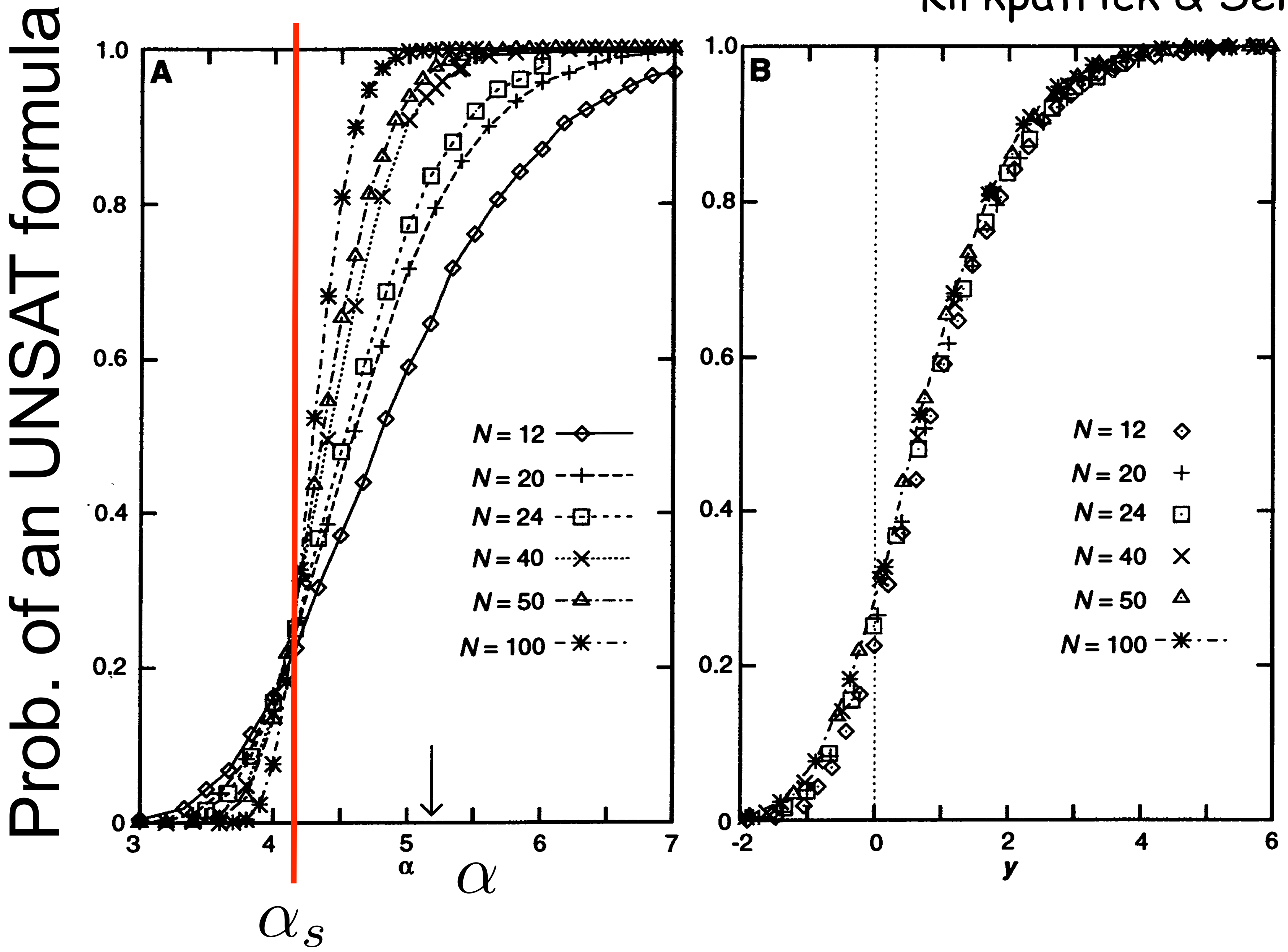
random
3-SAT

$$\alpha_s \sim 4.17$$

$$\nu \sim 1.5$$

SAT/UNSAT phase transition

Kirkpatrick & Selman, Science '94



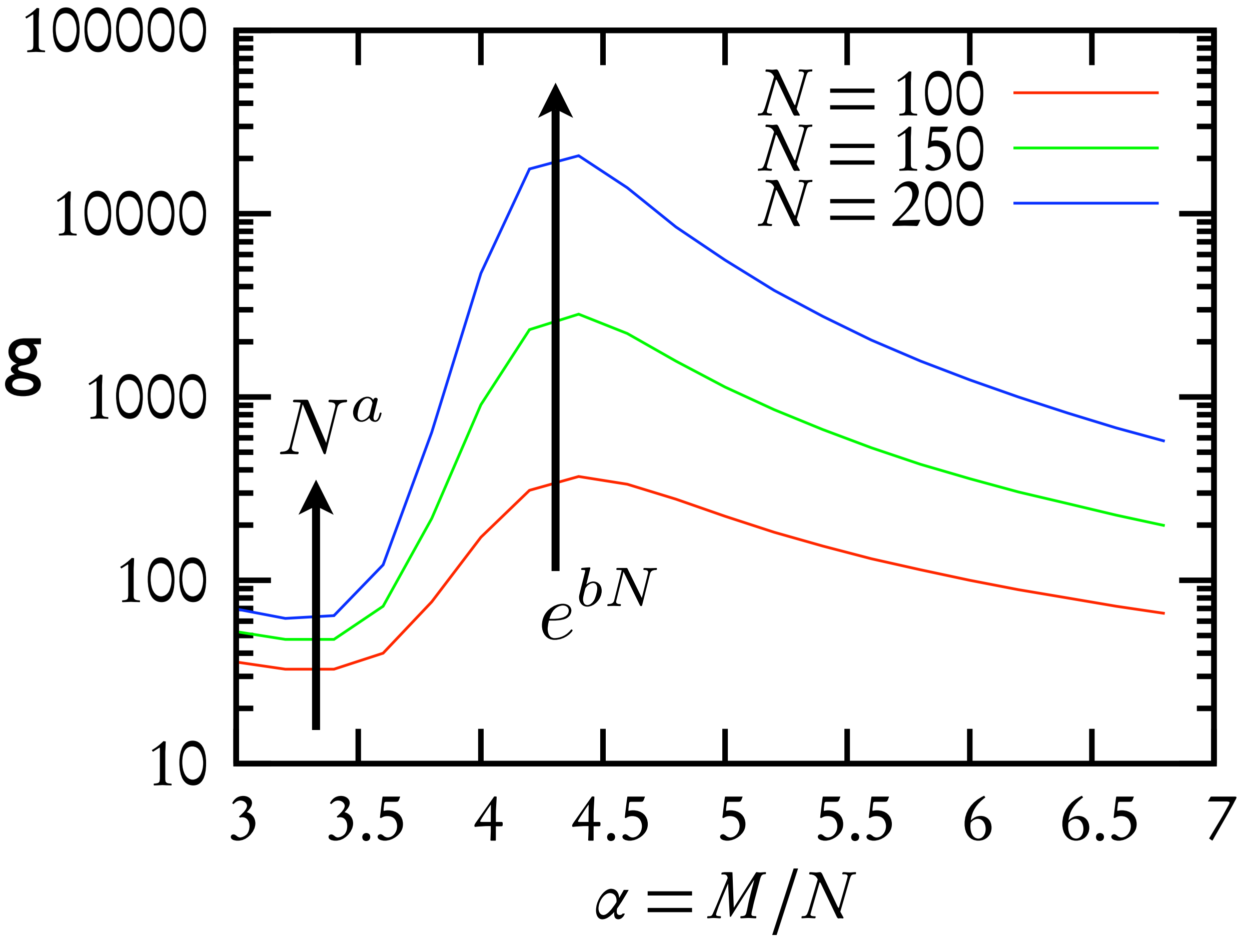
random
3-SAT

$$\alpha_s \sim 4.17$$

$$\nu \sim 1.5$$

Connection to computational complexity

Running times



Using a complete solving algorithm (DPLL)

A good ensemble is random K-SAT with *K*>2 close to the critical point.
QS: Why? General rules for producing hard instances? What happens by mixing *K* values?

Rigorous results

- Friedgut ('99): For any *K* there exist a sequence $\alpha_s(N)$ such that for $N \to \infty$

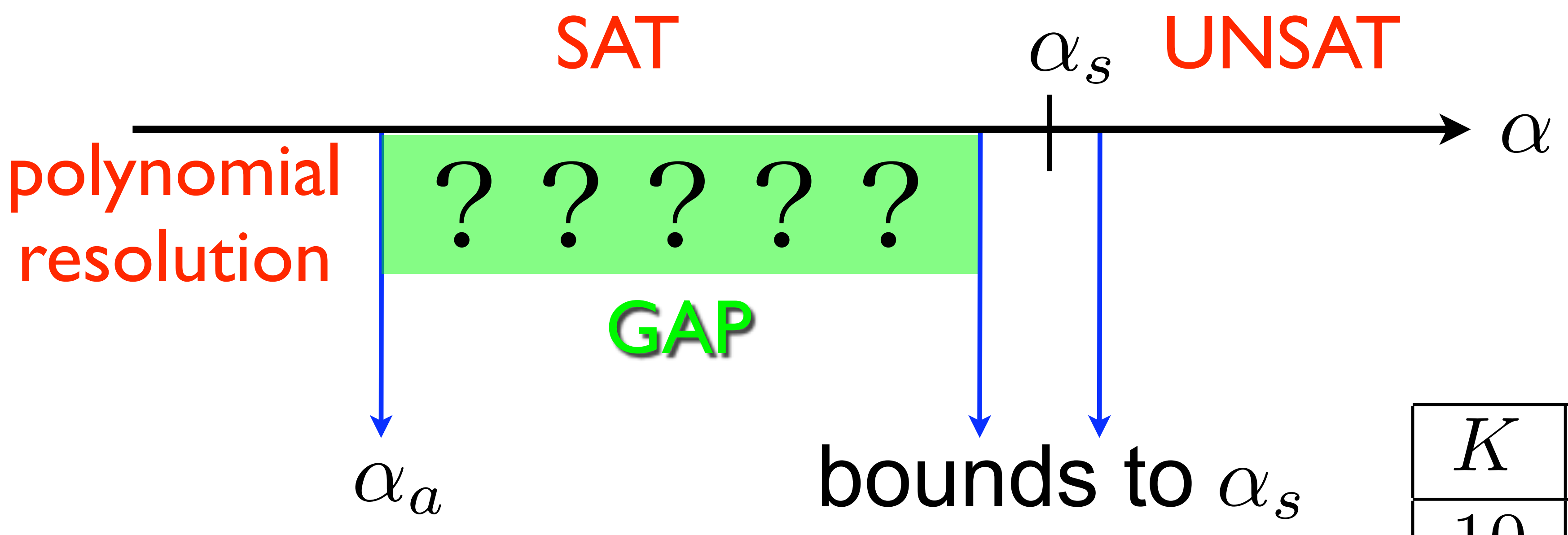
  $$\begin{aligned} &P_{\mathrm{SAT}}\big(M/N = \alpha_s(N) - \varepsilon\big) \to 1 \\ &P_{\mathrm{SAT}}\big(M/N = \alpha_s(N) + \varepsilon\big) \to 0 \end{aligned} \qquad \forall \varepsilon > 0$$

  Numerically $\alpha_s(N) \to \alpha_s$
  Rigorously only bounds to α_s are known.

- All provably linear time convergent algorithms stop working at some α_a , well before α_s
  E.g. for large *K*

  $$\alpha_a \leq \frac{\ln K}{K} 2^K \qquad \alpha_s \simeq 2^K$$

A big gap!



SAT

α_s UNSAT

α

polynomial resolution

$? ? ? ? ?$

GAP

α_a

bounds to α_s

K	α_a	α_s
10	172.65	707 ± 2
20	95263	726813 ± 4

QS: What happens in the gap?
Can we find an algorithm with $\alpha_a \simeq \alpha_s$?

Stat. Mech. approach

$$P_{\mathrm{GB}}(\vec{\sigma}) = \frac{e^{-\beta H(\vec{\sigma})}}{Z(\beta)} = \frac{1}{Z(\beta)} \prod_{a=1}^{M} \psi_a \left(\sigma_{i_a(1)}, \dots, \sigma_{i_a(k)}\right)$$

compatibility functions
(inference problems)

Limit $T \to 0, \beta \to \infty$

$$\mu(\vec{\sigma}) = \frac{1}{Z} \prod_{a=1}^{M} \mathbb{I}_a \left(\sigma_{i_a(1)}, \dots, \sigma_{i_a(k)}\right)$$

indicator functions

number of solutions

Stat. Mech. approach

- Compute the free energy $f(\beta)$ of $P_{\mathrm{GB}}(\vec{\sigma})$
- Find phase transitions lowering temperature
- Compute ground states of $H(\vec{\sigma})$
  in the limit $T \to 0, \beta \to \infty$
  - is $E_{\mathrm{GS}} = 0$?
  - structure of solutions space ?
  - phase transitions varying α ?
- Connections to computational complexity...

The role of the disorder

- $H(\vec{\sigma})$ depends on the RANDOM factor graph
  - annealed average $\ln\left(\overline{Z}\right)$
  - quenched average $\overline{\ln(Z)}$
- Annealed bound: $\overline{\mathcal{N}_{\mathrm{GS}}} = 2^N(1-2^{-K})^M$

  typical GS entropy$= \dfrac{1}{N}\overline{\ln \mathcal{N}_{\mathrm{GS}}} \leq \ln(2) + \alpha \ln(1-2^{-K})$

$$\alpha_s \leq \frac{\ln(2)}{|\ln(1-2^{-K})|}$$

Stat. Mech. approach

If the factor graph is locally tree-like,
we use Bethe approximation to compute

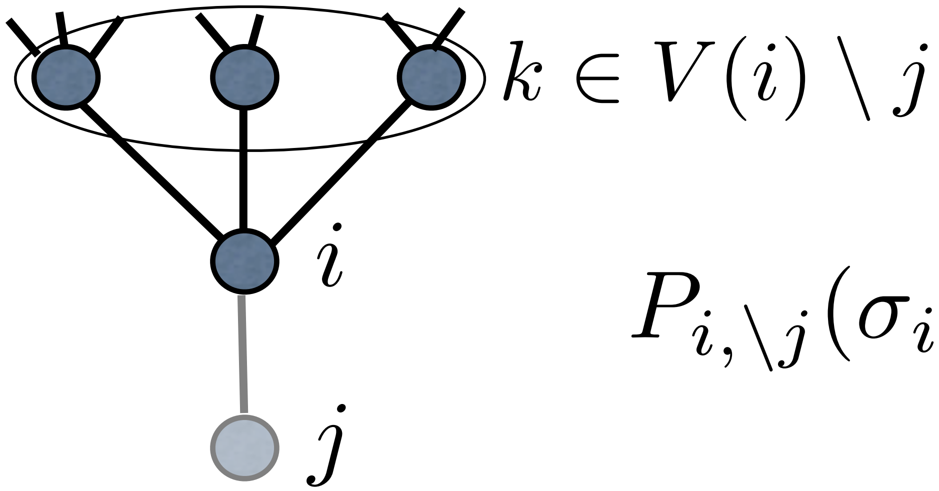
$$Z(\beta) = \sum_{\vec{\sigma}} e^{-\beta H(\vec{\sigma})} = e^{-\beta f N}$$

$$Z = \sum_{\vec{\sigma}} \prod_{a=1}^{M} \mathbb{I}_a \left(\sigma_{i_a(1)}, \ldots, \sigma_{i_a(k)}\right)$$

Cavity calculation

Mézard & Parisi, EPJB '01

Compute single variable marginals in the absence of a neighbor, $P_{i,\setminus j}(\sigma_i)$. For pairwise interactions



$$P_{i,\setminus j}(\sigma_i) \propto \sum_{\{\sigma_k\}} \prod_k P_{k,\setminus i}(\sigma_k)\psi(\sigma_i,\sigma_k)$$

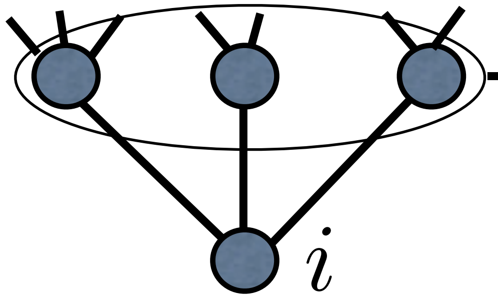
and full marginals by summing over $k \in V(i)$

$$P_i(\sigma_i) \propto \sum_{\{\sigma_k\}_{k\in V(i)}} \prod_{k\in V(i)} P_{k,\setminus i}(\sigma_k)\psi(\sigma_i,\sigma_k)$$

Estimate free-energy via Bethe aproximation

Why does cavity method work?

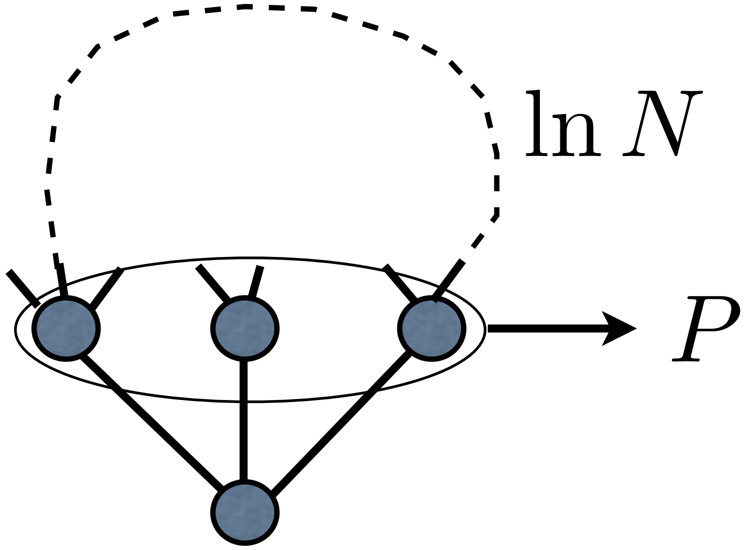
- Cavity equations are exact on a tree



$\longrightarrow$ independent without vertex i

$$P(\sigma_1, \sigma_2, \sigma_3) = P(\sigma_1)P(\sigma_2)P(\sigma_3)$$

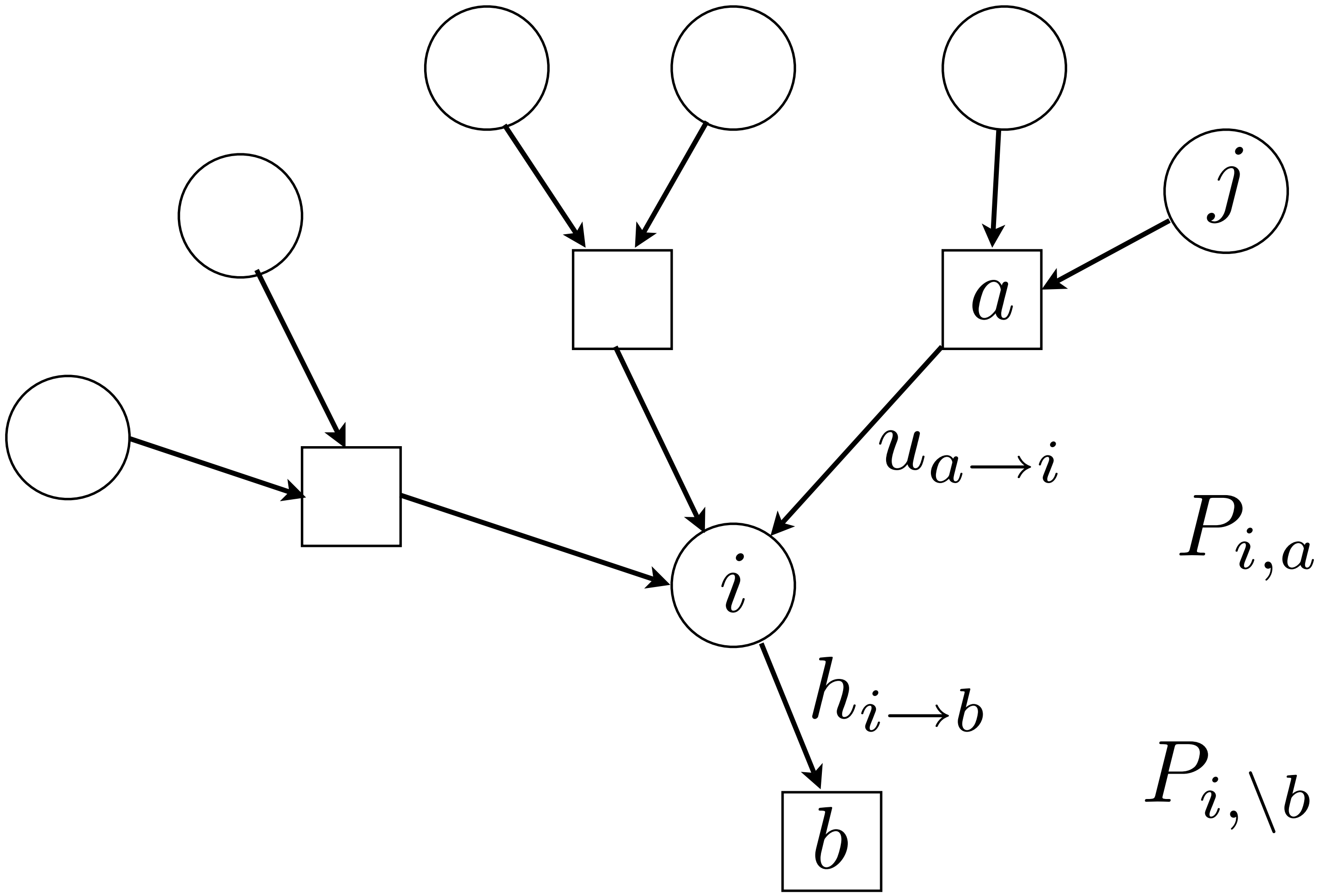
- Random structures are locally tree-like



$$\longrightarrow P(\sigma_1, \sigma_2, \sigma_3) \simeq P(\sigma_1)P(\sigma_2)P(\sigma_3)$$

When it is not true ⟶ RSB

Factor graph representation



○ variable nodes

□ check nodes (constraints)

$$P_{i,a}(\sigma_i) \propto \exp[\beta u_{a\to i}\sigma_i]$$

$$P_{i,\backslash b}(\sigma_i) \propto \exp[\beta h_{i\to b}\sigma_i]$$

$$P_i(\sigma_i) \propto \exp\left[\beta \sum_{a\in V(i)} u_{a\to i}\sigma_i\right]$$

RS cavity formalism

$$\begin{cases} h_{i\to b} & = & \displaystyle\sum_{a\in V(i)\setminus b} u_{a\to i} \\ u_{a\to i} & = & f(\{h_{j\to a}\}_{j\in V(a)\setminus i}; \vec{J}_a) \end{cases}$$

One equation
per link of the
factor graph

Free energy: Bethe approximation on the factor graph

The method works for a given instance!!

