Gibbs states and the set of solutions of random constraint satisfaction problems

Florent Krząkała[a], Andrea Montanari[b,c,d], Federico Ricci-Tersenghi[e], Guilhem Semerjian[c], and Lenka Zdeborová[f]

[a]Laboratoire de Physico-Chimie Théorique, Ecole Supérieure de Physique et de Chimie Industrielles, 75005 Paris, France, [b]Departments of Electrical Engineering and Statistics, Stanford University, CA 94305; [c]Laboratoire de Physique Théorique, Ecole Normale Supérieure, Université Pierre et Marie Curie, 75005 Paris, France; [e]Dipartimento di Fisica and Consiglio Nazionale delle Ricerche–Istituto Nazionale per la Fisica della Materia, Università di Roma ''La Sapienza,'' I-00185 Rome, Italy; and [f]Laboratoire de Physique Théorique et Modèles Statistiques, Université de Paris-Sud, 91405 Orsay, France

Communicated by Giorgio Parisi, University of Rome, Rome, Italy, April 25, 2007 (received for review November 29, 2006)

An instance of a random constraint satisfaction problem defines a random subset $\mathcal{S}$ (the set of solutions) of a large product space $\mathcal{X}^N$ (the set of assignments). We consider two prototypical problem ensembles (random *k*-satisfiability and *q*-coloring of random regular graphs) and study the uniform measure with support on $\mathcal{S}$. As the number of constraints per variable increases, this measure first decomposes into an exponential number of pure states (''clusters'') and subsequently condensates over the largest such states. Above the condensation point, the mass carried by the *n* largest states follows a Poisson-Dirichlet process. For typical large instances, the two transitions are sharp. We determine their precise location. Further, we provide a formal definition of each phase transition in terms of different notions of correlation between distinct variables in the problem. The degree of correlation naturally affects the performances of many search/sampling algorithms. Empirical evidence suggests that local Monte Carlo Markov chain strategies are effective up to the clustering phase transition and belief propagation up to the condensation point. Finally, refined message passing techniques (such as survey propagation) may also beat this threshold.

message passing algorithms | phase transitions | random graphs

Constraint satisfaction problems (CSPs) arise in a large spectrum of scientific disciplines. An instance of a CSP is said to be satisfiable if there exists an assignment of N variables $(x_1, x_2, \dots, x_N) \equiv \underline{x}$, $x_i \in \mathcal{X}$ ($\mathcal{X}$ being a finite alphabet), which satisfies all of the constraints within a given collection. The problem is in finding such an assignment or showing that the constraints are unsatisfiable. More precisely, one is given a set of functions $\psi_a : \mathcal{X}^k \to \{0, 1\}$, with $a \in \{1, \dots, M\} \equiv [M]$ and of k-tuples of indices $\{i_a(1), \dots, i_a(k)\} \subseteq [N]$, and has to establish whether there exists $\underline{x} \in \mathcal{X}^N$ such that $\psi_a(x_{i_a(1)}, \dots, x_{i_a(k)}) = 1$ for all as. In this article we shall consider two well known families of CSPs [both known to be NP-complete (1)]:

(*i*) k-satisfiability (k-SAT) with $k \geq 3$. In this case $\mathcal{X} = \{0, 1\}$. The constraints are defined by fixing a k-tuple $[z_a(1), \dots, z_a(k)]$ for each a, and setting $\psi_a(x_{i_a(1)}, \dots, x_{i_a(k)}) = 0$ if $(x_{i_a(1)}, \dots, x_{i_a(k)}) = (z_a(1), \dots, z_a(k))$ and $= 1$ otherwise.

(*ii*) q-coloring (q-COL) with $q \geq 3$. Given a graph G with N vertices and M edges, one is asked to assign colors $x_i \in \mathcal{X} \equiv \{1, \dots, q\}$ to the vertices in such a way that no edge has the same color at both ends.

The optimization (maximize the number of satisfied constraints) and counting (count the number of satisfying assignments) versions of this problems are defined straightforwardly. It is also convenient to represent CSP instances as factor graphs (2), i.e., bipartite graphs with vertex sets $[N]$, $[M]$ including an edge between node $i \in [N]$ and $a \in [M]$ if, and only if, the ith variable is involved in the ath constraint (compare Fig. 1). This representation allows one to define naturally a distance $d(i, j)$ between variable nodes.

Ensembles of random CSPs (rCSPs) were introduced (see e.g., ref. 3) with the hope of discovering generic mathematical phenomena that could be exploited in the design of efficient algorithms. Indeed several search heuristics, such as Walk-SAT (4) and ''myopic'' algorithms (5) have been successfully analyzed and optimized over rCSP ensembles. The most spectacular advance in this direction has probably been the introduction of a new and powerful message passing algorithm (survey propagation, SP) (6). The original justification for SP was based on the (nonrigorous) cavity method from spin glass theory. Subsequent work proved that standard message passing algorithms (such as belief propagation, BP) can indeed be useful for some CSPs (7–9). Nevertheless, the fundamental reason for the (empirical) superiority of SP in this context remains to be understood and is a major open problem in the field. Building on a refined picture of the solution set of rCSP, this article provides a possible (and testable) explanation. We consider two ensembles that have attracted the majority of work in the field: (*i*) random k-SAT: each k-SAT instance with N variables and $M = N\alpha$ clauses is considered with the same probability; (*ii*) q-COL on random graphs: the graph G is uniformly random among the ones over N vertices, with uniform degree l (the number of constraints is therefore $M = Nl/2$).

Phase Transitions in rCSP

It is well known that rCSPs may undergo phase transitions as the number of constraints per variable α is varied.[g] The best known of such phase transitions is the SAT-UNSAT one: as α crosses a critical value $\alpha_s(k)$ (that can, in principle, depend on N), the instances pass from being satisfiable to unsatisfiable with high probability[h] (10). For k-SAT, it is known that $\alpha_s(2) = 1$. A conjecture based on the cavity method was put forward in ref. 6 for all $k \geq 3$ that implied in particular the values presented in Table 1 and $\alpha_s(k) = 2^k \log 2 - (1 + \log 2)/2 + O(2^{-k})$ for large k (11). Subsequently, it was proved that $\alpha_s(k) = 2^k \log 2 - O(k)$, confirming this asymptotic behavior (12). An analogous conjecture for q-COL was proposed in ref. 13 yielding, for regular random graphs (14), the values reported in Table 1 and $l_s(q) = 2q \log q - \log q - 1 + o(1)$ for large q [according to our convention, random graphs are with high probability uncolorable if $l \geq l_s(q)$]. It was proved in refs. 12 and 15 that $l_s(q) = 2q \log q - O(\log q)$.

