Random 3-XORSAT

S. Cocco
O. Dubois
J. Mandler

N Boolean variables $x_i = 0, 1$

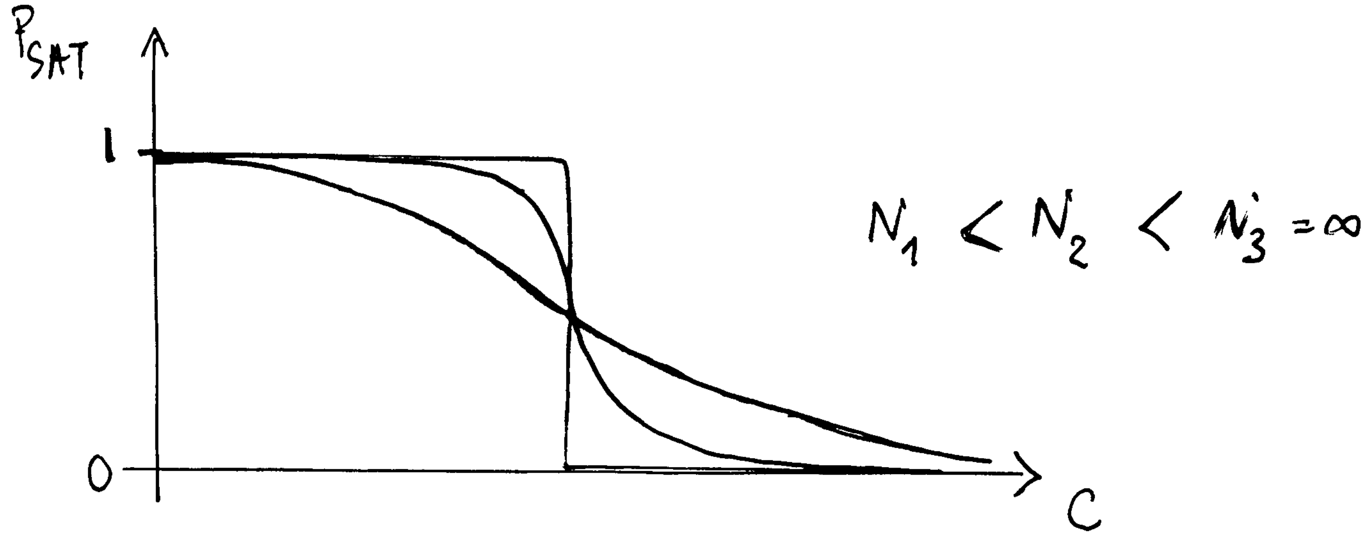
M clauses: $\begin{cases} i_m < j_m < k_m & (m = 1, \dots, M) \\ v_m = 0, 1 & \text{proba } 1/2, 1/2 \end{cases}$

$$x_{i_m} + x_{j_m} + x_{k_m} = v_m \qquad (\forall m)$$

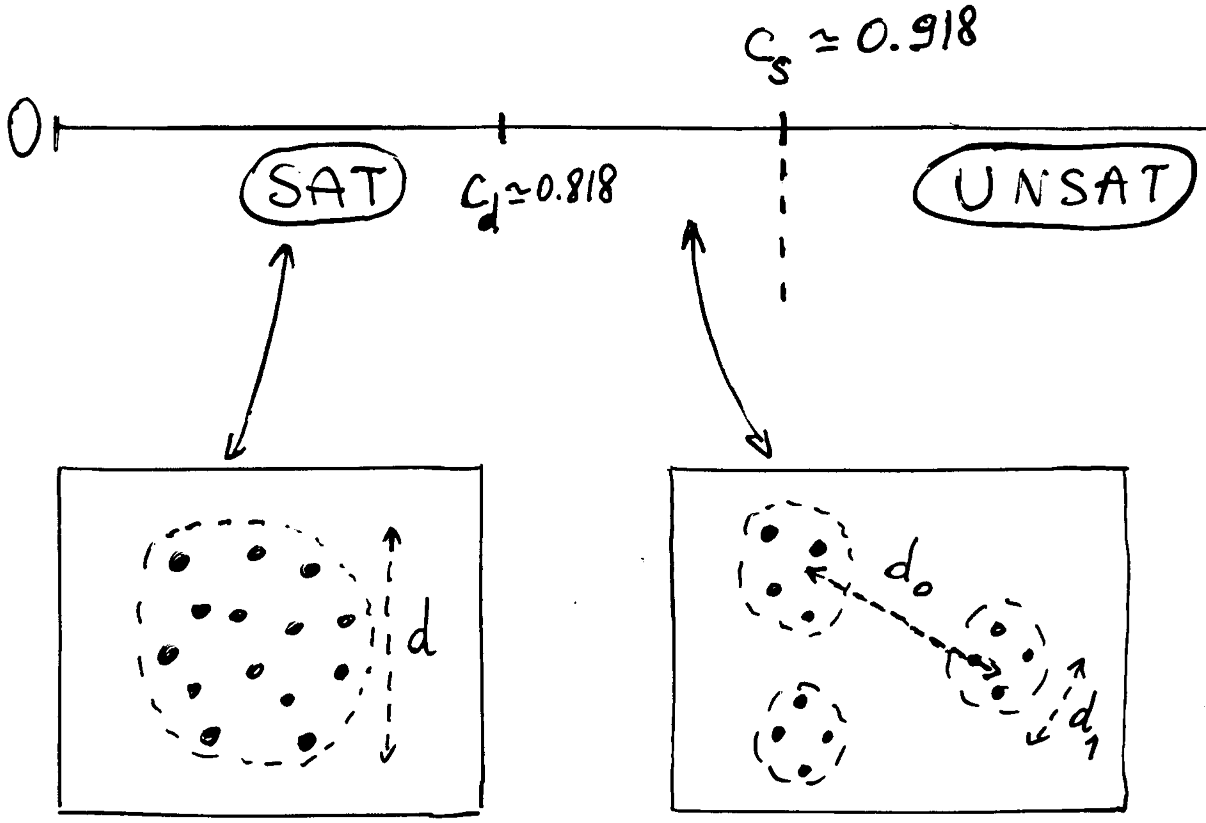
- Computationaly easy (P)
- SAT vs. UNSAT ?

