Clusters of solutions and replica symmetry breaking in random *k*-satisfiability

You may also be interested in:

On the cavity method for decimated random constraint satisfaction problems and theanalysis of belief propagation guided decimation algorithms
Federico Ricci-Tersenghi and Guilhem Semerjian

The large deviations of the whitening process in random constraint satisfaction problems
Alfredo Braunstein, Luca Dall'Asta, Guilhem Semerjian et al.

Constraint satisfaction problems with isolated solutions are hard
Lenka Zdeborová and Marc Mézard

Adversarial satisfiability problem
Michele Castellana and Lenka Zdeborová

Geometrical organization of solutions to random linear Boolean equations
Thierry Mora and Marc Mézard

The computational core and fixed point organization in Boolean networks
L Correale, M Leone, A Pagnani et al.

Next nearest neighbour Ising models on random graphs
Jack Raymond and K Y Michael Wong

Clusters of solutions and replica symmetry breaking in random k-satisfiability

Andrea Montanari[1,2]**, Federico Ricci-Tersenghi**[3] **and Guilhem Semerjian**[4]

[1] Department of Electrical Engineering, Stanford University, USA
[2] Department of Statistics, Stanford University, USA
[3] Dipartimento di Fisica and INFM-CNR, Università di Roma La Sapienza, Piazzale Aldo Moro 2, I-00185 Roma, Italy
[4] LPTENS, Unité Mixte de Recherche (UMR 8549) du CNRS et de l'ENS, associée à l'UPMC Université Paris 06, 24 Rue Lhomond, F-75231 Paris Cedex 05, France
E-mail: montanar@stanford.edu, Federico.Ricci@roma1.infn.it and guilhem@lpt.ens.fr

Abstract. We study the set of solutions of random k-satisfiability formulas through the cavity method. It is known that, for an interval of the clause-to-variables ratio, this decomposes into an exponential number of pure states (clusters). We refine substantially this picture by: (i) determining the precise location of the clustering transition; (ii) uncovering a second 'condensation' phase transition in the structure of the solution set for $k \geq 4$. These results both follow from computing the large deviation rate of the internal entropy of pure states. From a technical point of view our main contributions are a simplified version of the cavity formalism for special values of the Parisi replica symmetry breaking parameter m (in particular for $m = 1$ via a correspondence with the tree reconstruction problem) and new large-k expansions.

Contents

1. Introduction

An instance of k-satisfiability (k-SAT) consists in a Boolean formula in conjunctive normal form whereby each elementary clause is the disjunction of k literals (a Boolean variable or its negation). Solving it amounts to determining whether there exists an assignment of the variables such that at least one literal in each clause evaluates to true. The k-SAT problem plays a central role in the theory of computational complexity, being the first decision problem proven to be NP-complete [1] (for all $k \geq 3$). Its optimization

(minimize the number of unsatisfied clauses) and enumeration (count the number of optimal assignments) versions are defined straightforwardly and are also hard from the computational point of view.

Random k-satisfiability is the ensemble defined by drawing a uniformly random formula among all the ones involving Mk-clauses over N variables. Equivalently, each of the M clauses is drawn uniformly over the $2^k \binom{N}{k}$ possible ones, independently from the others. It was observed empirically earlier on [2] that, by tuning the clause density $\alpha = M/N$, this ensemble could produce formulas which were hard for known algorithms. Hardness was argued to be related to a sharp threshold in the satisfiability probability, emerging as $N \to \infty$ with α fixed. More precisely, it is believed that there exists a constant $\alpha_s(k)$ such that random formulas are with high probability[5] satisfiable if $\alpha < \alpha_s(k)$ and unsatisfiable if $\alpha > \alpha_s(k)$. The existence of a sharp threshold was proven in [3], with, however, a critical point $\alpha_s(k, N)$ which might not converge when $N \to \infty$. Despite important progress [4]–[6] the rigorous proof of the existence and determination of $\alpha_s(k)$ remains a major open problem (with the notable exception of $k = 2$ [7]).

The connection between threshold phenomena and phase transitions spurred a considerable amount of work [8]–[12] using techniques from the theory of mean-field spin glasses [13]. The main outcomes of this approach have been: (i) a precise conjecture on the location of the satisfiability threshold $\alpha_s(k)$ [10, 12]; (ii) the suggestion [9, 10] for $k \geq 3$ of another transition at $\alpha_d(k) < \alpha_s(k)$ affecting the geometry of the solutions space; (iii) most strikingly, the proposal of a new and extremely effective message passing algorithm, survey propagation (SP) [10, 11]. This exploits a detailed statistical picture of the solution space to efficiently find solutions.

According to statistical physics studies, in the intermediate regime $\alpha \in [\alpha_d(k), \alpha_s(k)]$ solutions tend to group themselves in clusters that are somehow disconnected. As α increases, the number of these clusters decreases. The satisfiability transition is thus due to the vanishing of a number of clusters, which still contain a large number of solutions just before $\alpha_s(k)$. The phase transition at $\alpha_d(k)$ has been referred to as 'clustering phase transition' or 'dynamic phase transition', depending on the feature emphasized. Its nature and location, as well as a refined description of the regime $\alpha \in [\alpha_d(k), \alpha_s(k)]$, will be the main topic of this paper. More precisely:

(i) we will argue that previous determinations of $\alpha_d(k)$ [10]–[12] have to be corrected when fluctuations of the cluster sizes are taken into account;

(ii) we will uncover (for $k \geq 4$) a new 'condensation' phase transition at $\alpha_c(k) \in [\alpha_d(k), \alpha_s(k)]$. For $\alpha \in [\alpha_d(k), \alpha_c(k)]$ the relevant clusters are exponentially numerous. For $\alpha \in [\alpha_c(k), \alpha_s(k)]$ most of the solutions are contained in a number of clusters that remain bounded as $N \to \infty$.

This paper is organized as follows. In section 2 we recall some general features of mean-field disordered models, emphasizing the notions of dynamical transitions and replica symmetry breaking. In section 3 we define more precisely the ensemble of random formulas studied and describe the replica symmetric (RS) and one step of replica symmetry breaking (1RSB) approach to this model. We then apply the program of section 2 to the

---

[5] Here and below 'with high probability' (w.h.p.) means with probability converging to 1 as $N \to \infty$.

random k-satisfiability problem and present our main results in section 4. For the sake of clarity some technicalities of the 1RSB treatment are presented shortly afterward, see section 5. To complement these results, which are partly based on a numerical resolution of integral equations, we present in section 6 an asymptotic expansion in the large-k limit which gives further credit to our theses. We draw our conclusions in section 7. Technical details are deferred to three appendices.

A short account of our results has been published in [14], and a detailed analysis of the related q-coloring problem in [15]. While the present work was being finished two very interesting papers confirmed the generality of the results of [14]. The first concerned 3-SAT [16] and the second the bi-coloring of random hypergraphs [17].

2. Mean-field disordered systems

The goal of this section is to provide a quick overview of the cavity method [13, 18]. We will further propose a more precise mathematical formulation of several notions that are crucial in the statistical physics approach.

2.1. Statistical mechanics and graphical models

Let us start by considering a general model defined by:

(1) A factor graph [19], i.e. a bipartite graph $G = (V, F, E)$. Here V, $|V| = N$, are 'variable nodes' corresponding to variables, F, $|F| = M$, are 'function (or factor) nodes' describing interactions among these variables, and E are edges between variables and factors. Given $i \in V$ (resp. $a \in F$), we shall denote by $\partial i = \{a \in F : (ia) \in E\}$ (resp. $\partial a = \{i \in V : (ia) \in E\}$) its neighborhood. Further, given $i, j \in V$, we let $d(i, j)$ be their graph theoretic distance (the minimal number of factor nodes encountered on a path between i and j).

(2) A space of configurations $\mathcal{X}^V$, with $\mathcal{X}$ a finite alphabet (a configuration will be denoted in the following as $\underline{\sigma} = (\sigma_1, \ldots, \sigma_N) \in \mathcal{X}^V$). For any set $A \subseteq V$, we let $\underline{\sigma}_A = \{\sigma_i : i \in A\}$.

(3) A set of non-negative weights $\{w_a : a \in F\}$, $w_a : \mathcal{X}^{\partial a} \to \mathbb{R}_+$, $\underline{\sigma}_{\partial a} \mapsto w_a(\underline{\sigma}_{\partial a})$. In the case of *constraint satisfaction problems*, these are often taken to be indicator functions (more details on this particular case will be given in section 2.4).

Given these ingredients, a measure over $\mathcal{X}^V$ is defined as

$$\mu_N(\underline{\sigma}) = \frac{1}{Z_N} w_N(\underline{\sigma}), \qquad w_N(\underline{\sigma}) = \prod_{a \in F} w_a(\underline{\sigma}_{\partial a}). \tag{1}$$

This is well defined only if there exists at least one configuration $\underline{\sigma}^*$ that makes all the weights strictly positive, namely $w_a(\underline{\sigma}_{\partial a}^*) > 0$ for each a. We will assume this to be the case throughout the paper (i.e. we focus on the 'satisfiable' phase). Further, it will be understood that we consider sequences of graphs (and weights) of diverging size N (although we shall often drop the subscript N).

An important role is played by the large-N behavior of the partition function Z_N. This is described by the free-entropy density[6]

$$\phi = \lim_{N\to\infty} \frac{1}{N} \log Z_N, \qquad Z_N = \sum_{\underline{\sigma}} w_N(\underline{\sigma}). \tag{2}$$

2.2. Pure states and replica symmetry breaking

The replica/cavity method allows us to compute a hierarchy of approximations to ϕ. This is thought to yield the exact value of ϕ itself in 'mean-field' models. The hierarchy is ordered according to the so-called number of steps of replica symmetry breaking (RSB). At each level the calculation is based on some hypotheses on the typical structure of μ, a pivotal role being played by the notion of *pure state*. Since this concept is only intuitively defined in the physics literature, we propose here two mathematically precise definitions. In both cases a pure state is a (sequence of) probability measures ρ_N on $\mathcal{X}^N$.

- *Definition of pure states through correlation decay.* We define the correlation function of ρ_N as

$$C_N(r) = \sup_{A,B:\, d(A,B)\geq r} \sum_{\underline{\sigma}_A,\underline{\sigma}_B} |\rho_N(\underline{\sigma}_A, \underline{\sigma}_B) - \rho_N(\underline{\sigma}_A)\rho_N(\underline{\sigma}_B)|, \tag{3}$$

  where the sup is taken over all subsets of variable nodes $A, B \subseteq V$ such that the distance between any pair of nodes $(i, j) \in A \times B$ is greater than r. Then ρ_N is a pure state if this correlation function decays at large r. Technically, we let $C_\infty(r) = \limsup_{N\to\infty} C_N(r)$ and require $C_\infty(r) \to 0$ as $r \to \infty$.

- *Definition of pure states through conductance.* We let the (ϵ, δ) conductance of ρ_N be

$$\mathfrak{F}_N(\epsilon, \delta) = \inf_{\mathcal{A}\subset\mathcal{X}^N} \left\{ \frac{\rho_N(\partial_\epsilon \mathcal{A})}{\rho_N(\mathcal{A})(1 - \rho_N(\mathcal{A}))} : \delta \leq \rho_N(\mathcal{A}) \leq 1 - \delta \right\}. \tag{4}$$

  Here the inf is taken over all subsets of the configuration space. Further, letting D denote the Hamming distance in $\mathcal{X}^N$, we defined the boundary of $\mathcal{A}$ as $\partial_\epsilon \mathcal{A} = \{\underline{\sigma} \in \mathcal{X}^N \setminus \mathcal{A} \,|\, D(\underline{\sigma}, \mathcal{A}) \leq N\epsilon\}$. With these definitions ρ_N is pure if its conductance is bounded below by an inverse polynomial in N for all ϵ and δ (while non-pure states have a conductance which typically decays exponentially with N).

These two definitions mimic the well-known ones on $\mathbb{Z}^d$ in terms of tail triviality and extremality [20]. Further, the second one is clearly related to the behavior of local Monte Carlo Markov chain dynamics. A small conductance amounts to a bottleneck in the distribution and hence to a large relaxation time. While we expect them to be equivalent for a large family of models, proving this is a largely open problem. Moreover we should emphasize that the heuristic cavity method followed in this paper never explicitly uses either of these definitions.

The hypotheses implicit in the cavity method can be expressed in terms of the *pure states decomposition* of μ. This is a partition of the configuration space (dependent on the

---

[6] One usually assumes that the limit exists. If the model is disordered, the almost sure limit can be used, or, equivalently, $\log Z_N$ is replaced by its expectation.

graph and weights) such that the measure μ constrained to each element of this partition is a pure state. More precisely, let us call $\{\mathcal{A}_\gamma\}_\gamma$ a partition of $\mathcal{X}^N$, and define

$$Z_\gamma = \sum_{\underline{\sigma} \in \mathcal{A}_\gamma} w(\underline{\sigma}), \qquad W_\gamma = \frac{Z_\gamma}{Z}, \qquad \mu_\gamma(\underline{\sigma}) = \frac{1}{Z_\gamma} w(\underline{\sigma}) \mathbb{I}(\underline{\sigma} \in \mathcal{A}_\gamma). \qquad (5)$$

Clearly μ can be written as the convex combination of the μ_γ with coefficients W_γ. This defines a pure state decomposition if: (i) each of the μ_γ is a pure state in the sense given above, (ii) this is the 'finest' such partition, in the sense that the μ_γ are no longer pure if any subset of them is replaced by their union.

Statistical physics calculations suggest that a wide class of mean-field models is described by one of the following 'universal behaviors'. The terminology used here is inherited from the literature on mean-field spin glasses [21, 22].

RS Most of the measure is contained in a single element of the partition, namely $W_{\max} = \max_\gamma W_\gamma \to 1$ as $N \to \infty$ (*replica symmetric*).

d1RSB Most of the measure is carried by $\mathcal{N} \doteq \mathrm{e}^{N\Sigma_*}$ pure states[7], each one with a weight $W_\gamma \doteq \mathrm{e}^{-N\Sigma_*}$ (*dynamical one-step replica symmetry breaking*).

1RSB The measure condensates on a subexponential number of pure states, namely, if $W_{[\gamma]}$ is the weight of the γth largest state, then $\lim_{n\to\infty} \lim_{N\to\infty} \sum_{\gamma=1}^n W_{[\gamma]} = 1$ (*one-step replica symmetry breaking*).

The reader will notice that this list does not include full replica symmetry breaking phases, in which pure states are organized according to an ultrametric structure. While this behavior is as generic as the previous ones, our understanding of it in sparse graph models is still rather poor.

We are mostly concerned with families of models of the type defined in equation (1) indexed by a continuous parameter α (such as the clause density in k-SAT). In this setting, the above behaviors often appear in sequence as listed above when the system becomes more and more constrained (e.g. as α is increased in k-SAT). The different regimes are then separated by phase transitions: the 'dynamical' or 'clustering' phase transition from RS to d1RSB (at α_d) and the 'condensation' phase transition between d1RSB and 1RSB (at α_c). The paradigmatic example of such transitions is the fully connected p-spin model [21, 22], where they are encountered upon lowering the temperature.

Let us stress that the above definitions are insensitive to what happens in a fraction of the space of configurations of vanishing measure. For instance, we neglect metastable states whose overall weight is exponentially small[8].

A convenient tool for distinguishing these various behaviors is the replicated free-entropy [23]:

$$\Phi(m) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E} \log \left\{ \sum_\gamma Z_\gamma^m \right\}, \qquad (6)$$

where m is an arbitrary real number (known as the Parisi replica symmetry breaking parameter) which allows us to weight differently the various pure states according to

---

[7] Here and in the following $\doteq$ means equality at the leading exponential order.

[8] In the fully connected models such metastable states are indeed seen as solutions of the Thouless–Anderson–Palmer equations, well above the dynamical phase transition.

their sizes. Suppose indeed that the number of pure states γ with internal free-entropy density $\phi_\gamma = (\log Z_\gamma)/N$ behave at leading order as $\exp\{N\Sigma(\phi_\gamma)\}$, where $\Sigma(\phi)$ is known as the complexity (or configurational entropy) of the states. The sum in (6) can then be computed by the Laplace method; if one assumes for simplicity that Σ is positive on an interval $[\phi_-, \phi_+]$, this leads to

$$\Phi(m) = \sup_{\phi \in [\phi_-, \phi_+]} [\Sigma(\phi) + m\phi]. \tag{7}$$

Provided Σ is concave, it can be reconstructed in a parametric way from $\Phi(m)$ by a Legendre inversion [23]:

$$\Sigma(\phi_{\text{int}}(m)) = \Phi(m) - m\Phi'(m), \qquad \phi_{\text{int}}(m) = \Phi'(m), \tag{8}$$

where m is such that the supremum in (7) lies in the interior of $[\phi_-, \phi_+]$, which defines a range $[m_-, m_+]$. Usually Σ vanishes continuously at ϕ_+. As explained below, when zero-energy states are concerned $\phi_{\text{int}}(m)$ coincides with the internal entropy of such states. Note that a given value of m selects the point of the curve $\Sigma(\phi)$ of slope $-m$; in particular, the value $m = 0$ corresponds to the maximum of the curve.