Even more interesting and challenging are phase transitions in the structure of the set $\mathcal{S} \subseteq \mathcal{X}^N$ of solutions of rCSP's (''structural'' phase transitions). Assuming the existence of solutions, a conve-

Author contributions: F.K., A.M., F.R.-T., G.S., and L.Z. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

Abbreviations: CSP, constraint satisfaction problem; rCSP, random CSP; k-SAT, k-satisfiability; q-COL, q-coloring; SP, survey propagation; BP, belief propagation; 1RSB, one-step replica symmetry breaking.

[d]To whom correspondence should be addressed. E-mail: montanari@stanford.edu.

[g]For coloring l-regular graphs, we can use $l = 2\alpha$ as a parameter. When considering a phase transition defined through some property $\mathcal{P}$ increasing in l, we adopt the convention of denoting its location through the smallest integer such that $\mathcal{P}$ holds.

[h]The term ''with high probability'' means with probability approaching one as $N \to \infty$.

www.pnas.org/cgi/doi/10.1073/pnas.0703685104

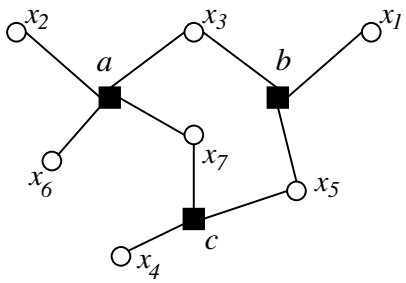

Fig. 1. The factor graph of a small CSP allows to define the distance $d(i, j)$ between variables x_i and x_j (filled squares are constraints and empty circles variables). Here, for instance, $d(6, 1) = 2$ and $d(3, 5) = 1$.

nient way of describing $\mathcal{S}$ is to introduce the uniform measure over solutions $\mu(\underline{x})$:

$$\mu(\underline{x}) = \frac{1}{Z}\prod_{a=1}^{M}\psi_a(x_{i_a(1)}, \dots, x_{i_a(k)}), \qquad [1]$$

where $Z \geq 1$ is the number of solutions. Let us stress that, since $\mathcal{S}$ depends on the rCSP instance, $\mu(\cdot)$ is itself random.

We shall now introduce a few possible "global" characterizations of the measure $\mu(\cdot)$. Each one of these properties has its counterpart in the theory of Gibbs measures, and we shall partially adopt that terminology here (17).

To define the first of such characterizations, we let $i \in [N]$ be a uniformly random variable index, denote as $\underline{x}_\ell$ the vector of variables whose distance from i is at least ℓ, and by $\mu(x_i|\underline{x}_\ell)$ the marginal distribution of x_i given $\underline{x}_\ell$. Then we say that the measure (Eq. **1**) satisfies the uniqueness condition if for any given $i \in [N]$,

$$\mathbb{E}\sup_{\underline{x}_\ell,\,\underline{x}'_\ell}\sum_{x_i\in\mathcal{X}}|\mu(x_i|\underline{x}_\ell) - \mu(x_i|\underline{x}'_\ell)| \to 0. \qquad [2]$$

as $\ell \to \infty$ (here and below the limit $N \to \infty$ is understood to be taken before $\ell \to \infty$). This expresses a "worst case" correlation decay condition. Roughly speaking: the variable x_i is (almost) independent of the far apart variables $\underline{x}_\ell$ irrespective of the instance realization and the variables distribution outside the horizon of radius ℓ. The threshold for uniqueness (above which uniqueness ceases to hold) was estimated in ref. 9 for random k-SAT, yielding $\alpha_u(k) = (2 \log k)/k\,[1 + o(1)]$ (which is asymptotically close to the threshold for the pure literal heuristics) and in ref. 18 for coloring implying $l_u(q) = q$ for q large enough (a "numerical" proof of the same statement exists for small q). Below such thresholds BP can be proved to return good estimates of the local marginals of the distribution (Eq. **1**).

Notice that the uniqueness threshold is far below the SAT-UNSAT threshold. Furthermore, several empirical studies (19, 20) pointed out that BP [as well as many other heuristics (4, 5)] is effective up to much larger values of the clause density. In a remarkable series of papers (6, 21), statistical physicists argued that a second structural phase transition is more relevant than the uniqueness one. Following this literature, we shall refer to this as the "dynamic phase transition" and denote the corresponding threshold as $\alpha_d(k)$ [or $l_d(q)$]. To precise this notion, we provide here two alternative formulations corresponding to two distinct intuitions. According to the first one, above $\alpha_d(k)$ the variables $(x_1, \dots, x_N)$ become globally correlated under $\mu(\cdot)$. The criterion in **2** is replaced by one in which far apart variables x_ℓ are themselves sampled from μ ("extremality" condition):

$$\mathbb{E}\sum_{\underline{x}_\ell}\mu(x_\ell)\sum_{x_i}|\mu(x_i|\underline{x}_\ell) - \mu(x_i)| \to 0. \qquad [3]$$

Table 1. Critical connectivities for the dynamic, condensation, and satisfiability transitions in *k*-SAT and *q*-COL

SAT				COL			
k	α_d	α_c	α_s (Ref. 11)	q	l_d (Ref. 16)	l_c	l_s (Ref. 14)
4	9.38	9.547	9.93	4	9	10	10
5	19.16	20.80	21.12	5	14	14	15
6	36.53	43.08	43.4	6	18	19	20

as $\ell \to \infty$. The infimum value of α (respectively l) such that this condition is no longer fulfilled is the threshold $\alpha_d(k)$ ($l_d(k)$). Of course this criterion is weaker than the uniqueness one [hence $\alpha_d(k) \geq \alpha_u(k)$].

