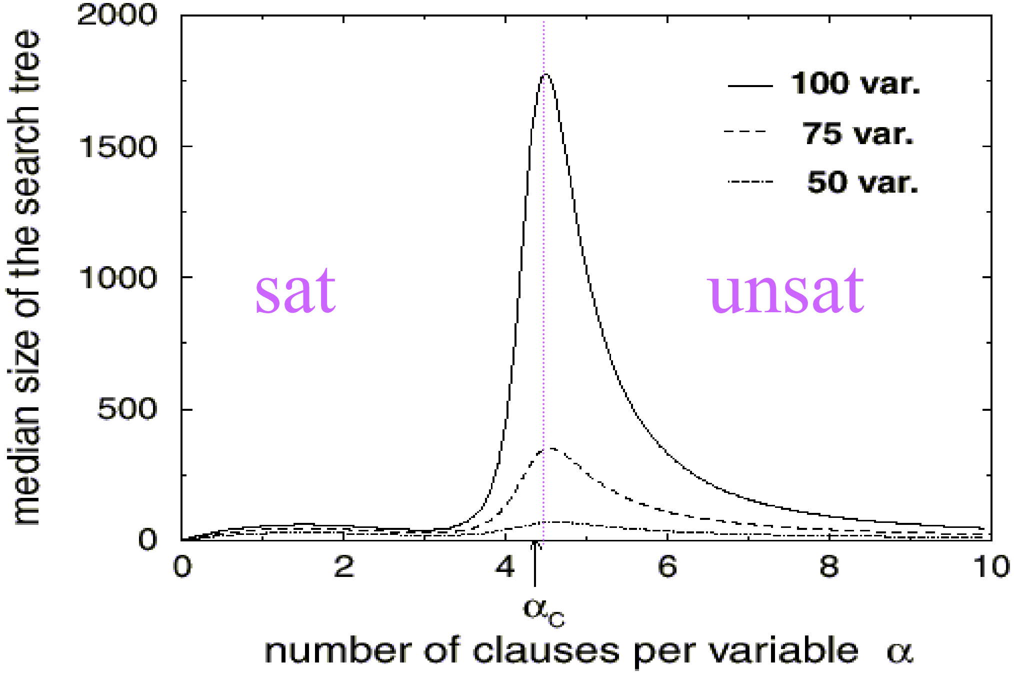
Typical resolution complexity of 3-SAT



easy-hard-less hard pattern

Rigorous results

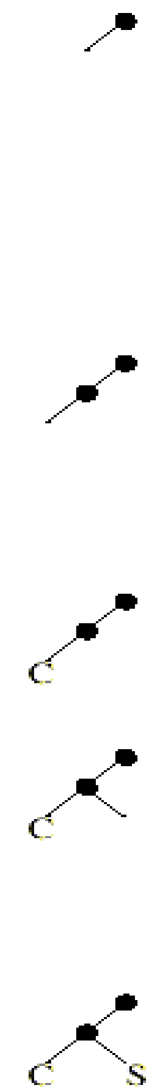
- linear if $\alpha << \alpha_C$
- exponential if $\alpha > \alpha_C$
- “time” < 1.51^N

How to solve 3-SAT?