→ same phenomenology as Random 3-SAT

$$c = \frac{M}{N}$$



Random 3-XORSAT : Results



configuration space (N-dim. hypercube)

2^{Ns} solutions

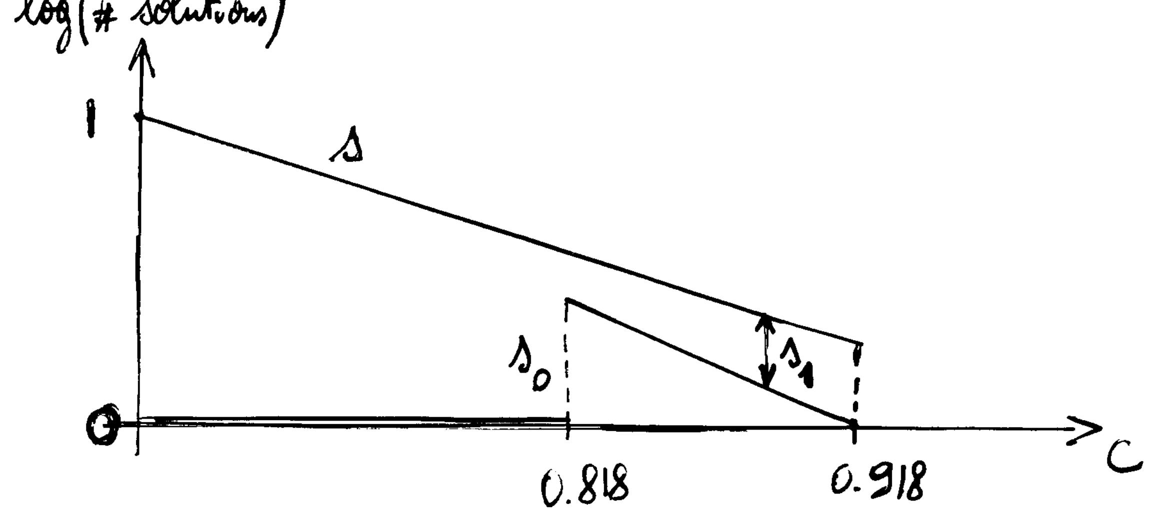
$s = 1 - c$

$d = 1/2$

2^{Ns_0} clusters with 2^{Ns_1} solutions each

$s_0 + s_1 = 1 - c$

$d_0 = 1/2 \quad , \quad d_1 \ll 1/2$



Replica approach (I)

$$\underset{\uparrow \atop \#\text{ solutions}}{\mathcal{N}} = \sum_{\{x_i=0,1\}} \prod_{\ell=1}^{M} \mathbb{1}_{\underbrace{x_{i_\ell} + x_{j_\ell} + x_{k_\ell} = v_\ell}_{\text{instance}}}$$

Compute moments of $\mathcal{N}$ ....

- $$E(\mathcal{N}) = \sum_{\{x_i=0,1\}} \underbrace{E\left(\prod_{\ell=1}^{M} \mathbb{1}_{x_{i_\ell}+x_{j_\ell}+x_{k_\ell}=v_\ell}\right)}_{\underbrace{\prod_{\ell=1}^{M} E\left(\mathbb{1}_{x_{i_\ell}+x_{j_\ell}+x_{k_\ell}=v_\ell}\right)}_{\left[E\left(\mathbb{1}_{x_i \pm x_j + x_k = v}\right)\right]^M}}$$

$$= \sum_{\{x_i=0,1\}} \left[\underbrace{\frac{1}{\binom{N}{3}} \sum_{i<j<k}}_{1} \underbrace{\frac{1}{2}\sum_{v=0,1} \mathbb{1}_{x_i+x_j+x_k=v}}_{\frac{1}{2}} \right]^M$$

$$= 2^N \cdot \left(\frac{1}{2}\right)^M = 2^{N(1-c)}$$

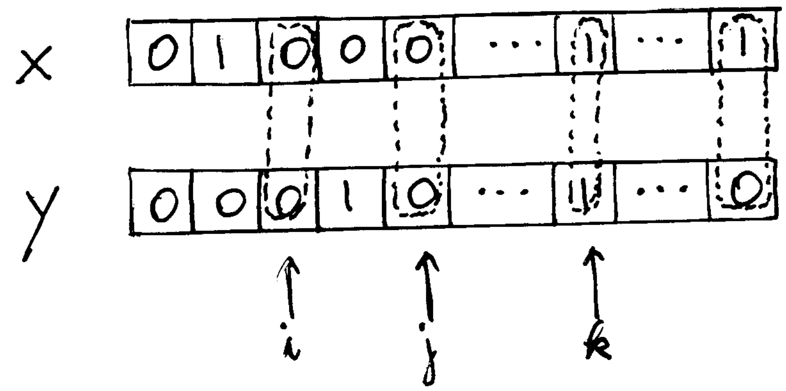
$c_1 = 1$ (upper bound)