The replica/cavity method at the level of one step of replica symmetry breaking allows us to compute the replicated free-entropy $\Phi(m)$ under an appropriate hypothesis on the organization of pure states. The various regimes can be distinguished through the behavior of this function, namely

RS $\Phi(m) = m\phi_*$, where ϕ_* is the contribution of the single dominant pure state, $Z_{[1]} \doteq \mathrm{e}^{N\phi_*}$.

d1RSB $\Phi(m)/m$ achieves its minimum for $m \in [0, 1]$ at $m = 1$, with $\Sigma_* = \Phi(1) - \Phi'(1) > 0$. Then the measure μ decomposes into approximately $\mathrm{e}^{N\Sigma_*}$ pure states of internal free-entropy $\Phi'(1)$.

1RSB $\Phi(m)/m$ achieves its minimum over the interval $[0, 1]$ at $m_{\text{s}} \in (0, 1)$. Then the ordered sequence of weights $W_{[1]} \geq W_{[2]} \geq W_{[3]} \geq \cdots$ keep fluctuating in the thermodynamic limit, and converges to a Poisson–Dirichlet process [24] of parameter m_{s}. The internal free-entropy of these states is $\Phi'(m_{\text{s}})$.

In all these cases the total free-entropy density is estimated by minimizing $\Phi(m)/m$ in the interval $[0, 1]$.

2.3. Cavity equations

We shall now recall the fundamental equations used within the 1RSB cavity method and propose a somehow original derivation. In the following we will be interested in factor graphs that converge locally[9] to trees in the thermodynamic limit.

As a consequence, let us first consider the case of a model of type (1) whose underlying factor graph is a tree, and discuss later how the long loops are taken into account by the cavity method. Tree factor graph models are easily solved by a 'message passing' procedure [19]. One associates to each directed edge from factor a to variable i (resp. from i to a) a 'message' $\eta_{a \to i}$ (resp. $\eta_{i \to a}$). Messages are probability measures on $\mathcal{X}$. On trees,

---

[9] More precisely, any finite neighborhood of a uniformly chosen random vertex converges to a tree.

they can be defined as the marginal law of σ_i with respect to the modified factor graph $G_{a \to i}$ (resp. $G_{i \to a}$) where all factor nodes in $\partial i \setminus a$ (resp. the factor node a) have been removed. Simple computations yield the following local equations between messages:

$$\eta_{a \to i} = f_{a \to i}(\{\eta_{j \to a}\}_{j \in \partial a \setminus i}),$$
$$f_{a \to i}(\{\eta_{j \to a}\})(\sigma_i) = \frac{1}{z_{a \to i}(\{\eta_{j \to a}\})} \sum_{\underline{\sigma}_{\partial a \setminus i}} w_a(\underline{\sigma}_{\partial a}) \prod_{j \in \partial a \setminus i} \eta_{j \to a}(\sigma_j), \tag{9}$$

$$\eta_{i \to a} = f_{i \to a}(\{\eta_{b \to i}\}_{b \in \partial i \setminus a}),$$
$$f_{i \to a}(\{\eta_{b \to i}\})(\sigma_i) = \frac{1}{z_{i \to a}(\{\eta_{b \to i}\})} \prod_{b \in \partial i \setminus a} \eta_{b \to i}(\sigma_i), \tag{10}$$

where the functions z are fixed by the normalization of the η's. As we consider a tree factor graph these equations have a unique solution, easily determined in a single sweep of updates from the leaves of the graph towards its inside. Moreover the free-entropy of the model follows from this solution and is

$$N\phi = \log Z = - \sum_{(i,a)} \log z_{ia}(\eta_{a \to i}, \eta_{i \to a}) + \sum_a \log z_a(\{\eta_{i \to a}\}_{i \in \partial a})$$
$$+ \sum_i \log z_i(\{\eta_{a \to i}\}_{a \in \partial i}). \tag{11}$$

Here the first sum runs over the undirected edges of the factor graph and the z's are given by

$$z_{ia} = \sum_{\sigma_i} \eta_{a \to i}(\sigma_i) \eta_{i \to a}(\sigma_i), \qquad z_a = \sum_{\underline{\sigma}_{\partial a}} w_a(\underline{\sigma}_{\partial a}) \prod_{i \in \partial a} \eta_{i \to a}(\sigma_i),$$
$$z_i = \sum_{\sigma_i} \prod_{a \in \partial i} \eta_{a \to i}(\sigma_i). \tag{12}$$

This computation is correct only on tree factor graphs. Nevertheless it is expected to yield good estimates of the marginals and free-entropy for a number of models on locally tree-like graphs. The belief propagation (BP) algorithm consists in iterating equations (9) and (10) in order to find an (approximate) fixed point. In particular, whenever the RS scenario holds, there should be one approximate solution of the above equations that yields the correct leading order of the free-entropy density in the thermodynamic limit. In any case, when dealing with random factor graphs, one can always turn this simple computation into a probabilistic one, defining a distribution of random messages by reading (9) and (10) in a distributional sense with random weight functions and variable degrees. The RS estimate of the average free-entropy is then obtained by averaging the various terms in (11) with respect to these random messages.

This approach can be refined in the d1RSB and 1RSB regimes. The BP equations (9) and (10) should be approximately valid if one computes the messages $\eta_{a \to i}$ and $\eta_{i \to a}$ as marginal laws of the measure μ_γ restricted to a single pure state γ. When the number of pure states is very large, one considers a distribution (with respect to the pure states γ with their weights W_γ) of messages on each directed edge of the factor graph.

A simple and suggestive derivation of the 1RSB equations goes as follows. Assume that the factor graph is a tree, and choose a subset B of the variable nodes that will act

as a boundary, for instance (but not necessarily) the leaves of the factor graph. Each configuration $\underline{\sigma}_B$ of the variables in B induces a conditional distribution $\mu^{\underline{\sigma}_B}$ on the remaining variables

$$\mu^{\underline{\sigma}_B}(\underline{\tau}) = \frac{1}{Z^{\underline{\sigma}_B}} w(\underline{\tau}) \mathbb{I}(\underline{\tau}_B = \underline{\sigma}_B), \tag{13}$$

where here and in the following $\mathbb{I}$ denotes the indicator function of an event and the normalizing factor $Z^{\underline{\sigma}_B}$ is the partition function restricted to the configurations coinciding with $\underline{\sigma}_B$ on the boundary.

Since the factor graph corresponding to $\mu^{\underline{\sigma}_B}$ is still a tree, the corresponding marginals and partition function $Z^{\underline{\sigma}_B}$ can be computed iterating the message passing equations (9) and (10), with an appropriate prescription for the messages $\eta_{i\to a}$ emerging from variables $i \in B$, namely $\eta_{i\to a}(\tau_i) = \delta_{\sigma_i, \tau_i}$. Let us denote by $\eta_{a\to i}^{\underline{\sigma}_B}$ and $\eta_{i\to a}^{\underline{\sigma}_B}$ the corresponding set of messages, solutions of (9) and (10) on all edges of the factor graph. Further define, for $m \in \mathbb{R}$, a probability measure on the boundary conditions as

$$\tilde{\mu}(\underline{\sigma}_B) = \frac{(Z^{\underline{\sigma}_B})^m}{\sum_{\underline{\sigma}'_B} (Z^{\underline{\sigma}'_B})^m}. \tag{14}$$

The idea is to mimic the pure states of a large, loopy factor graph model, by the boundary configurations of a tree model. Calling $P_{a\to i}$ (resp. $P_{i\to a}$) the distribution of the messages $\eta_{a\to i}^{\underline{\sigma}_B}$ (resp. $\eta_{i\to a}^{\underline{\sigma}_B}$) with respect to $\tilde{\mu}$,[10] a short reasoning reveals that

$$P_{a\to i}(\eta) = \frac{1}{Z[\{P_{j\to a}\}, m]} \int \prod_{j\in\partial a\setminus i} \mathrm{d}P_{j\to a}(\eta_{j\to a}) \, \delta(\eta - f_{a\to i}(\{\eta_{j\to a}\})) \, z_{a\to i}(\{\eta_{j\to a}\})^m, \tag{15}$$

$$P_{i\to a}(\eta) = \frac{1}{Z[\{P_{b\to i}\}, m]} \int \prod_{b\in\partial i\setminus a} \mathrm{d}P_{b\to i}(\eta_{b\to i}) \, \delta(\eta - f_{i\to a}(\{\eta_{b\to i}\})) \, z_{i\to a}(\{\eta_{b\to i}\})^m, \tag{16}$$

where the functions f and z are defined in equations (9) and (10), and the $Z[\cdots]$ are normalizing factors determined by the condition $\int \mathrm{d}P_{a\to i}(\eta) = \int \mathrm{d}P_{i\to a}(\eta) = 1$. Equations (15) and (16) coincide with the standard 1RSB equations with Parisi parameter m [25]. In addition the free-entropy density associated to the law $\tilde{\mu}$, $N\Phi(m) \equiv \log\{\sum_{\underline{\sigma}_B} (Z^{\underline{\sigma}_B})^m\}$ can be shown to be

$$N\Phi(m) = -\sum_{(i,a)\in E} \log Z_{ia}[P_{a\to i}, P_{i\to a}, m] + \sum_{a\in F} \log Z_a[\{P_{i\to a}\}_{i\in\partial a}, m]$$
$$+ \sum_{i\in V} \log Z_i[\{P_{a\to i}\}_{a\in\partial i}, m], \tag{17}$$

where the factors $Z_{\ldots}$ are fractional moments of the ones $z_{\ldots}$ defined in equation (12), namely

$$Z_{ia} = \int \mathrm{d}P_{a\to i}(\eta_{a\to i}) \mathrm{d}P_{i\to a}(\eta_{i\to a}) \, z_{ia}^m, \qquad Z_a = \int \prod_{i\in\partial a} \mathrm{d}P_{i\to a}(\eta_{i\to a}) \, z_a^m,$$
$$Z_i = \int \prod_{a\in\partial i} \mathrm{d}P_{a\to i}(\eta_{a\to i}) \, z_i^m. \tag{18}$$

---

[10] More precisely, with respect to the measure $\tilde{\mu}_{a\to i}$ (resp. $\tilde{\mu}_{i\to a}$) defined similarly for the factor graph $G_{a\to i}$ (resp. $G_{i\to a}$).

As in the RS case, one can heuristically apply (15) and (16) on any graph, even if it is not a tree. Of particular interest is the limit $B \to \emptyset$. Equations (15) and (16) may have two behaviors in this limit: (i) All the distributions $P_{i \to a}$, $P_{a \to i}$ become Dirac deltas in this limit. In this case a 'far-away' boundary has small influence on the system, and it is easily seen by comparing (11) and (17) that $\Phi(m) = m\phi$. (ii) These distributions remain non-trivial in the limit $B \to \emptyset$. This case is interpreted as a consequence of the existence of many pure states. In this situation, even a small boundary influences the system by selecting one of such states. We thus interpret the $B = \emptyset$ limit of $\Phi(m)$ as an estimate of the replicated potential (6).

In section 2.2 we emphasized the special role played by the value $m = 1$: the dynamical transition is signaled by the appearance of a non-trivial solution of the 1RSB equations with $m = 1$. This is particularly clear in the present derivation of the 1RSB equations. Indeed, the distribution $\tilde{\mu}$ of the boundary condition coincides in this case with the Boltzmann distribution μ.

The existence of a non-trivial solution of the 1RSB equations at $m = 1$ is thus related to a peculiar form of long range correlations under μ, as first pointed out in [26]. Such correlations can be measured through a point-to-set correlation function [27]–[29]. For concreteness let us give an expression of this correlation in the case of Ising spins. Given a variable node i and a set of variable nodes B, we let

$$C(i, B) \equiv \sum_{\underline{\sigma}_B} \mu(\underline{\sigma}_B) \left(\sum_{\sigma_i} \mu(\sigma_i | \underline{\sigma}_B) \, \sigma_i \right)^2 - \left(\sum_{\sigma_i} \mu(\sigma_i) \, \sigma_i \right)^2 . \tag{19}$$

The reader will recognize the analogy between this expression and the difference $q_1 - q_0$ of intra-and inter-state overlaps [30]. The Boltzmann measure has long range point-to-set correlations if $C(i, B)$ does not decay to 0 when $d(i, B)$ grows. Such correlations were shown in [31, 32] to imply a diverging relaxation time. We shall come back in the following to the actual computation of this quantity (see in particular section 5.1).

2.4. Application to constraint satisfaction problems

This short overview of the cavity method did not rely on any hypothesis on the form of the weight factors w_a in equation (1). We now comment briefly on the way this general formalism is applied to constraint satisfaction problems (CSP), in order to clarify the relationship of the present work with previous studies. In a CSP the factors a correspond to constraints, which can be either satisfied or not by the configuration of their adjacent variables, $\underline{\sigma}_{\partial a}$. For a satisfiable instance of a CSP one can take w_a to be the indicator function of the event 'constraint a is satisfied.' Then the law defined in (1) is the uniform distribution over the solutions of the CSP, the partition function counts the number of such solutions and the free-entropy reduces to the logarithm of the number of solutions. This 'entropic' method [33] is the most adequate to the study of the satisfiable phase.

This approach is, however, ill defined for unsatisfiable instances. The usual way to handle this case is to define a cost function $E(\underline{\sigma})$ on the space of configurations equal to the number of unsatisfied constraints under the assignment $\underline{\sigma}$. Following the traditional notations of statistical mechanics one introduces an inverse temperature β and weighs the configurations with $w(\underline{\sigma}) = \exp[-\beta E(\underline{\sigma})]$. Small temperatures (large β) favor low-energy configurations, in the limit $\beta \to \infty$ the measure μ concentrates on the optimal

configurations which maximize the number of satisfied constraints. Let us detail this approach, which was originally followed in [10, 12, 34]. At the 1RSB level the pure states are characterized by their energy density e and their entropy density s, with the free-entropy density given by $\phi = s - \beta e$. Defining the complexity $\Sigma(s, e)$ according to the number of pure states with these two characteristics, equation (7) becomes

$$\Phi(\beta, m) = \sup_{s,e}[\Sigma(s, e) + m(s - \beta e)]. \tag{20}$$

If one takes now the limit $\beta \to \infty$ and assumes $e > 0$, the entropic term becomes irrelevant; to obtain a finite result one has to take at the same time $m \to 0$ such that the product βm, usually denoted y, remains finite. One thus obtains

$$\Phi_{\rm e}(y) = \sup_e[\Sigma_{\rm e}(e) - ye], \qquad \Sigma_{\rm e}(e) \equiv \sup_s \Sigma(s, e). \tag{21}$$

In the unsatisfiable phase, the 'energetic' cavity approach allows us to characterize the minimal energy of the problem.

In the case of satisfiable problems, one has to perform a second limit $y \to \infty$ (after $\beta \to \infty$) to concentrate on the pure states with $e = 0$. It follows that the complexity thus computed is $\sup_s \Sigma(s, e = 0)$, i.e. the maximum of the entropic complexity. In other words the procedure $y \to \infty$ after $\beta \to \infty$ is equivalent to performing the entropic computation with a Parisi parameter $m = 0$, i.e. to weigh all the pure states in the same way, irrespective of their sizes. This is not a problem for the determination of the satisfiability threshold $\alpha_{\rm s}$, which corresponds to the disappearing of all zero-energy pure states, hence to the vanishing of the maximal complexity $\Sigma(m = 0)$. However, the value of $\alpha_{\rm d}$ in [10, 12] corresponds to the appearance of a solution of the 1RSB equations with $m = 0$, and not with $m = 1$ which we argued to be the relevant value for the definition of $\alpha_{\rm d}$.

In the rest of the paper we shall follow the entropic cavity method, i.e. we take (1) to be the uniform measure over the solutions of the CSP under study and keep a finite value for the Parisi parameter m. Before entering the details of this approach in the example of random k-satisfiability, let us mention that the existence of exponentially numerous pure states (called clusters in this context) for some values of α and k has been proved in [35, 36]. An intrinsic limitation of these works was that clusters were defined by much stricter conditions than the one exposed above (which thus implied limitations on α, k). The consequences of the existence of a distribution of cluster sizes have also been investigated in a toy model in [37].

We should also emphasize that, for the simpler CSP known as XORSAT [38, 39], a precise characterization of the clusters has been achieved through rigorous methods. A good part of the phenomena studied in the present paper is, however, absent in this simpler model. In particular, all clusters of XORSAT have the same size because of the linear structure of the constraints.