"Branch & bound" search algorithm

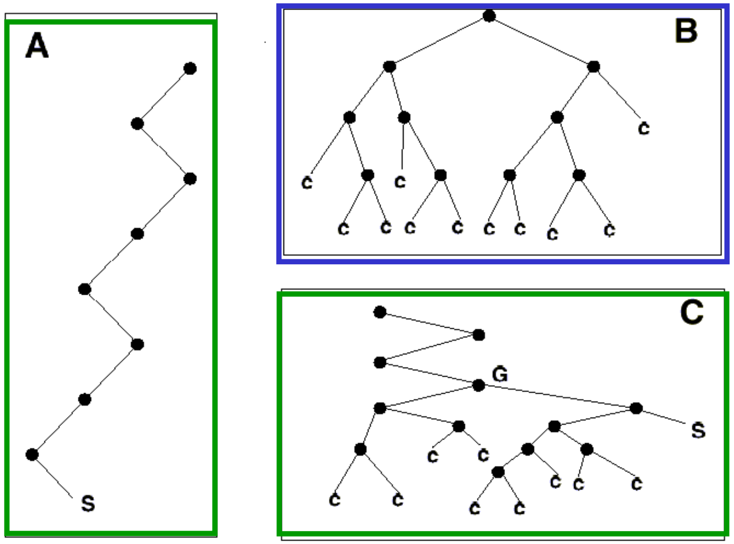
Davis, Putnam '60

step	clauses	search tree
0	$w \vee \bar{x} \vee y$ $\bar{w} \vee x \vee z$ $\bar{w} \vee \bar{x} \vee \bar{y}$ $\bar{w} \vee \bar{x} \vee y$ $x \vee y \vee \bar{z}$	
1	split : $w = T$	
2	$x \vee z$ $\bar{x} \vee \bar{y}$ $\bar{x} \vee y$ $x \vee y \vee \bar{z}$	
3	split : $x = T$	
4	$\bar{y}$ y	
5	propagation : $y = F$, $y = T$ contradiction	
6	backtracking to stage 1 : $x = F$	
7	z $y \vee \bar{z}$	
8	propagation : $z = T$, $y = T$ solution : $w = T$, $x = F$, $y = T$, $z = T$	



Backtrack algorithm, search tree and heuristic

Davis-Putnam algorithm = heuristic + backtracking
↓
search tree



A satisfiable instance (easy)

B unsatisfiable instance (hard)

C satisfiable instance (hard)

- Unit-Clause (UC): pick variable in 1-clause if any, or any unset variable
- Generalized unit-clause (GUC): pick variable in shortest clause
- Shortest Clause With Majority (SC_1): pick most frequent variable in 3-clauses

Chao, Franco '86, '90

Trajectories and the 2+p-SAT problem

clauses with 3 var.

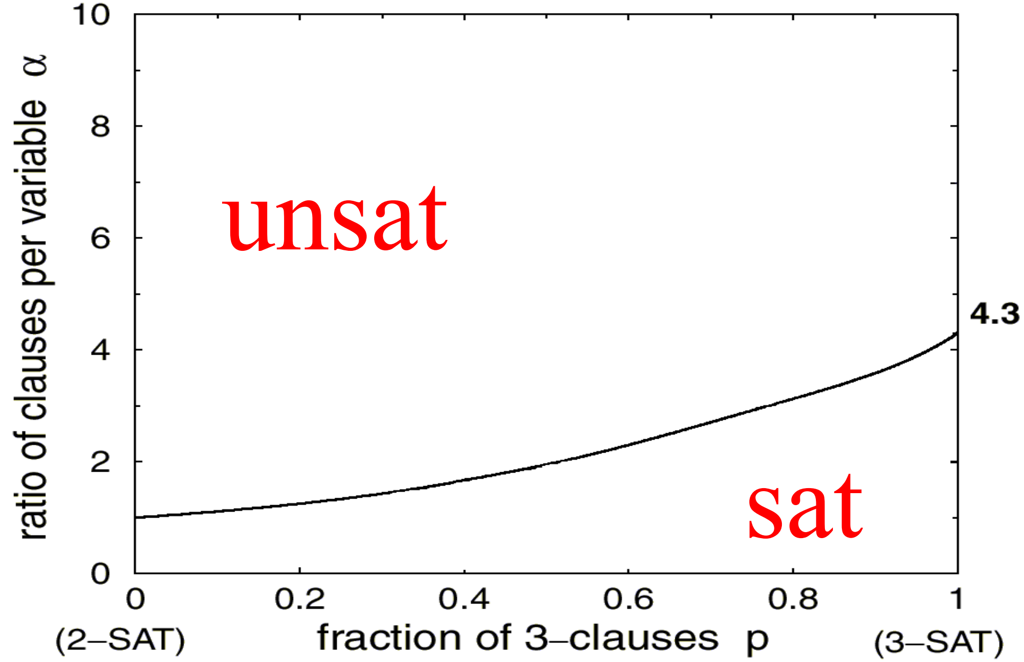
α

"dynamics" of the algorithm

clauses with 2 or 3 var.