RS cavity formalism

$$\begin{cases} h_{i\to b} & = \displaystyle\sum_{a\in V(i)\setminus b} u_{a\to i} \\ u_{a\to i} & = f(\{h_{j\to a}\}_{j\in V(a)\setminus i}; \vec{J}_a) \end{cases}$$

One equation per link of the factor graph

$$\downarrow$$

$$e^{\beta u_{a\to i}\sigma_i} \propto \sum_{\{\sigma_j\}} \prod_j e^{\beta h_{j\to a}\sigma_j} \psi_a(\sigma_i, \{\sigma_j\})$$

Free energy: Bethe approximation on the factor graph

The method works for a given instance!!

Lack of factorization

Even for $|i-j| \to \infty$

$$P_{ij}(\sigma_i, \sigma_j) \neq P_i(\sigma_i) P_j(\sigma_j)$$

because of many states.

E.g. ferromagnets for $T < T_c$

$$\langle \sigma_i \sigma_j \rangle = m_0^2 \neq \langle \sigma_i \rangle \langle \sigma_j \rangle = 0$$

$$\begin{aligned} P_{ij}(\sigma_i, \sigma_j) &= \frac{1}{2} P_{ij}(\sigma_i, \sigma_j | +) + \frac{1}{2} P_{ij}(\sigma_i, \sigma_j | -) = \\ &= \frac{1}{2} P_i(\sigma_i | +) P_j(\sigma_j | +) + \frac{1}{2} P_i(\sigma_i | -) P_j(\sigma_j | -) \end{aligned}$$

Cavity with many states

For $|i-j| \to \infty$

$$P_{ij}(\sigma_i, \sigma_j) \simeq \sum_\alpha w_\alpha P_i^\alpha(\sigma_i) P_j^\alpha(\sigma_j)$$

States are exponentially many

$\mathcal{N}(E) \equiv e^{N\Sigma(E/N)} \longrightarrow \Sigma(e)$ configurational entropy or complexity

Aim: compute $\Sigma(e)$

$\Sigma(0) < 0 \Longrightarrow$ UNSAT

Counting the states

Aim: compute $\Sigma_f(f,T)$ such that $\mathcal{N}(f,T) = e^{N\Sigma_f(f,T)}$

Counting the states

Aim: compute $\Sigma_f(f,T)$ such that $\mathcal{N}(f,T) = e^{N\Sigma_f(f,T)}$

Define the replicated free-energy $\Phi(m,T)$

$$e^{-\beta m \Phi(m,T) N} \equiv \sum_{\gamma} Z_{\gamma}^{m} = \int e^{-\beta m f N + N\Sigma_f(f,T)} \mathrm{d}f$$

Counting the states

Aim: compute $\Sigma_f(f,T)$ such that $\mathcal{N}(f,T) = e^{N\Sigma_f(f,T)}$

Define the replicated free-energy $\Phi(m,T)$

$$e^{-\beta m \Phi(m,T) N} \equiv \sum_\gamma Z_\gamma^m = \int e^{-\beta m f N + N\Sigma_f(f,T)} \mathrm{d}f$$

and by the Legendre transform

$$\Sigma_f(f,T) = \beta m f - \beta m \Phi(m,T)|_{f=\partial_m(m\Phi)}$$

Counting the states

Aim: compute $\Sigma_f(f,T)$ such that $\mathcal{N}(f,T) = e^{N\Sigma_f(f,T)}$

Define the replicated free-energy $\Phi(m,T)$

$$e^{-\beta m\Phi(m,T)N} \equiv \sum_\gamma Z_\gamma^m = \int e^{-\beta m f N + N\Sigma_f(f,T)} \mathrm{d}f$$

and by the Legendre transform

$$\Sigma_f(f,T) = \beta m f - \beta m \Phi(m,T)|_{f=\partial_m(m\Phi)}$$

For $T \to 0$ with $\beta m = \mu$

$$\Sigma_e(e) = \mu e - \mu\Phi(\mu)|_{e=\partial_\mu(\mu\Phi)}$$

m is the Parisi parameter

Populations of messages

- On each link of the factor graph:
  - one message u or h per state
  - many states ⟶ population of messages
- Extra re-weighting factors depending on m

Populations of messages

- On each link of the factor graph:
  - one message *u* or *h* per state
  - many states ⟶ population of messages
- Extra re-weighting factors depending on *m*

$$Q_{a\to i}(u) \propto \int \prod_{j\in V(a)\setminus i} dP_{j\to a}(h_j)\delta\big[u - f(\{h_j\}, \vec{J}_a)\big]$$

$$P_{i\to b}(h) \propto \int \prod_{a\in V(i)\setminus b} dQ_{a\to i}(u_a)\delta\big[h - \sum_a u_a\big] \times$$

$$\times e^{-\beta m(\sum_a |u_a| - |\sum_a u_a|)}$$

A simpler case: random K-XORSAT

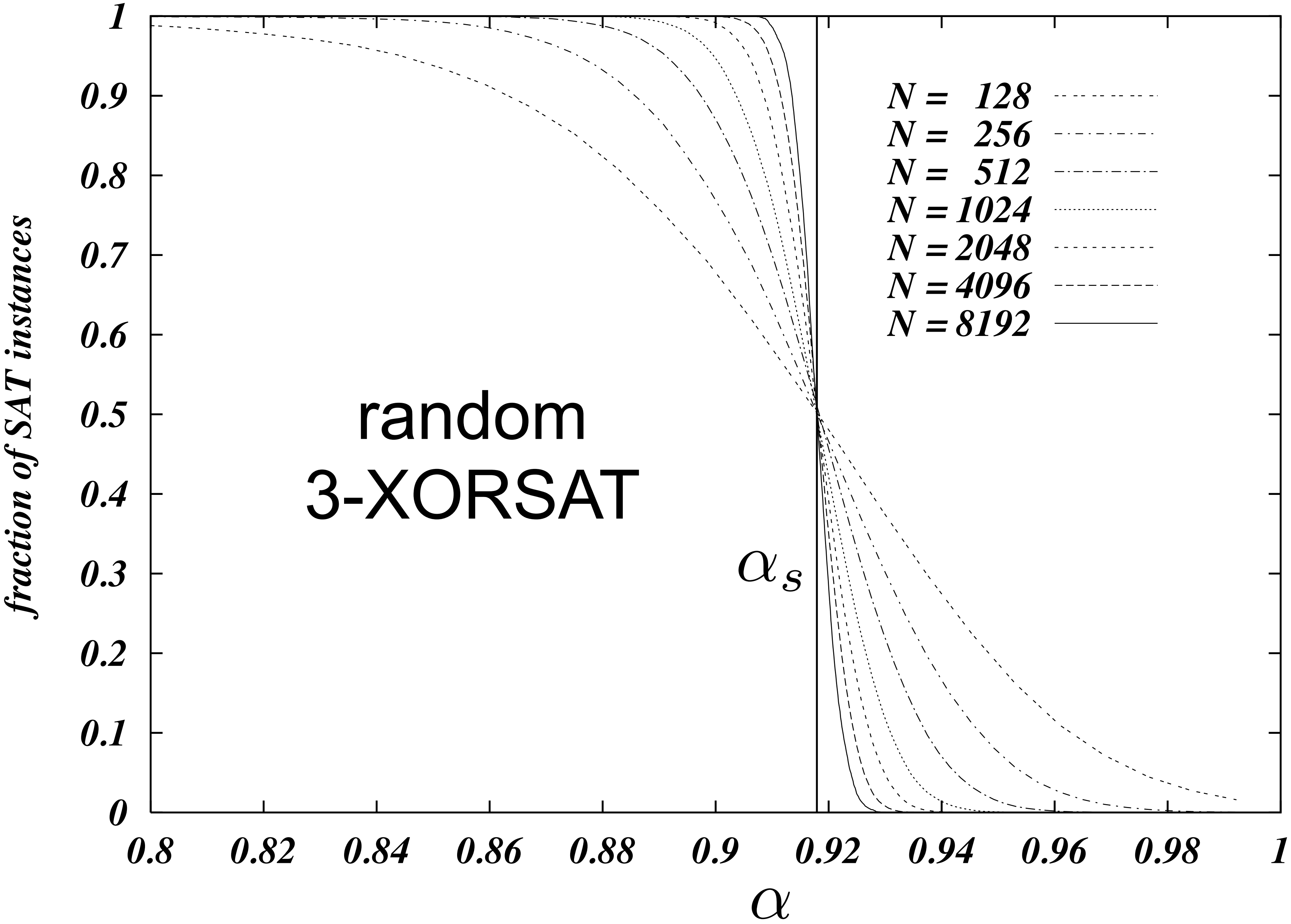
Ricci-Tersenghi, Zecchina & Weigt, PRE '01
Mézard, Ricci-Tersenghi & Zecchina, JSP '03
Cocco, Dubois, Mandler & Monasson, PRL '03

Like random K-SAT but replacing OR with XOR

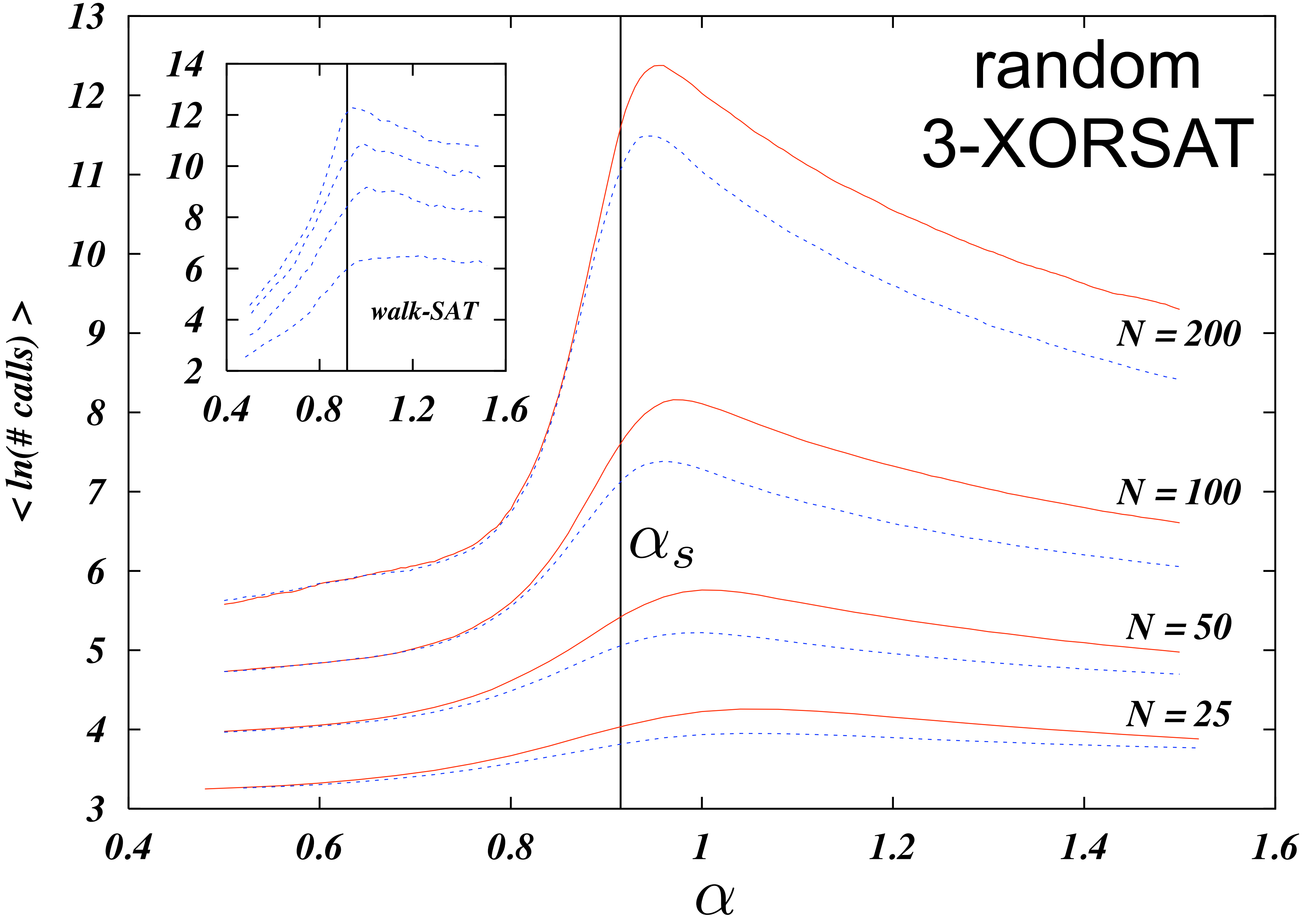
$$(\sigma_7 \oplus \bar{\sigma}_4 \oplus \sigma_{13}) \wedge (\sigma_{10} \oplus \bar{\sigma}_{13} \oplus \bar{\sigma}_2) \wedge \ldots$$

M parity checks over *N* variables

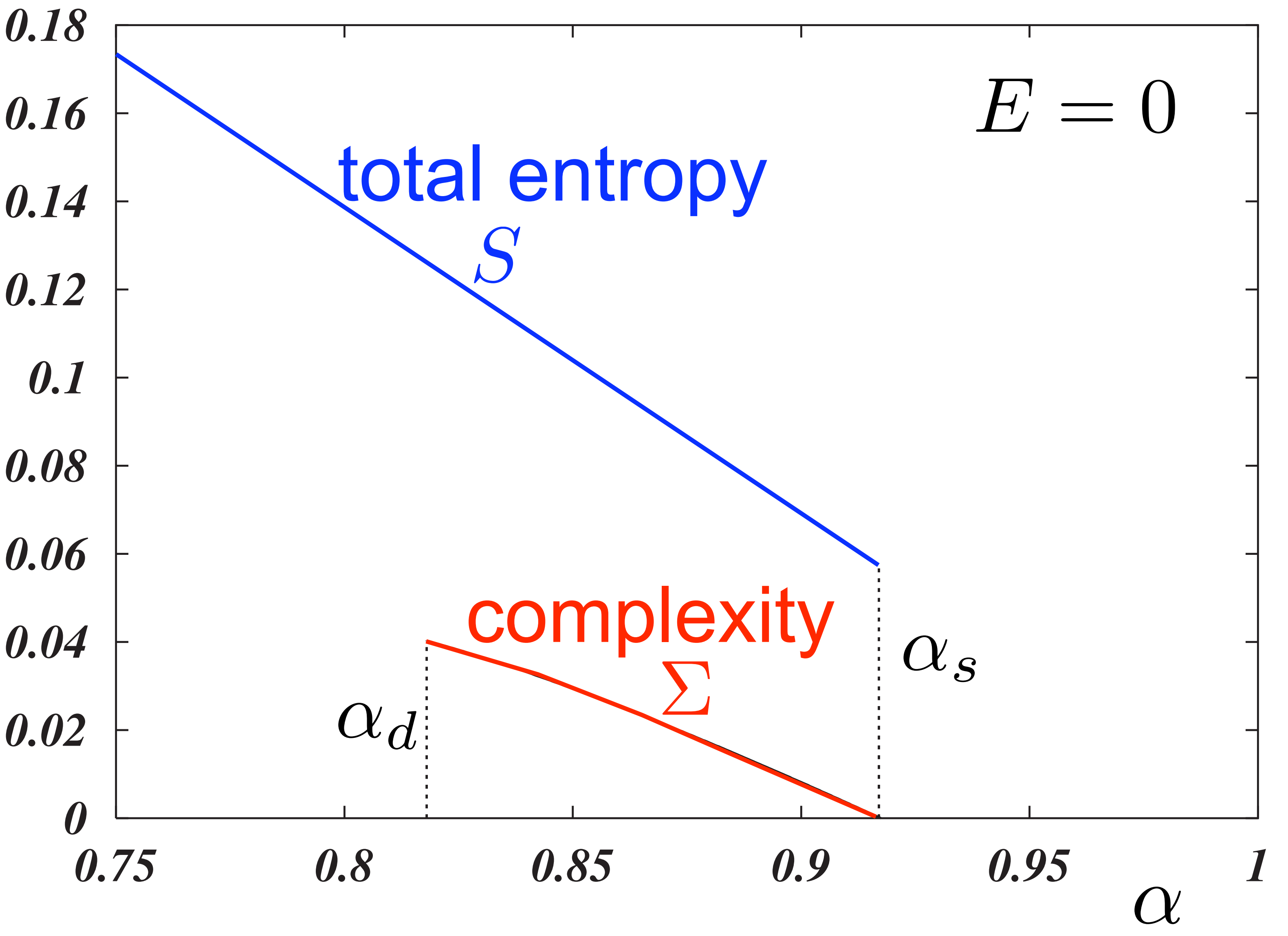
SAT/UNSAT phase transition in random K-XORSAT