3. The cavity method applied to the random k-SAT problem

3.1. Some definitions

In the application of the formalism to k-satisfiability, we use $\sigma_i \in \mathcal{X} = \{-1, +1\}$ to encode the Boolean variables. A constraint a on k variables $\underline{\sigma}_{\partial}a$ is satisfied by all the 2^k
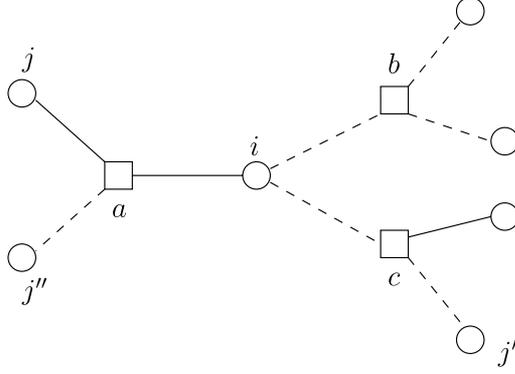
Figure 1. An example of the factor graph representation of a satisfiability formula for $k = 3$. The values J_i^a are encoded by drawing a solid (resp. dashed) edge between clause a and variable i if $\sigma_i = +1$ (resp. -1) satisfies clause a. The distances between some of the variable nodes are $d_{i,j} = d_{i,j'} = d_{i,j''} = 1$ and $d_{j,j'} = 2$. The neighborhoods are, for instance, $\partial i = \{a, b, c\}$, $\partial a = \{i, j, j'\}$, $\partial_+ i = \{a\}$, $\partial_- i = \{b, c\}$, $\partial_+ i(a) = \emptyset$, $\partial_- i(a) = \{b, c\}$, $\partial_+ i(b) = \{c\}$, $\partial_- i(b) = \{a\}$.

configurations except one, let us call it $\underline{J}^a = \{J_i^a : i \in \partial a\}$, in which all the literals of the clause are false. The weight factors are thus defined as $w_a(\underline{\sigma}_\partial a) = \mathbb{I}(\underline{\sigma}_\partial a \neq \underline{J}^a)$, the indicator function of the event 'clause a is satisfied.'

A formula is represented as a factor graph (cf. figure 1) whose edges are labeled by J_i^a. This suggests to refine the definition of the neighborhoods. Given a variable node i, $\partial_+ i$ (resp. $\partial_- i$) will denote the set of clauses which are satisfied by $\sigma_i = +1$ (resp. $\sigma_i = -1$). Further, given a clause $a \in \partial i$ we call $\partial_+ i(a)$ (resp. $\partial_- i(a)$) the set of clauses in $\partial i \setminus a$ which are satisfied by the same (resp. opposite) value of σ_i as is a.

For k-SAT formulas the general RS cavity equations (9) and (10) can be written in a pretty explicit form. As the variables take only two values the cavity probability messages $\eta_{a \to i}$ and $\eta_{i \to a}$ can be parameterized by a single real number, that we shall call respectively $u_{a \to i}$ and $h_{i \to a}$, and define by

$$\eta_{a \to i}(\sigma_i) = \frac{1 - J_i^a \sigma_i \tanh u_{a \to i}}{2}, \qquad \eta_{i \to a}(\sigma_i) = \frac{1 - J_i^a \sigma_i \tanh h_{i \to a}}{2}. \qquad (22)$$

With these conventions equations (9) and (10) take the form

$$u_{a \to i} = f(\{h_{j \to a}\}_{j \in \partial a \setminus i}), \qquad f(h_1, \ldots, h_{k-1}) = -\frac{1}{2} \log \left(1 - \prod_{i=1}^{k-1} \frac{1 - \tanh h_i}{2} \right), \qquad (23)$$

$$h_{i \to a} = \sum_{b \in \partial_+ i(a)} u_{b \to i} - \sum_{b \in \partial_- i(a)} u_{b \to i}. \qquad (24)$$

We are interested in the regime where the number M of uniformly chosen clauses and the number of variables N both diverge at fixed ratio $\alpha = M/N$. The random factor graphs thus generated enjoy properties reminiscent of the Erdös–Rényi random graphs $G(N, M)$ [40, 41]. In particular, for a uniformly random variable node i, the number of clauses in $\partial_+ i$ and $\partial_- i$ converges to two i.i.d. Poisson random variables of mean $\alpha k/2$.

The same statement is true for $\partial_+ i(a)$ and $\partial_- i(a)$ when (i, a) is an uniformly chosen edge of the factor graph. The degree distribution is a very local description of a graph, looking at one node or edge only. It is, however, easy to show that any bounded neighborhood of a uniformly random node i converges to a random (Galton–Watson) tree with the same degree distribution [41].

3.2. The RS description of the random formulas ensemble

The replica-symmetric treatment of the random k-SAT problem was first worked out using the replica formalism in [8]. In the cavity formulation one interprets the BP equations (9), (10), (23) and (24) in a probabilistic way. More precisely, we introduce the distributions of $u_{a \to i}$, $h_{i \to a}$ (over the choice of the random formula) and denote them as $\mathcal{P}_{(0)}(h)$ and $\mathcal{Q}_{(0)}(u)$. These distributions satisfy the distributional equations:

$$u \stackrel{\mathrm{d}}{=} f(h_1, \ldots, h_{k-1}), \qquad h \stackrel{\mathrm{d}}{=} \sum_{i=1}^{l_+} u_i^+ - \sum_{i=1}^{l_-} u_i^-. \tag{25}$$

In these expressions $h, \{h_i\}$ (resp. $u, \{u_i^\pm\}$) are independent copies of the random variable of distribution $\mathcal{P}_{(0)}(h)$ (resp. $\mathcal{Q}_{(0)}(u)$), the function f is defined in equation (23) and $l_\pm$ are two independent Poisson random variables of mean $\alpha k/2$. The symbol $\stackrel{\mathrm{d}}{=}$ denotes identity in distribution[11].

The RS prediction for the entropy is

$$\phi_{(0)} = -\alpha k \mathbb{E} \log z_1(u, h) + \alpha \mathbb{E} \log z_2(h_1, \ldots, h_k) + \mathbb{E} \log z_3(u_1^+, \ldots, u_{l_+}^+, u_1^-, \ldots, u_{l_-}^-), \tag{26}$$

where the expectations are over i.i.d. copies of the random variables u and h, and $l_\pm$ are as above. The various entropy shifts are obtained by rewriting the z's in equation (12) in terms of u and h:

$$z_1(u, h) = 1 + \tanh h \tanh u, \tag{27}$$

$$z_2(h_1, \ldots, h_k) = 1 - \prod_{i=1}^{k} \frac{1 - \tanh h_i}{2}, \tag{28}$$

$$z_3(u_1^+, \ldots, u_{l_+}^+, u_1^-, \ldots, u_{l_-}^-) = \prod_{i=1}^{l_+}(1 + \tanh u_i^+) \prod_{i=1}^{l_-}(1 - \tanh u_i^-)$$

$$+ \prod_{i=1}^{l_+}(1 - \tanh u_i^+) \prod_{i=1}^{l_-}(1 + \tanh u_i^-). \tag{29}$$

Similarly the RS overlap can be computed as

$$q_0 = \mathbb{E}[\tanh^2 h]. \tag{30}$$

Several equivalent expressions of the RS entropy can be found in the literature; the choice we made in (26) has the advantage of being variational. By this we mean that the

---

[11] More explicitly, given two random variables X and Y we write $X \stackrel{\mathrm{d}}{=} Y$ if the distributions of X and Y coincide. For instance, if X, X_1, X_2 are i.i.d. standard normal random variables, $X \stackrel{\mathrm{d}}{=} (X_1 + X_2)/\sqrt{2}$.

stationarity conditions of the function $\phi_{(0)}[\mathcal{P}, \mathcal{Q}, \alpha]$ with respect to $\mathcal{P}$ and $\mathcal{Q}$ are nothing but the self-consistency equations (25). Note also that the rigorous results of [42, 43] imply that[12] the entropy density ϕ is upper-bounded by the RS $\phi_{(0)}$ for any trial distribution $\mathcal{P}$, as long as $\mathcal{Q}$ is linked to $\mathcal{P}$ by the first equation in (25), for a regularized version of the model at finite temperature. Moreover the RS description was proven to be valid for small values of α in [44].

The numerical resolution of the equation on the order parameter is relatively easy. The distributions $\mathcal{P}_{(0)}$ and $\mathcal{Q}_{(0)}$ can indeed be represented by samples (or populations) of a large number $\mathcal{N}$ of representatives, $\{h_i\}_{i=1}^{\mathcal{N}}$ and $\{u_i\}_{i=1}^{\mathcal{N}}$. The fixed point condition stated in (25) is looked for by an iterative population dynamics algorithm [25, 41, 45].

We turn now to the cavity formalism at the 1RSB level, which assumes the organization of pure states described in section 2.2.

3.3. The 1RSB description of the random formulas ensemble

As in the RS case, when the underlying formula is random, the messages $P_{i \to a}$, $P_{a \to i}$ along a uniformly random edge become random variables, whose distributions are denoted as $\mathcal{P}_{(1)}[P]$, $\mathcal{Q}_{(1)}[Q]$. These distributions satisfy a couple of distributional equations that are the probabilistic version of equations (15) and (16):

$$Q(\bullet) \stackrel{\mathrm{d}}{=} \frac{1}{\mathcal{Z}_4[P_1, \ldots, P_{k-1}]} \int \prod_{i=1}^{k-1} \mathrm{d}P_i(h_i) \, \delta(\bullet - f(h_1, \ldots, h_{k-1})) z_4(h_1, \ldots, h_{k-1})^m, \qquad (31)$$

$$P(\bullet) \stackrel{\mathrm{d}}{=} \frac{1}{\mathcal{Z}_3[\{Q_i^+\}, \{Q_i^-\}]} \int \prod_{i=1}^{l_+} \mathrm{d}Q_i^+(u_i^+) \prod_{i=1}^{l_-} \mathrm{d}Q_i^-(u_i^-) \, \delta\left(\bullet - \sum_{i=1}^{l_+} u_i^+ + \sum_{i=1}^{l_-} u_i^-\right)$$
$$\times \, z_3(\{u_i^+\}_{i=1}^{l_+}, \{u_i^-\}_{i=1}^{l_-})^m, \qquad (32)$$

where the P's (resp. Q's) are i.i.d. from $\mathcal{P}_{(1)}$ (resp. $\mathcal{Q}_{(1)}$) and $l_\pm$ have the above-stated Poissonian distribution. The entropy shift z_3 used in equation (32) was defined in equation (29), while z_4 is given by

$$z_4(h_1, \ldots, h_{k-1}) = 2 - \prod_{i=1}^{k-1} \frac{1 - \tanh h_i}{2} = 1 + \mathrm{e}^{-2f(h_1, \ldots, h_{k-1})}. \qquad (33)$$

Finally, the 1RSB potential is obtained by taking the expectation of equation (17). One gets

$$\Phi(m) = -\alpha k \mathbb{E} \log \mathcal{Z}_1[Q, P] + \alpha \mathbb{E} \log \mathcal{Z}_2[P_1, \ldots, P_k]$$
$$+ \mathbb{E} \log \mathcal{Z}_3[Q_1^+, \ldots, Q_{l_+}^+, Q_1^-, \ldots, Q_{l_-}^-], \qquad (34)$$

where the factors $\mathcal{Z}_i$ are weighted averages of the corresponding entropy shifts:

$$\mathcal{Z}_1[Q, P] = \int \mathrm{d}P(h) \, \mathrm{d}Q(u) \, z_1(u, h)^m, \qquad (35)$$

---

[12] In [42, 43] this claim is made for k even. However, the proof holds verbatim for k odd as well. To the best of our knowledge, this was observed first by Elitza Maneva in 2005.

$$\mathcal{Z}_2[P_1, \ldots, P_k] = \int \prod_{i=1}^{k} \mathrm{d}P_i \, (h_i) z_2(h_1, \ldots, h_k)^m, \tag{36}$$

$$\mathcal{Z}_3[Q_1^+, \ldots, Q_{l_+}^+, Q_1^-, \ldots, Q_{l_-}^-] = \int \prod_{i=1}^{l_+} \mathrm{d}Q_i^+(u_i^+) \prod_{i=1}^{l_-} \mathrm{d}Q_i^-(u_i^-) z_3(u_1^+, \ldots, u_{l_+}^+, u_1^-, \ldots, u_{l_-}^-)^m. \tag{37}$$

The inter-and intra-state overlaps are given, respectively, by

$$q_0 = \mathbb{E}\left[\left(\int \mathrm{d}P(h) \tanh h\right)^2\right], \qquad q_1 = \mathbb{E}\left[\int \mathrm{d}P(h) \tanh^2 h\right]. \tag{38}$$

The variational property discussed at the RS level still applies to the 1RSB potential. This is of particular interest for the computation of the internal entropy of the states, given by a derivative with respect to m. This derivation can be applied to the explicit dependence only and yields

$$\phi_{\mathrm{int}}(m) = -\alpha k \mathbb{E}\left[\frac{\int \mathrm{d}P(h)\, \mathrm{d}Q(u) z_1(u,h)^m \log z_1(u,h)}{\mathcal{Z}_1(Q,P)}\right]$$

$$+ \alpha \mathbb{E}\left[\frac{\int \prod_{i=1}^{k}\mathrm{d}P_i(h_i)\; z_2(\{h_i\}_{i=1}^{k})^m \log z_2(\{h_i\}_{i=1}^{k})}{\mathcal{Z}_2[\{P_i\}_{i=1}^{k}]}\right]$$

$$+ \mathbb{E}\left[\frac{\int \prod_{i=1}^{l_+}\mathrm{d}Q_i^+(u_i^+) \prod_{i=1}^{l_-}\mathrm{d}Q_i^-(u_i^-)\; z_3(\{u_i^+\}_{i=1}^{l_+}, \{u_i^-\}_{i=1}^{l_-})^m \log z_3(\{u_i^+\}_{i=1}^{l_+}, \{u_i^-\}_{i=1}^{l_-})}{\mathcal{Z}_3[\{Q_i^+\}_{i=1}^{l_+}, \{Q_i^-\}_{i=1}^{l_-}]}\right]. \tag{39}$$

The rigorous results of [42, 43] also imply $\phi \leq \Phi(m)/m$ for any value of m in $(0, 1)$ and any trial order parameter $\mathcal{P}$ (with $\mathcal{Q}$ defined by equation (31)).

The numerical resolution of the 1RSB equations (31) and (32) is, in general, much harder than the one of their RS counterparts (compare with equation (25)). The population dynamics algorithm represents $\mathcal{P}_{(1)}$ by a sample of distributions $\{P_i\}_{i=1}^{\mathcal{N}}$, which themselves have to be encoded, for each i, by a finite set of cavity fields $\{h_{i,j}\}_{j=1}^{\mathcal{N}'}$. This drastically limits the sizes $\mathcal{N}$ and $\mathcal{N}'$, and hence the precision of the numerical results. Moreover generating one element, say Q_i, from $k-1$ P_i's is by itself a non-trivial task. The various fields representing Q_i are weighted in a non-uniform way because of the factor z_4^m in equation (31), which forces the use of delicate resampling procedures.

These equations can be greatly simplified analytically for two particular values of m, namely 0 and 1. For the sake of readability we postpone the discussion of these important simplifications until section 5 and proceed in the next section with the presentation and the interpretation of the results obtained either at arbitrary m with the full numerical procedure (whose implementation details are exposed in appendix A) or in $m = 0, 1$ with the simplified, more precise ones.
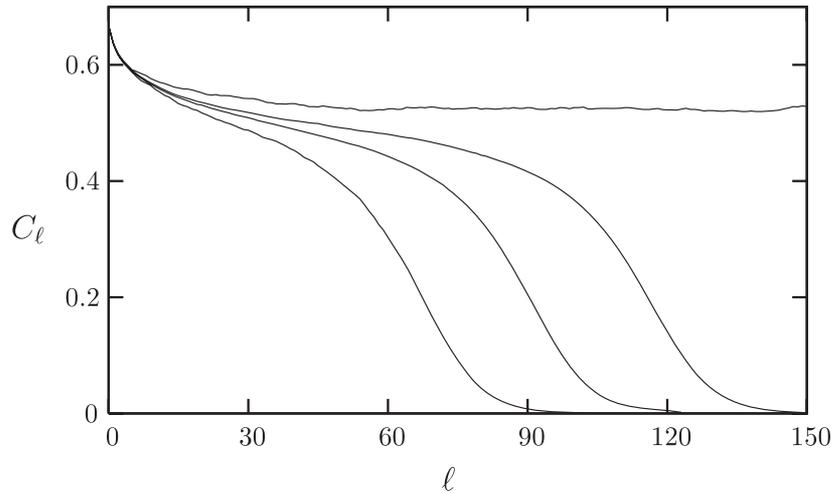
Figure 2. The point-to-set correlation function for $k = 4$; from left to right $\alpha = 9.30, 9.33, 9.35$ and 9.40.

4. Transitions in the satisfiable regime of random k-SAT

4.1. The dynamical, condensation and satisfiability transitions for $k \geq 4$

Let us begin our discussion of the satisfiable regime of random k-SAT by studying the case $k = 4$, the values $k \geq 4$ having the same qualitative behavior. On the other hand, the phenomenology of 3-SAT is different and we report on it in section 4.4.