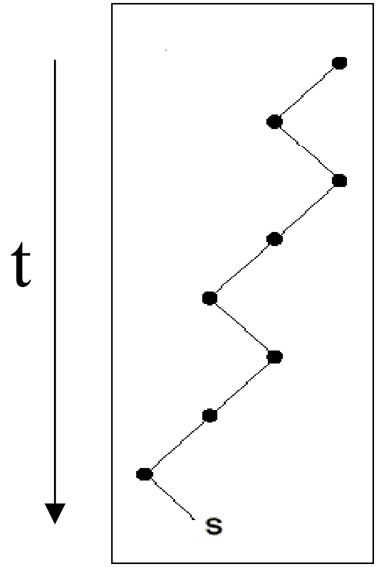
α , p

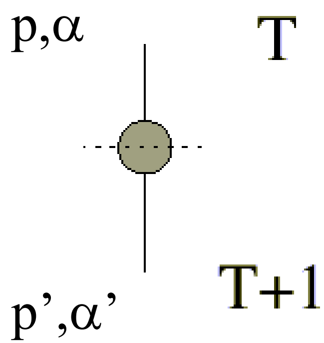
phase diagram of the 2+p-SAT model



Monasson, Zecchina, Kirpatrick, Selman, Troyansky '99
Achlioptas, Kirousis, Kranakis, Krizanc '01