According to the second intuition, above $\alpha_d(k)$, the measure (Eq. **1**) decomposes into a large number of disconnected "clusters." This means that there exists a partition $\{A_n\}_{n=1\dots\mathcal{N}}$ of $\mathcal{X}^N$ (depending on the instance) such that: (*i*) one cannot find n such that $\mu(A_n) \to 1$; (*ii*) denoting by $\partial_\epsilon A$ the set of configurations $\underline{x} \in \mathcal{X}^N\backslash A$ whose Hamming distance from A is at most $N\epsilon$, we have $\mu(\partial_\epsilon A_n)/\mu(A_n)(1 - \mu(A_n)) \to 0$ exponentially fast in N for all n and ϵ small enough. Notice that the measure μ can be decomposed as:

$$\mu(\cdot) = \sum_{n=1}^{\mathcal{N}} w_n\mu_n(\cdot), \qquad [4]$$

where $w_n \equiv \mu(A_n)$ and $\mu_n(\cdot) \equiv \mu(\,\cdot\,|A_n)$. We shall always refer to $\{A_n\}$ as the "finer" partition with these properties.

The above ideas are obviously related to the performance of algorithms. For instance, the correlation decay condition in Eq. **3** is likely to be sufficient for approximate correctness of BP on random formulae. Also, the existence of partitions as above implies exponential slowing down in a large class of Monte Carlo Markov chain sampling algorithms.[i]

Recently, some important rigorous results were obtained supporting this picture (22, 23). However, even at the heuristic level, several crucial questions remain open. The most important concern the distribution of the weights $\{w_n\}$: are they tightly concentrated (on an appropriate scale) or not? A (somewhat surprisingly) related question is: can the absence of decorrelation above $\alpha_d(k)$ be detected by probing a subset of variables bounded in N?

SP (6) can be thought as an inference algorithm for a modified graphical model that gives unit weight to each cluster (20, 24), thus tilting the original measure toward small clusters. The resulting performances will strongly depend on the distribution of the cluster sizes w_n. Further, under the tilted measure, $\alpha_d(k)$ is underestimated because small clusters have a larger impact. The correct value was never determined (but see ref. 16 for coloring). Mézard *et al.* (25) undertook the technically challenging task of determining the cluster size distribution without, however, clarifying several of its properties.

In this article we address these issues and unveil at least two unexpected phenomena. Our results are described in *Results and Discussion* and summarized in Fig. 2. Finally, we discuss the connection with the performances of SP. Some technical details of the calculation are collected in *Cavity Formalism*, *Tree Reconstruction*, *and SP*.

Results and Discussion

The formulation in terms of extremality condition (see Eq. **3**) allows for an heuristic calculation of the dynamic threshold $\alpha_d(k)$. Previous attempts were based instead on the cavity method, which is an

[i]One possible approach to the definition of a Monte Carlo Markov chain algorithm is to relax the constraints by setting $\psi_a(\cdots) = \epsilon$ instead of 0 whenever the ath constraint is violated. Glauber dynamics can then be used to sample from the relaxed measure $\mu_\epsilon(\cdot)$.

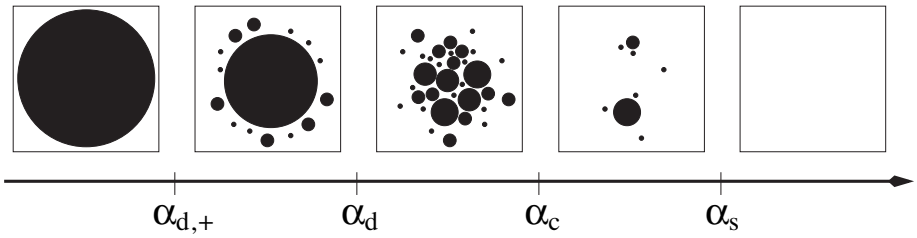

Fig. 2. Pictorial representation of the different phase transitions in the set of solutions of a rCSP. At $\alpha_{d,+}$ some clusters appear, but for $\alpha_{d,+} < \alpha < \alpha_d$ they comprise only an exponentially small fraction of solutions. For $\alpha_d < \alpha < \alpha_c$ the solutions are split among about $e^{N\Sigma_*}$ clusters of size e^{Ns_*}. If $\alpha_c < \alpha < \alpha_s$ the set of solutions is dominated by a few large clusters (with strongly fluctuating weights), and above α_s the problem does not admit solutions any more.

heuristic implementation of the definition in terms of pure state decomposition (see Eq. **4**). Generalizing the results of ref. 16, it is possible to show that the two calculations provide identical results. However, the first one is technically simpler and under much better control. As mentioned above we obtain, for all $k \geq 4$ a value of $\alpha_d(k)$ larger than the one quoted in refs. 6 and 11.

Further we determined the distribution of cluster sizes w_n, thus unveiling a third "condensation" phase transition at $\alpha_c(k) \geq \alpha_d(k)$ (strict inequality holds for $k \geq 4$ in SAT and $q \geq 4$ in coloring, see below). For $\alpha < \alpha_c(k)$ the weights w_n concentrate on a logarithmic scale [namely, $-\log w_n$ is $\Theta(N)$ with $\Theta(N^{1/2})$ fluctuations]. Roughly speaking, the measure is evenly split among an exponential number of clusters.

For $\alpha > \alpha_c(k)$ [and $< \alpha_s(k)$] the measure is carried by a subexponential number of clusters. More precisely, the ordered sequence $\{w_n\}$ converges to a well known Poisson-Dirichlet process $\{w_n^*\}$, first recognized in the spin glass context by Ruelle (26). This is defined by $w_n^* = x_n/\Sigma x_n$, where $x_n > 0$ are the points of a Poisson process with rate $x^{-1-m(\alpha)}$ and $m(\alpha) \in (0, 1)$. This picture is known in spin glass theory as one-step replica symmetry breaking (1RSB) and has been proven in ref. 27 for some special models. The Parisi 1RSB parameter $m(\alpha)$ is monotonically decreasing from 1 to 0 when α increases from $\alpha_c(k)$ to $\alpha_s(k)$ (see Fig. 3).

Remarkably, the condensation phase transition is also linked to an appropriate notion of correlation decay. If $i(1), \dots, i(n) \in [N]$ are uniformly random variable indices, then, for $\alpha < \alpha_c(k)$ and any fixed n:

$$\mathbb{E} \sum_{\{x_{i(\cdot)}\}} |\mu(x_{i(1)} \dots x_{i(n)}) - \mu(x_{i(1)}) \dots \mu(x_{i(n)})| \to 0 \qquad [5]$$

as $N \to \infty$. Conversely, the quantity on the left side of Eq. **5** remains positive for $\alpha > \alpha_c(k)$. It is easy to understand that this condition is even weaker than the extremality one (compare Eq. **3**) in that we probe correlations of finite subsets of the variables. In the next two sections we discuss the calculation of α_d and α_c.