Replica approach (II)

$$E(\mathcal{N}^2) = E\left(\left[\sum_{\{x_i=0,1\}} \prod_{\ell=1}^{M} \mathbb{1}_{x_{i_\ell}+x_{j_\ell}+x_{k_\ell}=v_\ell}\right]^2\right)$$

$$= E\left(\sum_{\{x_i=0,1\}} \sum_{\{y_i=0,1\}} \prod_{\ell=1}^{M} \mathbb{1}_{x_{i_\ell}+x_{j_\ell}+x_{k_\ell}=v_\ell} \mathbb{1}_{y_{i_\ell}+y_{j_\ell}+y_{k_\ell}=v_\ell}\right)$$

$$= \sum_{\{x_i=0,1\}} \sum_{\{y_i=0,1\}} \Big\{\underbrace{E\left(\mathbb{1}_{x_i+x_j+x_k=v} \mathbb{1}_{y_i+y_j+y_k=v}\right)}_{\frac{1}{2} E\left(\mathbb{1}_{x_i+x_j+x_k=y_i+y_j+y_k}\right)}\Big\}^M$$



$\Rightarrow$ depends upon Hamming distance $0 \leq d \leq 1$

$$= \sum_{\{x_i=0,1\}} \sum_{\{y_i=0,1\}} \left\{\frac{1}{4}\left(1-\left(2d(x,y)-1\right)^3\right)\right\}^M$$

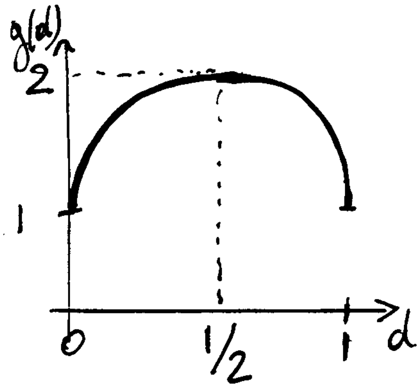
$$= 2^N \sum_{\{y_i=0,1\}} \left[\frac{1}{4}\left(1-\left(2d(0,y)-1\right)^3\right)\right]^M$$

$$= 2^N \sum_{d=0,\frac{1}{N},\dots,1} \left(\frac{1-(2d-1)^3}{4} \right)^M \mathcal{E}(d)$$

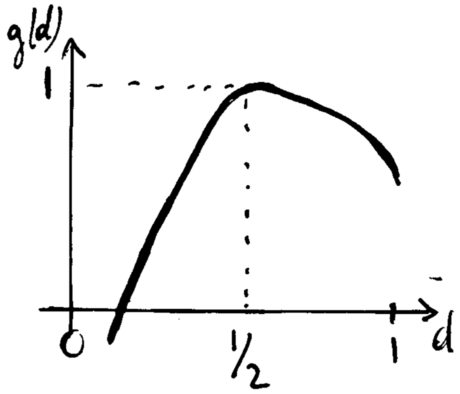
$$\mathcal{E}(d) = \#\,\text{configurations at distance } d \text{ from } (0,0,\dots,0) = \binom{N}{Nd}$$

$$E(\mathcal{N}^2) = \sum_{d=0,\frac{1}{N},\dots,1} 2^{N g(d)}$$

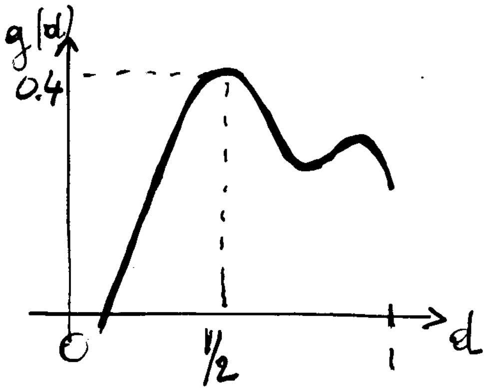
$$g(d) = 1 + c \cdot \log_2 \left(\frac{1-(2d-1)^3}{4} \right) - (1-d)\log_2(1-d) - d \log_2 d + o(1)$$