Satisfiable and easy instances $\alpha < 3.003$





$$p(0) = 1\,,\ \alpha(0) = \alpha_0$$

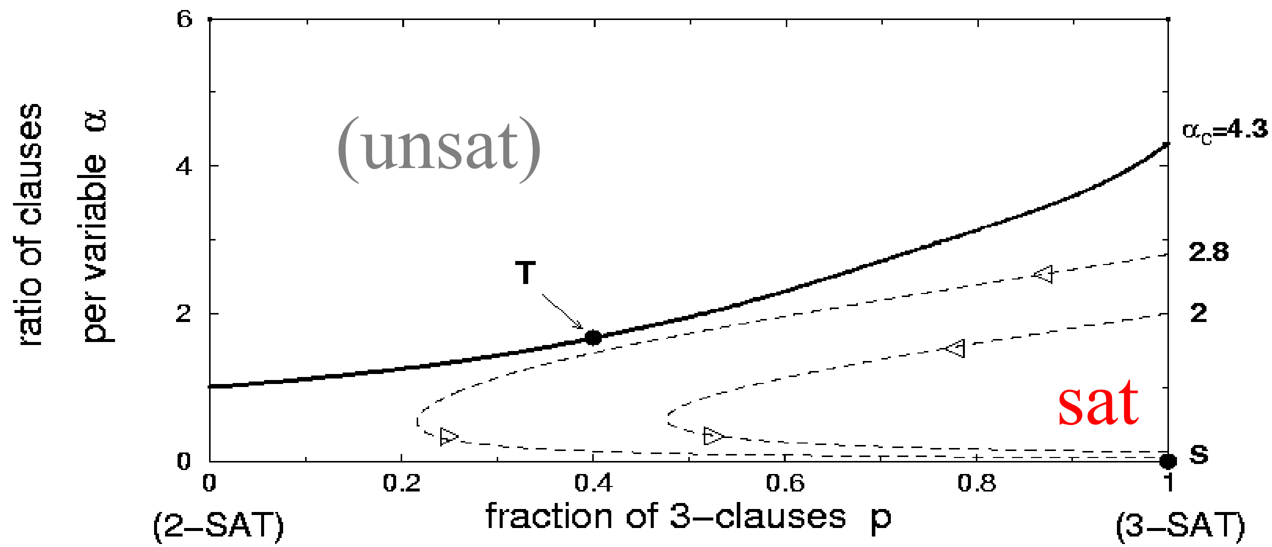
$$\frac{dp}{dt} = F_p(p,\alpha,t)$$

$$\frac{d\alpha}{dt} = F_\alpha(p,\alpha,t)$$

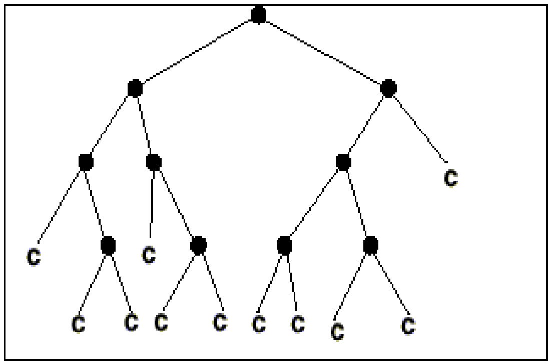
(ODE)