Dynamic Phase Transition and Gibbs Measure Extremality. A rigorous calculation of $\alpha_d(k)$ along any of the two definitions provided above (compare Eqs. **3** and **4**) remains an open problem. Each of the two approaches has, however, an heuristic implementation that we shall now describe. It can be proved that the two calculations yield equal results as further discussed in the last section.

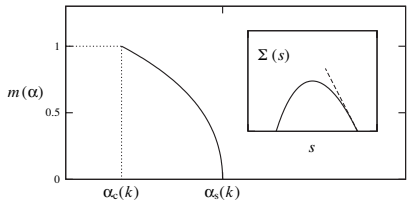

Fig. 3. The Parisi 1RSB parameter $m(\alpha)$ as a function of the constraint density α. In the *Inset*, the complexity $\Sigma(s)$ as a function of the cluster entropy for $\alpha = \alpha_s(k) - 0.1$ [the slope at $\Sigma(s) = 0$ is $-m(\alpha)$]. Both curves have been computed from the large k expansion.

The approach based on the extremality condition in Eq. **3** relies on an easy-to-state assumption and typically provides a more precise estimate. We begin by observing that, because of the Markov structure of $\mu(\cdot)$, it is sufficient for Eq. **3** to hold that the same condition is verified by the correlation between x_i and the set of variables at distance exactly ℓ from i, that we shall keep denoting as $\underline{x}_\ell$. The idea is then to consider a large yet finite neighborhood of i. Given $\bar{\ell} \geq \ell$, the factor graph neighborhood of radius $\bar{\ell}$ around i converges in distribution to the radius-$\bar{\ell}$ neighborhood of the root in a well defined random tree factor graph T.

For coloring of random regular graphs, the correct limiting tree model T is coloring on the infinite l-regular tree. For random k-SAT, T is defined by the following construction. Start from the root variable node and connect it to l new function nodes (clauses), l being a Poisson random variable of mean $k\alpha$. Connect each of these function nodes with $k-1$ new variables and repeat. The resulting tree is infinite with nonvanishing probability if $\alpha > 1/k(k-1)$. Associate a formula to this graph in the usual way, with each variable occurrence being negated independently with probability 1/2.

The basic assumption within the first approach is that the extremality condition in Eq. **3** can be checked on the correlation between the root and generation-ℓ variables in the tree model. On the tree, $\mu(\cdot)$ is defined to be a translation invariant Gibbs measure (17) associated to the infinite factor graph[j] T (which provides a specification). The correlation between the root and generation-ℓ variables can be computed through a recursive procedure (defining a sequence of distributions $\bar{P}_\ell$, see Eq. 15 below). The recursion can be efficiently implemented numerically yielding the values presented in Table 1 for k (*resp.* q) = 4, 5, 6. For large k (*resp.* q) one can formally expand the equations on $\bar{P}_\ell$ and obtain:

$$\alpha_d(k) = \frac{2^k}{k}\left[\log k + \log\log k + \gamma_d + O\left(\frac{\log\log k}{\log k}\right)\right] \qquad [6]$$

$$l_d(q) = q[\log q + \log\log q + \gamma_d + o(1)] \qquad [7]$$

with $\gamma_d = 1$ (under a technical assumption of the structure of P_ℓ).

The second approach to the determination of $\alpha_d(k)$ is based on the "cavity method" (6, 25). It begins by assuming a decomposition in pure states of the form **4** with two crucial properties: (*i*) if we denote by W_n the size of the nth cluster (and hence $w_n = W_n/\Sigma W_n$), then the number of clusters of size $W_n = e^{Ns}$ grows approximately as $e^{N\Sigma(s)}$; (*ii*) for each single-cluster measure $\mu_n(\cdot)$, a correlation decay condition of the form **3** holds.

The approach aims at determining the rate function $\Sigma(s)$, complexity: the result is expressed in terms of the solution of a distributional fixed point equation. For the sake of simplicity we

[j]More precisely $\mu(\cdot)$ is obtained as a limit of free boundary measures.

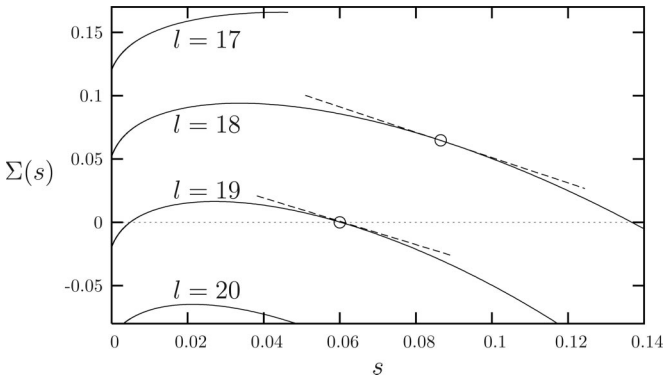

Fig. 4. The complexity function [the number of clusters with entropy density s is $e^{N\Sigma(s)}$] for the six-colorings of l-regular graphs with $l \in \{17, 18, 19, 20\}$. Circles indicate the dominating states with entropy s_*; the dashed lines have slopes $\Sigma'(s_*) = -1$ for $l = 18$ and $\Sigma'(s_*) = -0.92$ for $l = 19$. The dynamic phase transition is $l_d(6) = 18$, the condensation one $l_d(6) = 19$, and the SAT-UNSAT one $l_s(6) = 20$.

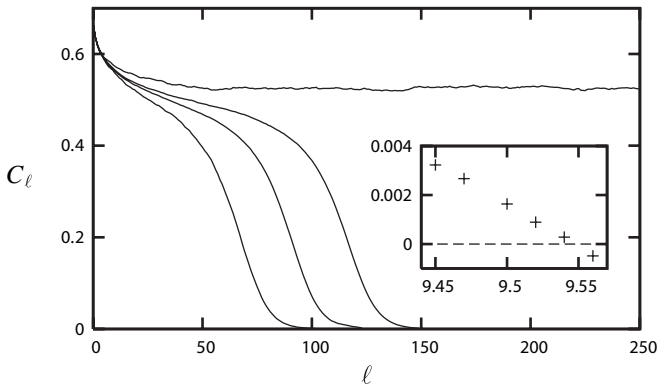

Fig. 5. Correlation function (**3**) between the root and generation ℓ variables in a random k-SAT tree formula. Here $k = 4$ and (from bottom to top) $\alpha = 9.30$, 9.33, 9.35, 9.40 [recall that $\alpha_d(4) \approx 9.38$]. In the *Inset*, the complexity $\Sigma(s_*)$ of dominant clusters as a function of α for 4-SAT.