Following the program of section 2 we first have to determine the value $\alpha_{\rm d}$ for the appearance of a non-trivial solution of the 1RSB equations with $m = 1$. To this aim we compute the point-to-set correlation function C_ℓ, which is the average of the correlation function (19) between a randomly chosen variable i and the set B of variables at distance ℓ from it (see section 5.1 for details of the computation). The plots of figure 2 show that for $\alpha \leq \alpha_{\rm d} \approx 9.38$ this correlation vanishes at large distance, while for larger values of α a strictly positive long range correlation sets in discontinuously. To distinguish between the d1RSB and 1RSB regime we then compute the complexity $\Sigma(m = 1)$. As demonstrated in figure 3 this is strictly positive at $\alpha_{\rm d}$, then decreases continuously until it vanishes at $\alpha_{\rm c} \approx 9.547$. Finally the satisfiability transition $\alpha_{\rm s}$ is found from the criterion of vanishing of $\Sigma(m = 0)$, i.e. the maximum of the entropic complexity curve (see figure 3): the value $\alpha_{\rm s} \approx 9.931$ is in agreement with [12] and we shall show in section 5.2 that this is indeed the same calculation.

To summarize, we find the three regimes RS, d1RSB and 1RSB described in section 2.2 occurring in this order, for the values of α in $[0, \alpha_{\rm d}]$, $[\alpha_{\rm d}, \alpha_{\rm c}]$ and $[\alpha_{\rm c}, \alpha_{\rm s}]$. We expect this pattern of transitions to be the same for all $k \geq 4$. This is supported by our numerical investigations for $k = 4, 5, 6$ (see table 1 for a summary of the numerical values of the thresholds) and by the large-k expansions presented in section 6.

The entropy density (see figure 3) is given by the RS formula both in the RS and d1RSB regimes. In the latter case it has to be understood as the sum of the complexity $\Sigma(m = 1)$ and of the internal entropy of the associated states, $\phi_{\rm int}(m = 1)$. In contrast, for $\alpha \in [\alpha_{\rm c}, \alpha_{\rm s}]$ it is necessary to compute the whole function $\Sigma(\phi)$ by varying m. The
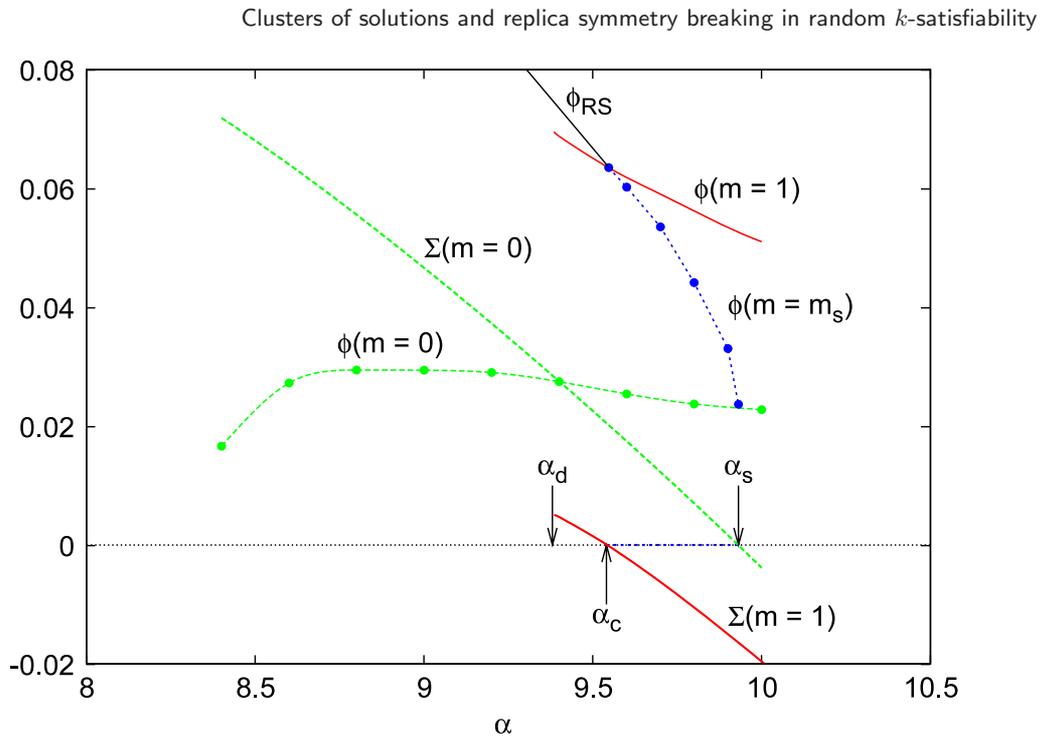
Figure 3. The complexity Σ and the internal entropy ϕ_{int} for the values $m = 0, 1$, and $m = m_{\text{s}}$ in the 1RSB regime, for $k = 4$.

Table 1. Numerical values of the various critical thresholds. For $k = 3$ we have formally $\alpha_{\text{c}} = \alpha_{\text{d}}$, see the text for details on the nature of the difference between $k = 3$ and $k \geq 4$.

k	α_{d}	α_{c}	α_{s} [12]	α_{f}
3	3.86	3.86	4.267	$*$
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 [46]

entropy density coincides with the one of dominant clusters and is given by the point where $\Sigma(\phi)$ vanishes.

4.2. The entropic complexity curves

The curves $\Sigma(\phi)$ are shown in figure 4 for several values of α. The symbols are obtained in a parametric way by solving the 1RSB equations for various values of m and plotting the point $(\phi_{\text{int}}(m), \Sigma(m))$. The lines in figure 4 are numerical interpolations, obtained by fitting not directly $\Sigma(\phi)$, but instead the data for $\Phi(m)$ with a generic smooth function[13] and then analytically deriving the fitting function to obtain the curves in figure 4. The

[13] We have tried different fitting functions and all provide equivalent and very good results thanks to the smoothness of $\Phi(m)$.

17
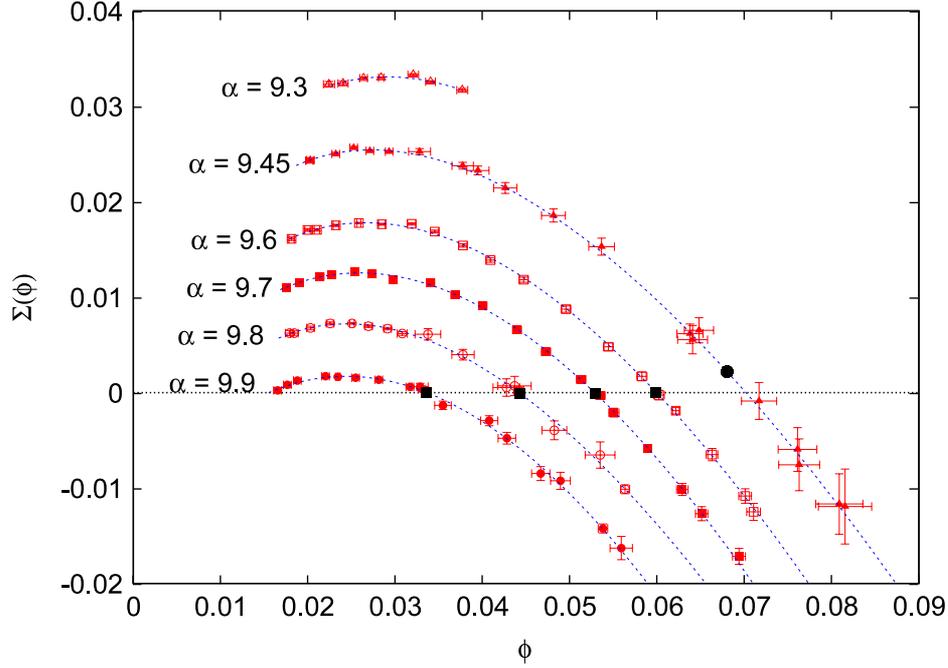
Figure 4. The complexity $\Sigma(\phi)$ for $k = 4$ and several values of α: from top to bottom $\alpha = 9.3, 9.45, 9.6, 9.7, 9.8$ and 9.9.

agreement of this fitting procedure with the parametric plot is excellent. The three regimes are clearly illustrated in this figure:

- For $\alpha < \alpha_\mathrm{d}$ a portion of the curve $\Sigma(\phi)$ can exist (for instance, there is a solution of the 1RSB equation with $m = 0$ for $\alpha \geq 8.297$ [12]), yet it has no point of slope $-m = -1$. The contribution of these clusters is negligible compared to the dominant RS cluster.

- For $\alpha \in [\alpha_\mathrm{d}, \alpha_\mathrm{c}]$ (see, e.g., $\alpha = 9.45$ data in figure 4) the complexity $\Sigma(m = 1)$ exists and is positive (it is marked by a black circle in the figure).

- For $\alpha \in [\alpha_\mathrm{c}, \alpha_\mathrm{s}]$ (see, e.g., $\alpha = 9.6, 9.7, 9.8, 9.9$ in figure 4) the complexity $\Sigma(m = 1)$ is negative and thus the $\Sigma(\phi)$ curve vanishes at $\phi(m_\mathrm{s})$ (marked with a black square), where the slope (in absolute value) is smaller than 1 and equals $m_\mathrm{s}(\alpha)$. The measure is dominated by a subexponential number of clusters of entropy $\phi(m_\mathrm{s})$, shown as a function of α in figure 3.

The value thus estimated of the Parisi parameter $m_\mathrm{s}(\alpha)$ in the 1RSB regime is plotted in figure 5 (it is identical to 1 in the d1RSB region). The curve close to the m_s data is not a fit, but instead an explicit approximate expression for $m_\mathrm{s}(\alpha)$ which becomes exact in the large-k limit (see section 6 for details). Indeed equation (78) (valid to leading order at large k) can be equivalently rewritten as

$$\frac{\alpha_\mathrm{s} - \alpha}{\alpha_\mathrm{s} - \alpha_\mathrm{c}} = \frac{1 - 2^m(1 - m \log 2)}{2 \log 2 - 1}, \tag{40}$$

and this gives an expression for $m_\mathrm{s}(\alpha)$ once values of α_c and α_s determined numerically for $k = 4$ are plugged into equation (40). Note that the solution to equation (40) is such
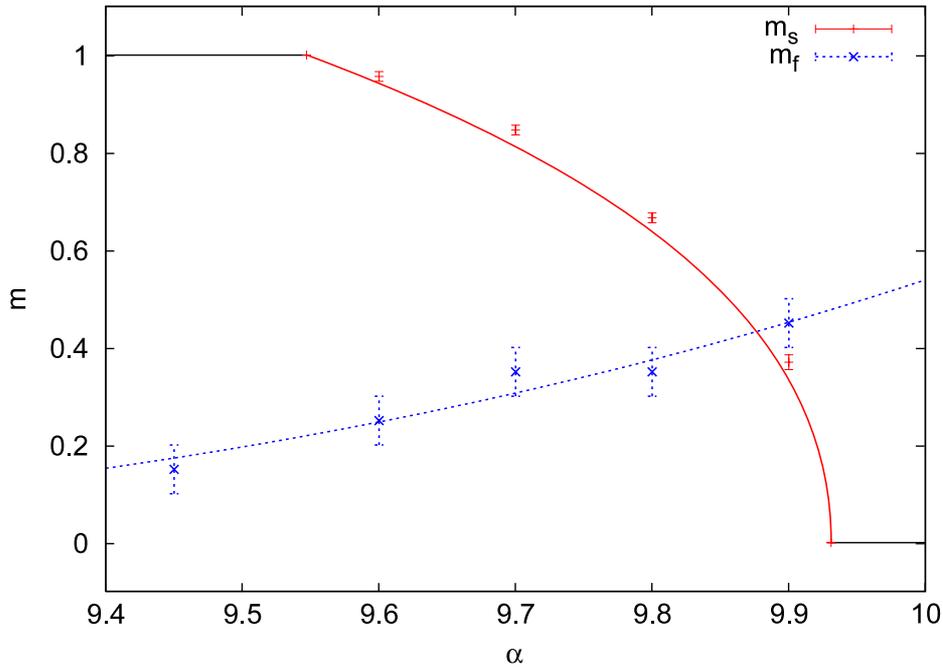
Figure 5. The value of the Parisi parameter m_s in the thermodynamically relevant pure states of the 1RSB regime in random 4-SAT, and the freezing transition m_f.

that (i) $m_s(\alpha_c) = 1$, (ii) $m_s(\alpha_s) = 0$ and (iii) m_s vanishes as a square root at α_s. The finite-k corrections to the expression (40) seem already small for $k = 4$, as can be inferred from the good agreement with the numerical data displayed in figure 5. This fact was also noticed for the coloring problem in [15].

Once we compute the optimal value m_s for each value of α, we can plot in figure 6 the overlap q_0 and q_1 of the dominating clusters as a function of α. Notice that the inter-state overlap q_0 is an increasing function of α for any fixed value of m, but becomes a decreasing function of α between α_c and α_s where we take $m = m_s(\alpha)$.

We did not attempt a complete determination of the portion of the plane (α, m) where non-trivial solutions of the 1RSB equations can be found. From our numerical investigations it seems that solutions with smaller values of m appear at smaller values of α, i.e. the threshold $\alpha_d(m)$ is an increasing function in the range of parameters we considered. In particular, solutions with negative m appear at rather small values of α. The limit of very large negative values of m is, however, difficult to study numerically, and more work could be done on this issue; the corresponding pure states are tiny because their variables are overconstrained, which plagues the numerical resolution of the 1RSB equations.

4.3. On the presence of frozen variables in clusters of solutions

Another characterization of the clusters of solutions, besides their internal entropy and self-overlap, is the presence or not of frozen variables, that is variables that take the same
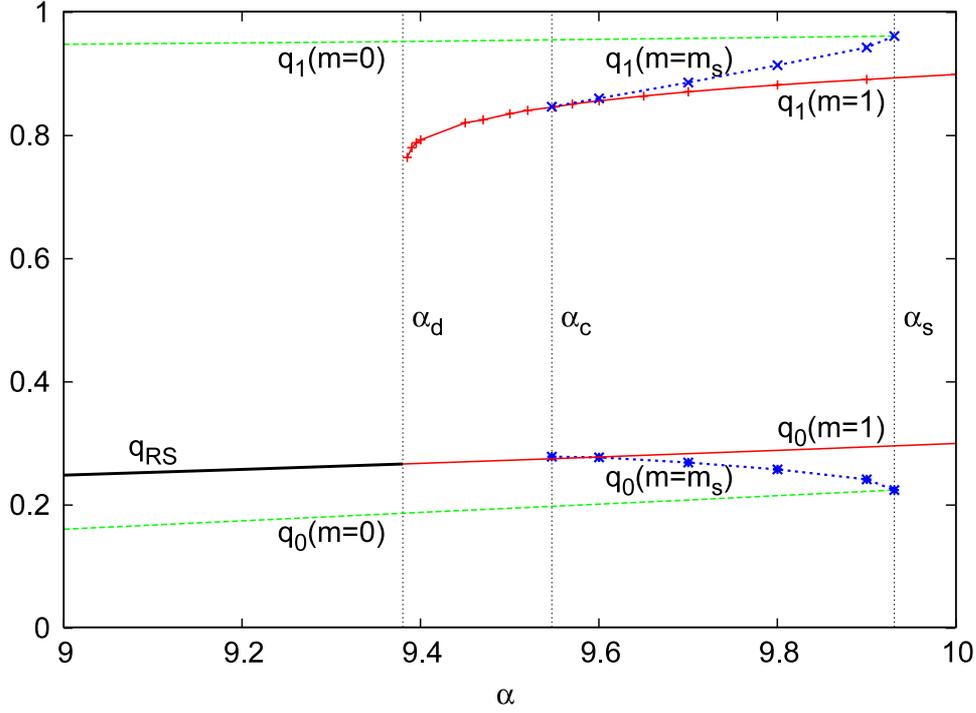
Figure 6. Intra-and inter-state overlaps for $k = 4$.

value in all the solutions of the cluster. In technical terms this corresponds to a non-vanishing weight on $\pm\infty$ in the 1RSB cavity field distributions $P(h)$ (see equation (59) below). Our data show that, given a value of α, there exists a threshold $m_{\mathrm{f}}(\alpha)$ such that clusters described by $m < m_{\mathrm{f}}$ do contain frozen variables, while those with $m > m_{\mathrm{f}}$ do not. This is consistent with the intuition: the freezing of variables is correlated with a smaller value of the internal entropy, and hence of m. Numerical estimates for the line $m_{\mathrm{f}}(\alpha)$ are plotted in figure 5 for $k = 4$. The large error bars are due to the fact that we have checked the presence of frozen variables only at m values which are multiples of 0.1 (and no interpolation can be done in between, since the property is just true or false). The interpolating curve is a fit to the $m_{\mathrm{f}}(\alpha)$ data with the function $A(x - 8.297)^B$ (for $m = 0$ the critical value of α is $\alpha_{\mathrm{d}}(m = 0) \approx 8.297$ [12]). The freezing transition α_{f} is defined by the appearance of frozen variables in dominating clusters, that is $m_{\mathrm{s}}(\alpha_{\mathrm{f}}) = m_{\mathrm{f}}(\alpha_{\mathrm{f}})$. From the crossing of these two lines in figure 5 we estimated the freezing threshold for $k = 4$ at $\alpha_{\mathrm{f}} \approx 9.88$.