Chao, Franco ‘90

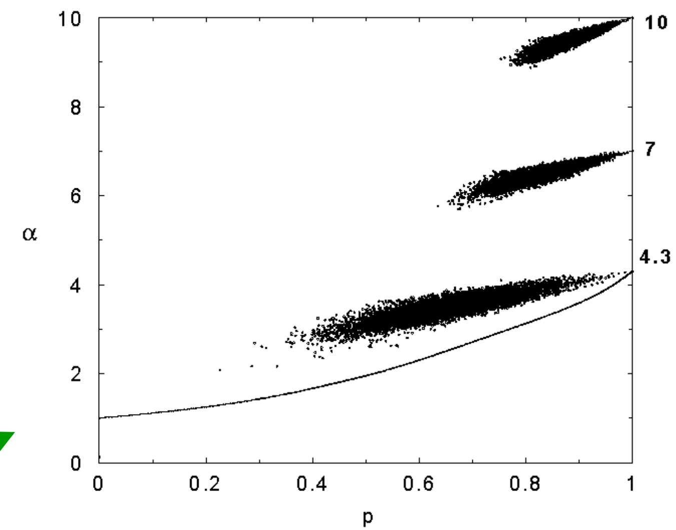
Frieze, Suen ‘96

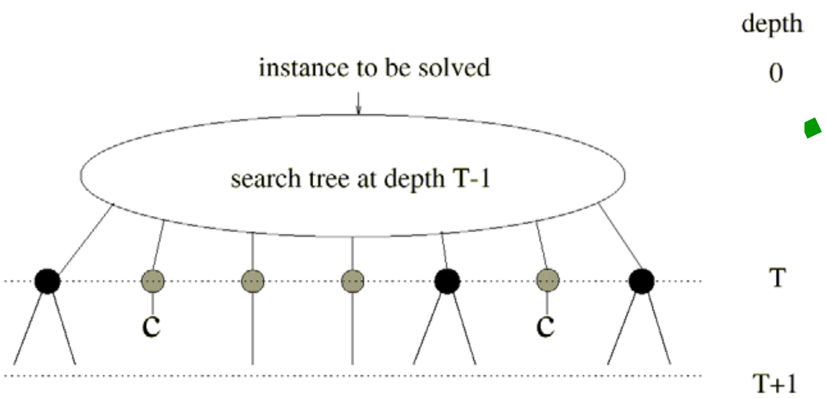


Unsatisfiable, hard instances $\alpha > 4.3$



DPLL induces a non Markovian evolution of the search tree

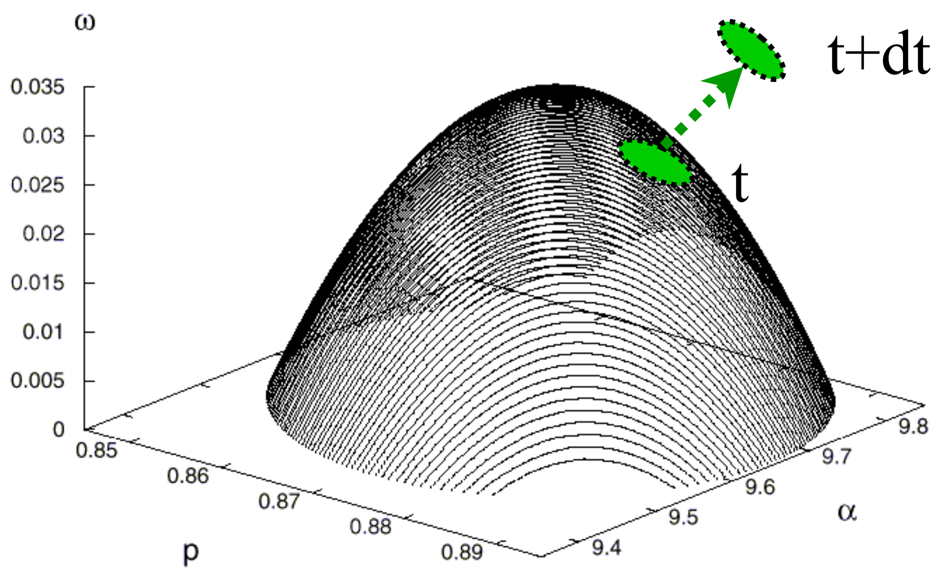




Imaginary, and parallel building up of the search tree

The search for solutions, a growth process

one branch: *p(t) , α(t)* → many branches: *ω (p,α,t)*



$$\frac{\partial \omega}{\partial t} = \mathcal{H}\left[p, \alpha, \frac{\partial \omega}{\partial p}, \frac{\partial \omega}{\partial \alpha}, t \right] \qquad \text{(PDE)}$$