describe here the simplest possible scenario[k] resulting from such a calculation (compare Fig. 4). For $\alpha < \alpha_{d,-\infty}(k)$ the cavity fixed point equation does not admit any solution: no clusters are present. At $\alpha_{d,-\infty}(k)$ a solution appears, eventually yielding, for $\alpha > \alpha_{d,+}$ a non-negative complexity $\Sigma(s)$ for some values of $s \in \mathbb{R}_+$. The maximum and minimum such values will be denoted by $s_{\max}$ and $s_{\min}$. At a strictly larger value $\alpha_{d,0}(k)$, $\Sigma(s)$ develops a stationary point (local maximum). It turns out that $\alpha_{d,0}(k)$ coincides with the threshold computed in refs. 6, 11, and 14. In particular, $\alpha_{d,0}(4) \approx 8.297$, $\alpha_{d,0}(5) \approx 16.12$, $\alpha_{d,0}(6) \approx 30.50$ and $l_{d,0}(4) = 9$, $l_{d,0}(5) = 13$, $l_{d,0}(6) = 17$. For large k (*resp.* q), $\alpha_{d,0}(k)$ admits the same expansion as in Eqs. **6** and **7** with $\gamma_{d,0} = 1 - \log 2$. However, up to the larger value $\alpha_d(k)$, the appearance of clusters is irrelevant from the point of view of $\mu(\cdot)$. In fact, within the cavity method it can be shown that $e^{N[s+\Sigma(s)]}$ remains exponentially smaller than the total number of solutions Z: most of the solutions are in a single cluster. The value $\alpha_d(k)$ is determined by the appearance of a point s_* with $\Sigma'(s_*) = -1$ on the complexity curve. Correspondingly, one has $Z \approx e^{N[\Sigma(s_*)+s_*]}$: most of the solutions are comprised in clusters of size about e^{Ns_*}. The entropy per variable $\phi = \lim_{N\to\infty} N^{-1} \log Z$ remains analytic at $\alpha_d(k)$.

Condensation Phase Transition. As α increases above α_d, $\Sigma(s_*)$ decreases: clusters of highly correlated solutions may no longer satisfy the newly added constraints. In Fig. 5 *Inset*, we show the α dependency of $\Sigma(s_*)$ for 4-SAT. In the large k limit, with $\alpha = \rho 2^k$ we get $\Sigma(s_*) = \log 2 - \rho - \log 2\, e^{-k\rho} + O(2^{-k})$, and $s_* = \log 2 e^{-k\rho} + O(2^{-k})$.

The condensation point $\alpha_c(k)$ is the value of α such that $\Sigma(s_*)$ vanishes: above $\alpha_c(k)$, most of the measure is contained in a subexponential number of large clusters[l]. Our estimates for $\alpha_c(k)$ are presented in Table 1 [see also Fig. 4 for $\Sigma(s)$ in the six-coloring] while in the large-k limit we obtain $\alpha_c(k) = 2^k \log 2 - 3/2 \log 2 + O(2^{-k})$ [recall that the SAT-UNSAT transition is at $\alpha_s(k) = 2^k \log 2 - (1 + \log 2)/2 + O(2^{-k})$] and $l_c(q) = 2q \log q - \log q - 2 \log 2 + o(1)$ [with the COL-UNCOL transition at $l_s(q) = 2q \log q - \log q - 1 + o(1)$]. Technically, the size of dominating clusters is found by maximizing $\Sigma(s) + s$ over the s interval on which $\Sigma(s) \geq 0$. For $\alpha \in [\alpha_c(k), \alpha_s(k)]$, the maximum is reached at $s_{\max}$, with $\Sigma(s_{\max}) = 0$ yielding $\phi = s_{\max}$. It turns out that the solutions are comprised within a finite number of clusters, with entropy $e^{Ns_{\max}+\Delta}$, where $\Delta = \Theta(1)$. The shifts Δ are asymptotically distributed according to a Poisson point process of rate $e^{-m(\alpha)\Delta}$ with $m(\alpha) = -\Sigma'(s_{\max})$. This leads to the Poisson Dirichlet distribution of weights discussed above. Finally, the entropy per variable ϕ is nonanalytic at $\alpha_c(k)$.

Let us conclude by stressing two points. First, we avoided the 3-SAT and three-coloring cases. These cases [as well as the three-coloring on Erdös-Rényi graphs (25)] are particular in that the dynamic transition point α_d is determined by a local instability [a Kesten-Stigum (28, 29) condition, see also ref. 21], yielding $\alpha_d(3) \approx 3.86$ and $l_d(3) = 6$ (the case $l = 5$, $q = 3$ being marginal). Related to this is the fact that $\alpha_c = \alpha_d$: throughout the clustered phase, the measure is dominated by a few large clusters [technically, $\Sigma(s_*) < 0$ for all $\alpha > \alpha_d$]. Second, we did not check the local stability of the 1RSB calculation. By analogy with ref. 30, we expect that an instability can modify the curve $\Sigma(s)$ but not the values of α_d and α_c.

Algorithmic Implications. Two message passing algorithms were studied extensively on random k-SAT: BP and SP (mixed strategies were also considered in refs. 19 and 20). A BP message $\nu_{u\to v}(x)$ between nodes u and v on the factor graph is usually interpreted as the marginal distribution of x_u (or x_v) in a modified graphical model. An SP message is instead a distribution over such marginals $P_{u\to v}(\nu)$. The empirical superiority of SP is usually attributed to the existence of clusters (6): the distribution $P_{u\to v}(\nu)$ is a survey of the marginal distribution of x_u over the clusters. As a consequence, according to the standard wisdom, SP should outperform BP for $\alpha > \alpha_d(k)$.

This picture, however, has several problems. Let us list two of them. First, it seems that essentially local algorithms (such as message passing ones) should be sensitive only to correlations among finite subsets of the variables[m], and these remain bounded up to the condensation transition. Recall in fact that the extremality condition in Eq. **3** involves a number of variables unbounded in N, while the weaker in Eq. **5** is satisfied up to $\alpha_c(k)$.