The fact that the freezing transition occurs after the condensation one for $k = 4$ is not generic; for $k \geq 6$ the threshold $m_{\mathrm{f}}(\alpha)$ reaches 1 at $\alpha_{\mathrm{f}} \leq \alpha_{\mathrm{c}}$ [46], hence in a part of the d1RSB regime the dominating clusters do contain frozen variables for these values of k.

Let us, however, emphasize that generally $\alpha_{\mathrm{d}} < \alpha_{\mathrm{f}}$, i.e. that in random k-satisfiability (and also in q-coloring [15]) clustering can occur without implying the freezing of variables. This fact has been obscured up to now because the energetic cavity method [10, 34] focused precisely on the fraction of frozen variables in the $m = 0$ solution of the 1RSB equations, and because in the simpler XORSAT model [38, 39] the freezing and clustering transitions coincide. We refer the reader to [46] for a more extensive study of the freezing transition, in
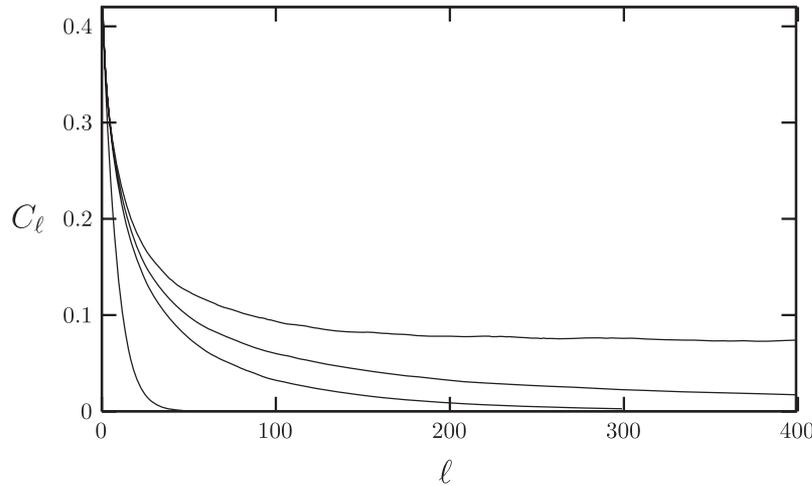
Figure 7. The point-to-set correlation function for $k = 3$; from left to right $\alpha = 3.60, 3.84, 3.86, 3.88$.

particular its interpretation in terms of the divergence of the minimal rearrangements [47] it induces, and to [36] where it has been proven that frozen variables exist in every cluster for $k \geq 9$ and α large enough.

4.4. $k = 3$, a special case

We turn now to the description of our numerical results in the particular case $k = 3$, recently investigated also in [16]. The onset of long range point-to-set correlations, displayed in figure 7 through the correlation function C_ℓ (details of the computation are deferred to section 5.1), is qualitatively different from $k = 4$ (compare with figure 2). The long range correlation $\lim_{\ell \to \infty} C_\ell$ grows indeed continuously from 0 at α_d (in qualitative agreement with the variational approximation of [9]). In fact, this transition coincides with a local instability of the RS solution with respect to 1RSB perturbations (this is a generic fact for all models with continuous dynamic transitions). A numerical procedure can be used to locate precisely this instability [48]–[50]. We get the estimate $\alpha_\mathrm{d} = \alpha_\mathrm{stab} \approx 3.86$. Please note that for $k \geq 4$ this local instability occurs after the discontinuous transition, for instance at $\alpha_\mathrm{stab} \approx 10.2$ for $k = 4$.

For $\alpha > \alpha_\mathrm{d}$ the complexity $\Sigma(m = 1)$ decreases continuously from 0 (see lowest curve in figure 8): there is no d1RSB regime for 3-SAT. We then turned to the resolution of the 1RSB equations for other values of m. In figure 8 we plotted the complexity as a function of α, for various values of m. According to the interpretation of the 1RSB regime of section 2.2, for each value of α we can find the Parisi parameter m_s such that $\Sigma = 0$, and obtain the 1RSB estimate of the entropy as the internal entropy of these states. We plot this quantity in figure 9, together with the replica symmetric (RS) estimate and the value obtained from the $m = 0$ solution.

We also present in figure 10 the entropic complexity curves for a few values of α. Note that these curves can seem incomplete; in fact, for some values of (α, m) we found only
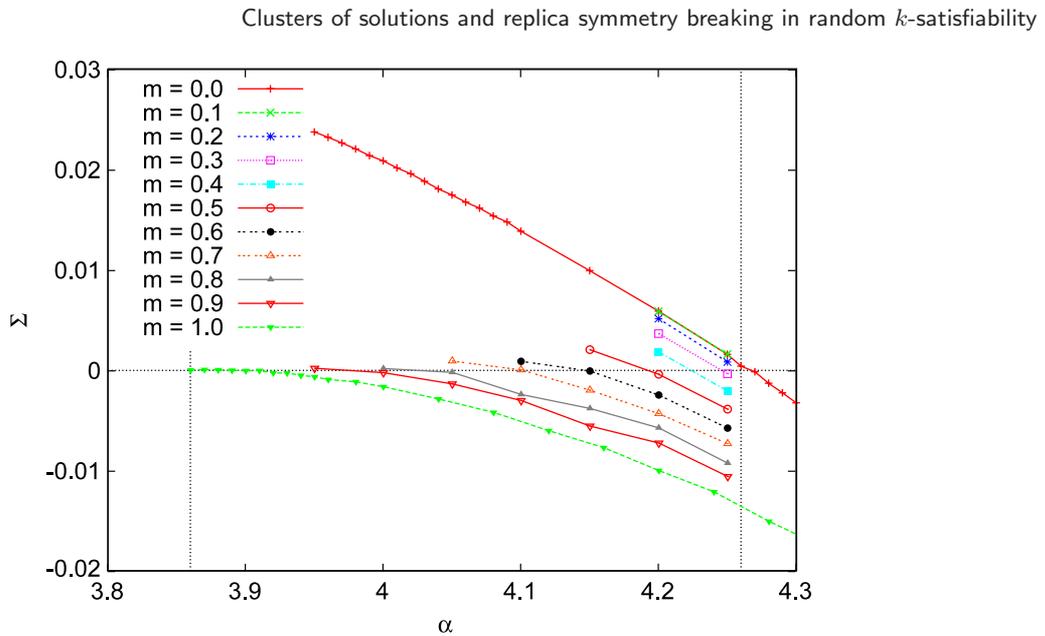
Figure 8. The complexity Σ for $k = 3$ and m from 0 (highest curve) to 1 (lowest curve). For $0 < m < 1$ the domain of existence of Σ may be slightly larger than the one shown in the plot (we have simulated only α values in multiples of 0.05).
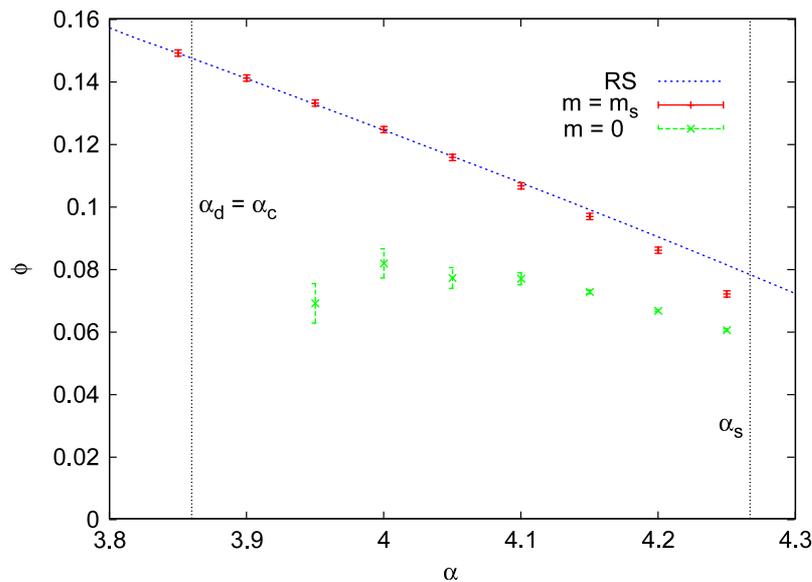
Figure 9. The 1RSB estimate for the entropy of random 3-SAT, compared to the replica symmetric (RS) estimate and to the internal entropy of the $m = 0$ solution, corresponding to the maximum of the $\Sigma(\phi)$ curve.

inconsistent solutions of the 1RSB equations, as is explained in more detail in appendix C. This might be related to an instability of the 1RSB solution toward higher levels of replica symmetry breaking [50]–[52].
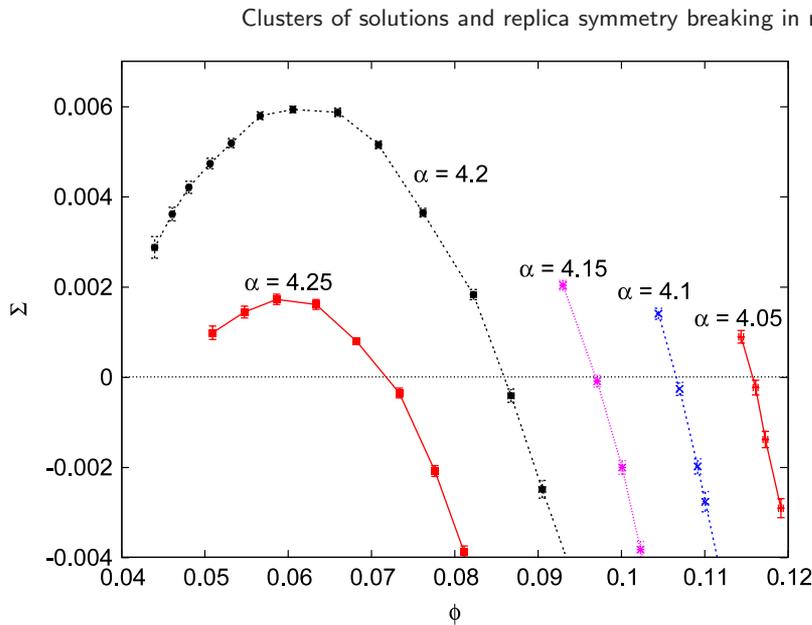
Figure 10. The complexity $\Sigma(\phi)$ in random 3-SAT, for several values of α.

5. Simplifications of the 1RSB equations

The numerical analysis of the 1RSB equations (31) and (32) is, in general, an extremely difficult task. Their analytical control is even more challenging. In this section we explain how the 1RSB approach simplifies in the two cases $m = 0$ and 1, allowing for a precise numerical calculation of the complexity and internal entropy in these points.

Because of the special role played by the value $m = 1$, see section 2, this enables us to estimate precisely the dynamical and condensation thresholds $\alpha_d(k)$ and $\alpha_c(k)$. The simplifications arising at $m = 0$ are, on the other hand, the reason for the efficiency of the SP algorithm [10]. Here we will show how the states entropy can be computed at a small extra cost with respect to the approach of [10].

For the sake of concreteness, we discuss these simplifications in the case of random k-satisfiability. They have, however, a much wider domain of validity. The same derivations do indeed hold for general mean-field models on sparse random graphs.

5.1. $m = 1$ and tree reconstruction

There is a strong connection between the 1RSB formalism with Parisi parameter $m = 1$ and the tree reconstruction problem (or computation of point-to-set correlation), as discussed in [26] and outlined in section 2.3. We follow here a somehow inverse perspective with respect to [26]: starting from the 1RSB equations we shall progressively simplify them. At the end we shall comment on their interpretation in terms of the tree reconstruction problem.

Let us first define the averaging functional $\overline{h}[P]$ (resp. $\overline{u}[Q]$) which associates to the distribution P (resp. Q) of cavity fields a single real through the relations

$$\tanh \overline{h}[P] = \int \mathrm{d}P(h) \tanh h, \qquad \tanh \overline{u}[Q] = \int \mathrm{d}Q(u) \tanh u. \qquad (41)$$

Consider now the right-hand side of equation (32) for $m = 1$. The normalization factor can be expressed in terms of these averaged fields:

$$\mathcal{Z}_3[Q_1^+, \ldots, Q_{l_+}^+, Q_1^-, \ldots, Q_{l_-}^-] = z_3[\overline{u}[Q_1^+], \ldots, \overline{u}[Q_{l_+}^+], \overline{u}[Q_1^-], \ldots, \overline{u}[Q_{l_-}^-]]. \tag{42}$$

Using this fact and denoting by $G[Q_1^+, \ldots, Q_{l_+}^+, Q_1^-, \ldots, Q_{l_-}^-]$ the right-hand side of equation (32) one can also show that

$$\overline{h}[G[Q_1^+, \ldots, Q_{l_+}^+, Q_1^-, \ldots, Q_{l_-}^-]] = \sum_{i=1}^{l_+} \overline{u}[Q_i^+] - \sum_{i=1}^{l_-} \overline{u}[Q_i^-]. \tag{43}$$

Treating similarly equation (31), whose rhs shall be denoted $F[P_1, \ldots, P_{k-1}]$, one obtains

$$\mathcal{Z}_4[P_1, \ldots, P_{k-1}] = z_4(\overline{h}[P_1], \ldots, \overline{h}[P_{k-1}]), \qquad \overline{u}[F[P_1, \ldots, P_{k-1}]] = f(\overline{u}[P_1], \ldots, \overline{u}[P_{k-1}]). \tag{44}$$

$\overline{h}$ (resp. $\overline{u}$) can be viewed as a random variable, induced by equation (41) with P (resp. Q) drawn from $\mathcal{P}_{(1)}$ (resp. $\mathcal{Q}_{(1)}$). The above remarks show that their distributions obey the RS self-consistency equation (25). Let us now define a conditional average of $\mathcal{P}_{(1)}$, focusing on the P's in the support of $\mathcal{P}_{(1)}$ with a prescribed value of $\overline{h}[P]$:

$$\overline{P}(h|\overline{h}) = \frac{1}{\mathcal{P}_{(0)}(\overline{h})} \int d\mathcal{P}_{(1)}[P] \; P(h) \; \delta(\overline{h} - \overline{h}[P]). \tag{45}$$

The conditional distribution $\overline{Q}(u|\overline{u})$ is defined analogously, with $\mathcal{P}_{(1)}[P]$ replaced by $\mathcal{Q}_{(1)}[Q]$.

Consider again the distributional equations (31) and (32). Once the normalization factors have been expressed in terms of the average fields $\overline{h}, \overline{u}$, the right-hand sides are multi-linear functions of the distributions P, Q. It is thus possible to take the conditional average as in equation (45). This yields closed equations on $\overline{P}$ and $\overline{Q}$:

$$\overline{Q}(u|\overline{u}) \mathcal{Q}_{(0)}(\overline{u}) = \int \prod_{i=1}^{k-1} d\mathcal{P}_{(0)}(\overline{h}_i) \delta(\overline{u} - f(\overline{h}_1, \ldots, \overline{h}_{k-1}))$$

$$\times \int \prod_{i=1}^{k-1} d\overline{P}(h_i|\overline{h}_i) \delta(u - f(h_1, \ldots, h_{k-1})) \times \frac{z_4(h_1, \ldots, h_{k-1})}{z_4(\overline{h}_1, \ldots, \overline{h}_{k-1})},$$

$$\overline{P}(h|\overline{h}) \mathcal{P}_{(0)}(\overline{h}) = \sum_{l_+, l_- = 0}^{\infty} \frac{e^{-\alpha k}(\alpha k/2)^{l_+ + l_-}}{l_+! l_-!} \int \prod_{i=1}^{l_+} d\mathcal{Q}_{(0)}(\overline{u}_i^+) \prod_{i=1}^{l_-} d\mathcal{Q}_{(0)}(\overline{u}_i^-) \tag{46}$$

$$\times \, \delta\left(\overline{h} - \sum_{i=1}^{l_+} \overline{u}_i^+ + \sum_{i=1}^{l_-} \overline{u}_i^-\right) \int \prod_{i=1}^{l_+} d\overline{Q}(u_i^+|\overline{u}_i^+) \prod_{i=1}^{l_-} d\overline{Q}(u_i^-|\overline{u}_i^-)$$

$$\times \, \delta\left(h - \sum_{i=1}^{l_+} u_i^+ + \sum_{i=1}^{l_-} u_i^-\right) \frac{z_3(u_1^+, \ldots, u_{l_+}^+, u_1^-, \ldots, u_{l_-}^-)}{z_3(\overline{u}_1^+, \ldots, \overline{u}_{l_+}^+, \overline{u}_1^-, \ldots, \overline{u}_{l_-}^-)}.$$