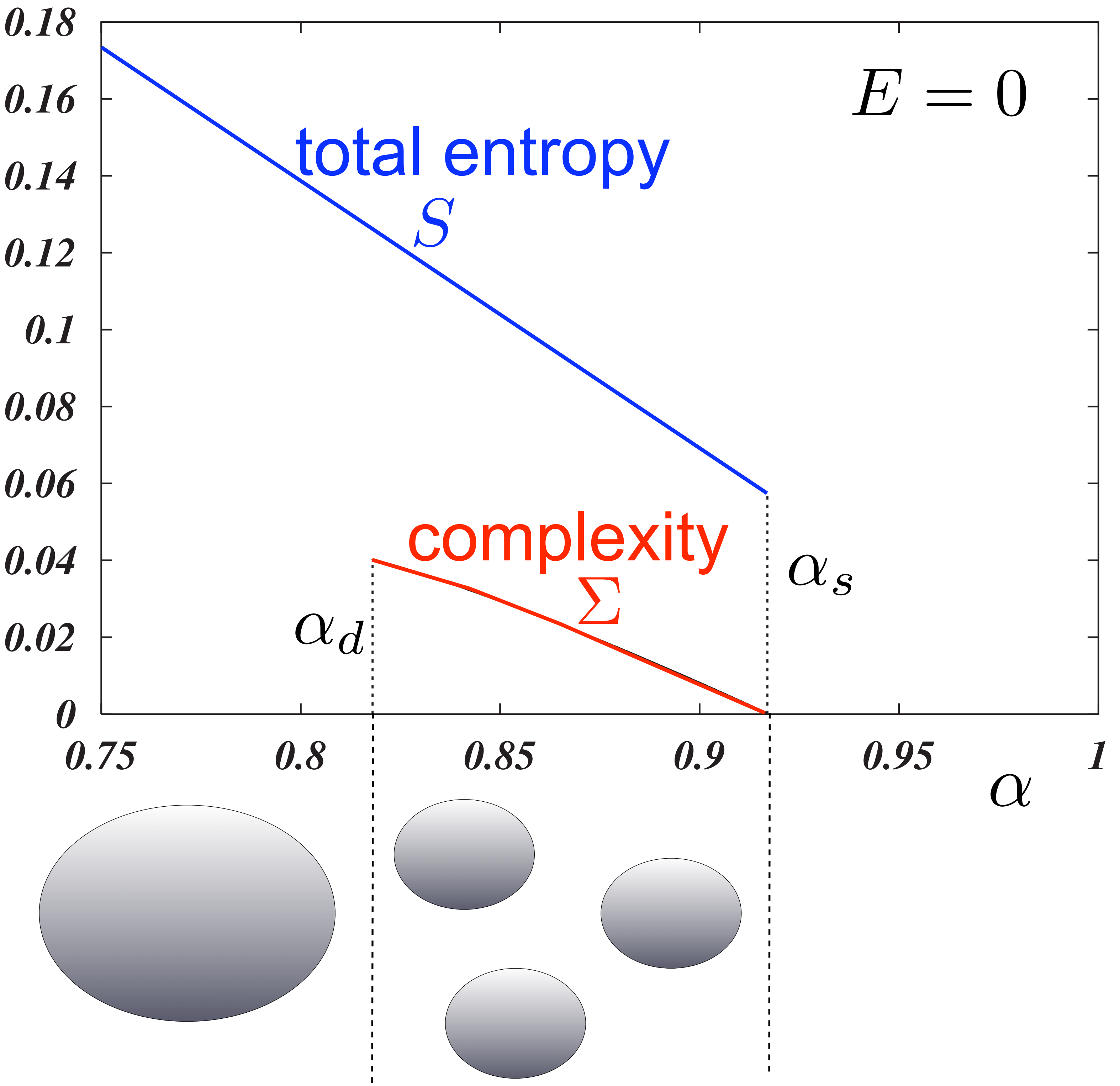
Increase of computing times



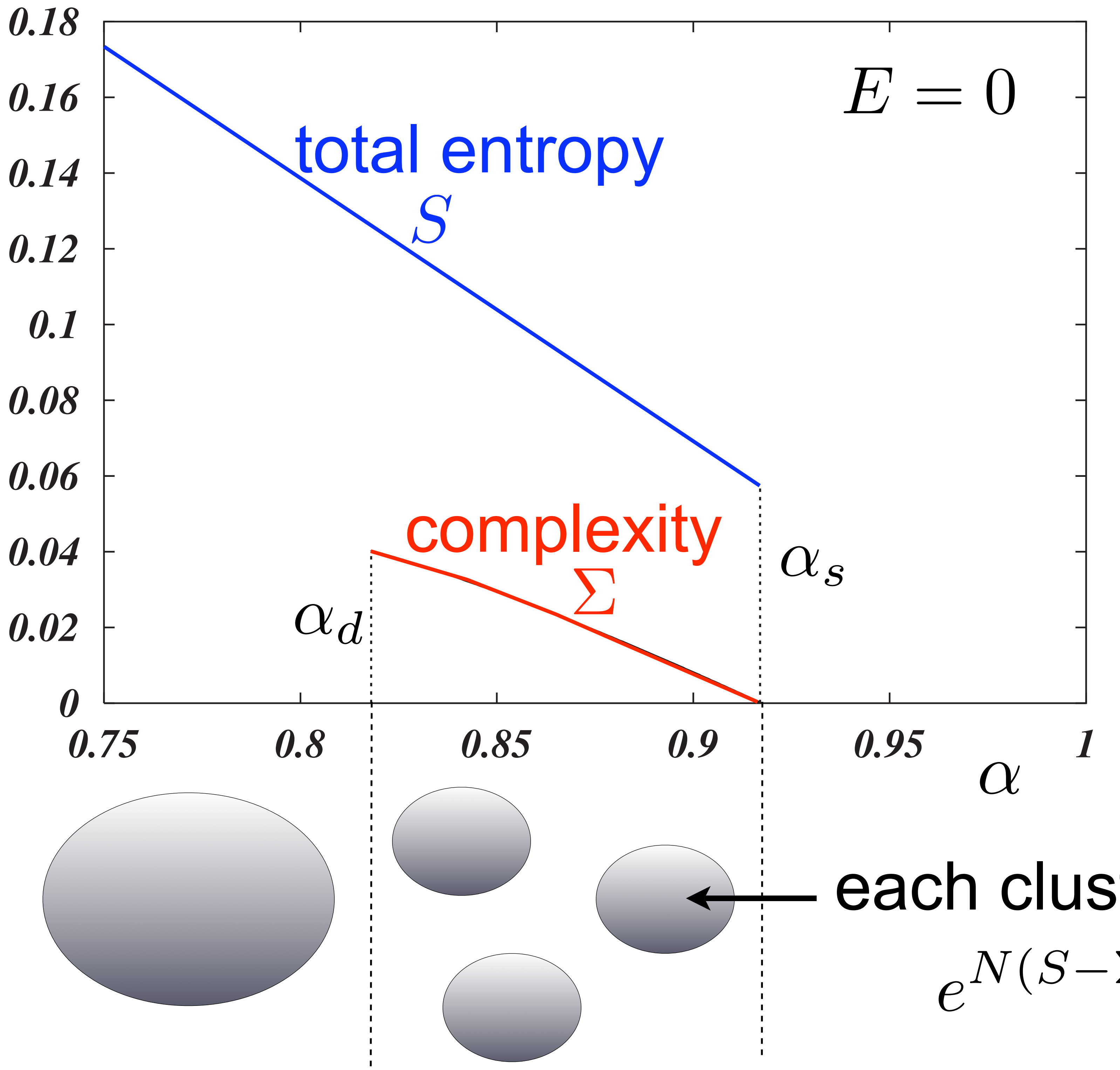
Structure of solutions space



Structure of solutions space



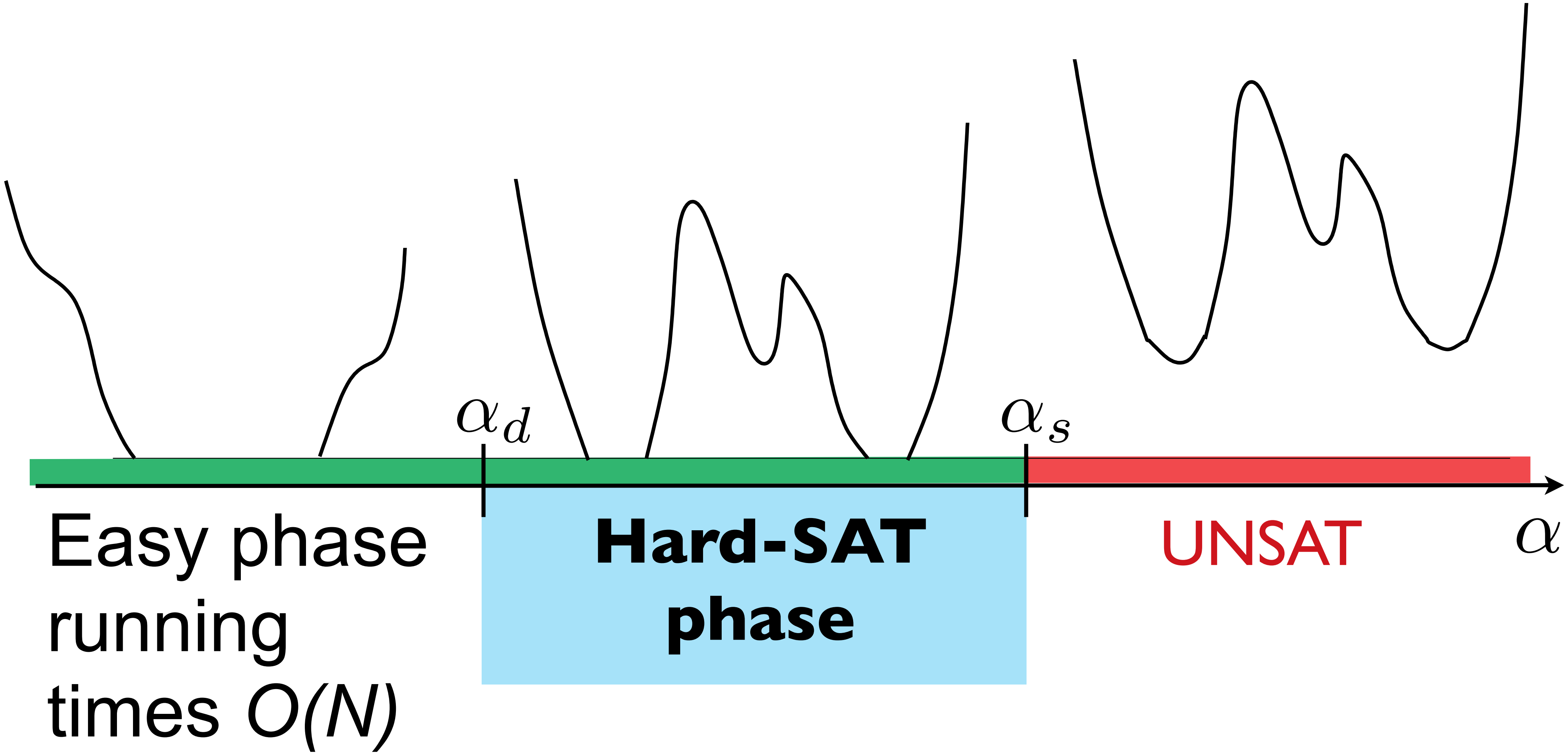
Structure of solutions space



each cluster contains

$e^{N(S-\Sigma)}$ solutions

Where are hard instances?



Easy phase running times *O(N)*

Hard-SAT phase

UNSAT

Leaf removal algorithm

- while (there exists a vertex of degree 1)
  remove it and the clause it belongs to

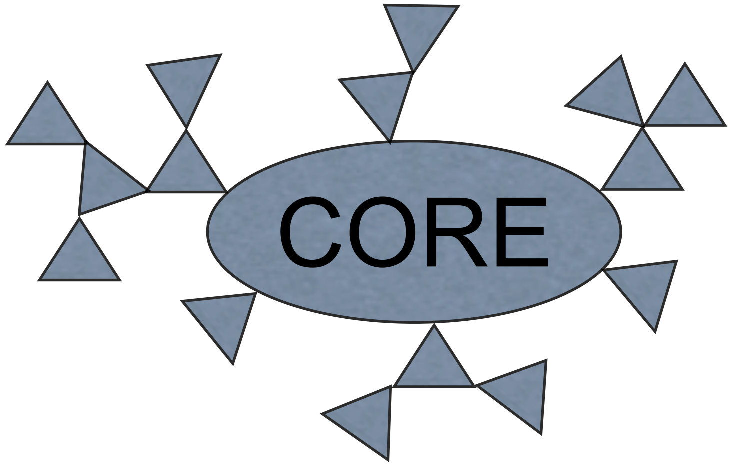
for $\alpha < \alpha_d$ $\qquad \mathcal{G} = (V, E) \to (V_c, \emptyset)$

for $\alpha \geq \alpha_d$ $\qquad \mathcal{G} = (V, E) \to (V_c, E_c)$

- reconstruction procedure for $\alpha < \alpha_d$:
  - assign to any value the variables in V_c
  - add clauses in the reverse order and assign the newly added variable to satisfy the clause

The core

For $\alpha \geq \alpha_d$

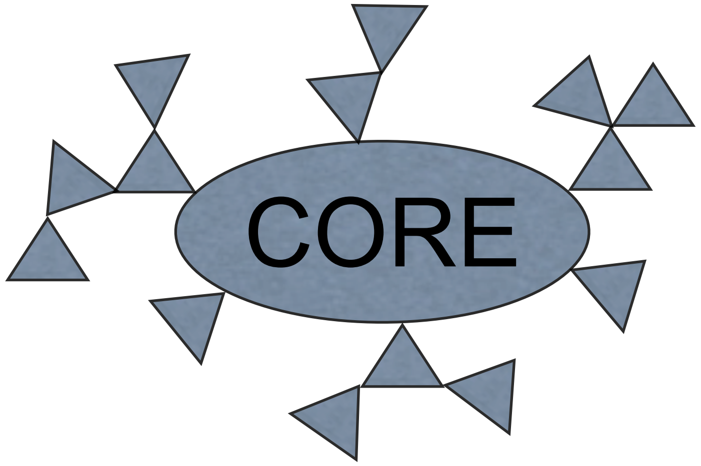


Minimum degree 2
Point-like clusters at distance *O(N)*
No sample-to-sample fluctuations:
  the annealed computation is exact
  a solution exist as long as $M_c \leq N_c$

Long range correlations: hard to find solution

The core

For $\alpha \geq \alpha_d$



Minimum degree 2

Point-like clusters at distance *O(N)*

No sample-to-sample fluctuations:

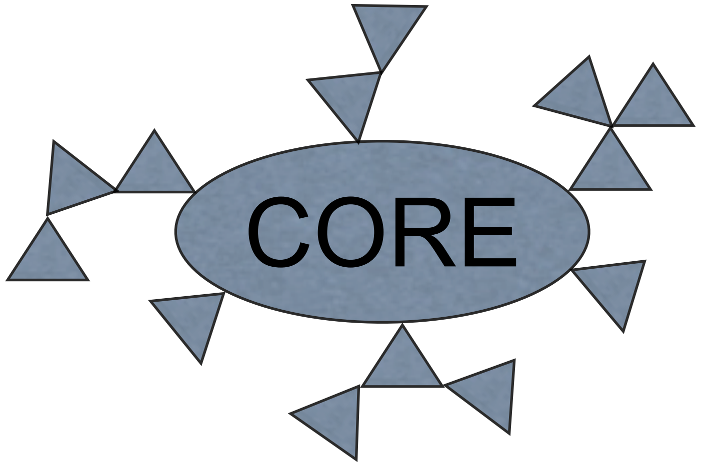
  the annealed computation is exact

  a solution exist as long as $M_c \leq N_c$

Long range correlations: hard to find solution

The core

For $\alpha \geq \alpha_d$

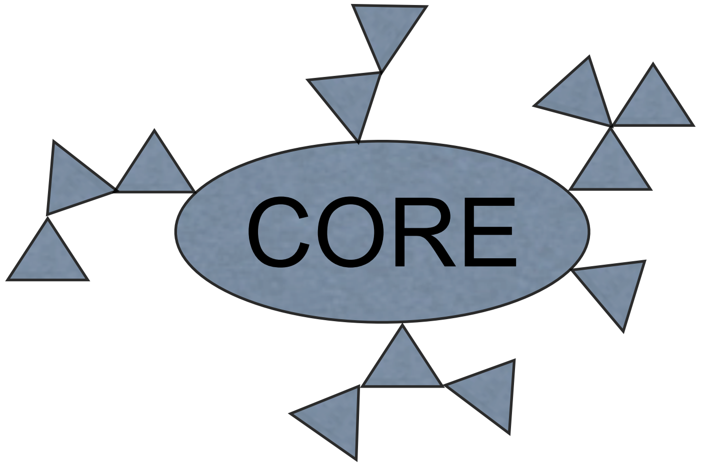


Minimum degree 2
Point-like clusters at distance *O(N)*
No sample-to-sample fluctuations:
  the annealed computation is exact
  a solution exist as long as $M_c \leq N_c$

Long range correlations: hard to find solution

The core

For $\alpha \geq \alpha_d$

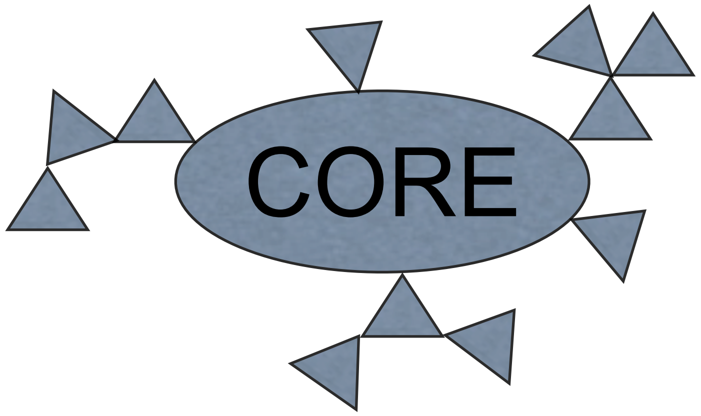


Minimum degree 2
Point-like clusters at distance *O(N)*
No sample-to-sample fluctuations:
  the annealed computation is exact
  a solution exist as long as $M_c \leq N_c$

Long range correlations: hard to find solution

The core

For $\alpha \geq \alpha_d$



Minimum degree 2
Point-like clusters at distance *O(N)*
No sample-to-sample fluctuations:
  the annealed computation is exact
  a solution exist as long as $M_c \leq N_c$

Long range correlations: hard to find solution

The core

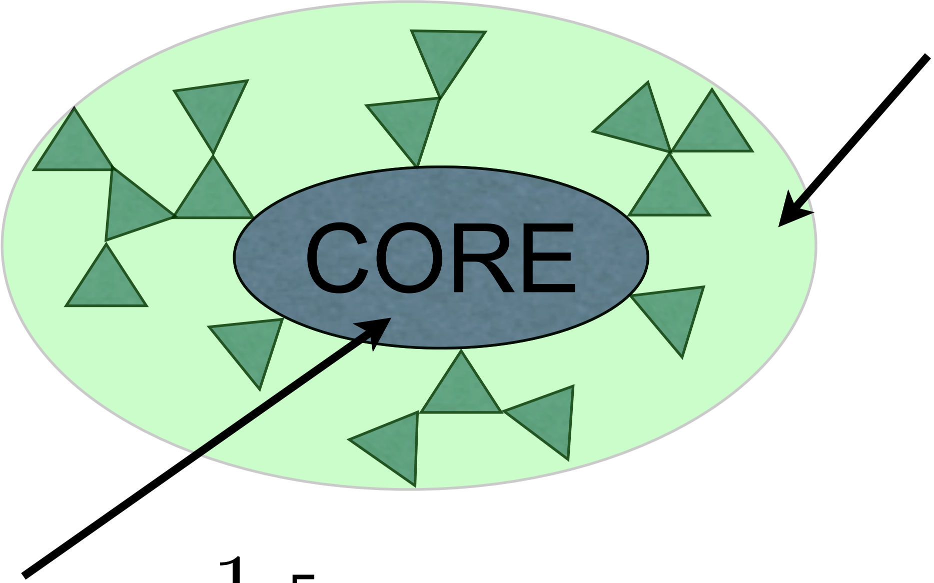
For $\alpha \geq \alpha_d$

CORE

Minimum degree 2
Point-like clusters at distance *O(N)*
No sample-to-sample fluctuations:
  the annealed computation is exact
  a solution exist as long as $M_c \leq N_c$

Long range correlations: hard to find solution

The cavity solution

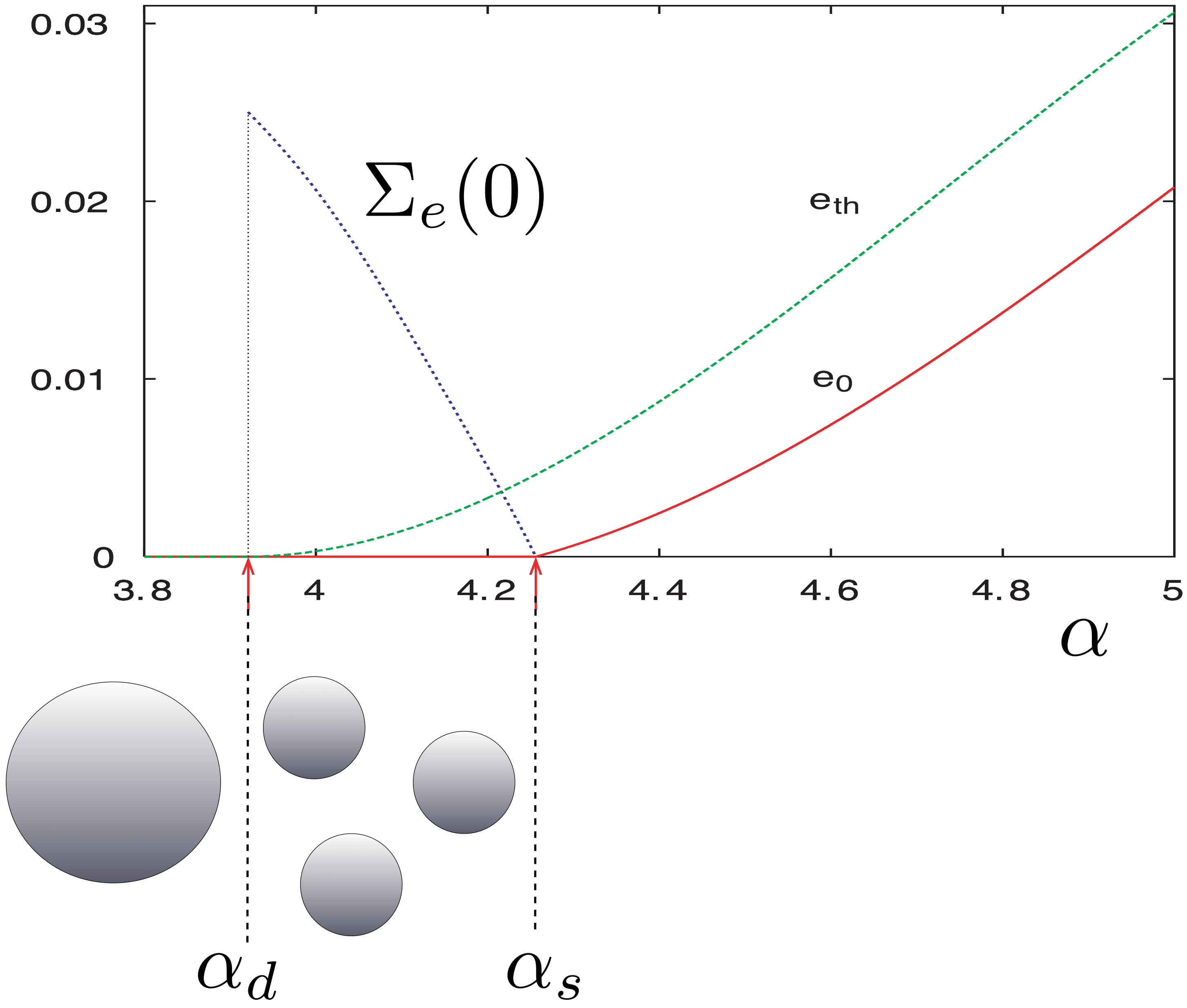


$$Q(u) = \delta(u)$$

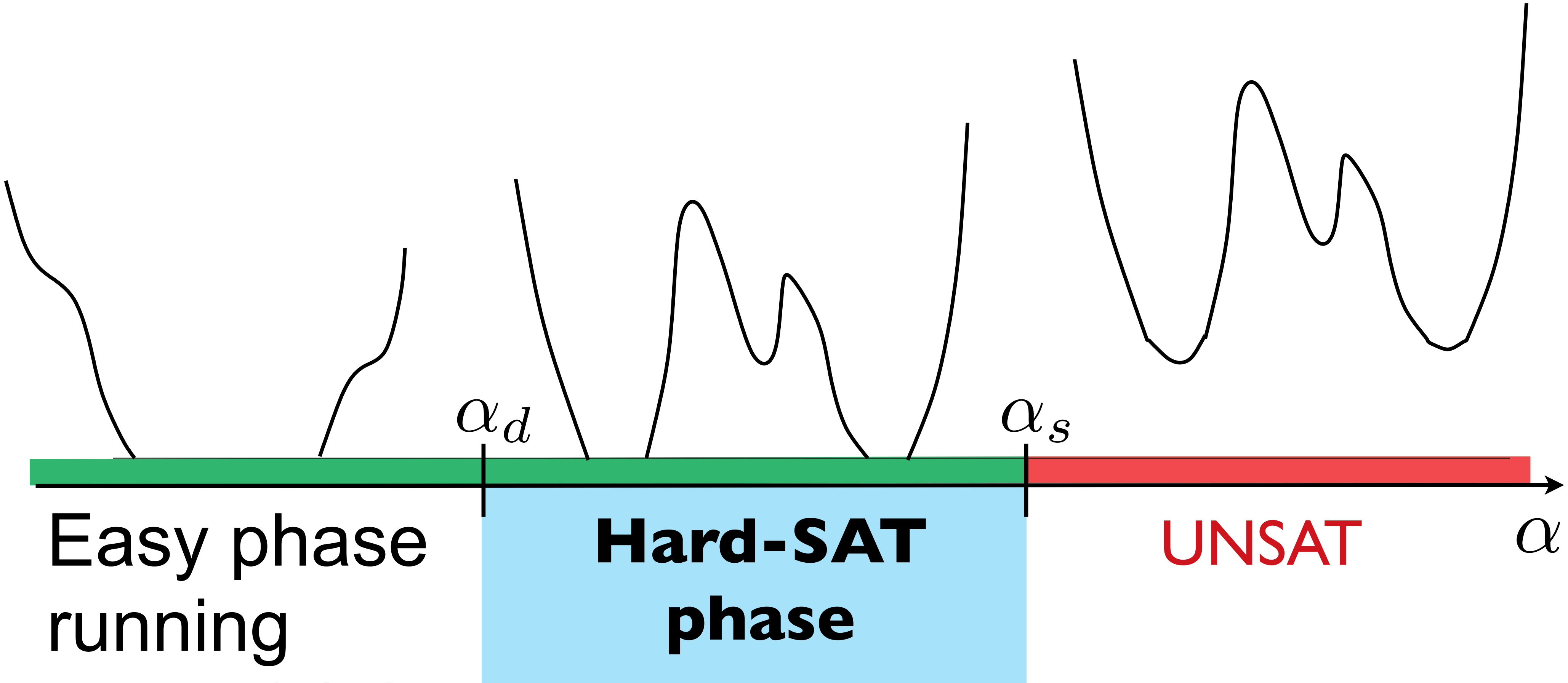
$$Q(u) = \frac{1}{2}\Big[\delta(u-1) + \delta(u+1)\Big]$$