Comparison to numerical experiments

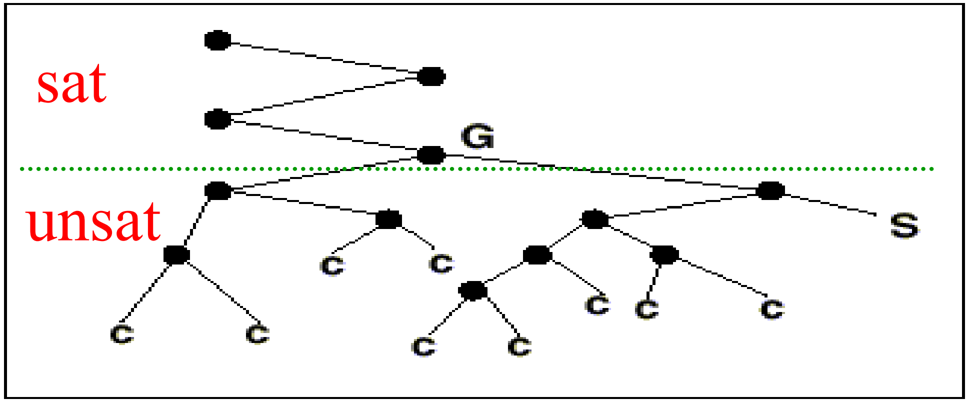
$$Q = 2^{N\omega}$$

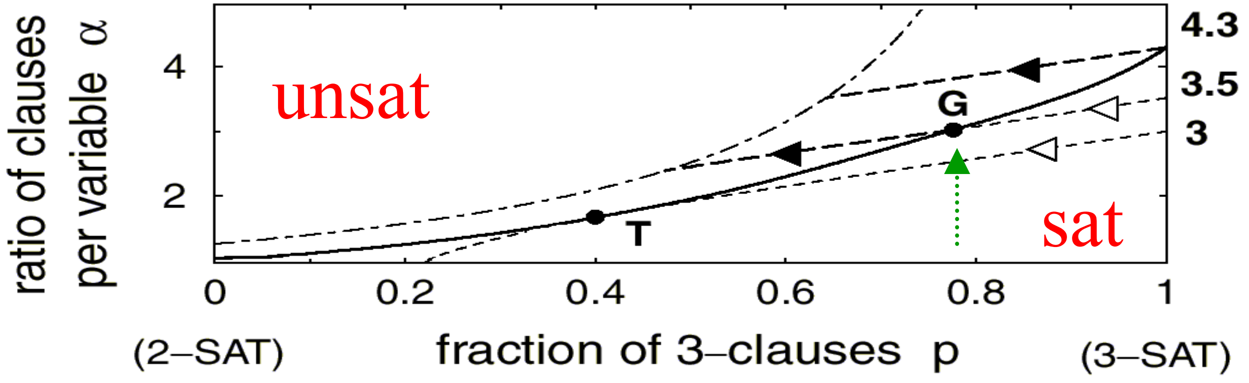
	Initial Ratio α_0	Experiments $\log_2 Q$	$\log_2 B$	Theory $\hat{\omega}$
unsat	20	0.0153 ± 0.0002	0.0151 ± 0.0001	0.0152
	15	0.0207 ± 0.0002	0.0206 ± 0.0001	0.0206
	10	0.0320 ± 0.0005	0.0317 ± 0.0002	0.0319
	7	0.0482 ± 0.0005	0.0477 ± 0.0005	0.0477
	4.3	0.089 ± 0.001	0.0895 ± 0.001	0.0875
sat	3.5	0.034 ± 0.003		0.035
	G	0.040 ± 0.002	0.041 ± 0.003	0.044
		(nodes)	(leaves)	

$$\omega = \frac{3+\sqrt{5}}{6\ln 2}\left[\ln\left(\frac{1+\sqrt{5}}{2}\right)\right]^2 \frac{1}{\alpha} \approx \frac{0.292}{\alpha}$$

Beame, Karp, Pitassi, Saks ‘98

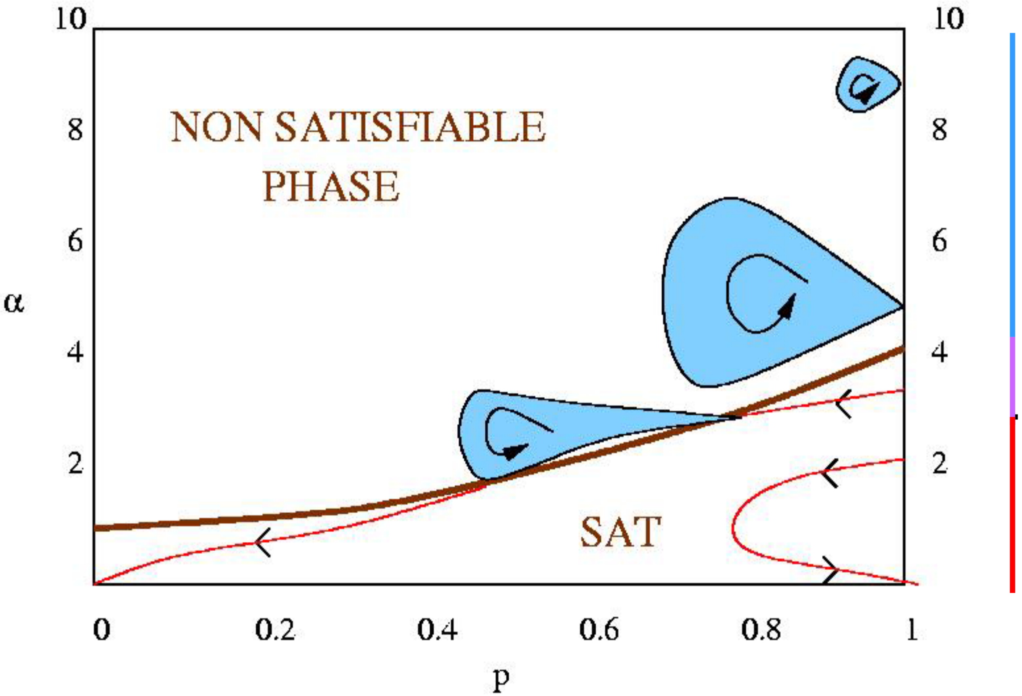
Satisfiable, hard instances 3.003< α < 4.3
(which could made be easier?)