These equations are definitely simpler than the original ones (31) and (32). In particular $\overline{P}(h|\overline{h})\mathcal{P}_{(0)}(\overline{h})$ can be viewed as a joint distribution of $(h, \overline{h})$ and represented by a population of couples $\{(h_i, \overline{h}_i)\}_{i=1}^{\mathcal{N}}$. The presence of the reweighting factors still represents

a difficulty that we shall now get rid of by a further simplification. Before proceeding, let us emphasize the identities

$$\int \mathrm{d}\overline{P}(h|\overline{h}) \, \tanh h = \tanh \overline{h}, \qquad \int \mathrm{d}\overline{Q}(u|\overline{u}) \, \tanh u = \tanh \overline{u}, \tag{47}$$

which follow directly from the definition (45) and which are indeed preserved by equations (46). We define now, for $\sigma = \pm 1$,

$$\overline{P}_\sigma(h|\overline{h}) = \frac{1 + \sigma \tanh h}{1 + \sigma \tanh \overline{h}} \, \overline{P}(h|\overline{h}). \tag{48}$$

Using property (47), one can check that for any $\overline{h}$ and any σ $\overline{P}_\sigma(\bullet|\overline{h})$ is well normalized, and that

$$\overline{P}(h|\overline{h}) = \sum_\sigma \frac{1 + \sigma \tanh \overline{h}}{2} \overline{P}_\sigma(h|\overline{h}). \tag{49}$$

Similar definitions and properties hold for $\overline{Q}_\sigma(u|\overline{u})$. Inserting these definitions in equation (46), one obtains

$$\overline{Q}_\sigma(u|\overline{u})\mathcal{Q}_{(0)}(\overline{u}) = \int \prod_{i=1}^{k-1} \mathrm{d}\mathcal{P}_{(0)}(\overline{h}_i)\delta(\overline{u} - f(\overline{h}_1, \ldots, \overline{h}_{k-1}))$$

$$\times \sum_{\sigma_1, \ldots, \sigma_{k-1}} \mu(\sigma_1, \ldots, \sigma_{k-1}|\sigma, \overline{h}_1, \ldots, \overline{h}_{k-1})$$

$$\times \int \prod_{i=1}^{k-1} \mathrm{d}\overline{P}_{\sigma_i}(h_i|\overline{h}_i)\delta(u - f(h_1, \ldots, h_{k-1})), \tag{50}$$

where the summation runs over the 2^{k-1} configurations of the Ising spins $\sigma_1, \ldots, \sigma_{k-1}$ with probabilities given by

$$\mu(\sigma_1, \ldots, \sigma_{k-1}|+, \overline{h}_1, \ldots, \overline{h}_{k-1}) = \prod_{i=1}^{k-1} \frac{1 + \sigma_i \tanh \overline{h}_i}{2}, \tag{51}$$

$$\mu(\sigma_1, \ldots, \sigma_{k-1}|-, \overline{h}_1, \ldots, \overline{h}_{k-1}) = \frac{(1 - \mathbb{I}(\sigma_1 = \cdots = \sigma_{k-1} = -))}{1 - \prod_{i=1}^{k-1}(1 - \tanh \overline{h}_i)/(2)} \prod_{i=1}^{k-1} \frac{1 + \sigma_i \tanh \overline{h}_i}{2}. \tag{52}$$

The second of the equations in (46) yields

$$\overline{P}_\sigma(h|\overline{h})\mathcal{P}_{(0)}(\overline{h}) = \sum_{l_+, l_- = 0}^{\infty} \frac{\mathrm{e}^{-\alpha k}(\alpha k/2)^{l_+ + l_-}}{l_+! l_-!} \int \prod_{i=1}^{l_+} \mathrm{d}\mathcal{Q}_{(0)}(\overline{u}_i^+) \prod_{i=1}^{l_-} \mathrm{d}\mathcal{Q}_{(0)}(\overline{u}_i^-)$$

$$\times \delta\left(\overline{h} - \sum_{i=1}^{l_+} \overline{u}_i^+ + \sum_{i=1}^{l_-} \overline{u}_i^-\right) \int \prod_{i=1}^{l_+} \mathrm{d}\overline{Q}_\sigma(u_i^+|\overline{u}_i^+) \prod_{i=1}^{l_-} \mathrm{d}\overline{Q}_{-\sigma}(u_i^-|\overline{u}_i^-)$$

$$\times \delta\left(h - \sum_{i=1}^{l_+} u_i^+ + \sum_{i=1}^{l_-} u_i^-\right). \tag{53}$$

The equations (50) and (53) are particularly convenient for numerical resolution. This can be obtained through an appropriate generalization of the population dynamics algorithm, that employs two population of triples $\{(\overline{h}_i, h_i^+, h_i^-) : i = 1, \ldots, \mathcal{N}\}$ and $\{(\overline{u}_j, u_j^+, u_j^-) : j = 1, \ldots, \mathcal{N}\}$. In the actual implementation it is actually more convenient to store the hyperbolic tangent of these quantities, e.g. $\tanh \overline{h}_i$, $\tanh h_i^+$, etc. These populations are updated recursively according to the pseudocode below.

POPULATION DYNAMICS $m = 1$ (size $\mathcal{N}$, iterations $t_{\max}$)

1:    For all $i \in \{1, \ldots, \mathcal{N}\}$:
2:        Set $h_i^\pm = \pm\infty$ and draw $\overline{h}_i$ from $\mathcal{P}_{(0)}$;
3:    For all $t \in \{1, \ldots, t_{\max}\}$:
4:        For all $j \in \{1, \ldots, \mathcal{N}\}$ generate a new triple $(\overline{u}_j, u_j^+, u_j^-)$:
5:            Choose $k-1$ indices $i_1 \ldots i_{k-1}$ uniformly in $[\mathcal{N}]$;
6:            Compute $\overline{u}_j = f(\overline{h}_{i_1}, \ldots, \overline{h}_{i_{k-1}})$;
7:            Generate a configuration $\sigma_1 \ldots \sigma_{k-1}$ with the law $\mu(\cdots | +, \overline{h}_{i_1} \ldots \overline{h}_{i_{k-1}})$ in equation (51);
8:            Compute $u_j^+ = f(h_{i_1}^{\sigma_1}, \ldots, h_{i_{k-1}}^{\sigma_{k-1}})$;
9:            Generate a second configuration of spins with the law (52);
10:           Set $u_j^- = f(h_{i_1}^{\sigma_1}, \ldots, h_{i_{k-1}}^{\sigma_{k-1}})$;
11:       End-For;
12:       For all $i \in \{1, \ldots, \mathcal{N}\}$ generate a new triple $(\overline{h}_i, h_i^+, h_i^-)$:
13:           Draw two independent Poisson random variables l_+ and l_- of mean $\alpha k/2$;
14:           Draw $l_+ + l_-$ i.i.d. indices $i_1^+, \ldots, i_{l_+}^+, i_1^-, \ldots, i_{l_-}^-$ uniformly random in $[\mathcal{N}]$;
15:           Set $\overline{h}_j = \sum_{m=1}^{l_+} \overline{u}_{i_m^+} - \sum_{m=1}^{l_-} \overline{u}_{i_m^-}$, $h_j^\pm = \sum_{m=1}^{l_+} u_{i_m^+}^\pm - \sum_{m=1}^{l_-} u_{i_m^-}^\mp$;
16:   End-For;

The justification of the initialization will be given below. After a moment of thought one can convince oneself that the above update rules are the correct discretization of equations (50) and (53). More precisely, if the triples $(\overline{h}_i, h_i^+, h_i^-)$ are i.i.d. and the two pairs $(\overline{h}_i, h_i^+)$, $(\overline{h}_i, h_i^-)$ have distributions (respectively) $\overline{P}_+(h^+|\overline{h})\mathcal{P}_{(0)}(\overline{h})$ and $\overline{P}_-(h^-|\overline{h})\mathcal{P}_{(0)}(\overline{h})$, then the pairs $(\overline{u}_j, u_j^+)$, $(\overline{u}_j, u_j^-)$ resulting from the above update have distributions $\overline{Q}_+(u^+|\overline{u})\mathcal{Q}_{(0)}(\overline{u})$, $\overline{Q}_-(u^-|\overline{u})\mathcal{Q}_{(0)}(\overline{u})$. An analogous statement holds for the update from the triples $(\overline{u}_j, u_j^+, u_j^-)$ to $(\overline{h}_i, h_i^+, h_i^-)$.[14]

Most relevant observables can be written as expectations with respect to the distributions $\overline{P}_\pm(h_\pm|\overline{h})\mathcal{P}_{(0)}(\overline{h})$, $\overline{Q}_\pm(u_\pm|\overline{u})\mathcal{Q}_{(0)}(\overline{u})$ and hence estimated from these populations of triplets.

Notice that, by definition, the 1RSB potential computed at $m = 1$ is equal to the RS free-entropy, $\Phi(m = 1) = \phi_{(0)}$. The internal entropy can be expressed in terms of $\overline{P}(h|\overline{h})$ and $\overline{Q}(u|\overline{u})$ by integrating over $\mathcal{P}_{(1)}$, $\mathcal{Q}_{(1)}$ in equation (39). These conditional distributions can be further replaced by $\overline{P}_\sigma$ and $\overline{Q}_\sigma$ thanks to equation (49), yielding finally

$$\phi_{\text{int}}(m = 1) = -\alpha k \int d\mathcal{P}_{(0)}(\overline{h}) d\mathcal{Q}_{(0)}(\overline{u}) \sum_\sigma \frac{1 + \sigma \tanh(\overline{u} + \overline{h})}{2} \int d\overline{P}_\sigma(h|\overline{h}) d\overline{Q}_\sigma(u|\overline{u})$$

$$\times \log z_1(u, h) + \alpha \int \prod_{i=1}^k d\mathcal{P}_{(0)}(\overline{h}_i) \sum_{\sigma_1, \ldots, \sigma_k} \mu(\sigma_1, \ldots, \sigma_k | \overline{h}_1, \ldots, \overline{h}_k)$$

---

[14] Notice that it would be wrong to claim that $(\overline{h}_i, h_i^+, h_i^-)$ is distributed according to $\overline{P}_+(h^+|\overline{h})\overline{P}_-(h_-|\overline{h})\mathcal{P}_{(0)}(\overline{h})$: the update rules used in the algorithm induce correlations between (for instance) the fields h^+ and h^- inside the same triplet. These correlations do not spoil our claim.

$$\times \int \prod_{i=1}^{k} \mathrm{d}\overline{P}_{\sigma_i}(h_i|\overline{h}_i) \log z_2(h_1,\ldots,h_k) + \sum_{l_+,l_-=0}^{\infty} \frac{\mathrm{e}^{-\alpha k}(\alpha k/2)^{l_++l_-}}{l_+!l_-!}$$

$$\times \int \prod_{i=1}^{l_+} \mathrm{d}\mathcal{Q}_{(0)}(\overline{u}_i^+) \prod_{i=1}^{l_-} \mathrm{d}\mathcal{Q}_{(0)}(\overline{u}_i^-) \sum_{\sigma} \frac{1 + \sigma \tanh\left(\sum_{i=1}^{l_+} \overline{u}_i^+ - \sum_{i=1}^{l_-} \overline{u}_i^-\right)}{2}$$

$$\times \int \prod_{i=1}^{l_+} \mathrm{d}\overline{Q}_\sigma(u_i^+|\overline{u}_i^+) \prod_{i=1}^{l_-} \mathrm{d}\overline{Q}_{-\sigma}(u_i^-|\overline{u}_i^-) \log z_3(u_1^+,\ldots,u_{l_+}^+, u_1^-,\ldots,u_{l_-}^-). \quad (54)$$

In the second term the distribution of the configuration $(\sigma_1,\ldots,\sigma_k)$ is

$$\mu(\sigma_1,\ldots,\sigma_k|\overline{h}_1,\ldots,\overline{h}_k) = \frac{(1 - \mathbb{I}(\sigma_1 = \cdots = \sigma_k = -))}{1 - \prod_{i=1}^{k}(1 - \tanh \overline{h}_i)/2} \prod_{i=1}^{k}(1 + \sigma_i \tanh \overline{h}_i)/(2). \quad (55)$$

This expression of the internal free-entropy is readily evaluated by sampling from the population of triplets defined above. The complexity of the $m = 1$ states is then finally expressed as $\Sigma(m = 1) = \Phi(m = 1) - \phi_{\mathrm{int}}(m = 1)$.

Consider now the definition of the overlaps given in equation (38). The inter-state one q_0 is easily seen to be equal to the RS one. Moreover q_1 can be written as

$$q_1 = \int \mathrm{d}\mathcal{P}_{(0)}(\overline{h}) \int \mathrm{d}\overline{P}(h|\overline{h}) \tanh^2 h. \quad (56)$$

To rewrite q_1 in terms of the distribution $\overline{P}_\sigma$, note that $\tanh^2 h = (\tanh h)\sum_\sigma \sigma(1 + \sigma \tanh h)/2$ and use (48) to obtain

$$q_1 = \int \mathrm{d}\mathcal{P}_{(0)}(\overline{h}) \sum_\sigma \sigma \frac{1 + \sigma \tanh \overline{h}}{2} \int \mathrm{d}\overline{P}_\sigma(h|\overline{h}) \tanh h. \quad (57)$$

These expressions allow us to estimate q_0, q_1 from the population of triples $\{(\overline{h}_i, h_i^+, h_i^-)\}$.

In figures 2 and 7 we followed this approach to plot the difference $q_1(\ell) - q_0$ for several values of α and $k = 3, 4$, whereby the population $\{(\overline{h}_i, h_i^+, h_i^-)\}$ is obtained after ℓ iterations of the above algorithm. For $\alpha < \alpha_{\mathrm{d}}(k)$, $q_1(\ell) - q_0 \xrightarrow{\ell} 0$, while for $\alpha > \alpha_{\mathrm{d}}(k)$ it is bounded away from 0. Let us emphasize the great simplification achieved: equations (50) and (53) are much simpler than the original 1RSB equations: they can be solved using a simple population of triples, instead of a population of populations. Further, the initialization used in the pseudocode above is the *correct* one, in the following sense. If equations (50) and (53) admit a non-trivial solution, then their iteration converges to a non-trivial solution under such an initialization.

The last statement follows from the interpretation of the order parameters in terms of tree reconstruction. Consider an infinite tree k-satisfiability formula rooted at variable node i. The tree is random with distribution defined by letting each variable to be directly (resp. negated) in l_+ (resp. l_-) clauses, where $l_\pm$ are independent random Poisson random variables with mean $\alpha k/2$. One can define a (uniform) free boundary Gibbs measure μ over SAT assignments of such a tree. Imagine now we generate a solution from this measure, conditional on the root value being σ, and denote by $\underline{\sigma}_B$ the values of variables

at a distance at least ℓ from the root. Define the fields $\overline{h}$, h_ℓ^σ by

$$\mu(\sigma_i) \equiv \frac{1 + \sigma_i \tanh \overline{h}}{2}, \qquad \mu(\sigma_i | \underline{\sigma}_B) \equiv \frac{1 + \sigma_i \tanh h_\ell^\sigma}{2}. \tag{58}$$

Notice that both are random quantities, $\overline{h}$ because of the tree randomness and $h_\ell^\pm$ both because of the tree and of the random configuration $\underline{\sigma}_B$. Let $\overline{P}_\sigma^\ell(h|\overline{h})$ be the conditional distribution of h_ℓ^σ given $\overline{h}$.

It is not hard to show that $\overline{P}_\sigma^\ell(h|\overline{h})$ is the distribution obtained by iterating (50) and (53) ℓ times with initial condition $\overline{P}_\pm^\ell(h|\overline{h}) = \mathcal{P}_{(0)}(\overline{h})\delta(h \mp \infty)$. This corresponds indeed to the initialization we used in the population dynamics algorithm. It follows from the arguments in [26] that this is the *correct* initialization, in the sense described above. Further, under the usual assumptions of the cavity method and for $\alpha < \alpha_c(k)$, the quantity $q_1(\ell) - q_0$ plotted in figures 2 and 7 coincides with the correlation function (19) in the large N limit.

5.2. $m = 0$: survey propagation and the associated internal entropy

We turn now to the second particular case for which a simplified treatment of the 1RSB formalism is possible, namely at $m = 0$.

To begin with, let us consider the structure of the distributions $P(h)$ (resp. $Q(u)$) in the support of $\mathcal{P}_{(1)}$ (resp. $\mathcal{Q}_{(1)}$) for an arbitrary value of m. A moment of thought reveals the possibility of 'hard fields' $h = \pm\infty$ that strictly constrains a variable to take the same value in all configurations of a cluster of solutions. We can take care explicitly of this possibility by denoting

$$P(h) = x^- \delta(h + \infty) + x^+ \delta(h - \infty) + (1 - x^- - x^+)\widetilde{P}(h),$$
$$Q(u) = y\,\delta(u - \infty) + (1 - y)\widetilde{Q}(u), \tag{59}$$

where $\widetilde{P}$ and $\widetilde{Q}$ have their support on finite values of the fields, that shall be called 'soft' or 'evanescent'. Rewriting the right-hand side of (31) with these notations yields

$$Q(\bullet) \overset{\mathrm{d}}{=} \frac{1}{\mathcal{Z}_4[P_1, \ldots, P_{k-1}]} \left[\left(\prod_{i=1}^{k-1} x_i^- \right) \delta(\bullet - \infty) + 2^m \left(1 - \prod_{i=1}^{k-1}(1 - x_i^+) \right) \delta(\bullet) \right.$$
$$+ \sum_{|I| \geq 1} \prod_{i \in I}(1 - x_i^+ - x_i^-) \prod_{i \notin I} x_i^- \int \prod_{i \in I} \mathrm{d}\widetilde{P}_i(h_i)(1 + \mathrm{e}^{-2\bullet})^m$$
$$\left. \times \, \delta\left(\bullet + \frac{1}{2} \log\left(1 - \prod_{i \in I} \frac{1 - \tanh h_i}{2} \right) \right) \right], \tag{60}$$

where the summation on I is over the non-empty subsets of $\{1, \ldots, k-1\}$.