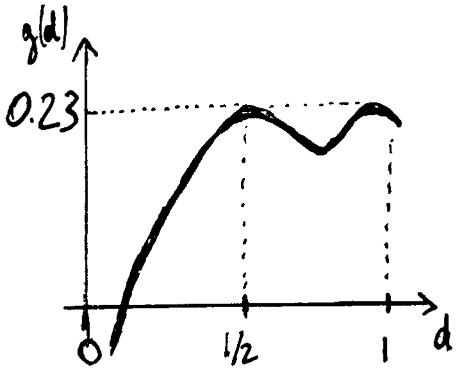
$c = 0$



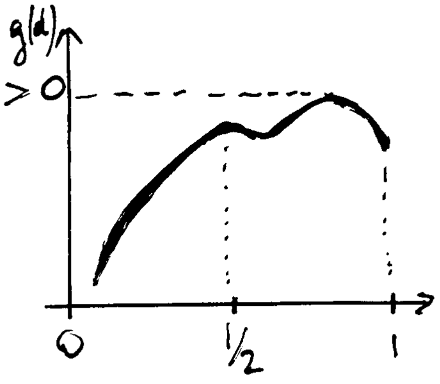
$c = 1/2$



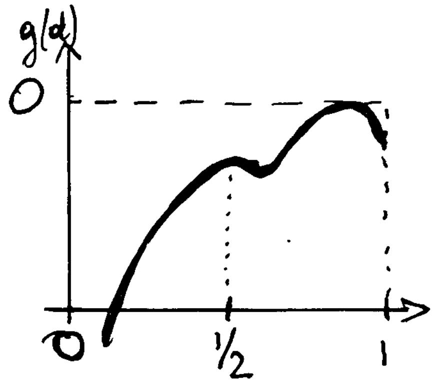
$c = 0.8$



$c = 0.889$

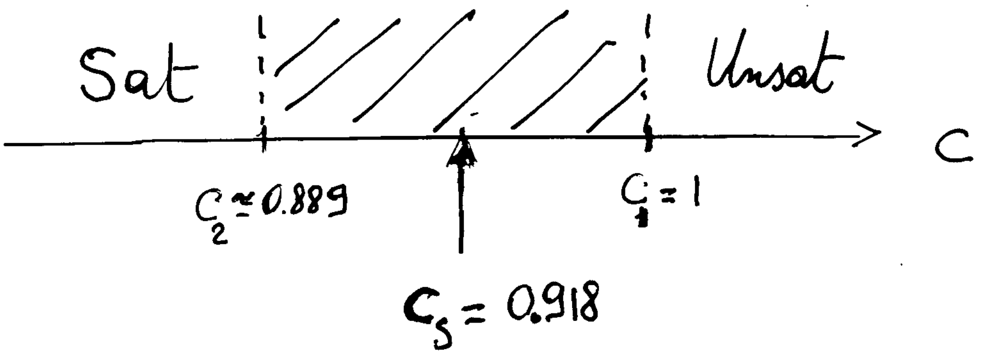


$c = 0.95$



$c = 1.06$

1^{st} and 2^{nd} moments methods

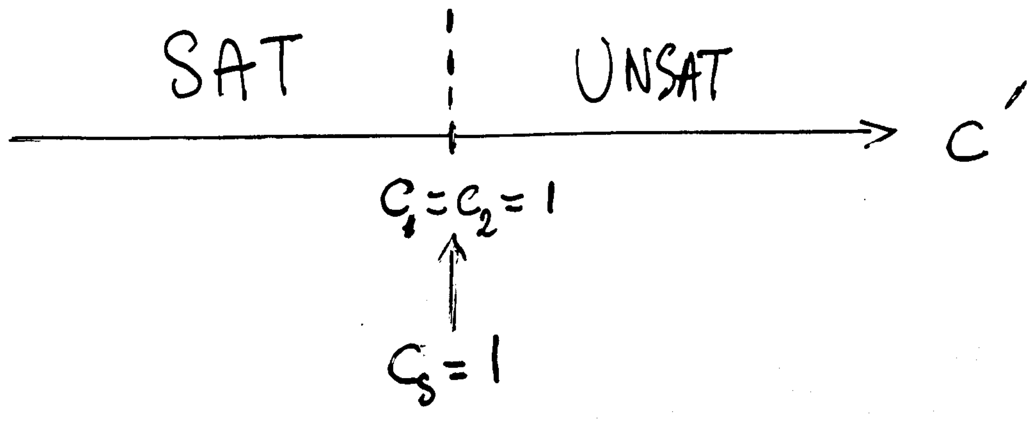


"Random 3-XORSAT" ensemble

Dubois, Mandler
Cocco, R.M.

pure variable algorithm

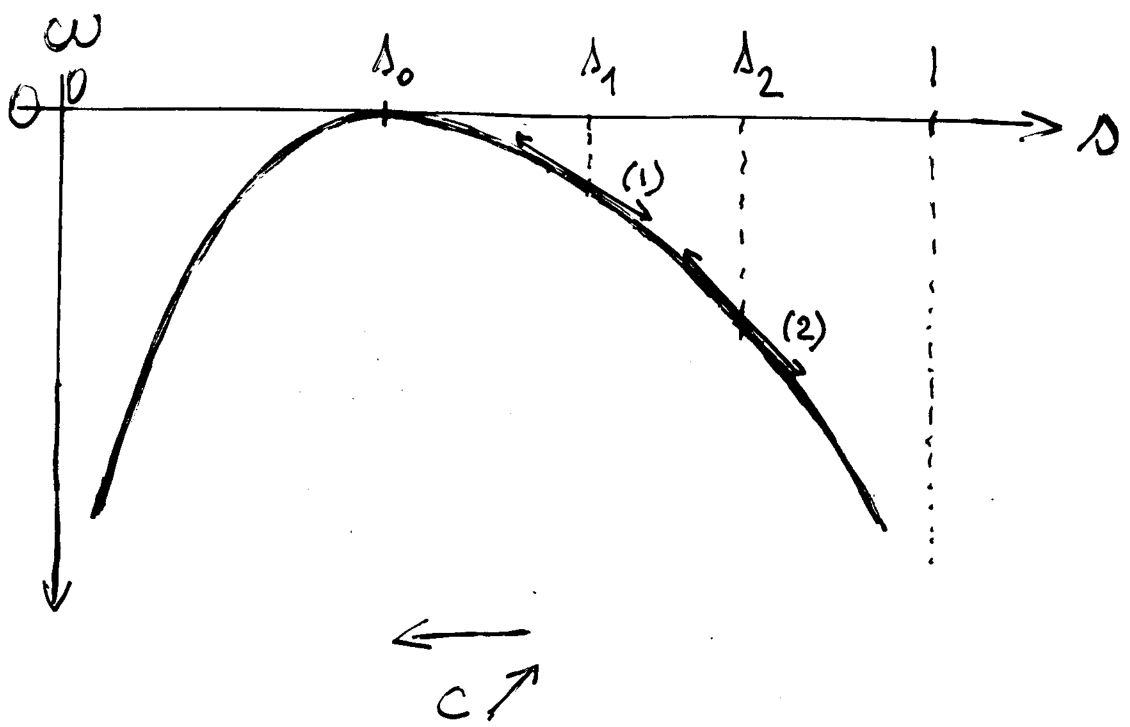
"Constrained Random 3-XORSAT" ensemble



Replicas and Large deviations

$$\mathcal{N} = 2^{Ns}$$

$$\mathcal{P}(\mathcal{N}) = 2^{N\omega(s)}$$



$$E(\mathcal{N}^q) = \sum_{\mathcal{N}=1}^{2^N} \mathcal{P}(\mathcal{N})\,\mathcal{N}^q$$

$$\sim \int_0^1 ds\; 2^{N(\omega(s) + qs)}$$

Laplace method : maximize : $\omega(s) + qs \; (= g(q))$

$$\left.\frac{\partial \omega}{\partial s}\right|_{s^*} = -q$$

Typical case : $q \to 0$

Calculation of $E(\mathcal{N}^q)$

- $E(\mathcal{N})$: immediate
- $E(\mathcal{N}^2)$: $\{x^1, x^2\} \longrightarrow d(x^1, x^2) = d^{12}$

....