Second, it would be meaningful to weight uniformly the solutions when computing the surveys $P_{u\to v}(\nu)$. In the cavity method jargon, this corresponds to using a 1RSB Parisi parameter $r = 1$ instead of $r = 0$ as is done in ref. 6. It is a simple algebraic fact of the cavity formalism that for $r = 1$ the means of the SP surveys satisfy the BP equations. Since the means are the most important statistics used by SP to find a solution, BP should perform roughly as SP. Both

[k]The precise picture depends on the value of k (*resp.* q) and can be somewhat more complicated.

[l]Notice that for q-COL, since l is an integer, the "condensated" regime [$l_c(q)$, $l_s(q)$] may be empty. This is the case for $q = 4$. On the contrary, $q = 5$ is always condensated for $l_d < l < l_s$.

[m]This paradox was noticed independently by Dimitris Achlioptas (personal communication).

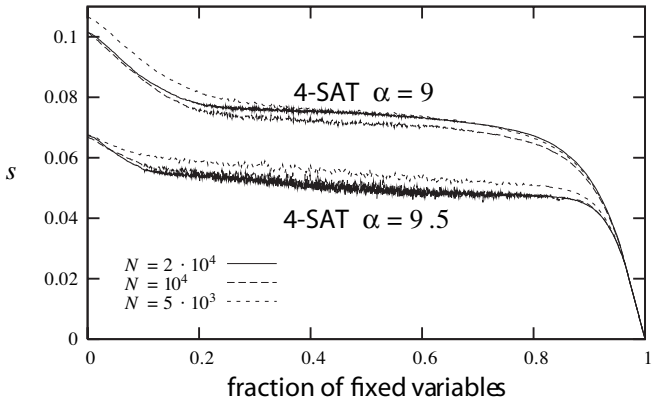

Fig. 6. Performance of BP heuristics on random 4-SAT formulae. The residual entropy per spin $N^{-1}\log Z$ (here we estimate it within Bethe approximation) as a function of the fraction of fixed variables. $t_{\max} = 20$ in these experiments.

arguments suggest that BP should perform well up to the condensation point $\alpha_c(k)$. We tested this conclusion on 4-SAT at $\alpha = 9.5 \in [\alpha_d(4), \alpha_c(4)]$, through the following numerical experiment (compare Fig. 6). (*i*) Run BP for $t_{\max}$ iterations. (*ii*) Compute the BP estimates $\nu_i(x)$ for the single-bit marginals and choose the one with largest bias. (*iii*) Fix $x_i = 0$ or 1 with probabilities $\nu_i(0)$, $\nu_i(1)$. (*iv*) Reduce the formula accordingly (i.e., eliminate the constraints satisfied by the assignment of x_i and reduce the ones violated). This cycle is repeated until a solution is found or a contradiction is encountered. If the marginals $\nu_i(\cdot)$ were correct, this procedure would provide a satisfying assignment sampled uniformly from $\mu(\cdot)$. In fact, we found a solution with finite probability ($\approx$0.4), despite the fact that $\alpha > \alpha_d(4)$. The experiment was repeated at $\alpha = 9$ with a similar fraction of successes.

Above the condensation transition, correlations become too strong and the BP fixed point no longer describes the measure μ. Indeed the same algorithm proved unsuccessful at $\alpha = 9.7 \in [\alpha_c(4), \alpha_s(4)]$. As mentioned above, SP can be regarded as an inference algorithm in a modified graphical model that weights preferentially small clusters. More precisely, it selects clusters of size $e^{N\bar{s}}$ with $\bar{s}$ maximizing the complexity $\Sigma(s)$. With respect to the new measure, the weak correlation condition in Eq. **5** still holds and allows one to perform inference by message passing.

Within the cavity formalism, the optimal choice would be to take $r \approx m(\alpha) \in [0, 1)$. Any parameter corresponding to a non-negative complexity $r \in [0, m(\alpha)]$ should, however, give good results. SP corresponds to the choice $r = 0$ that has some definite computational advantages, since messages have a compact representation in this case (they are real numbers).

Cavity Formalism, Tree Reconstruction, and SP

This section provides some technical elements of our computation. The reader not familiar with this topic should consult refs. 6, 11, 25, and 32 for a more extensive introduction. The expert reader will find a new derivation and some hints of how we overcame technical difficulties.

On a tree factor graph, the marginals of $\mu(\cdot)$, Eq. **1** can be computed recursively. The edge of the factor graph from variable node i to constraint node a (respectively from a to i) carries "message" $\bar{\eta}_{i\to a}$ ($\bar{\nu}_{a\to i}$), a probability measure on $\mathcal{X}$ defined as the marginal of x_i in the modified graphical model obtained by deleting constraint node a (*resp.* all constraint nodes around i apart from a). The messages are determined by the equations:

$$\bar{\eta}_{i\to a}(x_i) = \frac{1}{z_{i\to a}(\{\bar{\nu}_{b\to i}\})} \prod_{b\in\partial i\setminus a} \bar{\nu}_{b\to i}(x_i), \qquad [8]$$

$$\bar{\nu}_{a\to i}(x_i) = \frac{1}{\hat{z}_{a\to i}(\{\bar{\eta}_{j\to a}\})} \sum_{\underline{x}_{\partial a\setminus i}} \psi_a(\underline{x}_{\partial a}) \prod_{j\in\partial a\setminus i} \bar{\eta}_{j\to a}(x_j), \qquad [9]$$

where ∂u is the set of nodes adjacent to u, $\setminus$ denotes the set subtraction operation, and $\underline{x}_A = \{x_j : j \in A\}$. These are just the BP equations for the model (**1**). The constants $z_{i\to a}$, $\hat{z}_{a\to i}$ are uniquely determined from the normalization conditions $\Sigma_{x_i}\ \bar{\eta}_{i\to a}(x_i) = \Sigma_{x_i}\ \bar{\nu}_{a\to i}(x_i) = 1$. In the following we refer to these equations by introducing functions $f_{i\to a}(\cdot)$, $f_{a\to i}(\cdot)$ such that:

$$\bar{\eta}_{i\to a} = f_{i\to a}(\{\bar{\nu}_{b\to i}\}_{b\in\partial i\setminus a}), \quad \bar{\nu}_{a\to i} = f_{a\to i}(\{\bar{\eta}_{j\to a}\}_{j\in\partial a\setminus i}). \qquad [10]$$

The marginals of μ are then computed from the solution of these equations. For instance $\mu(x_i)$ is a function of the messages $\bar{\nu}_{a\to i}$ from neighboring function nodes.