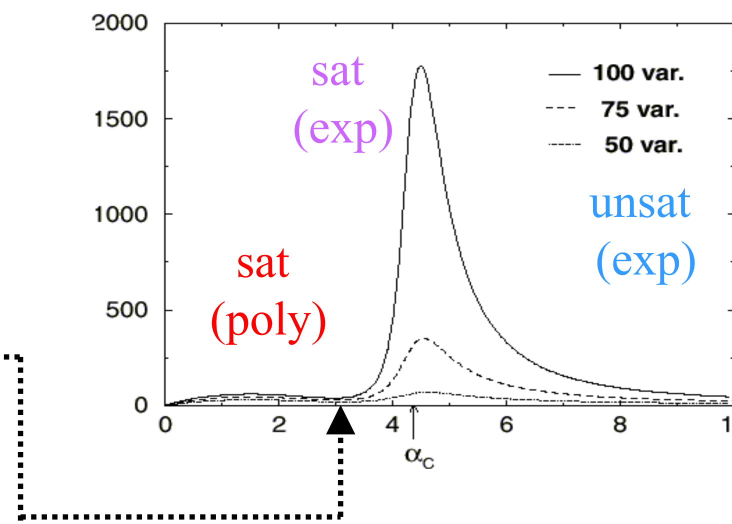




The complexity of 3-SAT solving is strongly affected by the phase transitions of 2+p-SAT !

The polynomial/exponentiel crossover





"dynamical" transition
(depends on the heuristic)

$$\begin{aligned} UC &: \quad 2.667 \\ GUC &: \quad 3.003 \\ M &: \quad \approx 3 \end{aligned}$$

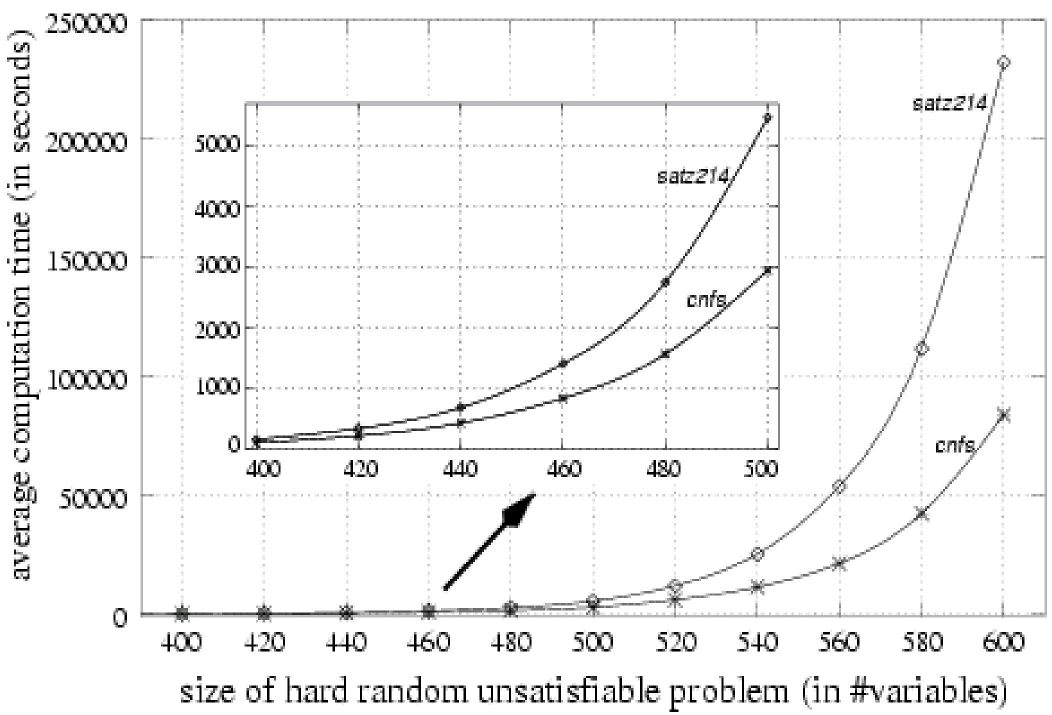
but

$$p_T = \frac{2}{5}, \alpha_T = \frac{5}{3}$$

T is largely heuristic independent …. (and close to tricritical point!)