- $E(\mathcal{N}^q)$: $\{x^1, x^2, ..., x^q\} \longrightarrow \begin{cases} d(x^a, x^b) = d^{ab} \\ d(x^a, x^b, x^c) = d^{abc} \\ d(x^a, x^b, x^c, x^d) = d^{abcd} \\ \cdots \end{cases}$

$2^q - q - 1$ order parameters

① maximize $g[d^{ab}, d^{abc}, d^{abcd}, ...]$

$$\Rightarrow \quad \frac{\partial g}{\partial d^{a_1 \ldots a_\ell}} = 0 \quad , \quad 2 \leq \ell \leq q$$

② get $g(q)$, then $\omega(s) = \min_q [g(q) - qs]$

and $s_0 = \left.\frac{dg}{dq}\right|_{q \to 0}$

⚠

* q integer-valued , then $q \to 0$.

* $E(\mathcal{N}^q) \leq (2^N)^q \leadsto$ moment theorem applies

but here $N \to \infty$ first , $q \to 0$ next

$\Rightarrow$ No unicity of $q \to 0$ continuation.

Replica Symmetric Theory:

Kacz (1967) (ex Dyson, 1953)
Edwards - Anderson (1974)
Sherrington - Kirkpatrick (1975)

$$\begin{cases} d^{a_1,a_2} = d_2 \\ d^{a_1,a_2,a_3} = d_3 \\ d^{a_1,a_2,a_3,a_4} = d_4 \\ \cdots \end{cases} \qquad \forall\, a_1, a_2, a_3, a_4, \ldots$$

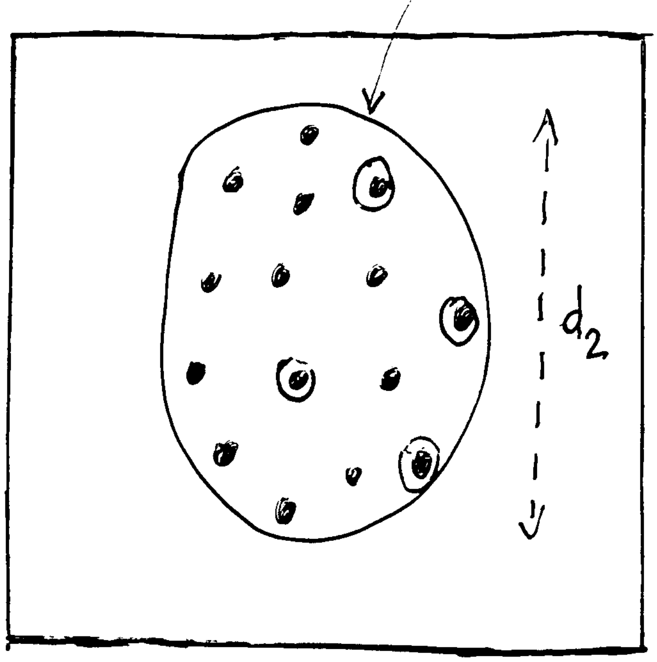
"replicas cannot be distinguished from each other"

Optimize over $d_2, d_3, d_4, \ldots$
Then, send $q \to 0$.

Interpretation:

solution space

Configuration Space



example of $q = 4$ replicas

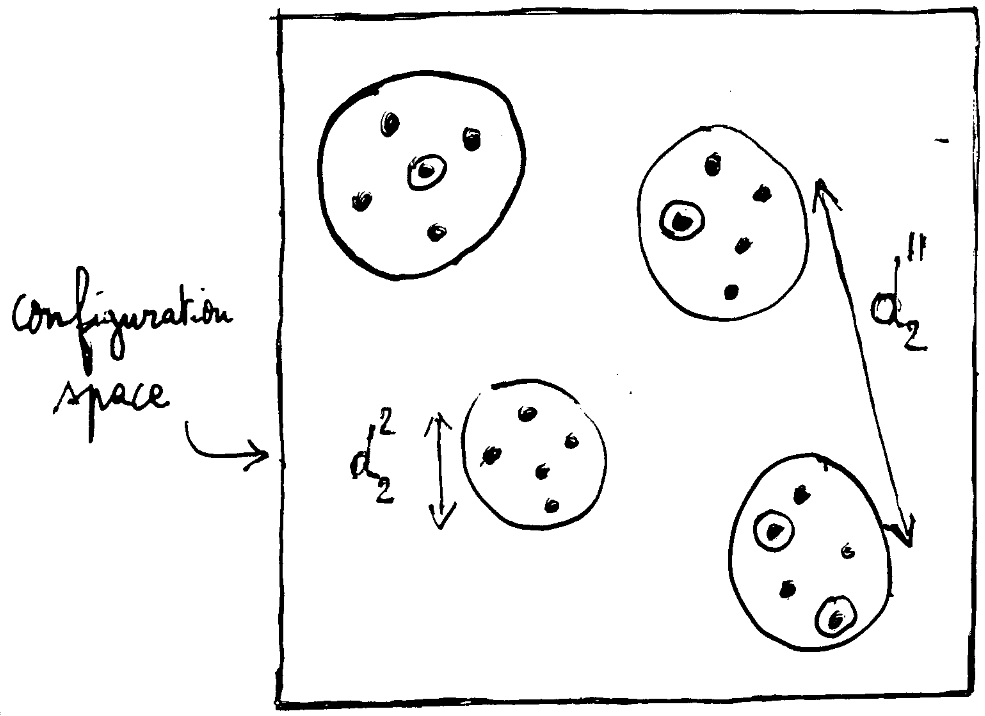
Replica symmetry broken theory

Thouless, Andersson, Palmer 1977
$\Longrightarrow$ Parisi 1979, 1980

$$\begin{cases} d^{a_1,a_2} = d_2^{11},\ d_2^{2} \\ d^{a_1 a_2 a_3} = d_3^{111},\ d_3^{21},\ d_3^{3} \\ d^{a_1 a_2 a_3 a_4} = d_4^{1111},\ d_4^{211},\ d_4^{22},\ d_4^{31},\ d_4^{4} \\ \ldots \end{cases}$$

$\Leftrightarrow$ $d^{a_1,a_2,\ldots,a_\ell}$ has as many values as the number of partitions of ℓ.