Cavity solution for random K-SAT

Mézard, Parisi & Zecchina, Science '02



A problem with the standard picture



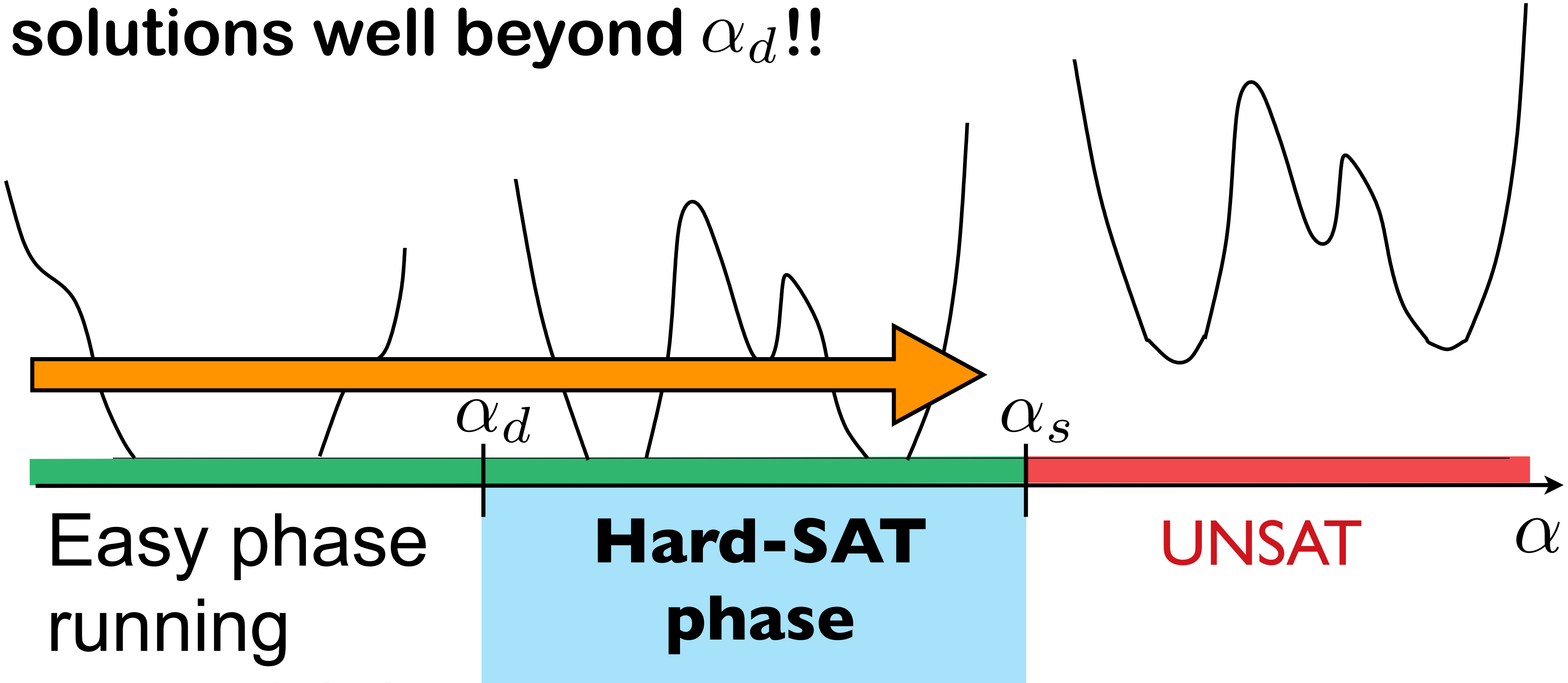
Easy phase
running
times *O(N)*

Hard-SAT
phase

UNSAT

A problem with the standard picture

Very simple algorithms can find solutions well beyond α_d!!



Easy phase running times *O(N)*

Entropic effects at very low temperatures

- taking first the limit $T \to 0$

  then $f = e - Ts$

  $Z_\gamma = e^{-\beta N f_\gamma} \simeq e^{-\beta N e_\gamma}$

  ok if $e_\gamma > 0$

  but if $e_\gamma = 0$

  $Z_\gamma = 1$ always!

Entropic effects at very low temperatures

- taking first the limit $T \to 0$

  then $f = e - \cancel{Ts}$

  $$Z_\gamma = e^{-\beta N f_\gamma} \simeq e^{-\beta N e_\gamma}$$

  ok if $e_\gamma > 0$

  but if $e_\gamma = 0$

  $Z_\gamma = 1$ always!

Entropic effects at very low temperatures

- taking first the limit $T \to 0$

  then $f = e - \cancel{Ts}$

  $$Z_\gamma = e^{-\beta N f_\gamma} \simeq e^{-\beta N e_\gamma}$$

  ok if $e_\gamma > 0$

  but if $e_\gamma = 0$
  $Z_\gamma = 1$ always!

- consider only solutions ($e_\gamma = 0$)

  $$f = -Ts \qquad Z_\gamma = e^{-\beta N f_\gamma} \simeq e^{N s_\gamma}$$

  larger clusters count more!

New replicated potential

Mézard, Palassini & Rivoire, PRL '05

Krzakala, Montanari, Ricci-Tersenghi, Semerjian, Zdeborova, PNAS '07

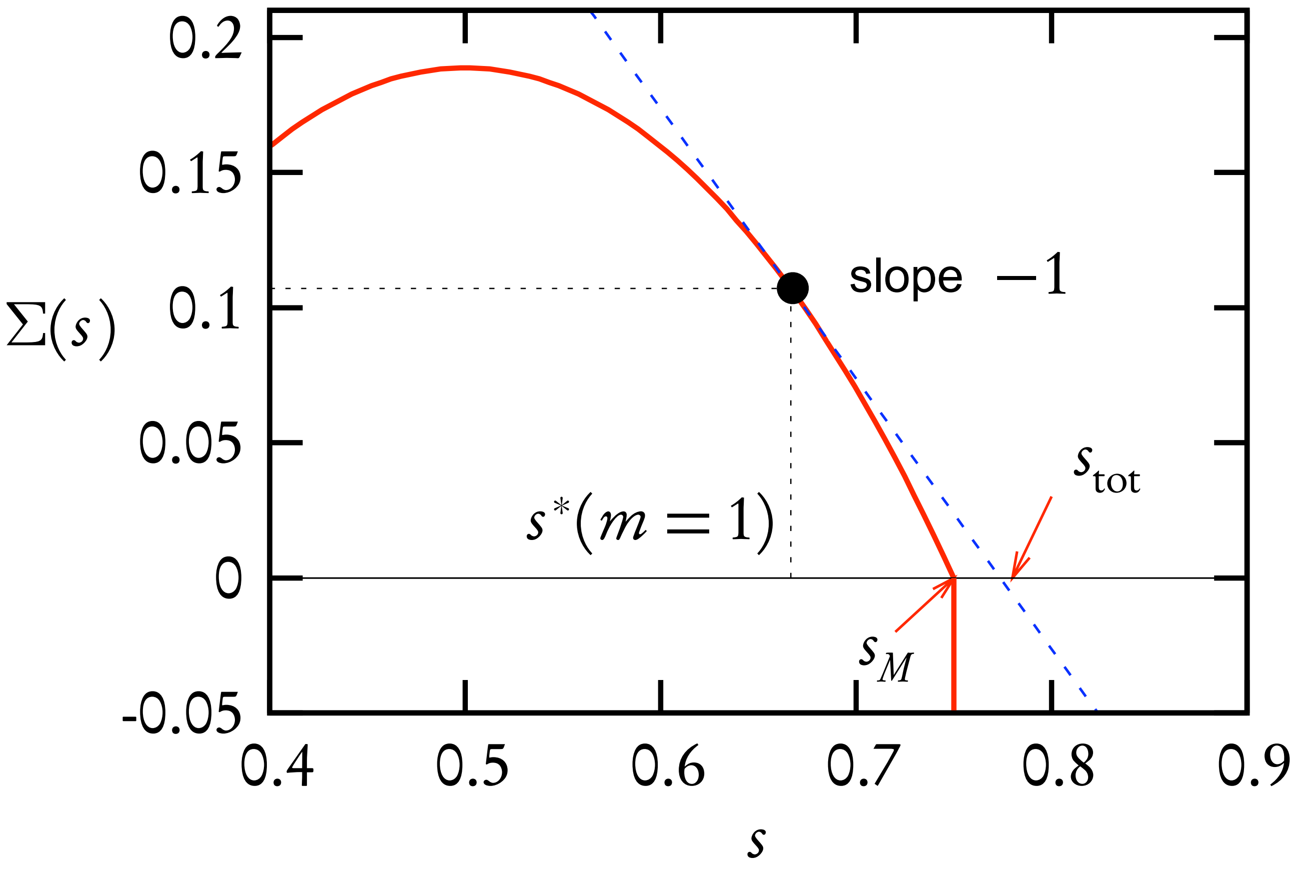
$$e^{N\Psi(m)} = \sum_{\gamma} e^{mNs_\gamma + N\Sigma_s(s_\gamma)}$$

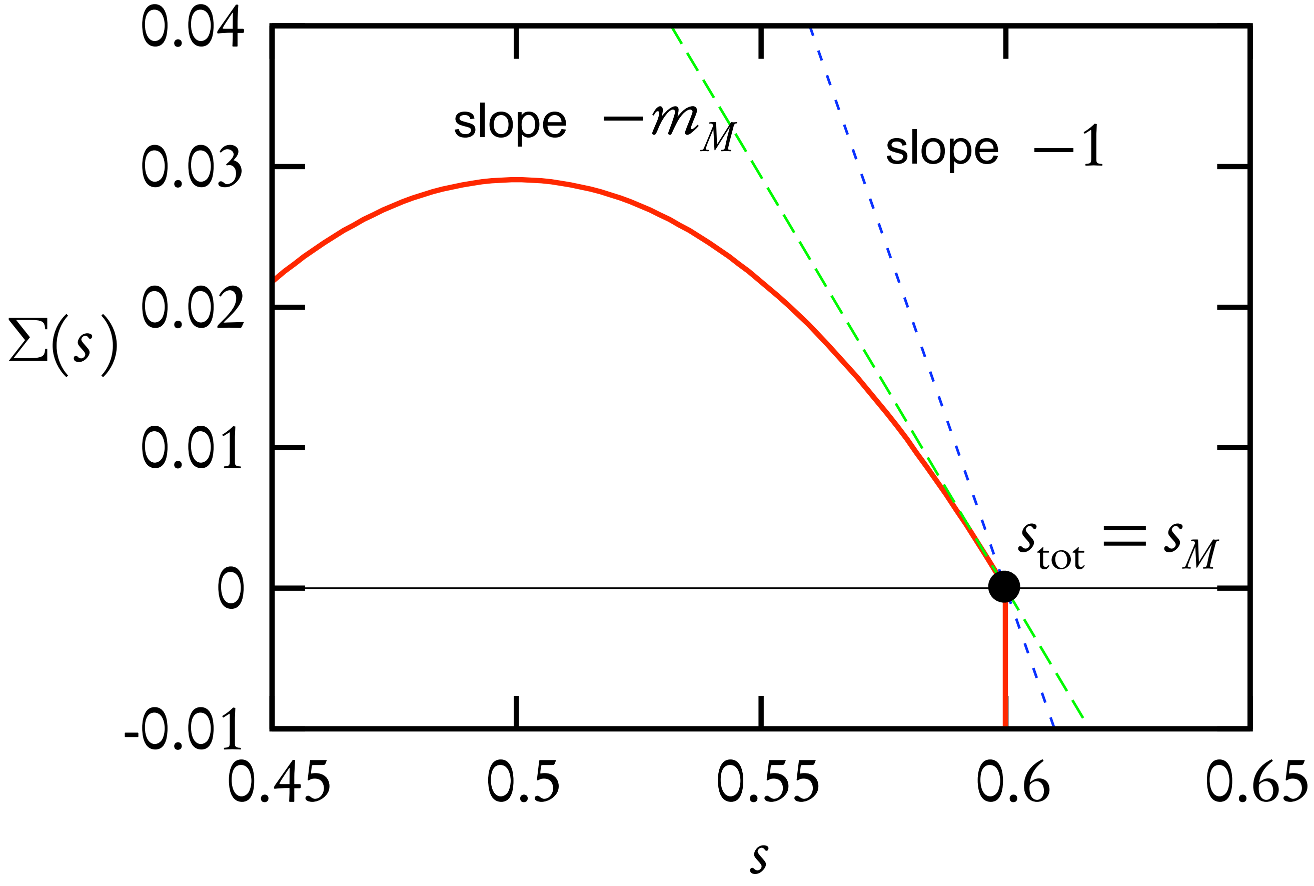
$$\Psi(m) = \max_s \left[\Sigma_s(s) + ms\right]$$

$m = 0 \longrightarrow$ most numerous clusters
(like with the energetic method)

$m = 1 \longrightarrow$ clusters dominating the measure
(if they exists, i.e. have $\Sigma > 0$)

How to compute dominating clusters of solutions

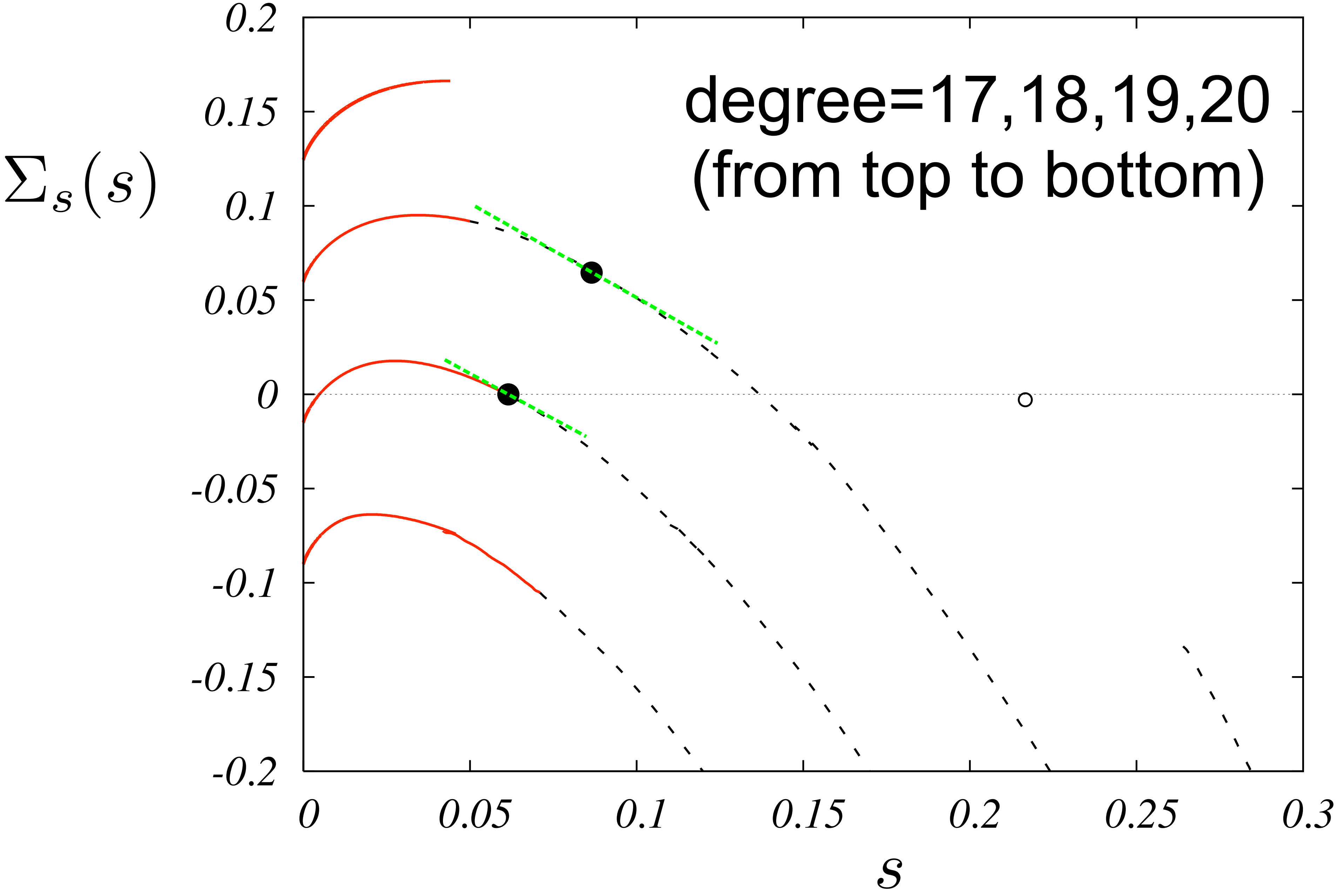


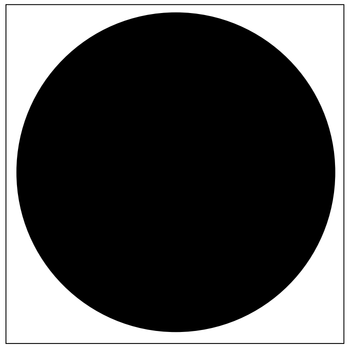


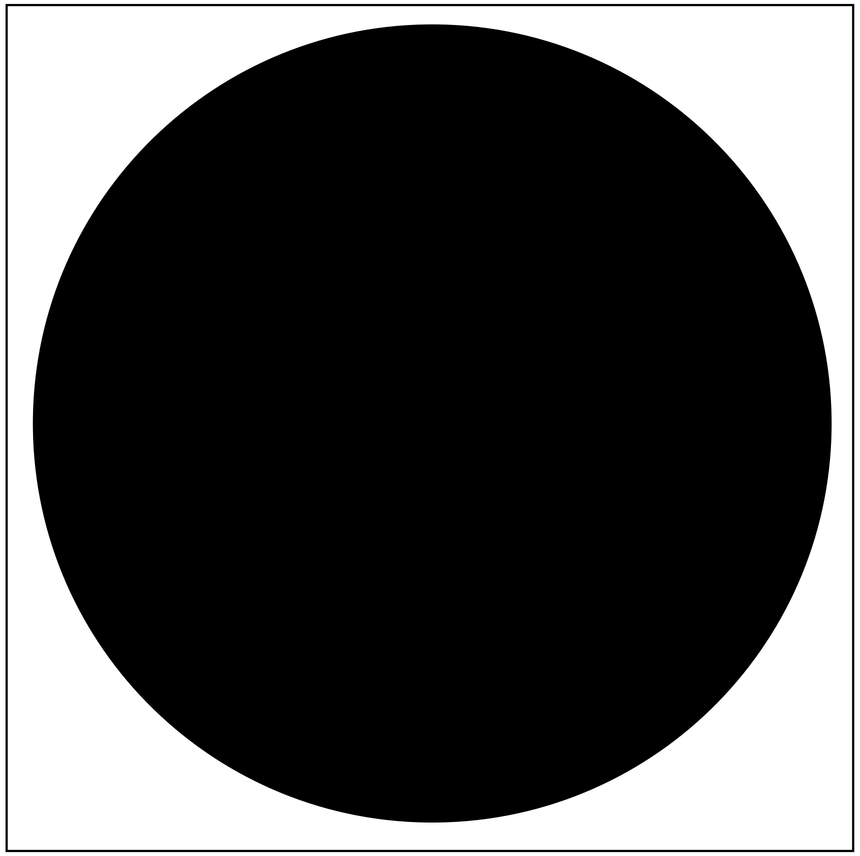
$$\frac{1}{N}\log Z = \Psi(1) = \max_{s:\Sigma(s)\geq 0}\left[\Sigma_s(s) + s\right]$$