Application
I. Search heuristic and backbone

Heuristic to assign variables : Pick up variable that eliminates the largest number of clauses.



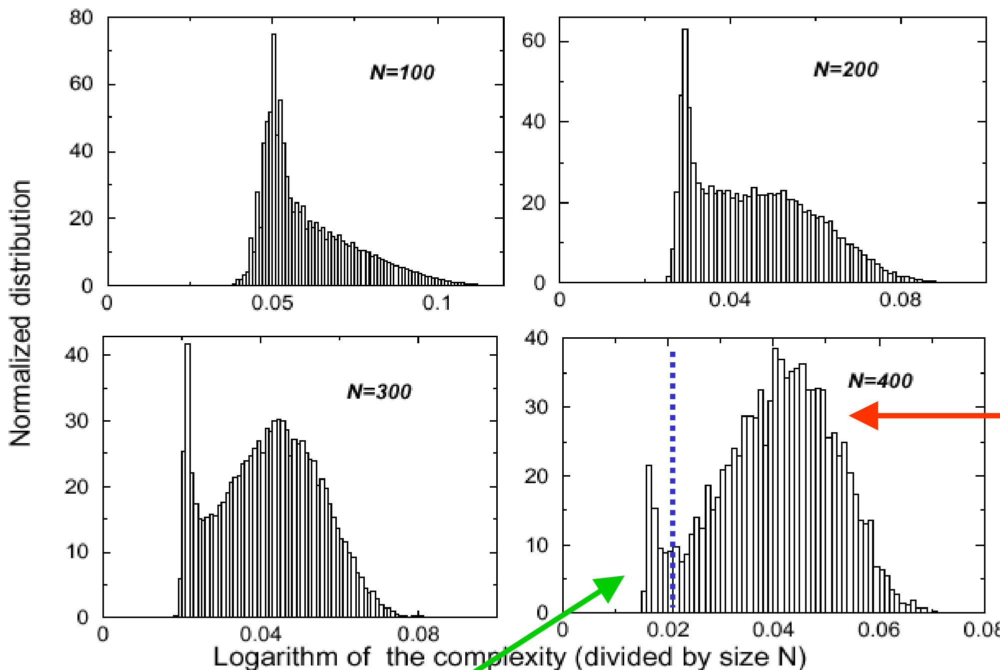
Dubois, Dequen '00

Choose a variable likely to be in the backbone

Application II. *Fluctuations and restarts*

Histograms of solving times

$\alpha=3.5$



Exponential regime
Complexity $= 2^{0.035\,N}$

Linear regime
Very rare! frequence $= 2^{-0.011\,N}$

Resolution through systematic stop-and-restart of the search:

- stop algorithm after time N;
- restart until a solution is found.

Cocco, R.M.
Montanari, Zecchina '01

Time of resolution : $2^{0.035\,N}$ → $2^{0.011\,N}$

Conclusions

- Computational problems can be studied with statistical physics concepts and techniques

  (replica method, phase diagram, dynamical trajectories, growth processes, ...)

- General framework for the probabilistic analysis of hard decision or optimization problems for both static and dynamic properties

  (Traveling Salesman Problem, Vertex Cover, Graph Coloring, ...)

- Open Issues:

  - *robustness to instance perturbation*

    (replica symmetry breaking vs. droplet theory)
  - *study of approximation algorithm*
  - *question of probabilistic analysis (in physics too?)*
    - * *more realistic distributions*
    - * *analysis of algorithm for a given instance (thermal vs. quenched disorder)*