Interpretation:



example of $q=4$ replicas (d_4^{211})

higher levels of symmetry breaking ...

Replica predictions for 3-XORSAT

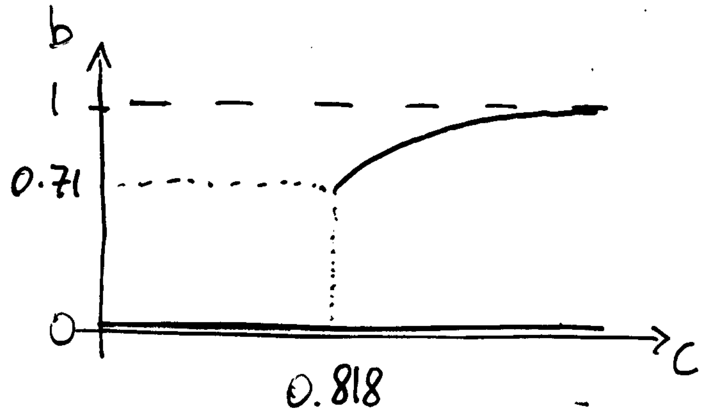
Franz, Ricci, Leone, Weigt, Zecchina (2001)

$$\begin{cases} s_0 = b - 3cb^2 + 2cb^3 & (\log \#\ \text{clusters}) \\ s_1 = 1 - c - s_0 & (\log \#\ \text{solutions in each cluster}) \end{cases}$$

with:

$$b = 1 - e^{-3cb^2}$$

local backbone (in a cluster)



$$\longrightarrow \begin{cases} d_2' = \dfrac{1-b}{2} \\ d_2'' = \dfrac{1}{2} \end{cases}$$

Hand-waving argument

variable x

$$\begin{cases} x + \dots + \dots = \dots \\ x + \dots + \dots = \dots \\ x + \dots + \dots = \dots \end{cases}$$

$$1 - b = \sum_{\ell=0}^{\infty} e^{-3c} \frac{(3c)^\ell}{\ell!} \left(1 - b^2\right)^\ell$$

proba(x free)

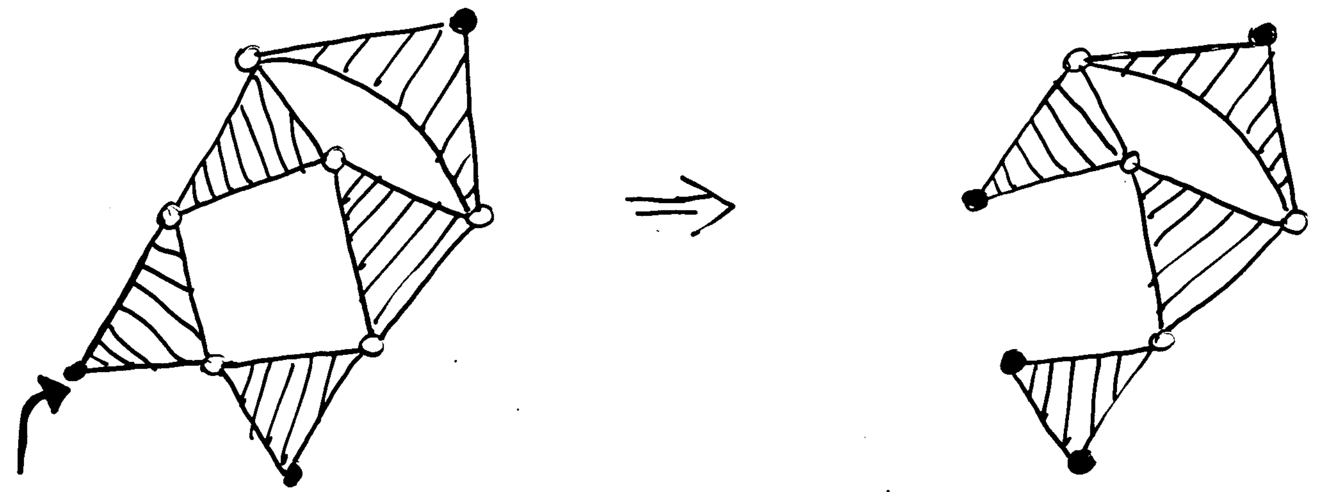
"Pure variable" algorithm

Broder et al. 1983

Franco 1980's

Pittel, Spencer, Wormald 1996

= 2-core percolation on hypergraphs



after T steps (removals):