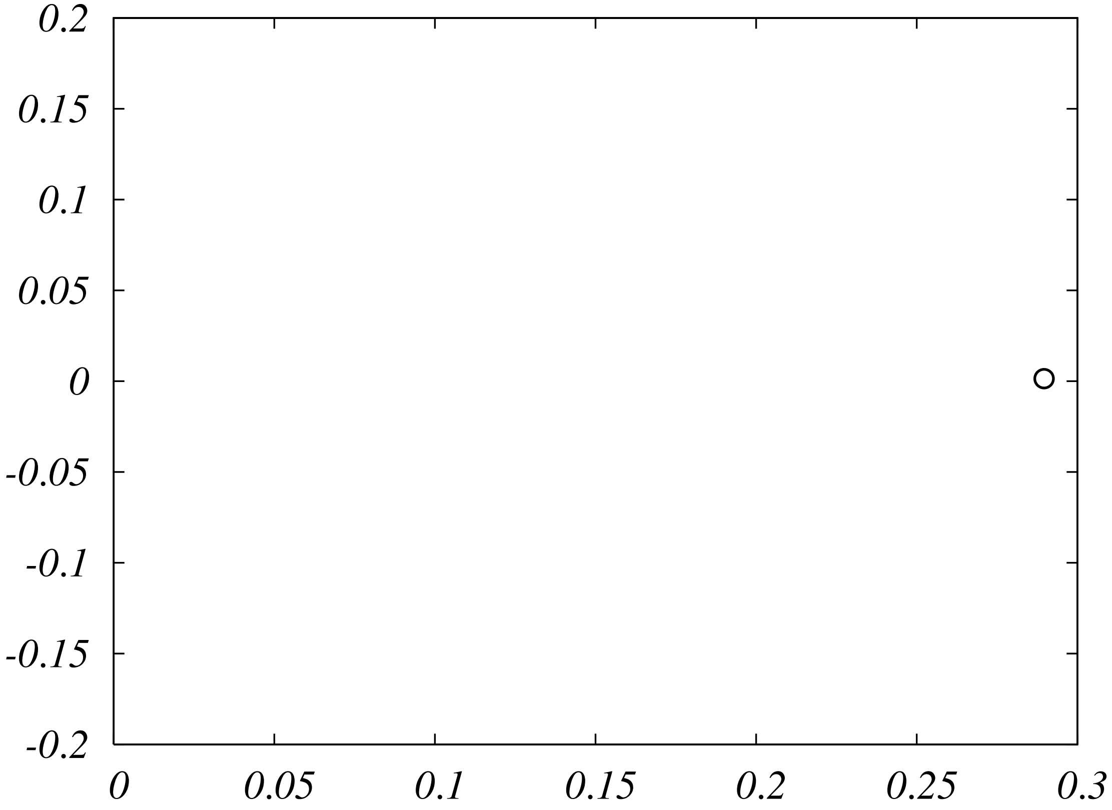
6-coloring random regular graphs

Krzakala, Montanari, Ricci-Tersenghi, Semerjian, Zdeborova, PNAS '07
Krzakala, Zdeborova, PRE '07