The log number of solutions, $\log Z$, can be expressed as a sum of contributions that are local functions of the messages that solve Eqs. **8** and **9**:

$$\log Z = \sum_a \log z_a(\{\bar{\eta}_{i\to a}\}) + \sum_i \log z_i(\{\bar{\nu}_{a\to i}\}) + - \sum_{(ai)} \log z_{ai}(\bar{\eta}_{i\to a}, \bar{\nu}_{a\to i}) \qquad [11]$$

where the last sum is over undirected edges in the factor graph and

$$z_a \equiv \sum_{\underline{x}_{\partial a}} \psi_a(\underline{x}_{\partial a}) \prod_{i\in\partial a} \bar{\eta}_{i\to a}(x_i),$$
$$z_i \equiv \sum_{x_i} \prod_{a\in\partial i} \bar{\nu}_{a\to i}(x_i), \qquad z_{ai} \equiv \sum_{x_i} \bar{\eta}_{i\to a}(x_i)\bar{\nu}_{a\to i}(x_i).$$

Each term z gives the change in the number of solutions when merging different subtrees (for instance, $\log z_i$ is the change in entropy when the subtrees around i are glued together). This expression coincides with the Bethe free-energy (31) as expressed in terms of messages.

To move from trees to loopy graphs, we first consider an intermediate step in which the factor graph is still a tree but a subset of the variables, $\underline{x}_B = \{x_j : j \in B\}$ is fixed. We are therefore replacing the measure $\mu(\cdot)$ (compare Eq. **1**), with the conditional one $\mu(\cdot\,|\underline{x}_B)$. In physics terms, the variables in $\underline{x}_B$ specify a boundary condition.

Notice that the measure $\mu(\cdot\,|\underline{x}_B)$ still factorizes according to (a subgraph of) the original factor graph. As a consequence, the conditional marginals $\mu(x_i|\underline{x}_B)$ can be computed along the same lines as above. The messages $\eta^{\underline{x}_B}_{i\to a}$ and $\nu^{\underline{x}_B}_{a\to i}$ obey Eq. **10**, with an appropriate boundary condition for messages from B. Also, the number of solutions that take values $\underline{x}_B$ on $j \in B$ [call it $Z(\underline{x}_B)$] can be computed by using Eq. **11.**

Next, we want to consider the boundary variables themselves as random variables. More precisely, given $r \in \mathbb{R}$, we let the boundary to be $\underline{x}_B$ with probability

$$\tilde{\mu}(\underline{x}_B) = \frac{Z(\underline{x}_B)^r}{\mathcal{Z}(r)}, \qquad [12]$$

where $\mathcal{Z}(r)$ enforces the normalization $\Sigma_{\underline{x}_B}\ \tilde{\mu}(\underline{x}_B) = 1$. Define $P_{i\to a}(\eta)$ as the probability density of $\eta^{\underline{x}_B}_{i\to a}$ when $\underline{x}_B$ is drawn from $\tilde{\mu}$, and similarly $Q_{a\to i}(\nu)$. One can show that Eq. **8** implies the following relation between messages distributions:

$$P_{i\to a}(\eta) = \frac{1}{\mathcal{Z}_{i\to a}} \int \prod_{b\in\partial i\setminus a} \mathrm{d}Q_{b\to i}(\nu_b)\,\delta[\eta - f_{i\to a}(\{\nu_b\})]\, z_{i\to a}(\{\nu_b\})^r, \qquad [13]$$

where $f_{i\to a}$ is the function defined in Eq. **10**, $z_{i\to a}$ is determined by Eq. **8**, and $\mathcal{Z}_{i\to a}$ is a normalization. A similar equation holds for $Q_{a\to i}(\nu)$. These coincide with the 1RSB equations with Parisi parameter r. SP corresponds to a particular parameterization of Eq. **13** (and the analogous one expressing $Q_{a\to i}$ in terms of the Ps) valid for $r = 0$.

The log-partition function $\Phi(r) = \log \mathcal{Z}(r)$ admits an expression that is analogous to Eq. **11**,

$$\log \mathcal{Z}(r) = \sum_a \log \mathcal{Z}_a(\{P_{i\to a}\}) + \sum_i \log \mathcal{Z}_i(\{Q_{a\to i}\}) + - \sum_{ai} \log \mathcal{Z}_{ai}(P_{i\to a}, Q_{a\to i}) \qquad [14]$$

where the shifts $\mathcal{Z}(\cdots)$ are defined through moments of order r of the zs, and sums run over vertices not in B. For instance $\mathcal{Z}_{ai}$ is the expectation of $z_{ai}(\eta, \nu)^r$ when η, ν are independent random variables with distribution (respectively) $P_{i\to a}$ and $Q_{a\to i}$. The (Shannon) entropy of the distribution $\tilde{\mu}$ is given by $\Sigma(r) = \Phi(r) - r\Phi'(r)$.

As mentioned, the above derivation holds for tree factor graphs. Nevertheless, the local recursion equations **10** and **13** can be used as an heuristics on loopy factor graphs as well. Further, although we justified Eq. **13** through the introduction of a random boundary condition $\underline{x}_B$, we can take $B = \varnothing$ and still look for nondegenerate solutions of such equations.

Starting from an arbitrary initialization of the messages, the recursions are iterated until an approximate fixed point is reached. After convergence, the distributions $P_{i\to a}$, $Q_{a\to i}$ can be used to evaluate the potential $\Phi(r)$ (compare Eq. **14**). From this we compute the complexity function $\Sigma(r) \equiv \Phi(r) - r\Phi'(r)$ that gives access to the decomposition of $\mu(\cdot)$ in pure states. More precisely, $\Sigma(r)$ is the exponential growth rate of the number of states with internal entropy $s = \Phi'(r)$. This is how curves such as in Fig. 4 are traced.