To achieve the same task for equation (32) it is advisable to introduce some more compact notations:

$$\pi_\sigma = \prod_{i=1}^{l_\sigma}(1 - y_i^\sigma), \qquad S_\sigma = \prod_{i=1}^{l_\sigma}(1 + \tanh u_i^\sigma), \qquad T_\sigma = \prod_{i=1}^{l_\sigma}(1 - \tanh u_i^\sigma), \tag{61}$$

in terms of which we have, for instance, $z_3 = S_+ T_- + T_+ S_-$. We shall also denote $\mathsf{E}[\bullet]$ the average over the $u_i^\pm$ drawn from the $Q_i^\pm$, and $\widetilde{\mathsf{E}}$ similarly using $\widetilde{Q}_i^\pm$. We then obtain

$$
\begin{aligned}
P(\bullet) \overset{\mathrm{d}}{=} \frac{1}{\mathcal{Z}_3[\{Q_i^+\},\{Q_i^-\}]} &\Bigg[\pi_+ \widetilde{\mathsf{E}}[T_+^m]\Big(\mathsf{E}[S_-^m] - \pi_- \widetilde{\mathsf{E}}[S_-^m]\Big)\delta(\bullet + \infty) \\
&+ \pi_- \widetilde{\mathsf{E}}[T_-^m]\Big(\mathsf{E}[S_+^m] - \pi_+ \widetilde{\mathsf{E}}[S_+^m]\Big)\delta(\bullet - \infty) \\
&+ \pi_+ \pi_- \int \prod_{i=1}^{l_+} \mathrm{d}\widetilde{Q}_i^+(u_i^+) \prod_{i=1}^{l_-} \mathrm{d}\widetilde{Q}_i^-(u_i^-) \\
&\times \delta\Big(\bullet - \sum_{i=1}^{l_+} u_i^+ + \sum_{i=1}^{l_-} u_i^- \Big)(S_+ T_- + T_+ S_-)^m \Bigg].
\end{aligned} \tag{62}
$$

Analogously, the replicated free-entropy $\Phi(m)$ and its derivative can be rewritten by making explicit the distinction between hard and soft fields.

Consider now the previous equations with $m = 0$. As we have explicitly removed all the contradictory terms which had a strictly vanishing reweighting factor in the original relations (31) and (32), all the terms raised to the power m in equations (60) and (62) are strictly positive, hence these factors go to 1 when m vanishes. Two important consequences are to be underlined: the normalization factors $\mathcal{Z}_3$ and $\mathcal{Z}_4$ do not depend on the evanescent distributions $\widetilde{P}$, $\widetilde{Q}$. In fact, $\mathcal{Z}_3 = \pi_+ + \pi_- - \pi_+ \pi_-$ and $\mathcal{Z}_4 = 1$. Moreover the equations on the intensity of the hard fields peaks decouple from the evanescent part when m goes to 0 (60) and (62) yielding for them

$$
y \overset{\mathrm{d}}{=} \prod_{i=1}^{k-1} x_i^-, \qquad (x^+, x^-) \overset{\mathrm{d}}{=} \left(\frac{(1 - \pi_+)\pi_-}{\pi_+ + \pi_- - \pi_+ \pi_-}, \frac{(1 - \pi_-)\pi_+}{\pi_+ + \pi_- - \pi_+ \pi_-} \right), \tag{63}
$$

which are nothing other than the probabilistic form of the survey propagation equations [12]. For future use we denote $Q_{\mathrm{SP}}(y)$ and $P_{\mathrm{SP}}(x_+, x_-)$ the distributions of these random variables. The complexity at $m = 0$ is $\Sigma(m = 0) = \Phi(m = 0)$ and can then be expressed from equation (34) as

$$
\begin{aligned}
\Sigma(m = 0) = \Phi(m = 0) = &-\alpha k \mathbb{E}[\log(1 - x^- y)] + \alpha \mathbb{E}\left[\log\left(1 - \prod_{i=1}^{k} x_i^- \right) \right] \\
&+ \mathbb{E}[\log(\pi_+ + \pi_- - \pi_+ \pi_-)],
\end{aligned} \tag{64}
$$

where the average is done with respect to P_{SP} and Q_{SP}.

By focusing on the intensity of the hard fields this 'energetic' version of the cavity method [10, 12] lost the information contained in the evanescent field distributions $\widetilde{P}$, $\widetilde{Q}$, which is necessary to obtain the internal entropy of the states, $\Phi'(m = 0)$. This quantity can, however, be obtained in a rather simple way. We shall indeed define $\widetilde{Q}(u|y)$ as the average of the evanescent part of Q drawn from $\mathcal{Q}_{(1)}$, conditioned on the value of the hard field delta peak, and similarly $\widetilde{P}(h|x^+, x^-)$. As the right-hand sides of (60) and (62) are linear functionals of these evanescent distributions when $m = 0$, closed equations on this conditional average can be obtained. We shall write them in terms of the joint

distributions $\widetilde{Q}(u,y) = \widetilde{Q}(u|y)Q_{\mathrm{SP}}(y)$ and $\widetilde{P}(h,x^+,x^-) = \widetilde{P}(h|x^+,x^-)P_{\mathrm{SP}}(x^+,x^-)$:

$$
\begin{aligned}
\widetilde{Q}(u,y) = \int \prod_{i=1}^{k-1} \mathrm{d}h_i\, \mathrm{d}x_i^+\, \mathrm{d}x_i^-\, \widetilde{P}(h_i,x_i^+,x_i^-)\delta\left(y - \prod_{i=1}^{k-1} x_i^-\right) \Bigg[\frac{1 - \prod_{i=1}^{k-1}(1 - x_i^+)}{1-y}\delta(u) \\
+ \frac{\sum_{p=1}^{k-1} \binom{k-1}{p} \prod_{i=1}^{p}(1 - x_i^+ - x_i^-) \prod_{i=p+1}^{k-1} x_i^-}{1-y} \\
\times \delta\left(u + \frac{1}{2}\log\left(1 - \prod_{i=1}^{p} \frac{1-\tanh h_i}{2}\right)\right)\Bigg],
\end{aligned}
\tag{65}
$$

$$
\begin{aligned}
\widetilde{P}(h,x^+,x^-) = \sum_{l_+,l_-=0}^{\infty} \frac{\mathrm{e}^{-\alpha k}(\alpha k/2)^{l_++l_-}}{l_+!l_-!} \int \prod_{i=1}^{l_+} \mathrm{d}u_i^+\, \mathrm{d}y_i^+\, \widetilde{Q}(u_i^+,y_i^+) \prod_{i=1}^{l_-} \mathrm{d}u_i^-\, \mathrm{d}y_i^-\, \widetilde{Q}(u_i^-,y_i^-) \\
\times \delta\left(x^+ - \frac{(1-\pi_+)\pi_-}{\pi_+ + \pi_- - \pi_+\pi_-}\right)\delta\left(x^- - \frac{(1-\pi_-)\pi_+}{\pi_+ + \pi_- - \pi_+\pi_-}\right) \\
\times \delta\left(h - \sum_{i=1}^{l_+} u_i^+ + \sum_{i=1}^{l_-} u_i^-\right).
\end{aligned}
\tag{66}
$$

A solution of these equations can be obtained through a simple population dynamics algorithm, encoding $\widetilde{Q}(u,y)$ as a population of couples $\{(u_i,y_i)\}_{i=1}^{\mathcal{N}}$ and $\widetilde{P}(h,x^+,x^-)$ as $\{(h_i,x_i^+,x_i^-)\}_{i=1}^{\mathcal{N}}$. The update rules of the algorithm can be deduced from (65) and (66): a new element (h,x^+,x^-) is obtained drawing two Poisson random variables $l_\pm$ of mean $\alpha k/2$, $l_+ + l_-$ elements of the population $\{(u_i,y_i)\}$ and combining them according to (66). The translation of (65) is only slightly more complicated. After extracting $k-1$ elements at random from the population $\{(h_i,x_i^+,x_i^-)\}$ one obtains y as the product of the $k-1$ elements x_-. One then draws a configuration $(s_1,\dots,s_{k-1}) \in \{-1,0,+1\}^{k-1}$, each 'spin' s_i being ± 1 with probability $x_i^\pm$ and 0 with probability $1 - x_i^+ - x_i^-$, conditional on $(s_1,\dots,s_{k-1}) \neq (-1,\dots,-1)$. If at least one of the s_i is equal to $+1$ the new value of u is taken to 0, otherwise $u = -\log(1 - \prod(1 - \tanh h_i)/2)/2$, the product being taken on the indices i such that $\sigma_i = 0$.

The internal entropy of the $m = 0$ pure states can be obtained from the solution of these equations, simplifying equation (39) into

$$
\begin{aligned}
\phi_{\mathrm{int}}&(m=0) \\
={}& -\alpha k\mathbb{E}\left[\frac{x^+ y\log 2 + (1-y)(x^-\log(1-\tanh u) + x^+\log(1+\tanh u))}{1 - x^- y}\right] \\
& -\alpha k\mathbb{E}\left[\frac{(1 - x^+ - x^-)(y\log(1+\tanh h) + (1-y)\log(1 + \tanh h \tanh u))}{1 - x^- y}\right] \\
& + \alpha\mathbb{E}\left[\sum_{p=1}^{k} \binom{k}{p} \prod_{i=1}^{p}(1 - x_i^+ - x_i^-) \prod_{i=p+1}^{k} x_i^- \log\left(1 - \prod_{i=1}^{p} \frac{1-\tanh h_i}{2}\right)\right] \\
& + \mathbb{E}\left[\pi_+\pi_-\log(S_+T_- + T_+S_-) + \pi_-(1-\pi_+)\log(S_+T_-) + \pi_+(1-\pi_-)\log(T_+S_-)\right],
\end{aligned}
\tag{67}
$$

where the expectation is over independent copies of elements drawn from $\widetilde{P}(h, x^+, x^-)$ and $\widetilde{Q}(u, y)$, and in the last line (where we used the shorthand notations defined in (61)) over the Poissonian random variables $l_\pm$. This quantity was plotted for $k = 4$ in figure 3.

Let us emphasize the major numerical simplification with respect to the general 1RSB equations: we have to deal here with populations of couples (or triplets) of fields, not populations of populations. Yet we manage to extract not only the complexity, which was the one computed in the probabilistic version of survey propagation, but also the associated internal entropy.

6. Large k results

To complement the numerical resolution of the 1RSB equations, we present in this section analytic expansions of the various thresholds and thermodynamic quantities for large k. Some technical details of these computations are deferred to appendix B.

6.1. Dynamical transition regime

A non-trivial solution of the 1RSB equations appears in the regime defined by

$$\alpha_{\mathrm{d}} = \frac{2^k}{k} \left[\log k + \log \log k + \gamma + \mathrm{O}\left(\frac{\log \log k}{\log k}\right) \right], \tag{68}$$

with γ finite as $k \to \infty$. In this regime the 1RSB distributional order parameters $\mathcal{P}_{(1)}$, $\mathcal{Q}_{(1)}$ are supported on cavity field distributions of the form (59) with $\widetilde{P}(\cdot)$, $\widetilde{Q}(\cdot)$ supported on finite fields. The weights of the hard fields are deterministic to leading order, with

$$
\begin{aligned}
x^\pm &= \frac{1}{2} - \frac{\delta(\gamma, m)}{2k \log k} + \mathrm{O}\left(\frac{1}{k(\log k)^2}\right), \\
y &= \frac{1}{2^k} 2^{1-m} \left\{ 1 - \frac{\delta(\gamma, m)}{\log k} + \mathrm{O}\left(\frac{1}{(\log k)^2}\right) \right\}.
\end{aligned}
\tag{69}
$$

A set of coupled equations can also be written for the averages of $\widetilde{P}$, $\widetilde{Q}$, in terms of which one computes a function $\Lambda(\delta, m)$ that finally determines $\delta(\gamma, m)$ as a function of γ by solving the following equation:

$$\gamma = \delta + \log \frac{1}{2\delta} + \Lambda(\delta, m). \tag{70}$$

Both the expressions for $\Lambda(\delta, m)$ and the equations for the averages of $\widetilde{P}$, $\widetilde{Q}$ are quite involved and we report them in appendix B. In any case the right-hand side of equation (70) diverges for $\delta \to 0$ and $\delta \to \infty$. As a consequence a pair of solutions[15] appears for $\gamma \geq \gamma_{\mathrm{d}}(m)$, where $\gamma_{\mathrm{d}}(m)$ is obtained by minimizing the above expression over δ. For $m = 0, 1$ the formulas simplify yielding $\Lambda(\delta, m = 0) = 0$ and $\Lambda(\delta, m = 1) = \log 2$ independently of δ, whence the minimum takes place at $\delta = 1$ for these two values of m.

To summarize this yields the following estimate for the dynamical threshold:

$$\alpha_{\mathrm{d}}(k, m) = \frac{2^k}{k} \left[\log k + \log \log k + \gamma_{\mathrm{d}}(m) + \mathrm{O}\left(\frac{\log \log k}{\log k}\right) \right], \tag{71}$$

---

[15] Consistency arguments imply that the one with smaller δ must be selected.

with $\gamma_{\mathrm{d}}(m = 1) = 1$ and $\gamma_{\mathrm{d}}(m = 0) = 1 - \log 2$. Notice that the transition at $m = 1$ occurs slightly after the one at $m = 0$, in agreement with what is found numerically for small values of $k \geq 4$.

6.2. Intermediate regime

Consider now the limit $k \to \infty$ with $\alpha = 2^k \widehat{\alpha}$ for some fixed $\widehat{\alpha} > 0$. On this scale the SAT/UNSAT phase transition occurs at $\widehat{\alpha}_{\mathrm{s}} = \log 2 + \mathrm{O}(2^{-k})$ [6, 12]. We shall therefore assume $\widehat{\alpha} \in (0, \log 2)$. In this regime it is convenient to use again the decomposition (59), with at leading order $\widetilde{P}(h) = \delta(h)$ and $x^{\pm} = (1/2)(1 - \widehat{x}^{\pm} \mathrm{e}^{-\widehat{\alpha} k})$. From this ansatz one finds that $\widehat{x}^{\pm} = 2^{m-1}$; then it follows that the 1RSB potential is asymptotically

$$\Phi(m) = \log 2 - \widehat{\alpha} + \mathrm{e}^{-\widehat{\alpha} k}(2^{m-1} - 1) + \mathrm{O}(2^{-k}). \tag{72}$$

By derivation of this expression one obtains the internal entropy

$$\phi_{\mathrm{int}}(m) = \mathrm{e}^{-\widehat{\alpha} k} 2^{m-1} \log 2 + \mathrm{O}(2^{-k}), \tag{73}$$

and defining a reduced quantity σ by $\phi_{\mathrm{int}} = \mathrm{e}^{-\widehat{\alpha} k}(\log 2)\sigma$, we get the complexity function explicitly:

$$\Sigma(\sigma) = \log 2 - \widehat{\alpha} + \mathrm{e}^{-\widehat{\alpha} k}\widetilde{\Sigma}(\sigma) + \mathrm{O}(2^{-k}), \qquad \widetilde{\Sigma}(\sigma) = \sigma(1 - \log 2) - \sigma \log \sigma - 1. \tag{74}$$

Notice that, for large k, the internal entropy of states is exponentially smaller (in k) than the complexity. Further, to leading order, the complexity vanishes at $\widehat{\alpha} = \log 2$, independently on m.

6.3. Condensation regime

In order to resolve the separation between the condensation and satisfiability phase transitions we must let $k \to \infty$ with $\alpha \simeq 2^k \log 2$. More precisely, we define $\alpha = 2^k \log 2 - \zeta$ and take $k \to \infty$ with ζ fixed. Again, we use the ansatz (59) with, at leading order $\widetilde{P}(h) = \delta(h)$ and $x^{\pm} = (1/2)(1 - \widehat{x}^{\pm} 2^{-k})$.