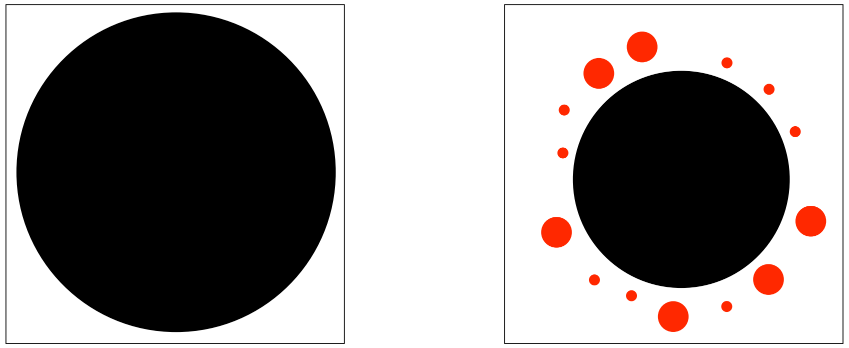


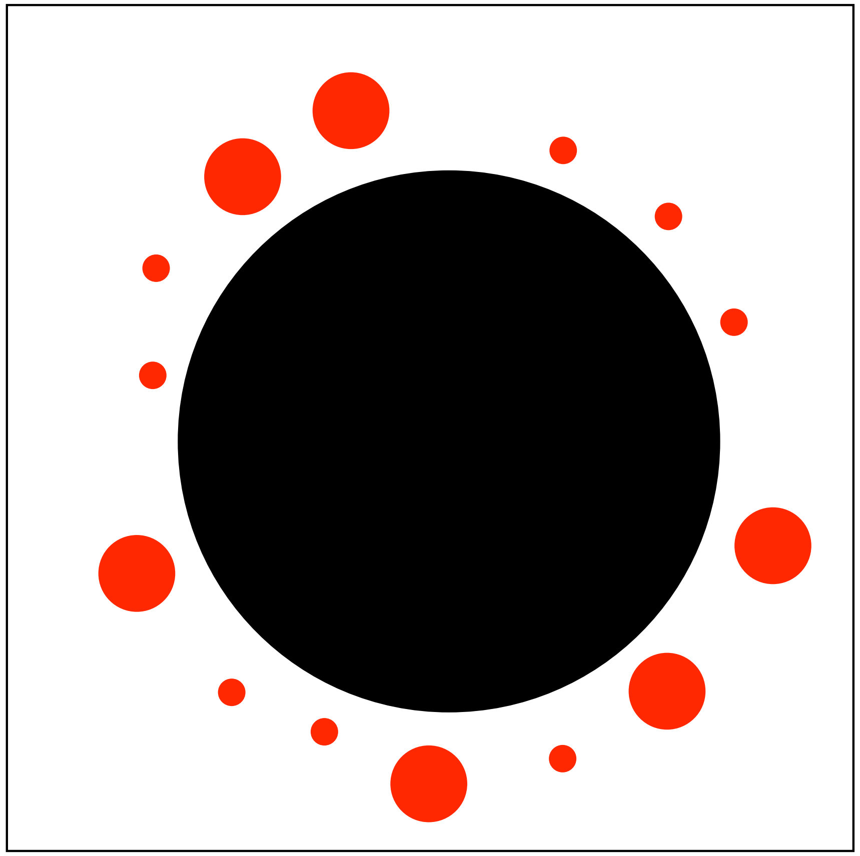


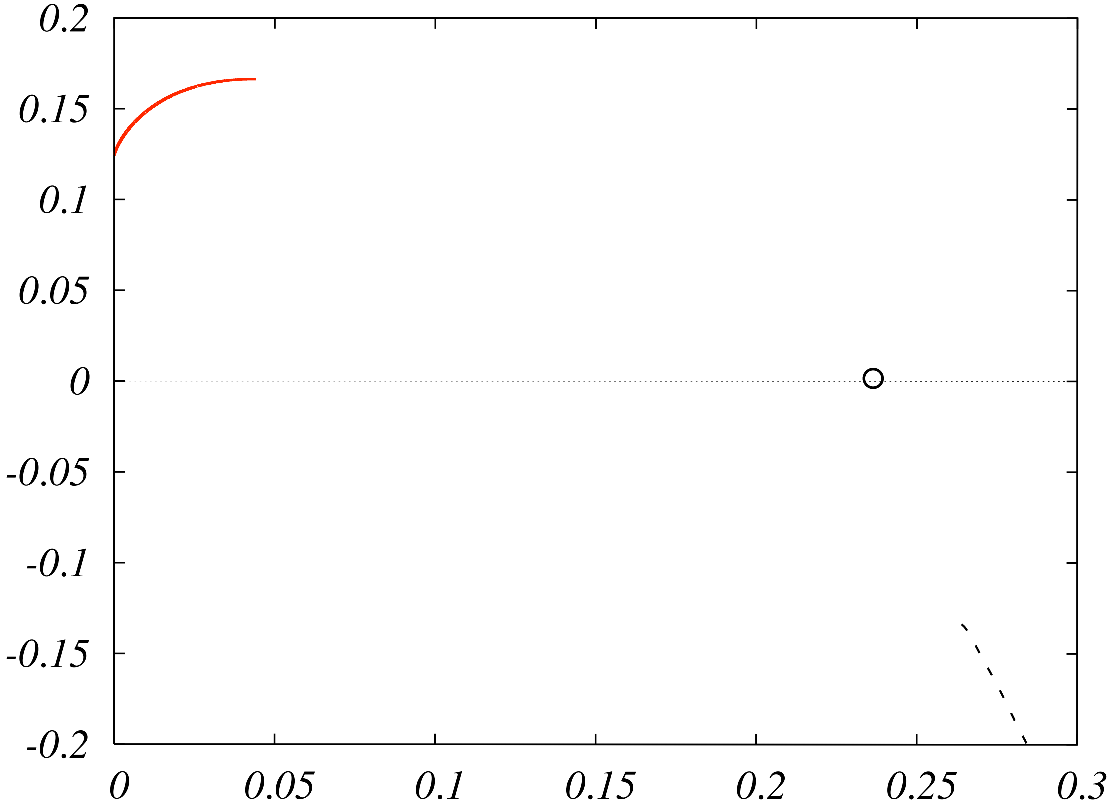


6 coloring of regular random graph

very low connectivity

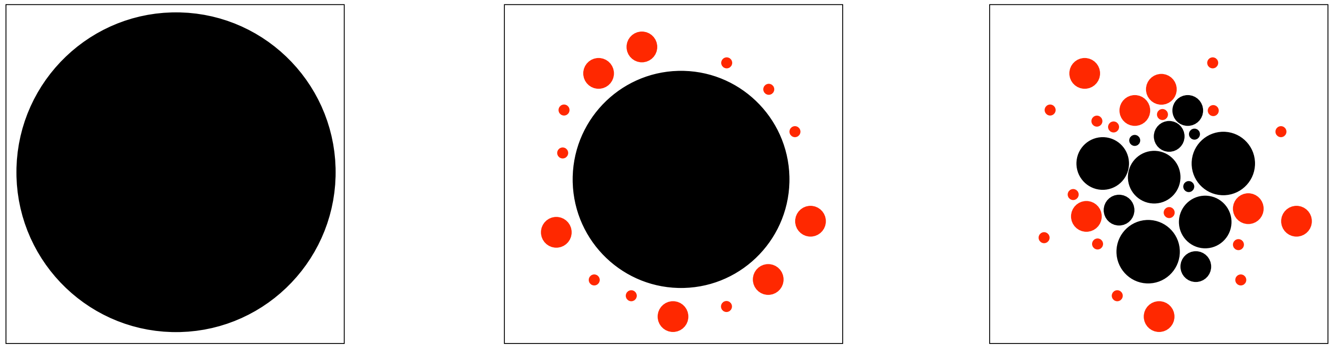


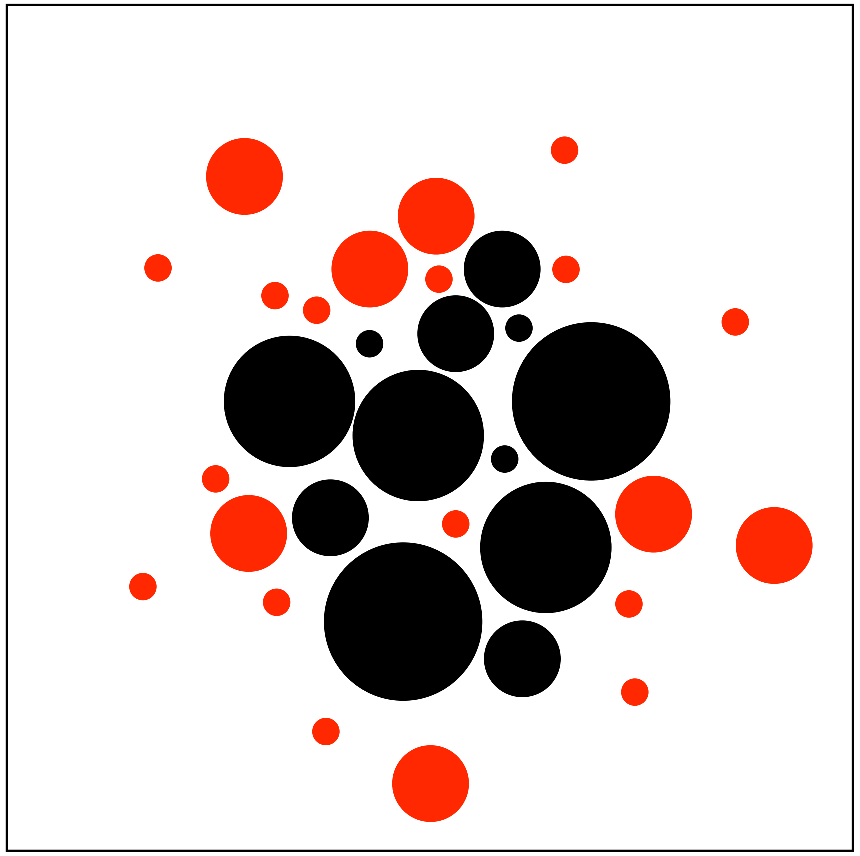


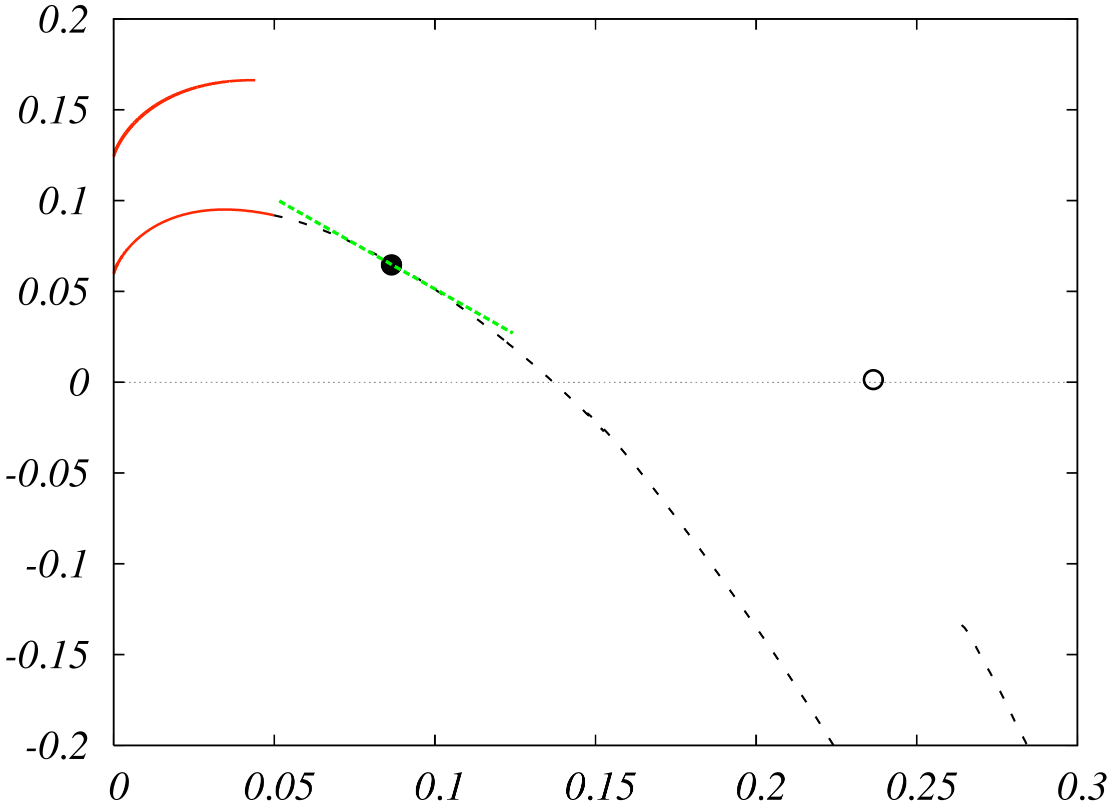


6 coloring of regular random graph

connectivity c=17

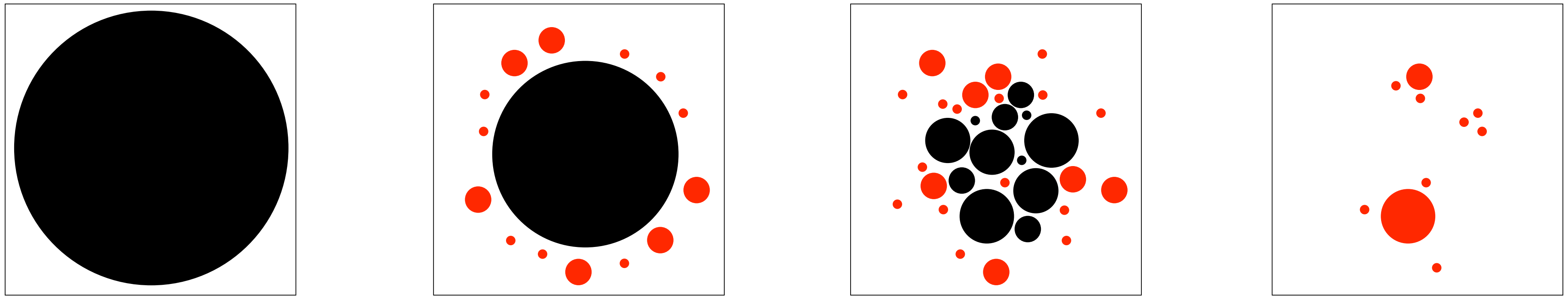


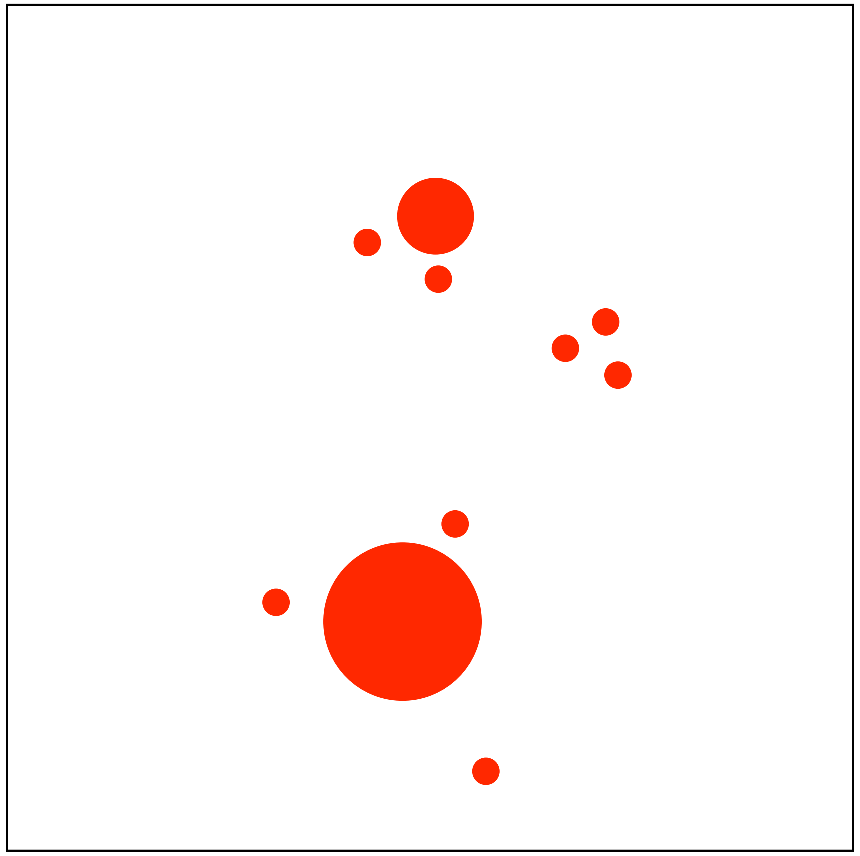


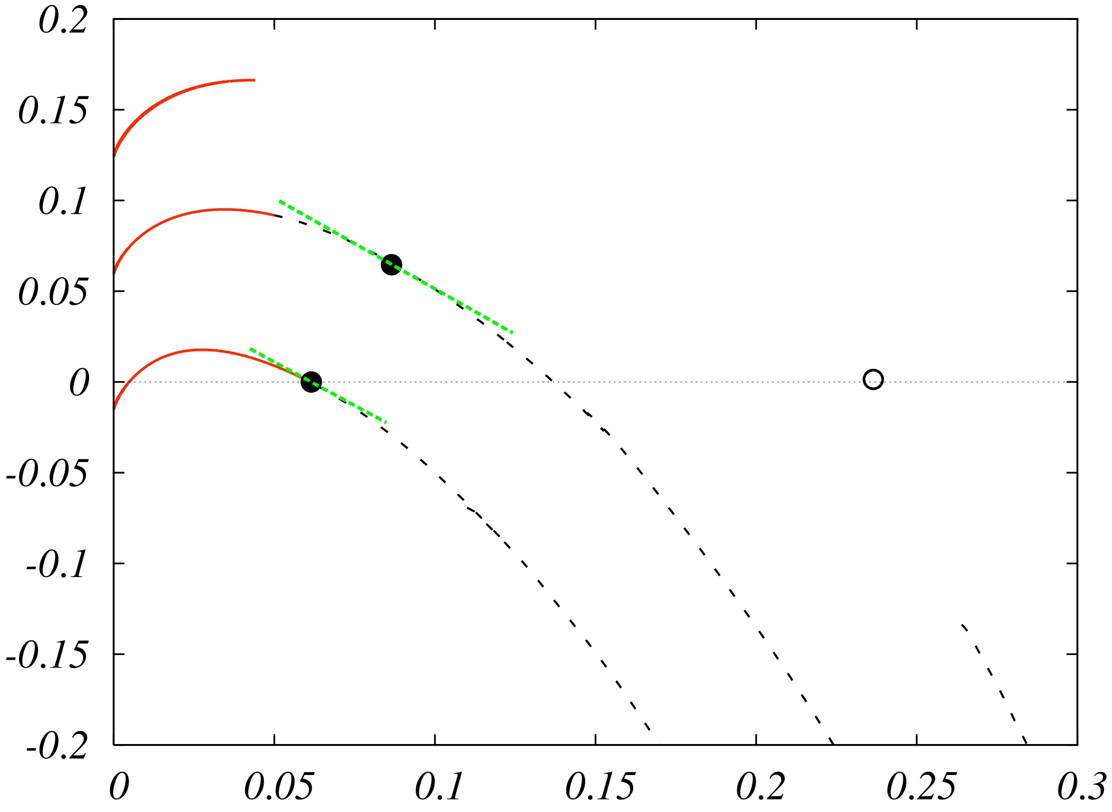


6 coloring of regular random graph

connectivity c=18

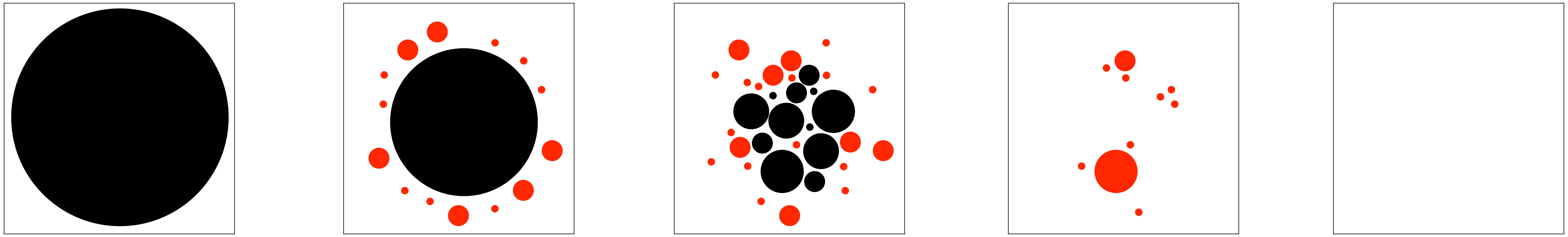


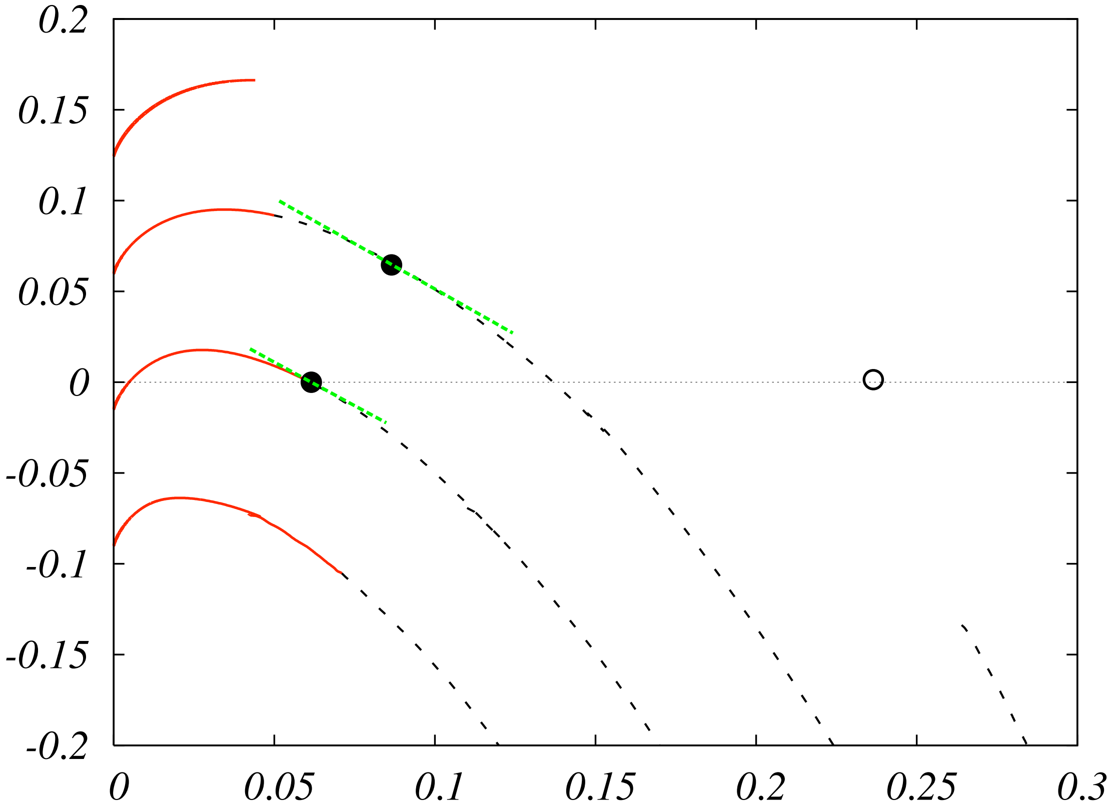




6 coloring of regular random graph

connectivity c=19





6 coloring of regular random graph

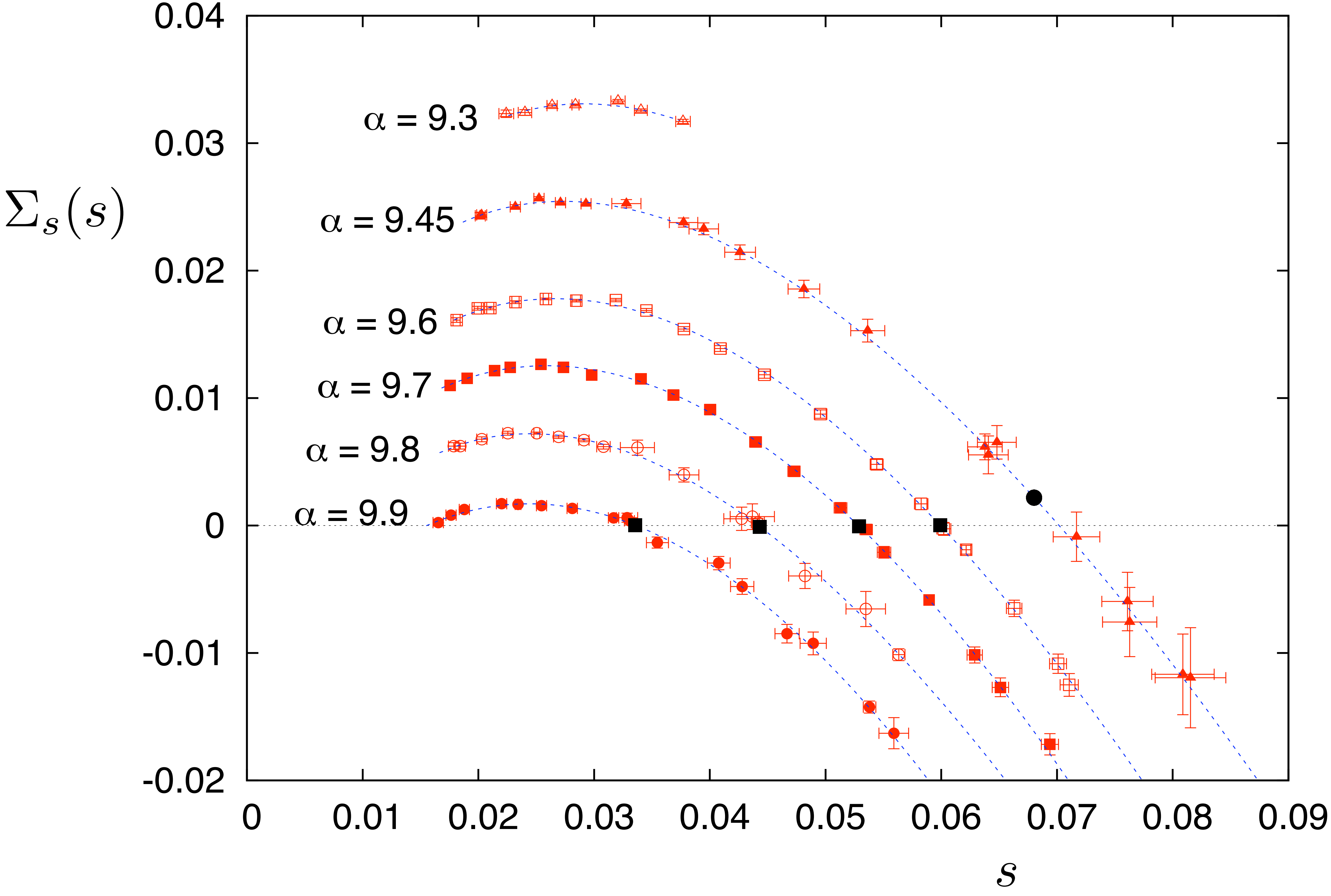
connectivity c=20

Random K-SAT revised

Krzakala, Montanari, Ricci-Tersenghi, Semerjian, Zdeborova, PNAS '07
Montanari, Ricci-Tersenghi, Semerjian, JSTAT '08

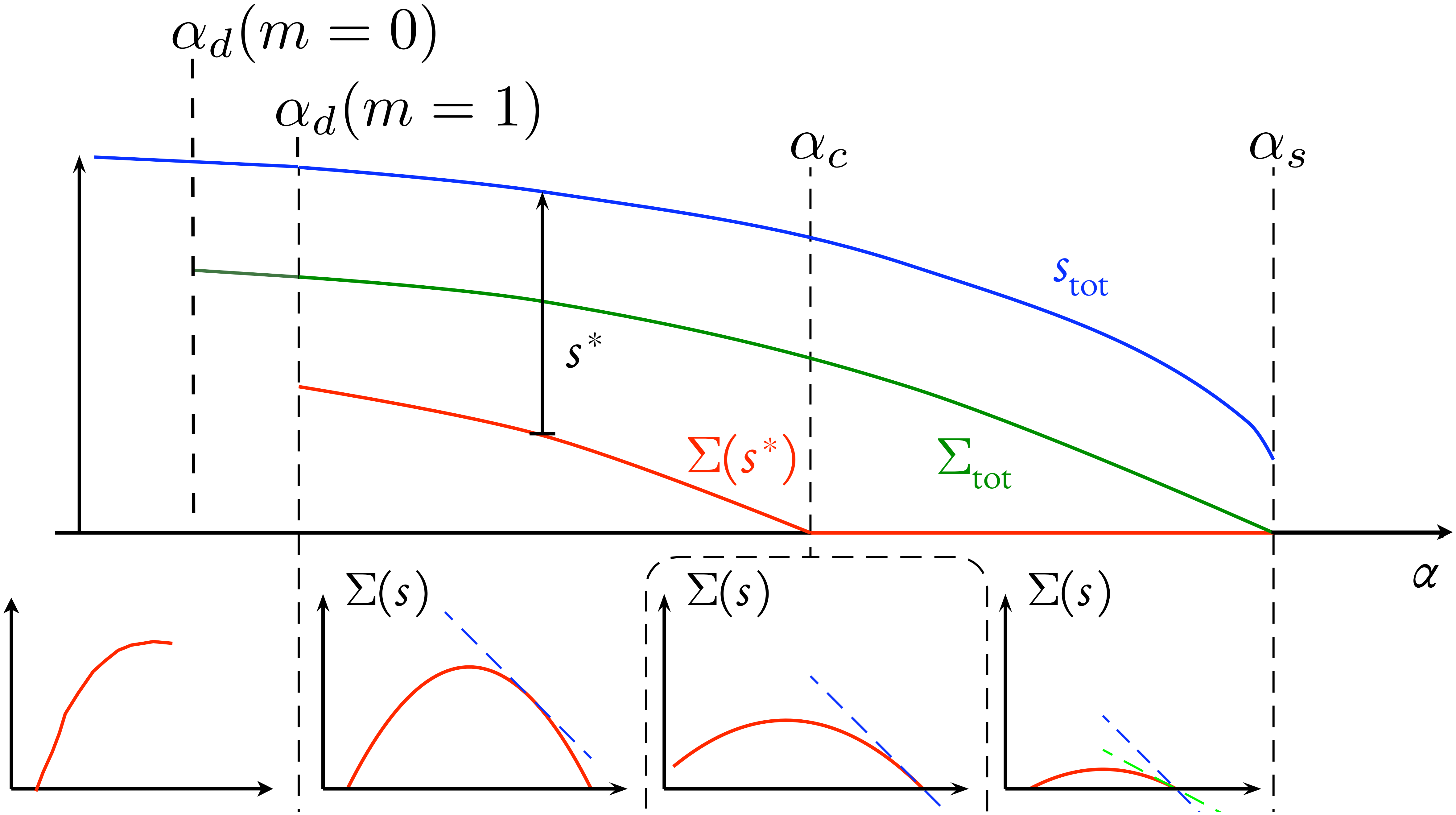
- We have computed $\Sigma_s(s,\alpha)$ for *K*=3 and *K*=4
- It is numerically very demanding: on each link there is a population of messages, to be updated and re-weighted at each iteration step, until convergence.
- For *m*=0 and *m*=1 equations simplify a lot
  - simpler messages (couple or triples) per link

random 4-SAT

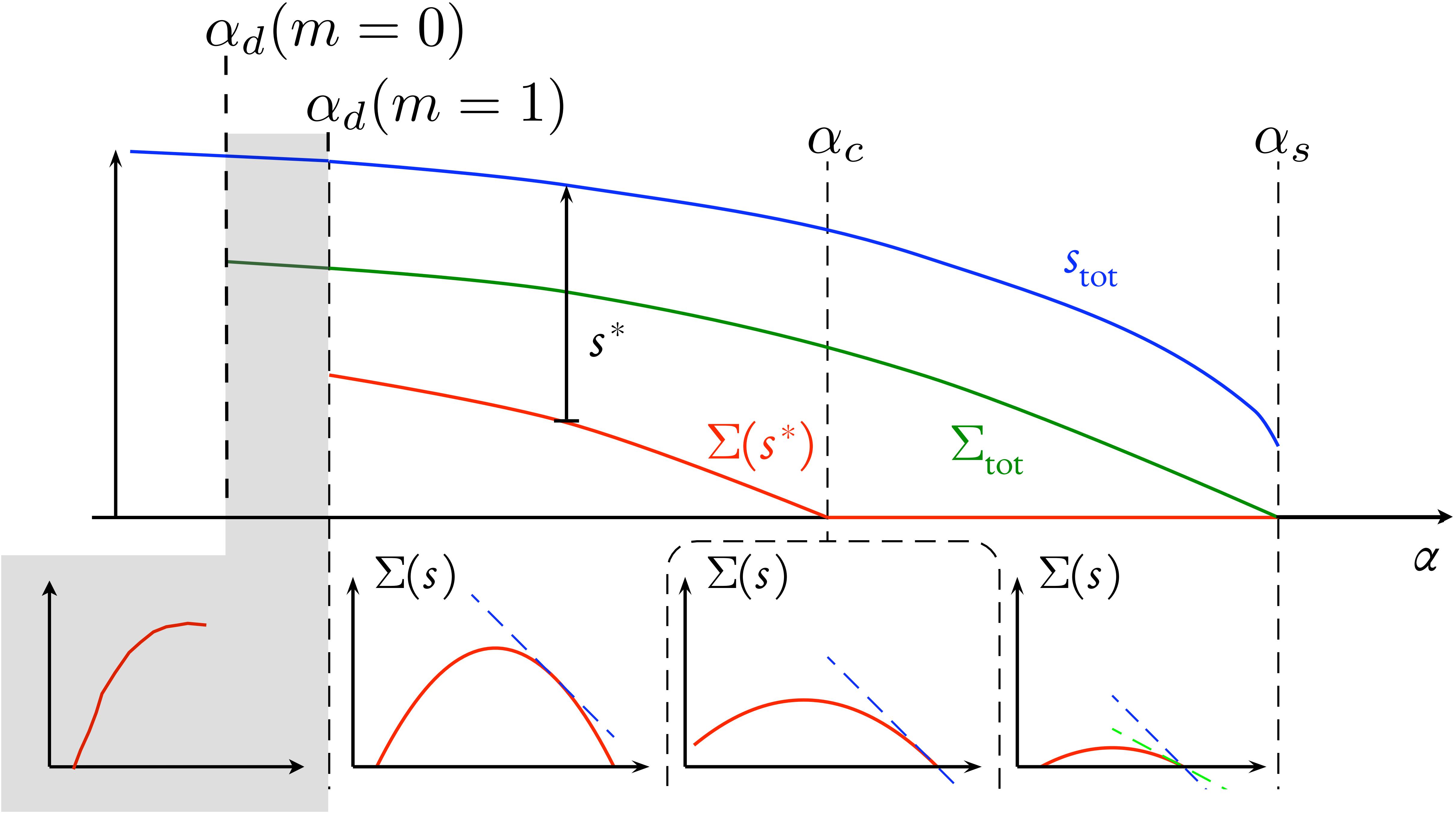


$\Sigma_s(s)$
0.04
0.03
0.02
0.01
0
-0.01
-0.02
α = 9.3
α = 9.45
α = 9.6
α = 9.7
α = 9.8
α = 9.9
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
s

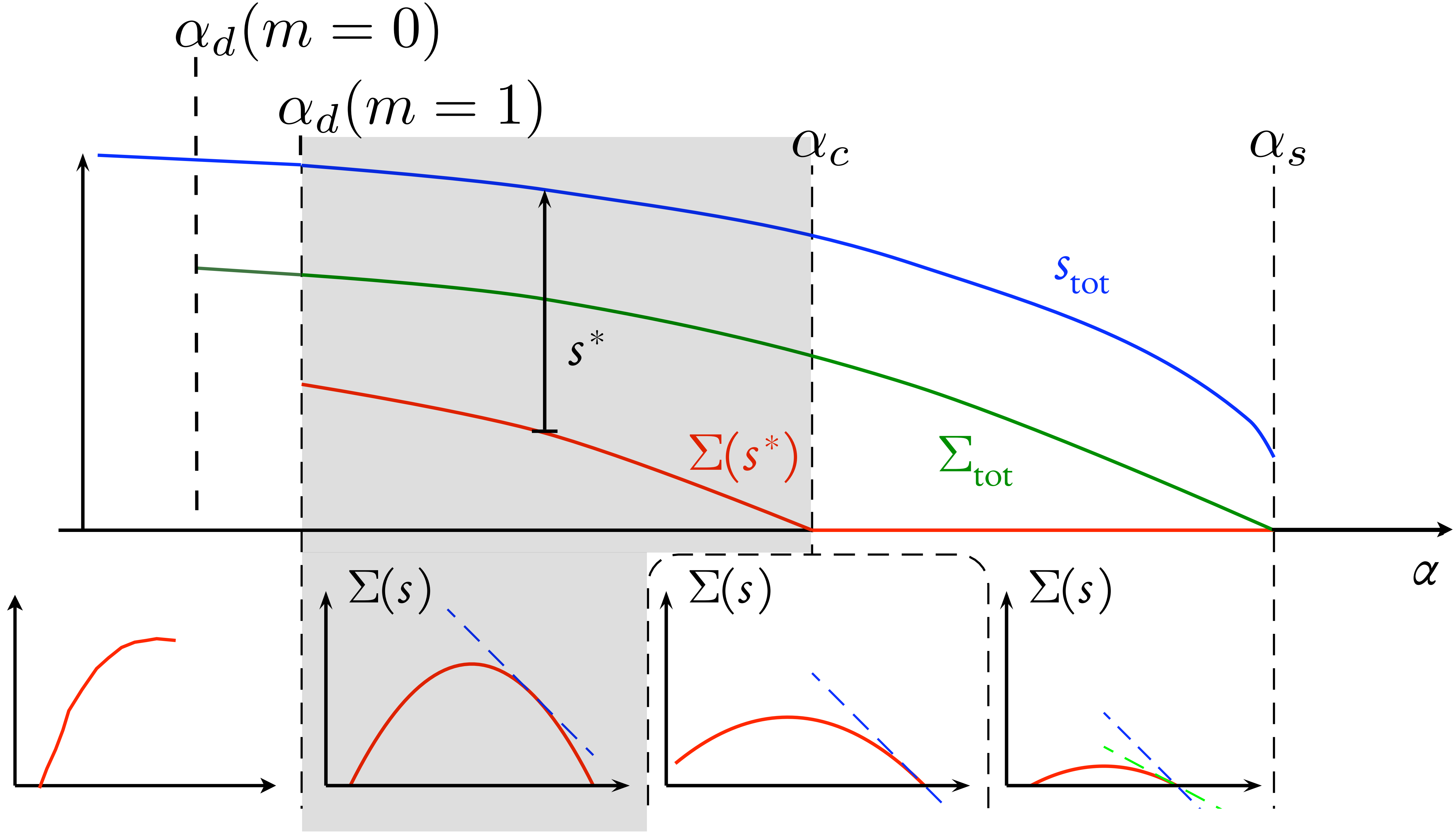
Results


$\alpha_d(m=0)$
$\alpha_d(m=1)$
α_c
α_s
s_{tot}
s^*
$\Sigma(s^*)$
Σ_{tot}
α
$\Sigma(s)$
$\Sigma(s)$
$\Sigma(s)$