- $N_\ell(T)$ = # variables with ℓ occurences
- $M(T) = M - T$ = # clauses

$$E\left(N_\ell(T+1) - N_\ell(T) \mid \mathcal{H} \right) = -\mathbb{1}_{\ell,1} + \mathbb{1}_{\ell,0} + 2\,p_{\ell+1} - 2\,p_\ell$$

$$p_j = \frac{j\,E(N_j(T))}{3\,M(T)}$$

Analysis of the algorithm

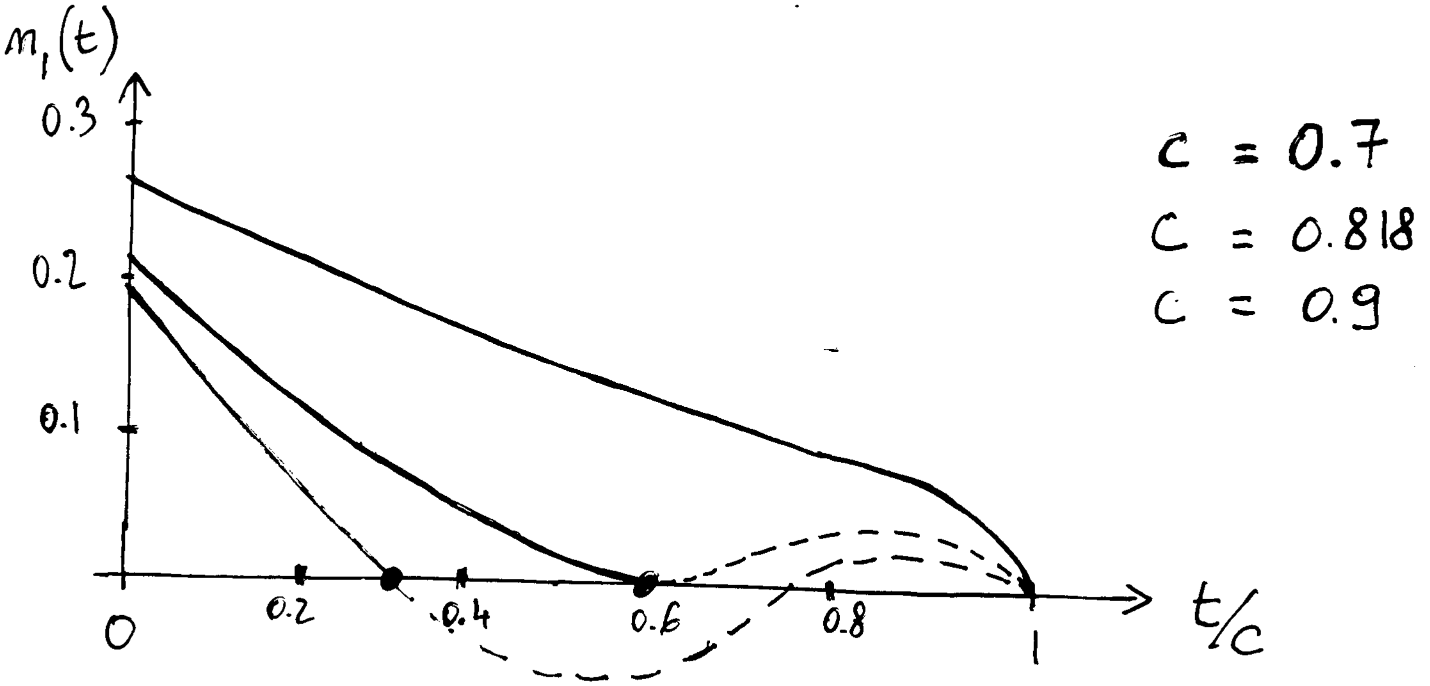
$$N_\ell(T) \longrightarrow N .\; n_\ell\left(t = \frac{T}{N}\right)$$

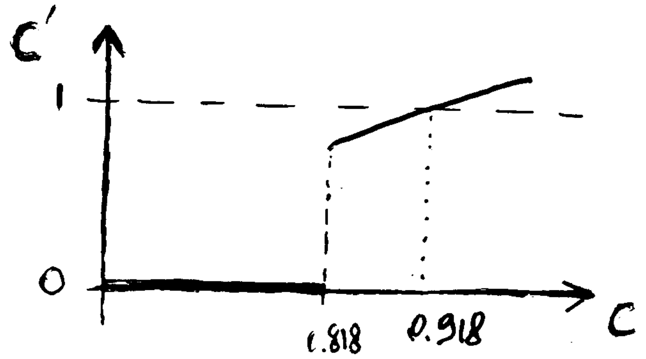
$$\begin{cases} \dfrac{dn_\ell}{dt} = \dfrac{2}{3(c-t)}\Big((\ell+1)\, n_{\ell+1} - \ell\, n_\ell\Big) - \mathbb{1}_{\ell,1} + \mathbb{1}_{\ell,0} \\ n_\ell(0) = e^{-3c}\,\dfrac{(3c)^\ell}{\ell!} \end{cases}$$

solution :

$$\begin{cases} n_1(t) = 3c\,b(t)^2\left(e^{-3c\,b(t)^2} + b(t) - 1\right) \\ n_\ell(t) = e^{-3c\,b(t)^2}\,\dfrac{(3c\,b^2(t))^\ell}{\ell!} \qquad (\ell \geq 2) \end{cases}$$

where $\quad b(t) = \left(1 - \dfrac{t}{c}\right)^{1/3}$





Constrained random 3-XORSAT

$$\mathcal{F} \longrightarrow \mathcal{F}' \subset \mathcal{F}$$

each variable appears at least twice.

$$c' = \frac{M'}{N'}$$

- $\overline{\mathcal{N}} = 2^{N'(1-c')} \Rightarrow c'_s \leq 1$
- $\overline{\mathcal{N}^2}$ more complicated but can be done (Dubois, Mandler 2002)

$$\frac{\overline{\mathcal{N}}^2}{\overline{\mathcal{N}^2}} > 0 \text{ if } c<1 \Rightarrow c'_s \geq 1$$

$$c'_s = 1$$

- solutions of $\mathcal{F}'$:

$$d_0 = \frac{1}{2}$$

$$s_0 = \log_2 \mathcal{N} = 1 - c'$$
$$= b - 2cb^2 + 3cb^3$$

$\Updownarrow$

clusters !