In practice, it can be convenient to consider the distributions of messages $P_{i\to a}$, $Q_{a\to i}$ with respect to the graph realization. This approach is sometimes referred to as density evolution in coding theory. If one considers a uniformly random directed edge $i \to a$ (or $a \to i$) in a rCSP instance, the corresponding message will be a random variable. After t parallel updates according to Eq. **13**, the message distribution converges (in the $N \to \infty$ limit) to a well defined law $\mathcal{P}_t$ (for variable to constraint messages) or $\mathcal{Q}_t$ (for constraint to variable). As $t \to \infty$, these converge to a fixed point $\mathcal{P}$, $\mathcal{Q}$ that satisfies the distributional equivalent of Eq. **13.**

To be definite, let us consider the case of graph coloring. Since the compatibility functions are pairwise in this case (i.e., $k = 2$ in Eq. **1**), the constraint-to-variable messages can be eliminated and Eq. **13** takes the form:

$$P_{i\to j}(\eta) \propto \int \prod_{l\in\partial i\setminus j} \mathrm{d}P_{l\to i}(\eta_l)\ \delta[\eta - f(\{\eta_l\})]z(\{\eta_l\})^r,$$

where f is defined by $\eta(x) = z^{-1}\ \Pi_l(1 - \eta_l(x))$ and z by normalization. The distribution of $P_{i\to j}$ is then assumed to satisfy a distributional version of the last equation. In the special case of random regular graphs, a solution is obtained by assuming that $P_{i\to j}$ is indeed independent of the graph realization and i, j. One has therefore simply to set $P_{i\to j} = P$ in the above and solve it for P.

In general, finding messages distributions $\mathcal{P}$, $\mathcal{Q}$ that satisfy the distributional version of Eq. **13** is an extremely challenging task, even numerically. We adopted the population dynamics method (32), which represents the distributions by samples (this is closely related to particle filters in statistics). For instance, one represents $\mathcal{P}$ by a sample of Ps, each encoded as a list of ηs. Since computer memory drastically limits the samples size, and thus the precision of the results, we worked in two directions: (*i*) we analytically solved the distributional equations for large k (in the case of k-SAT) or q (q-Col); and (*ii*) we identified and exploited simplifications arising for special values of r.

Let us briefly discuss the second point. Simplifications emerge for $r = 0$ and $r = 1$. The first case corresponds to SP. Refs. 6 and 11 showed how to compute efficiently $\Sigma(r = 0)$ through population dynamics. Building on this, we could show that the clusters internal entropy $s(r = 0)$ can be computed at a small supplementary cost.

The value $r = 1$ corresponds instead to the tree reconstruction problem (33): In this case $\tilde{\mu}(\underline{x}_B)$ (compare Eq. **12**), coincides with the marginal of μ. Averaging Eq. **13** (and the analogous one for $Q_{a\to i}$) one obtains the BP equations (**8** and **9**), e.g., $\int \mathrm{d}P_{i\to a}(\eta)\ \eta = \bar{\eta}_{i\to a}$. These remarks can be used to show that the constrained averages:

$$\bar{P}(\eta,\bar{\eta}) = \int \mathrm{d}\mathcal{P}[P]\ P(\eta)\delta\left(\bar{\eta} - \int \mathrm{d}P(\eta')\,\eta'\right), \qquad [15]$$

and $\bar{Q}(\nu, \bar{\nu})$ (defined analogously) satisfy closed equations that are much easier to solve numerically.

We thank J. Kurchan and M. Mézard for stimulating discussions. This work has been partially supported by the European Commission under Contracts EVERGROW and STIPCO.

1. Garey MR, Johnson DS (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, New York).
2. Kschischang FR, Frey BJ, Loeliger H-A (2001) *IEEE Trans Inform Theory* 47:498–519.
3. Franco J, Paull M (1983) *Discrete Appl Math* 5:77–87.
4. Selman B, Kautz HA, Cohen B (1994) in *Proceedings of the Twelfth National Conference on Artificial Intelligence* (AAAI Press, Seattle WA), pp 337–343.
5. Achlioptas D, Sorkin GB (2000) in *Proceedings of the 41st Annual Symposium on the Foundations of Computer Science* (IEEE Press, Los Alamitos, CA), pp 590–600.
6. Mézard M, Parisi G, Zecchina R (2002) *Science* 297:812–815.
7. Bayati M, Shah D, Sharma M (2006) in *Proceedings of the IEEE International Symposium on Information Theory* (IEEE Press, Los Alamitos, CA), pp 557–561.
8. Gamarnik D, Bandyopadhyay A (2006) in *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms* (ACM Press, New York), pp 890–899.
9. Montanari A, Shah D (2007) in *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms* (ACM Press, New York), pp 1255–1264.
10. Friedgut E (1999) *J Am Math Soc* 12:1017–1054.
11. Mertens S, Mézard M, Zecchina R (2006) *Rand Struct Alg* 28:340–373.
12. Achlioptas D, Naor A, Peres Y (2005) *Nature* 435:759–764.
13. Mulet R, Pagnani A, Weigt M, Zecchina R (2002) *Phys Rev Lett* 89:268701.
14. Krzakala F, Pagnani A, Weigt M (2004) *Phys Rev E* 70:046705.
15. Luczak T (1991) *Combinatorica* 11:45–54.
16. Mézard M, Montanari A (2006) *J Stat Phys* 124:1317–1350.
17. Georgii HO (1988) *Gibbs Measures and Phase Transitions* (De Gruyter, Berlin).
18. Jonasson J (2002) *Stat Prob Lett* 57:243–248.
19. Aurell E, Gordon U, Kirkpatrick S (2004) *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA), pp 49–56.
20. Maneva EN, Mossel E, Wainwright MJ (2005) in *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms* (ACM Press, New York, 1089–1098.
21. Biroli G, Monasson R, Weigt M (2000) *Eur Phys J B* 14:551–568.
22. Mézard M, Mora T, Zecchina R (2005) *Phys Rev Lett* 94:197205.
23. Achlioptas D, Ricci-Tersenghi F (2006) in *Proceedings of the 38th Annual ACM Symposium on Theory of Computing* (ACM Press, New York), pp 130–139.
24. Braunstein A, Zecchina R (2004) *J Stat Mech* P06007.
25. Mézard M, Palassini M, Rivoire O (2005) *Phys Rev Lett* 95:200202.
26. Ruelle D (1987) *Commun Math Phys* 108:225–239.
27. Talagrand M (2000) *C R Acad Sci Paris Ser I* 331:939–942.
28. Kesten H, Stigum BP (1966) *Ann Math Stat* 37:1463–1481.
29. Kesten H, Stigum BP (1966) *J Math Anal Appl* 17:309–338.
30. Montanari A, Ricci-Tersenghi F (2004) *Phys Rev B* 70:134406.
31. Yedidia J, Freeman WT, Weiss Y (2005) *IEEE Trans Inform Theory* 51:2282–2312.
32. Mézard M, Parisi G (2001) *Eur Phys J B* 20:217–233.
33. Mossel E, Peres Y (2003) *Ann Appl Probab* 13:817–844.