Results

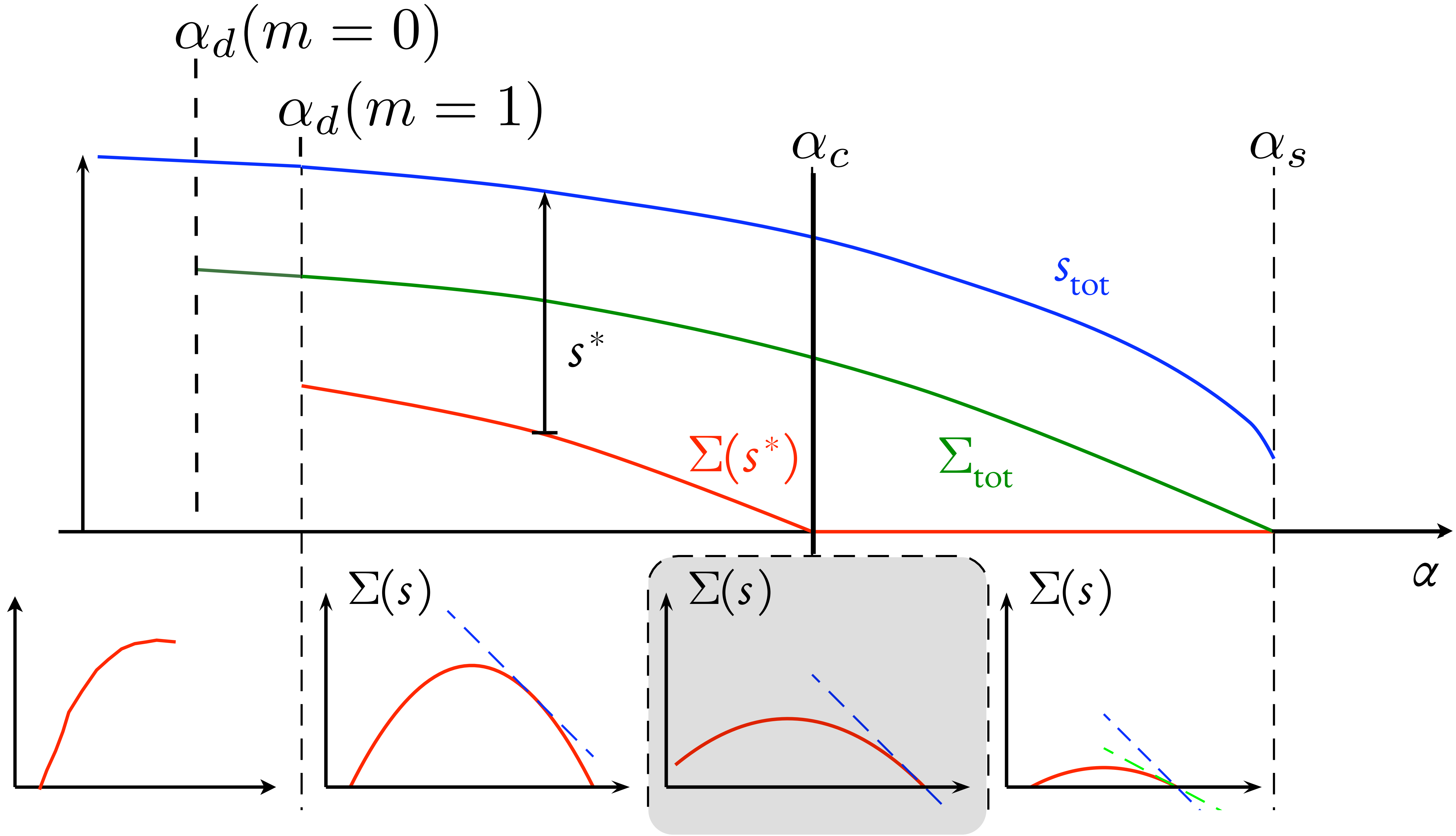

$\alpha_d(m=0)$
$\alpha_d(m=1)$
α_c
α_s
s_{tot}
s^*
$\Sigma(s^*)$
Σ_{tot}
α
$\Sigma(s)$
$\Sigma(s)$
$\Sigma(s)$

Results

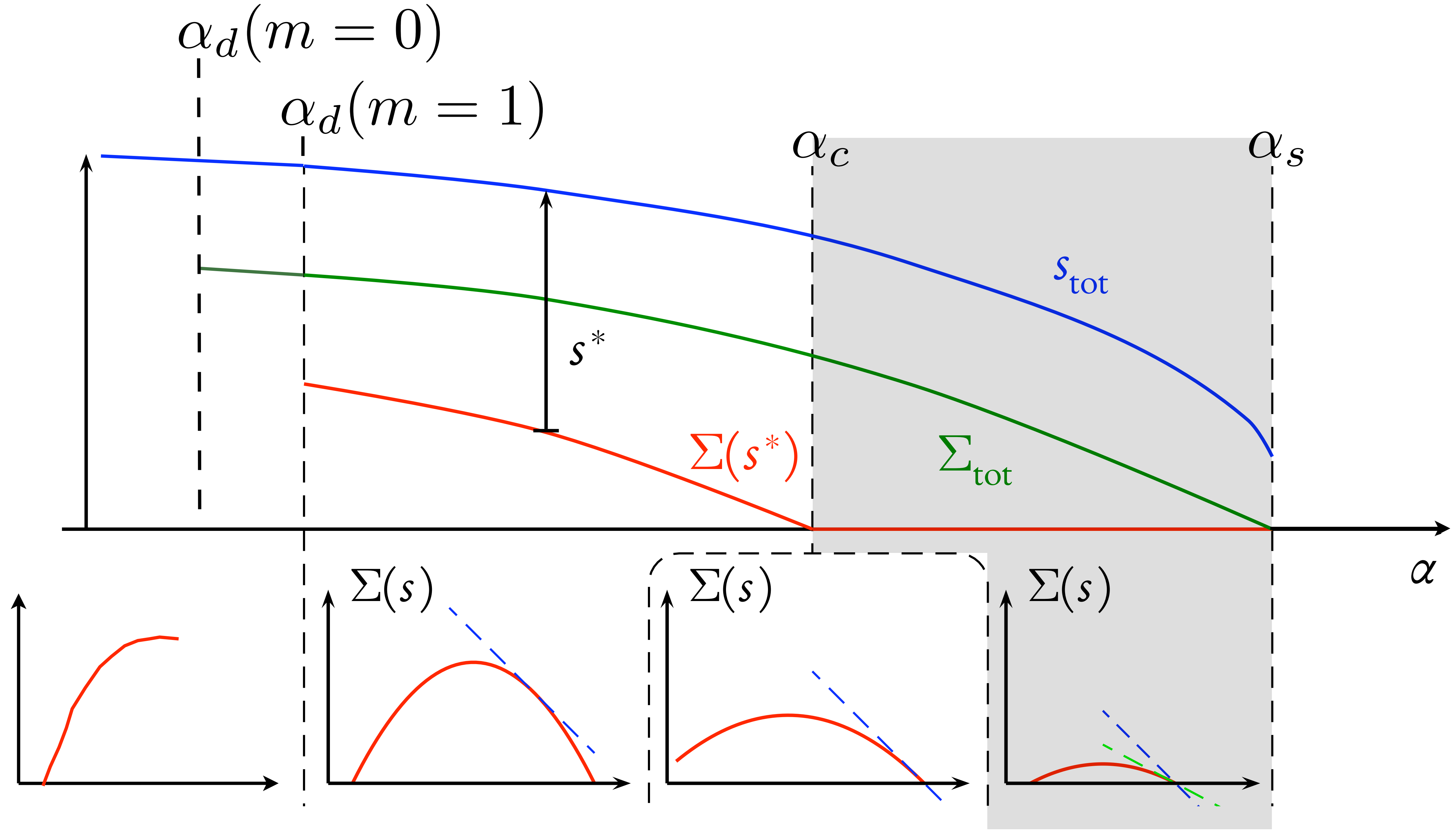


$\alpha_d(m=0)$
$\alpha_d(m=1)$
α_c
α_s
s_{tot}
s^*
$\Sigma(s^*)$
Σ_{tot}
α
$\Sigma(s)$
$\Sigma(s)$
$\Sigma(s)$

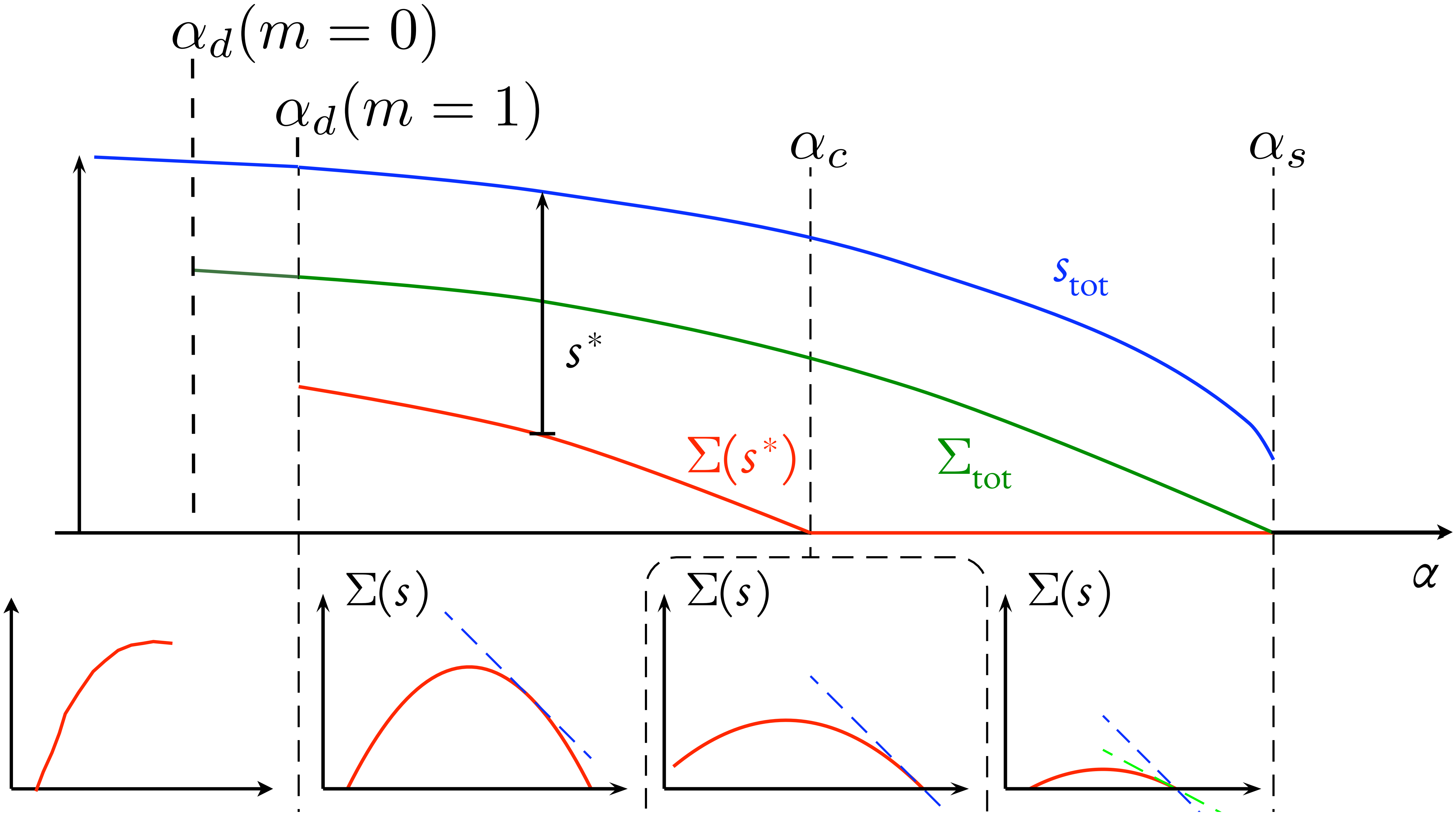
Results


$\alpha_d(m=0)$
$\alpha_d(m=1)$
α_c
α_s
s_{tot}
s^*
$\Sigma(s^*)$
Σ_{tot}
α
$\Sigma(s)$
$\Sigma(s)$
$\Sigma(s)$

Results

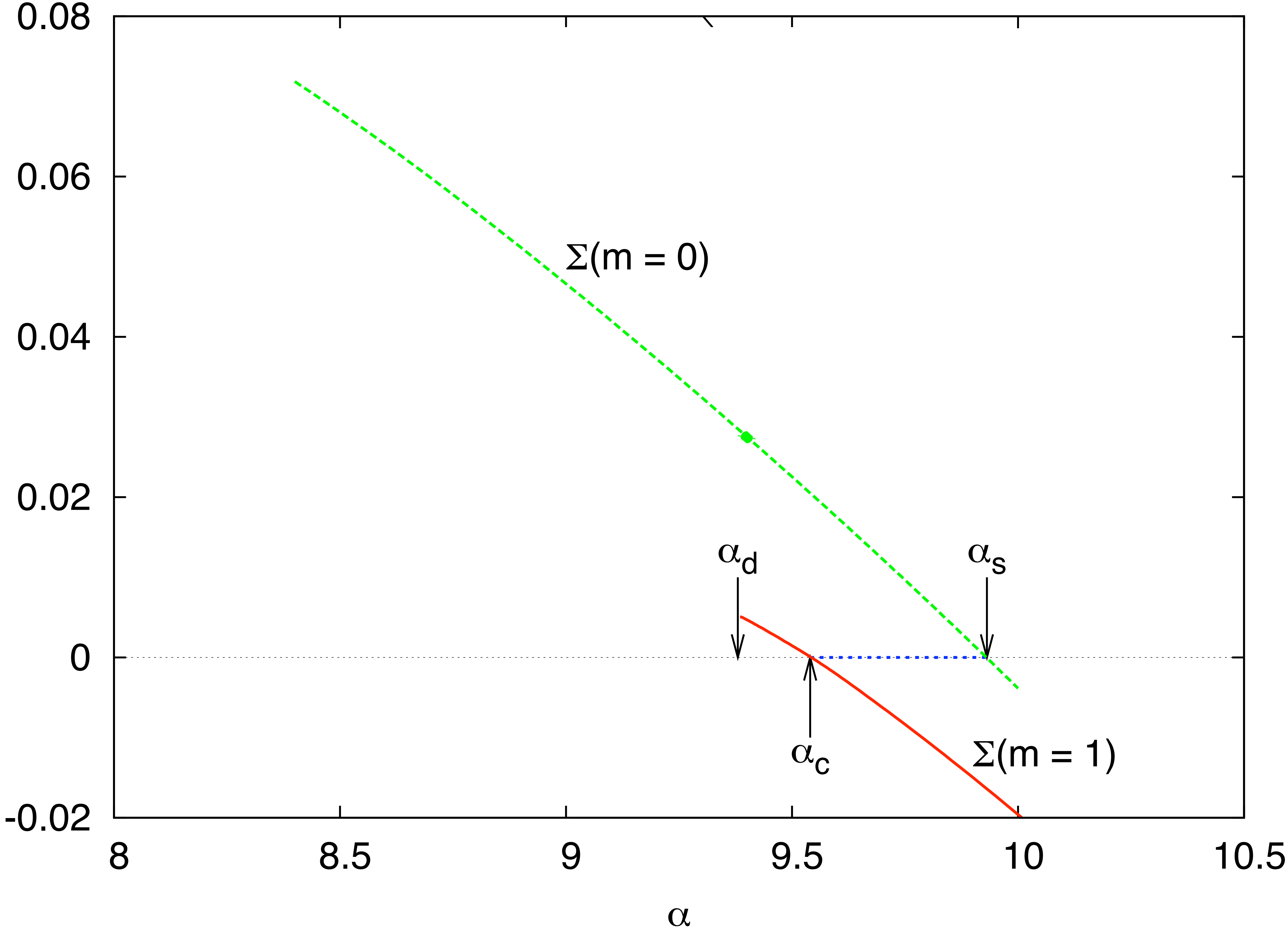

$\alpha_d(m=0)$
$\alpha_d(m=1)$
α_c
α_s
s_{tot}
s^*
$\Sigma(s^*)$
Σ_{tot}
α
$\Sigma(s)$
$\Sigma(s)$
$\Sigma(s)$

Results

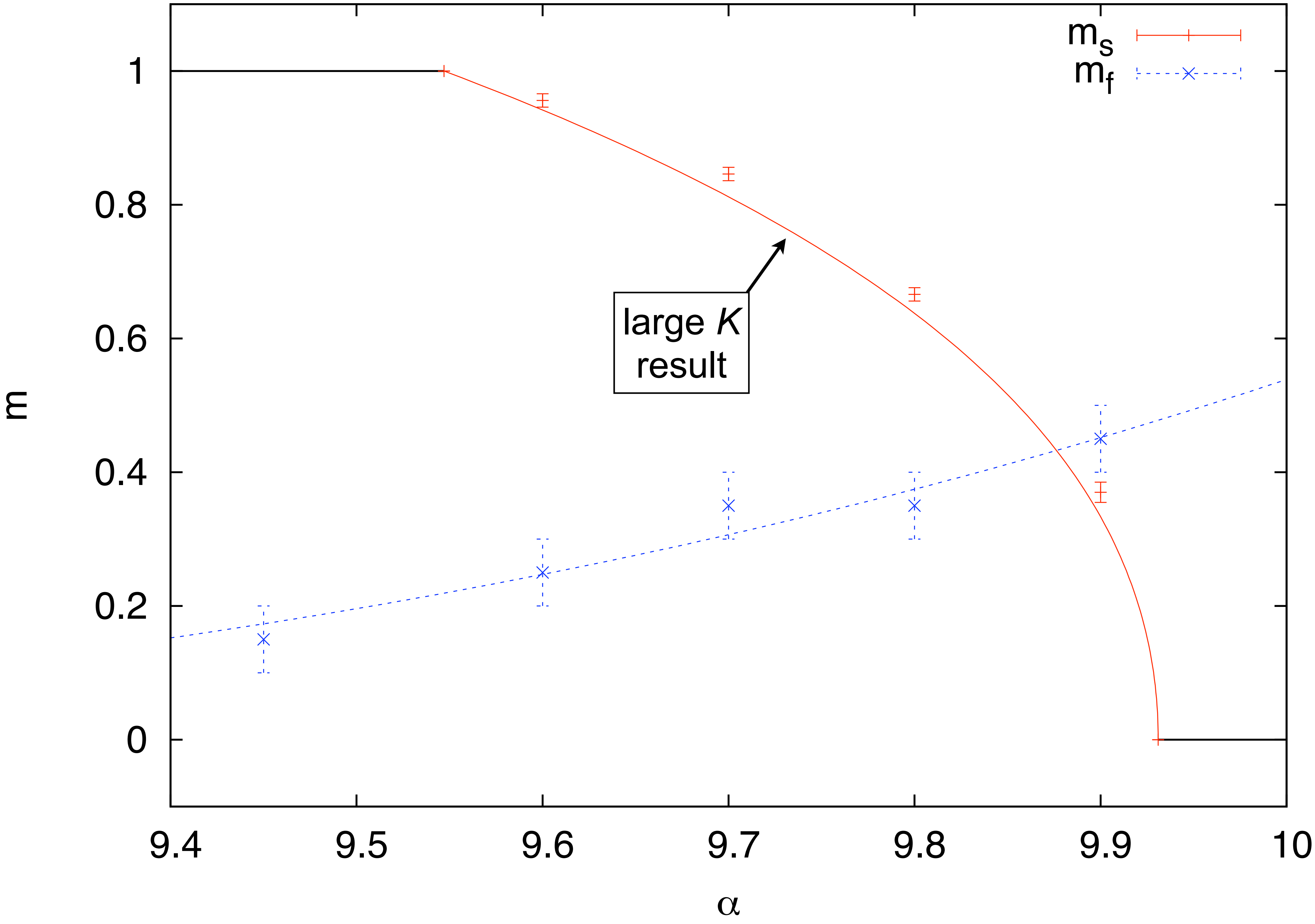


$\alpha_d(m=0)$
$\alpha_d(m=1)$
α_c
α_s
s_{tot}
s^*
$\Sigma(s^*)$
Σ_{tot}
α
$\Sigma(s)$
$\Sigma(s)$
$\Sigma(s)$