Reconstructing solutions

$$
\begin{tikzcd}
\mathcal{F} \arrow[rr, bend left, "\text{removal}", "\text{pure literal } (= \text{clause})"'] & & \mathcal{F}' \arrow[d] \\
\text{cluster of solutions} \arrow[u, dotted] & & \text{pick a solution} \arrow[ll, bend left, "\text{add back clauses}"]
\end{tikzcd}
$$

solution of $\mathcal{F}'$ + clause $(x + y + z)$

pure literal

- $$s_1 = 2 \int_0^{t_{stop}} d\tau \; \rho_1(\tau) \; + \; e^{-3c}$$

  variables that never appear

- $$\underbrace{\underbrace{1 \text{ solution of } \mathcal{F}'}_{\text{seed of the cluster}} + \text{frozen var.}}_{\text{backbone } b} + \text{free var.}$$

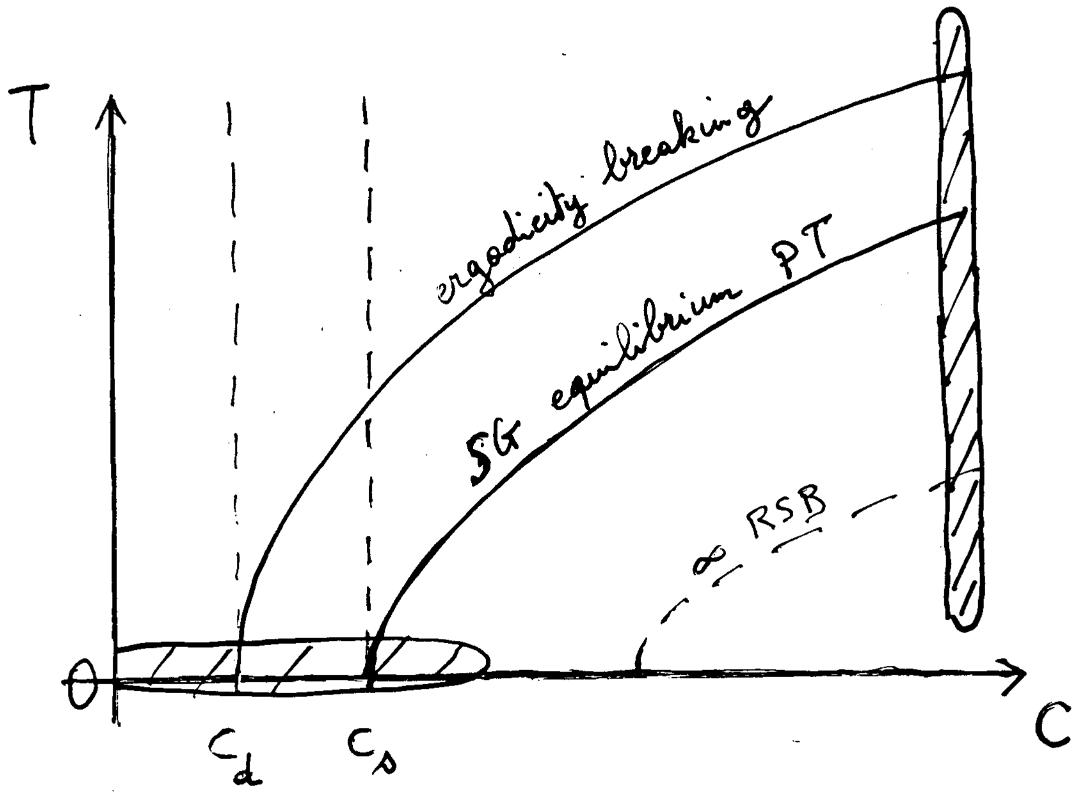
  $$\downarrow$$

  $$d_1 = \frac{1-b}{2}$$

Finite temperature

$$\underbrace{\mathcal{F} + \text{clause with pure literal}}_{Z} \longrightarrow \underbrace{\mathcal{F}}_{Z'}$$

$$Z = 2^{V-1}\left(e^{-\beta} + e^{\beta}\right) \cdot Z'$$



- easy to generalize to larger $K \geqslant 4$
- case $K = 2$.

Back to random K-SAT

$$x_i = 0,1 \longleftrightarrow S_i = -1,+1$$

3-XORSAT : $$\text{Energy} = -\frac{1}{2}\sum_{\ell=1}^{M} J_\ell\, S_{i_\ell} S_{j_\ell} S_{k_\ell} + \frac{M}{2}$$

3-SAT : $$\text{Energy} = +\sum_{\ell=1}^{M}\left(\frac{1+\sigma_{1\ell}S_{i\ell}}{2}\right)\left(\frac{1+\sigma_{2\ell}S_{j\ell}}{2}\right)\left(\frac{1+\sigma_{3\ell}S_{k\ell}}{2}\right)$$

$$= -\sum_\ell \underbrace{J_\ell}_{-\frac{1}{8}\sigma_{i\ell}\sigma_{j\ell}\sigma_{k\ell}} S_{i\ell}S_{j\ell}S_{k\ell} + SS + S + \text{cst}$$

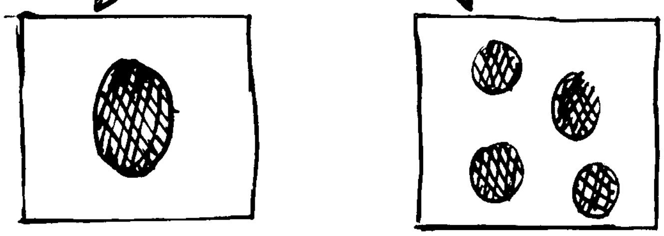
SAT UNSAT

$0 \longrightarrow \alpha$

α_d α_c

3.96 Biroli, R.M., Weigt '99

3.92 Mézard, Parisi, Zecchina '02



$b=0$	$b_\ell>0$, $b_g=0$	$b_g>0$
	α_d	α_c