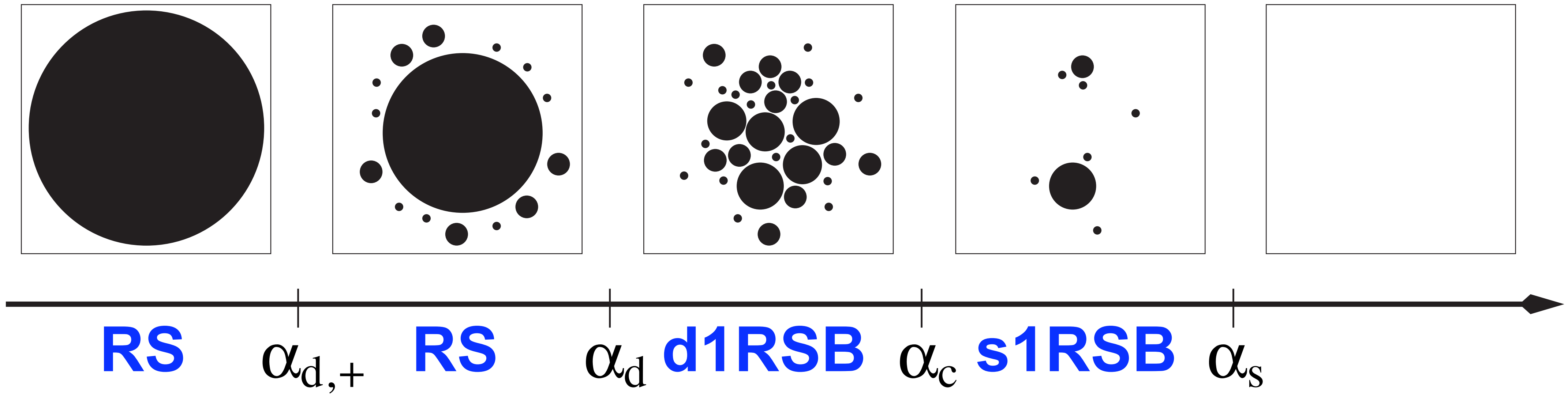
random 4-SAT


0.08
0.06
0.04
0.02
0
-0.02
8
8.5
9
9.5
10
10.5
α
Σ(m = 0)
Σ(m = 1)
α_d
α_s
α_c

random 4-SAT



Summary



liquid (paramagnet)

liquid with sub-dominating clusters

clustered liquid

clustered condensed

UNSAT no solution

RS $\alpha_{\mathrm{d},+}$ **RS** α_{d} **d1RSB** α_{c} **s1RSB** α_{s}

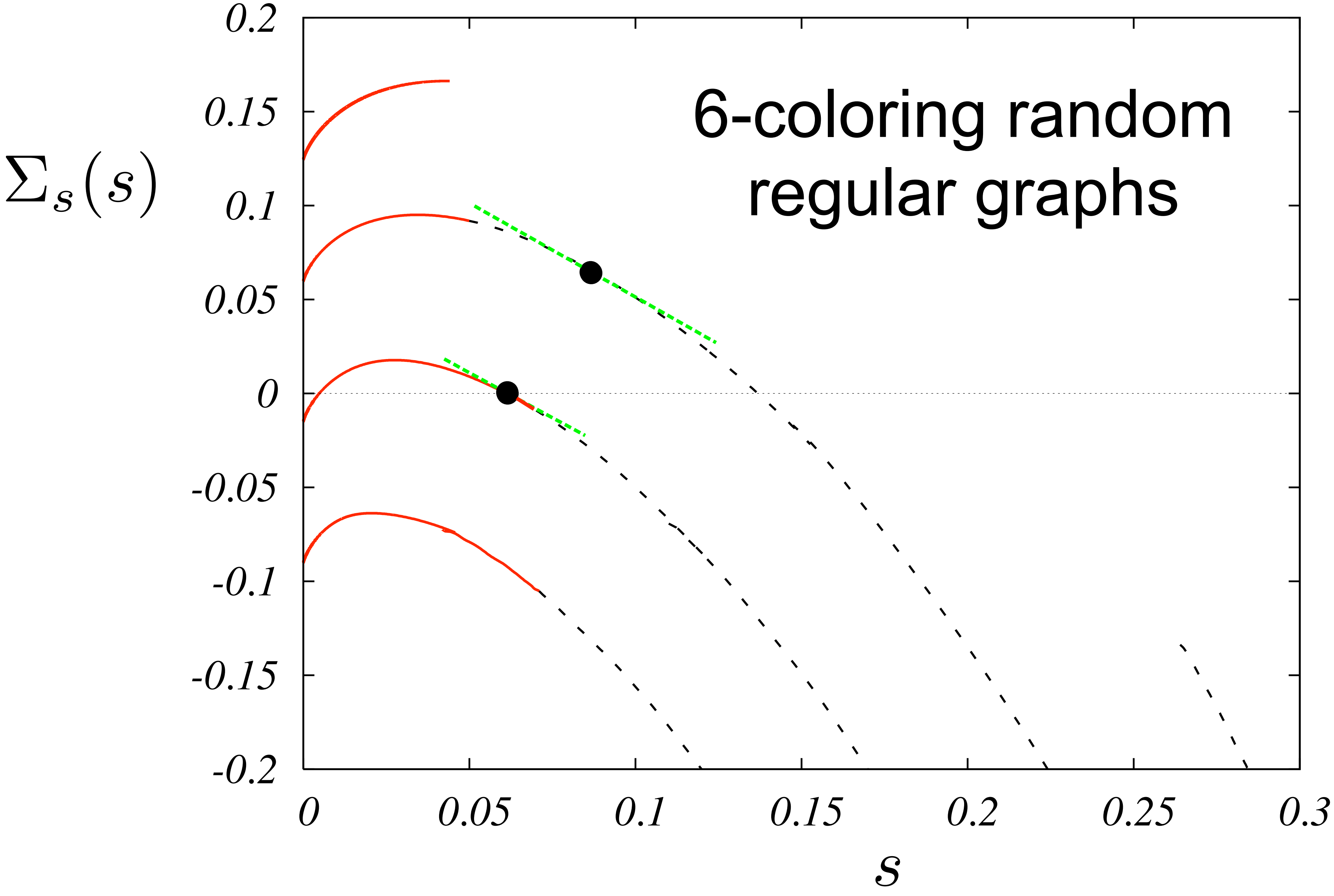
k	α_{d}	α_{c}	α_s
3	3.86	3.86	4.267
4	9.38	9.547	9.931
5	19.16	20.80	21.117
6	36.53	43.08	43.37

the largest for large *K*

$$\alpha_d \sim \frac{\log(k)}{k} 2^k$$

$$\alpha_c \sim \alpha_s \sim 2^k$$

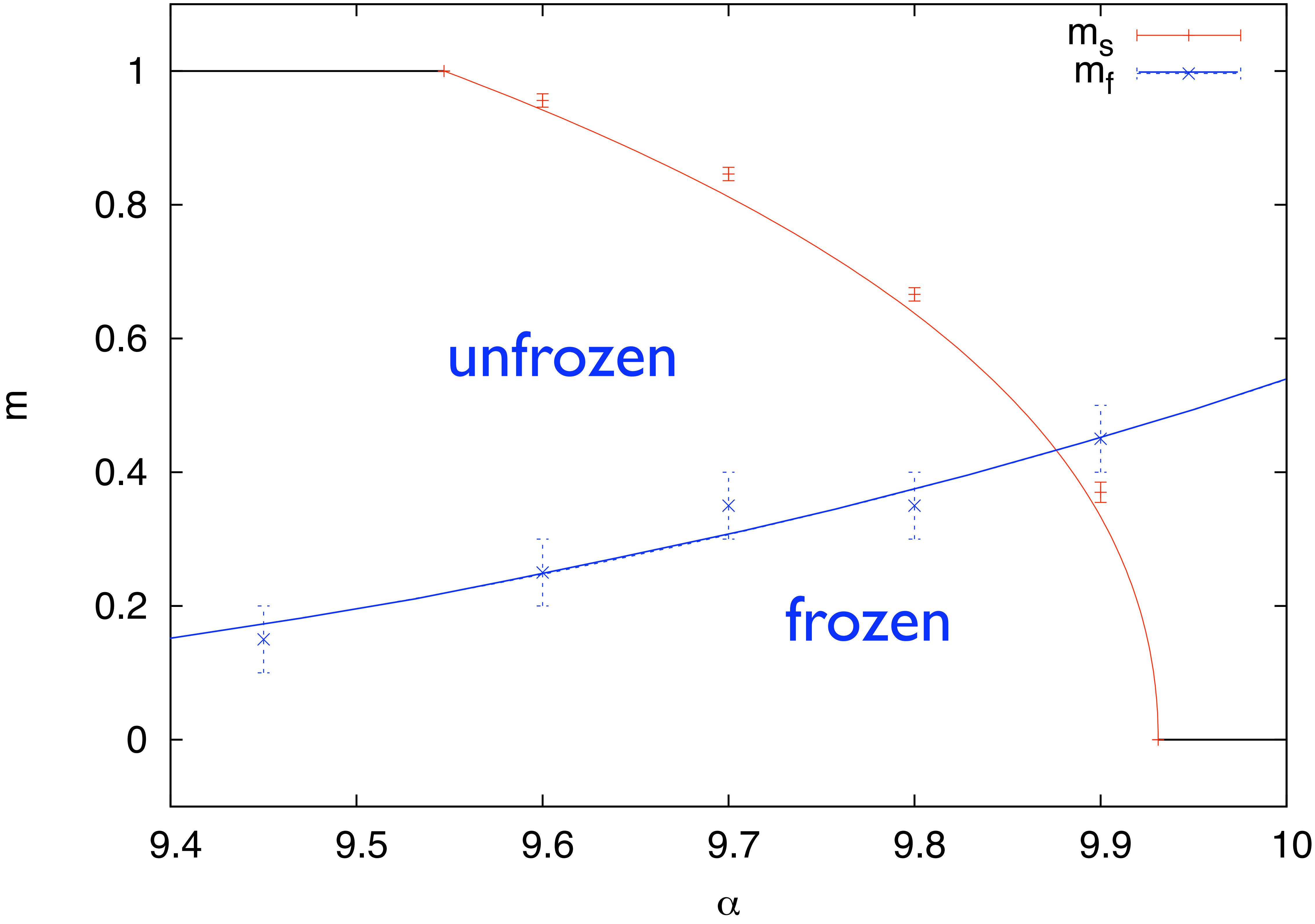
Frozen variables



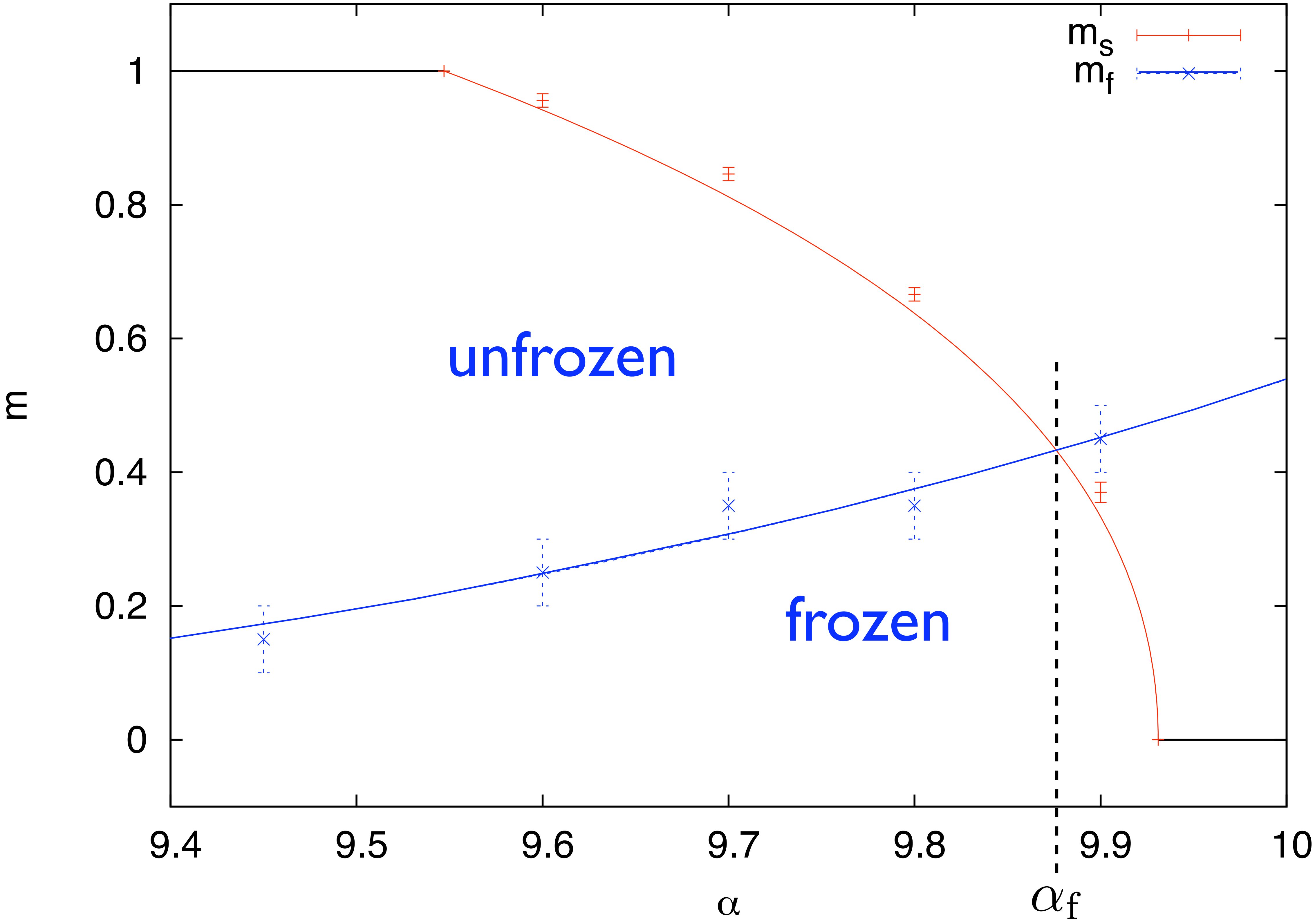
— dominating clusters have frozen variables

--- dominating clusters do not have frozen variables

random 4-SAT

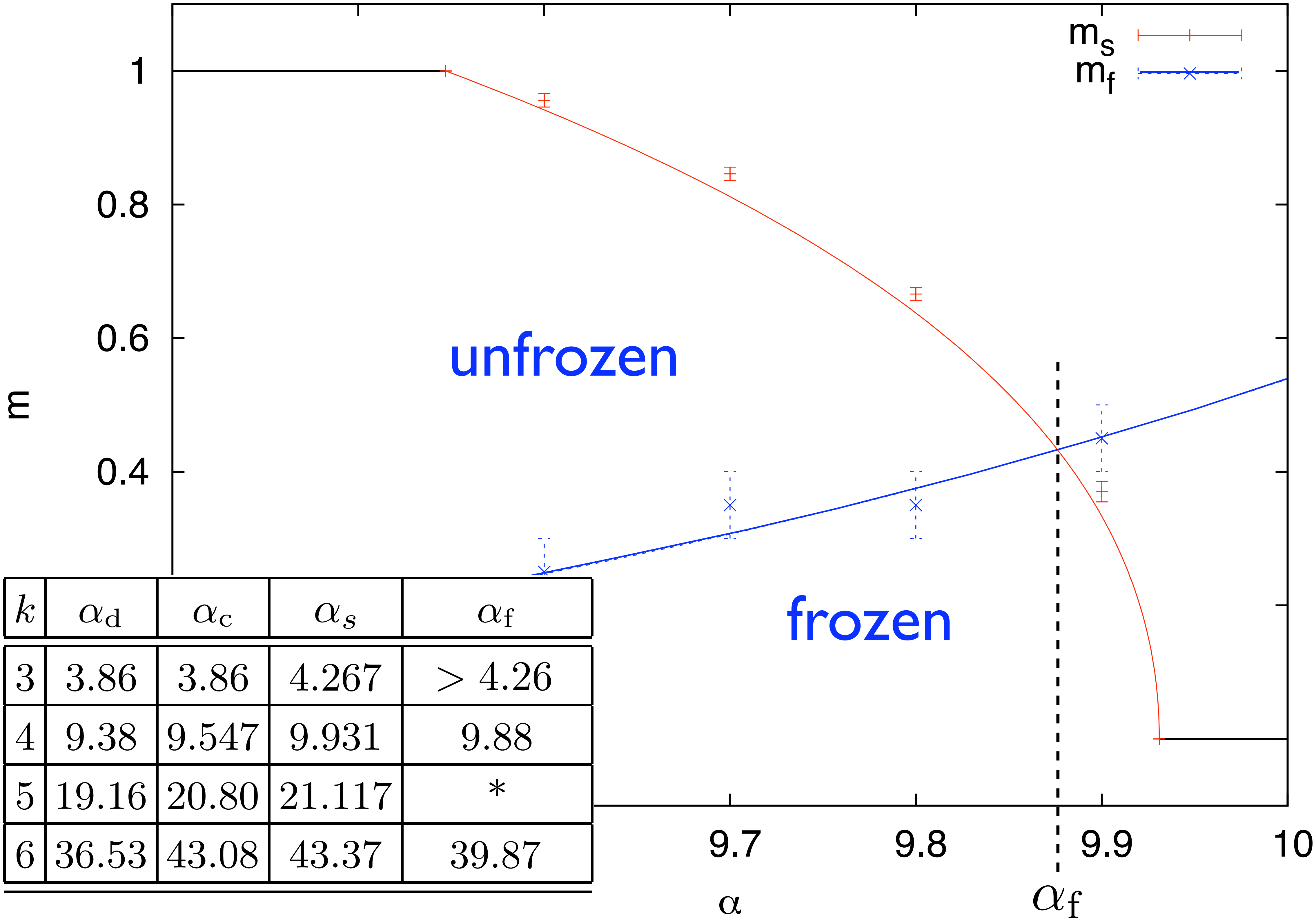


random 4-SAT



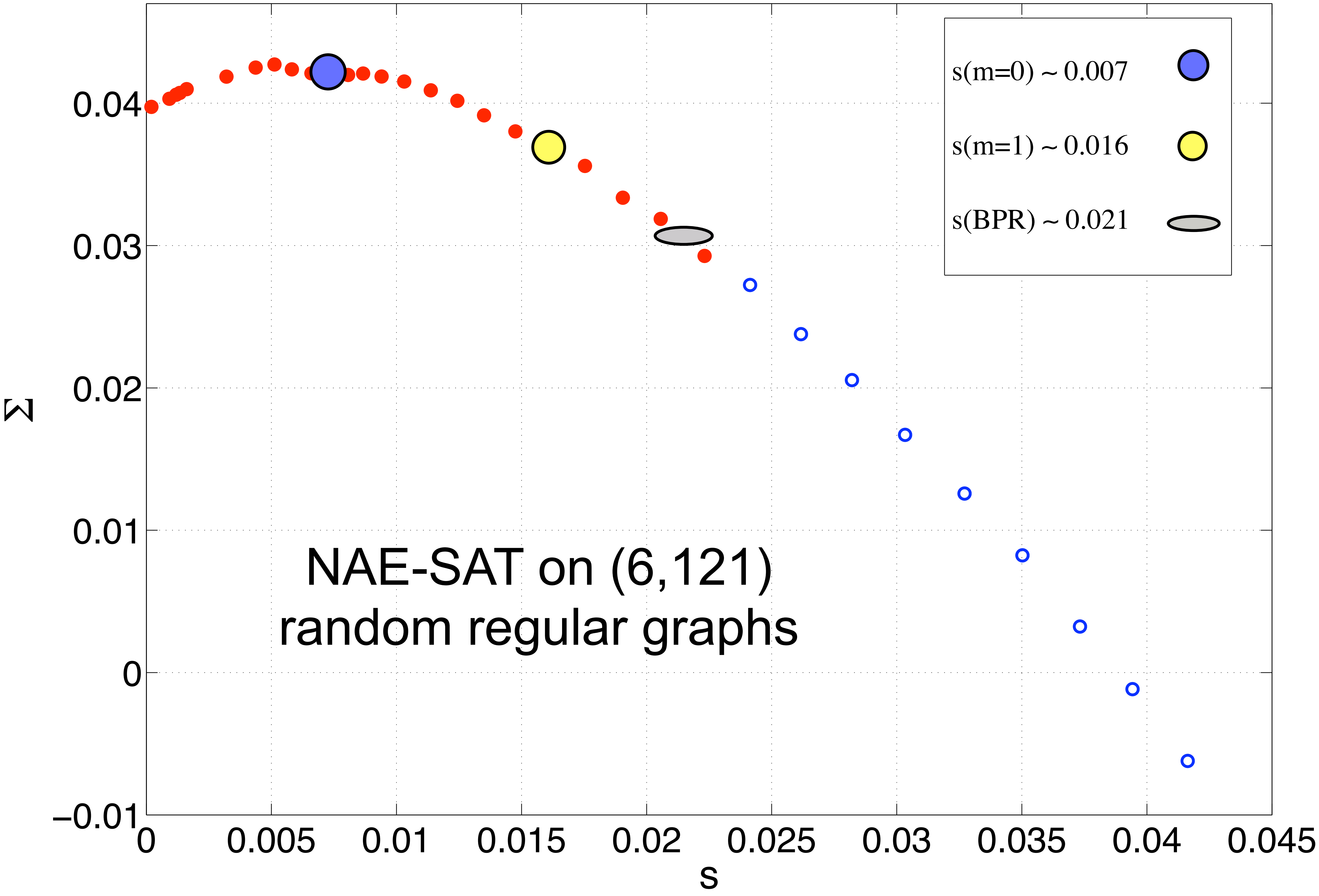
m_s
m_f
1
0.8
0.6
0.4
0.2
0
m
9.4
9.5
9.6
9.7
9.8
9.9
10
α
α_f
unfrozen
frozen

random 4-SAT



k	α_d	α_c	α_s	α_f
3	3.86	3.86	4.267	> 4.26
4	9.38	9.547	9.931	9.88
5	19.16	20.80	21.117	*
6	36.53	43.08	43.37	39.87

Dall'Asta, Ramezanpour & Zecchina, PRE '08



NAE-SAT on (6,121) random regular graphs

Frozen variables for large *k*

Achlioptas, Ricci-Tersenghi, STOC '06

For $k \geq 9$ and $\tilde{\alpha}_{\rm f} \leq \alpha < \alpha_{\rm s}$ **every** cluster has

at least $\left(1 - \frac{2}{k}\right) N$ frozen variables.

For $k \to \infty$, $\tilde{\alpha}_{\rm f} \sim \frac{4}{5} 2^k \log(2)$

Is $\tilde{\alpha}_{\rm f}$ (which is a bound to $\alpha_{\rm f}$) a threshold for algorithms? Maybe yes, but... $\alpha_{\rm f} \sim 2^k \log(2)/k$

Main open problems

- Stability of 1RSB solutions (technical point)
- Closing the gap between algorithmic threshold and SAT/UNSAT threshold
  - improvements in analysis of algorithms (decimation, reinforcement, etc.)
- Non-random structures, like those present in real world problems
  - beyond Bethe approximation (effects of loops, Cluster Variation Method, etc.)