We then get the expansion of the potential

$$\Phi(m) = \frac{1}{2^k}\{\zeta - \zeta_{\mathrm{s}} + (2^m - 1)/2\} + \mathrm{O}(2^{-2k}), \tag{75}$$

with $\zeta_{\mathrm{s}} \equiv \frac{1}{2}(1 + \log 2)$. The entropy can be determined by deriving the above with respect to m; defining the reduced entropy density through $\phi_{\mathrm{int}} = 2^{-k}(\log 2)\sigma$, the complexity is in this regime

$$\Sigma(\sigma) = \frac{1}{2^k}\left\{\zeta - \zeta_{\mathrm{s}} + \sigma(1 - \log 2) - \sigma \log \sigma - \frac{1}{2}\right\} + \mathrm{O}(2^{-2k}). \tag{76}$$

The condensation and satisfiability transition are located by determining ζ such that $\Sigma(m) = 0$ for (respectively) $m = 1$ and 0. We get

$$\alpha_{\mathrm{c}}(k) = 2^k \log 2 - \frac{3 \log 2}{2} + \mathrm{O}(2^{-k}), \qquad \alpha_{\mathrm{s}}(k) = 2^k \log 2 - \frac{1 + \log 2}{2} + \mathrm{O}(2^{-k}). \tag{77}$$

The thermodynamic value $m_{\mathrm{s}}(\zeta)$ of the Parisi parameter between these two thresholds is obtained by minimizing $\Phi(m)/m$. At the order of the expression of $\Phi(m)$ given above
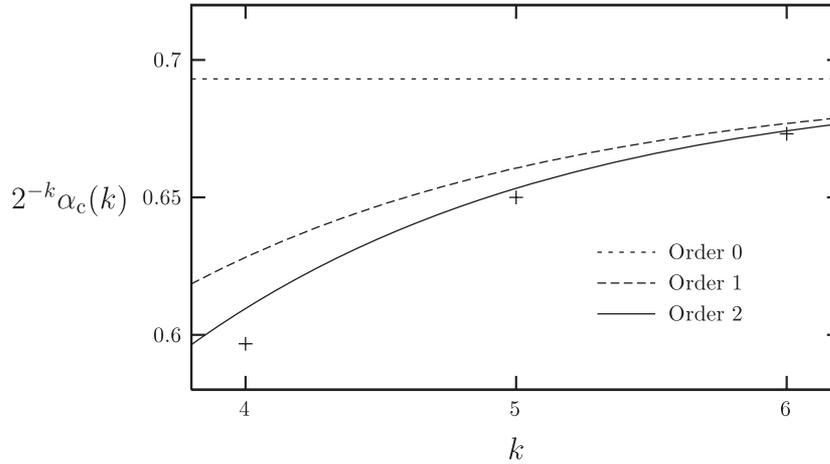
Figure 11. Condensation threshold in reduced units, $2^{-k}\alpha_{\mathrm{c}}(k)$. Symbols: numerical determination by population dynamics algorithm, see table 1. Lines: analytical large-k expansion, truncated at the three first orders, see equation (80).

$m_{\mathrm{s}}(\zeta)$ is a solution of

$$\zeta - \zeta_{\mathrm{s}} = 2^{m-1}(2^{-m} - 1 + m\log 2). \tag{78}$$

In particular one finds close to the satisfiability transition

$$m_{\mathrm{s}}(\zeta) \simeq \frac{2}{\log 2}\sqrt{\zeta - \zeta_{\mathrm{s}}}. \tag{79}$$

A systematic expansion in powers of 2^{-k} of the satisfiability threshold $\alpha_{\mathrm{s}}(k)$ has been performed up to seventh order in [12]. The corresponding expansion for the condensation threshold $\alpha_{\mathrm{c}}(k)$ is slightly more difficult, because of the necessary control of the corrections to the evanescent field distributions. We thus contented ourselves with the computation of the next order in the expansion:

$$
\begin{aligned}
\alpha_{\mathrm{c}}(k) = {} & 2^k \log 2 - \frac{3\log 2}{2} \\
& - \left[\frac{6(\log 2)(\log 3) - 7(\log 2)^2}{4}k^2 + \frac{5(\log 2)^2 - 3(\log 2)(\log 3)}{2}k - \frac{5\log 2}{12}\right]\frac{1}{2^k} \\
& + \mathrm{O}\left(\mathrm{poly}(k)\frac{1}{2^{2k}}\right).
\end{aligned}
\tag{80}
$$

This expression is compared in figure 11 with the numerical results for small k.

7. Conclusion

The set of solutions of random k-satisfiability formulas exhibits a surprisingly rich structure that has been explored in a series of statistical mechanics studies [8]–[10]. Either implicitly or explicitly, these studies are based on defining a probability distribution over the solutions, and then analyzing its properties. While the most natural choice is the uniform measure, the authors of [10] achieved a great simplification (and a wealth of

exact results) by implicitly weighting each solution inversely to the size of the 'cluster' it belongs to. Since clusters sizes are exponential in the number of variables, and have large deviations, this amounts to focusing on an exponentially small subset of solutions.

In this paper we resumed the (technically more challenging) task of studying the uniform measure and obtained the first complete phase diagram (including replica symmetry breaking) in this setting. While we confirmed several of the predictions in [10], our analysis unveiled a number of new phenomena:

(1) There exists a critical value $\alpha_d(k)$ of the clause density that can be characterized in several equivalent ways: (i) divergence of autocorrelation time under Glauber dynamics; (ii) divergence of point-to-set correlation length; (iii) appearance of bottlenecks between 'sizable' subsets of solutions. The value of $\alpha_d(k)$ is bigger than the value obtained with the method of [10] (except for $k = 3$, where it is smaller).

(2) While $\alpha_d(k)$ does not correspond to an actual thermodynamic phase transition, such a phase transition takes place at a second threshold $\alpha_c(k) < \alpha_s(k)$ ($\alpha_s(k)$ being the satisfiability threshold). This manifests in two-point correlations, as well as in the overlap distribution.

(3) The phase diagram is qualitatively different for $k \geq 4$ and $k = 3$. The latter value has been most commonly used in numerical simulations. This difference had not been recognized before because it does not show up in the behavior of the maximal complexity $\Sigma(m = 0)$ investigated up to now.

A number of research directions are suggested by this refined understanding:

(a) We kept ourselves to 1RSB: it would be extremely interesting to investigate whether more complex hierarchical (FRSB) structures can arise in the set of solutions. A first step in this direction would be to analyze the stability [50]–[52] of the 1RSB ansatz, in particular to clarify our numerical findings for $k = 3$. For $k \geq 4$ we believe that our determination of α_d and α_c is not affected by FRSB, yet it might be that the pure states, for some values of their internal entropy, are to be described by an FRSB structure.

(b) The dynamical threshold $\alpha_d(k)$ is expected to affect algorithms that satisfy detailed balance with respect to the uniform measure over solutions (or its positive temperature version). Let us stress that it is likely not to have any relation to more general local search algorithms [53]–[55]. It is an open problem to generalize the static computations performed here to obtain meaningful predictions in those cases.

(c) Finally, the discovery of the condensation phase transition at $\alpha_c(k)$ suggests that belief propagation might be effective in computing marginals up to this threshold, as the average of the 1RSB equations with $m = 1$ corresponds to BP. The possible use of this information in constructing solutions is discussed in [56]–[59].

Acknowledgments

Appendix A. On the numerical resolution of the 1RSB cavity equations

In this section we discuss some issues related to the numerical resolution of equations (31) and (32). As already mentioned, the 1RSB order parameter $\mathcal{P}_{(1)}[P]$ is approximated by a sample of $\mathcal{N}$ populations, each composed of $\mathcal{N}'$ elements $h_{i,j}$, $i \in [\mathcal{N}]$, $j \in [\mathcal{N}']$. The numerical results presented in this work have been obtained with $\mathcal{N} = 10^4$ and $\mathcal{N}' = 10^3$.

The solution to the 1RSB cavity equations is found by an iterative procedure: starting from a 'good' initial guess for the fixed point solution, we iterate a sampled version of equations (31) and (32). After some iterations the sample of populations converges to a stationary state with fluctuations of order $\mathrm{O}(1/\sqrt{\mathcal{N}}, 1/\sqrt{\mathcal{N}'})$. Convergence to the stationary regime is usually fast and may take around 10^2 iterations in the worst cases we encountered. Once in the stationary regime, we keep iterating for at least 10^4 steps. Meanwhile we take averages (over the populations and over the time evolution) of the quantities of interest. This considerably reduces statistical errors.

Our actual numerical implementation makes use of two transformations with respect to equations (31) and (32). First, we make a change of variables into

$$\varphi = \mathrm{e}^{-2u}, \qquad \psi = \frac{1 + \tanh(h)}{2} , \tag{A.1}$$

both taking values in $[0, 1]$ (note that the variable u is defined non-negative, see the definition of the function $f(h_1, \ldots, h_{k-1})$ in equation (23)). Moreover we exploit the fact that the reweighting term $z_4(h_1, \ldots, h_{k-1})$ in equation (31) is a function of $u = f(h_1, \ldots, h_{k-1})$ (cf. equation (33)). This allows us to transfer all the effects of reweighting to the other equation. Denoting $\widehat{Q}(\varphi)$ and $\widehat{P}(\psi)$ the new distributions, these two transformations lead to

$$\widehat{Q}(\bullet) \stackrel{\mathrm{d}}{=} \int \prod_{i=1}^{k-1} \mathrm{d}\widehat{P}_i(\psi_i) \, \delta \left[\bullet - 1 + \prod_i (1 - \psi_i) \right], \tag{A.2}$$

$$\widehat{P}(\bullet) \stackrel{\mathrm{d}}{=} \frac{1}{Z} \int \prod_{i=1}^{l_+} \mathrm{d}\widehat{Q}_i^+(\varphi_i^+) \prod_{i=1}^{l_-} \mathrm{d}\widehat{Q}_i^-(\varphi_i^-) \, \delta \left[\bullet - \frac{\prod_i \varphi_i^-}{\prod_i \varphi_i^+ + \prod_i \varphi_i^-} \right] \left(\prod_i \varphi_i^+ + \prod_i \varphi_i^- \right)^m , \tag{A.3}$$

where Z in the last equation is obtained by normalization.

One delicate issue in solving this kind of equation is how to represent faithfully the left-hand side of equation (A.3) by a sample of $\mathcal{N}'$ representative elements of $\widehat{P}$, because of the reweighting term $(\prod_i \varphi_i^+ + \prod_i \varphi_i^-)^m$. A possible solution [25] consists in first generating a larger number, say $5\mathcal{N}'$, of outgoing fields, storing them along with the associated weights, and then performing a resampling step to extract $\mathcal{N}'$ elements from this intermediate population. This approach has the advantage of having complexity independent of the distributions $\widehat{Q}_i$. Unhappily, if the weights are strongly concentrated on a small subset of the $5\mathcal{N}'$ fields, the resampled population will have many copies of these elements. This leads to a deterioration of the sample.

We adopted a different strategy whose running time depends on how strong is the reweighting. For $m \geq 0$, we generate fields sequentially and include them in the new population with probability proportional to the reweighting factor (divided by the

normalization factor 2^m). This procedure becomes slower when m grows, but it ensures that no repetitions appear in the new sample.

Solving the equations for $m < 0$ is instead much easier and no particular care is needed. For the sake of simplicity we have used the same algorithm as for $m \geq 0$ (which now produces many repetitions in the populations) and we have simply checked the validity of our results by changing the number and size of populations.

As explained in section 5.2 the cavity field distributions can have a positive weight on 'hard' fields, i.e. on fields that constrain a variable to take either value $+1$ or -1 in all solutions of the cluster. This corresponds to $\varphi = 0$, or $\psi \in \{0, 1\}$. This would show up into a positive fraction of the sample taking value, say, $\varphi = 0$, thus leading to an inefficient representation.

In order to circumvent this problem, we kept track explicitly of the weights on $\varphi = 0$ and $\psi \in \{0, 1\}$, in analogy with equations (60) and (62). This also allows us to locate more precisely the appearance of a positive fraction of hard fields in the distributions, as discussed in section 4.3. There is unfortunately one drawback to this approach. Consider equation (A.3) and suppose that all the distributions $\widehat{Q}$ of the right-hand sides are supported on $\varphi_i^{\pm} \in (0, 1]$. By definition the fields ψ thus generated are also strictly positive. However, the degrees $l_{\pm}$ are of order αk (i.e. around 40 for 4-SAT in the 1RSB regime). As a consequence, it may happen that the product of the l_{-} fields φ_i^{-} is smaller than the smallest number in the computer representation used (using 64 bits and the denormalized floating point notation this limit is roughly $\psi_{\min} \approx 5 \times 10^{-324}$). How should one treat such cases? We have adopted the solution of ignoring, that is not including it in the population, any number below $\psi_{\min}$. This solution is equivalent to saying that we are describing with a finite population of numbers the distribution $\widehat{P}(\psi)$ not on the domain $\psi \in (0, 1]$, but on the domain $\psi \in (\psi_{\min}, 1]$.

A different solution could be to convert all the numbers smaller than $\psi_{\min}$ to zero. We have tried this procedure, but it seems to be unstable, and to introduce systematic errors. In particular, one obtains a positive weight for $\psi = 0$, even for values of the parameters for which this is inconsistent.

The last point we would like to discuss is the problem of how to initialize the population dynamics algorithm. It is clear that an iterative procedure does, in general, lead to different solutions depending on the starting point of the iterations. For instance, the RS solution, where the distributions $P(h)$ in $\mathcal{P}_{(1)}$ are concentrated on a single value of h, is always a fixed point of the 1RSB equations. In the case $m = 1$, the interpretation in terms of tree reconstruction [26] leads to a clear prescription for this initialization, as explained in more detail in section 5.1. One can follow the same procedure for other values of m, namely initialize the populations with essentially only hard fields. This is crucial in particular for $k = 3$, where softer initial conditions lead to an unphysical fixed point, cf. appendix C.

Appendix B. Large k analysis: some technical details

In this appendix we provide the complete formulas for the dynamical transition regime of section 6.1. To leading order one can write a set of coupled equations for the average of $\widetilde{P}(\cdot)$, $\widetilde{Q}(\cdot)$ over the 1RSB order parameters $\mathcal{P}_{(1)}$, $\mathcal{Q}_{(1)}$. With a slight abuse of notation

we shall keep denoting by $\widetilde{P}$, $\widetilde{Q}$ such averages. In terms of this quantities we have

$$\Lambda(\delta, m) = -\log \left\{ \int \mathrm{d}\widetilde{P}(h) \left(\frac{1 + \tanh h}{2} \right)^m \right\}, \tag{B.1}$$

where the dependence on δ is through $\widetilde{P}$. Notice that $\Lambda(\delta, m = 0) = 0$ independently of $\widetilde{P}$. For $m = 1$ one can use the fact that, by symmetry $\int \mathrm{d}\widetilde{P}(h) \, (\tanh h) = 0$, to deduce $\Lambda(\delta, m = 1) = \log 2$.

For $m \neq 0, 1$ one has to determine the distributions $\widetilde{P}$ and $\widetilde{Q}$. It turns out that

$$\widetilde{Q}(u) = \left(1 - \frac{w}{2^k \log k} \right) \delta(u) + \frac{w}{2^k \log k} \widetilde{Q}'(u),$$
$$w \equiv 2^{2-m} \delta \int \mathrm{d}\widetilde{P}(h) \left(\frac{3 + \tanh h}{2} \right)^m. \tag{B.2}$$

The distributions $\widetilde{P}$, $\widetilde{Q}'$ are solutions of the coupled equations

$$\widetilde{Q}'(u) = \frac{2^{2-m}\delta}{w} \int \mathrm{d}\widetilde{P}(h) \left(\frac{3 + \tanh h}{2} \right)^m \delta \left(u - \frac{1}{2} \log(1 + \mathrm{e}^{-2h}) \right), \tag{B.3}$$

$$\widetilde{P}(h) = \frac{1}{Z} \mathbb{E}_{l_\pm} \int \prod_{i=1}^{l_+} \mathrm{d}\widetilde{Q}'(u_i^+) \int \prod_{i=1}^{l_-} \mathrm{d}\widetilde{Q}'(u_i^-) \, z_3(u_1^+, \ldots, u_{l_+}^+, u_1^-, \ldots, u_{l_-}^-)^m$$
$$\times \delta \left(h - \sum_{i=1}^{l_+} u_i^+ + \sum_{i=1}^{l_-} u_i^- \right), \tag{B.4}$$

where in the second equation Z is a normalizing factor and $l_\pm$ are two independent Poisson random variables of mean $w/2$.

Appendix C. Non-uniqueness of solutions of the 1RSB equations for $k = 3$

This appendix provides further details on the numerical solution of the 1RSB equations for $k = 3$. A difficulty that arises in this case is the presence, for some values of α and m, of at least two distinct non-trivial solutions of the 1RSB equations (this has been already noticed in [16] for $\alpha = 4.2$, and in [15] for the related coloring problem). As a consequence the initial conditions of the iterative resolution play an important role in selecting the fixed point that shall be reached.

One can justify the existence of multiple solutions as follows. As mentioned in the main text, the continuous dynamical transition at $\alpha_\mathrm{d} \approx 3.86$ corresponds to a local instability of the RS solution with respect to 1RSB perturbations. It is important to underline that this instability condition is independent of the value of m, that is at α_d a new solution of the 1RSB equations should grow continuously away from the RS one, for all values of m. This is illustrated in figure C.1, where the overlaps q_0 and q_1 meet at α_d for various values of m. By continuity these solutions do not contain hard fields in the neighborhood of α_d. In contrast it is known since [10] that another solution of the $m = 0$ equations, with a finite weight on hard fields, arises discontinuously at $\alpha \approx 3.92$.
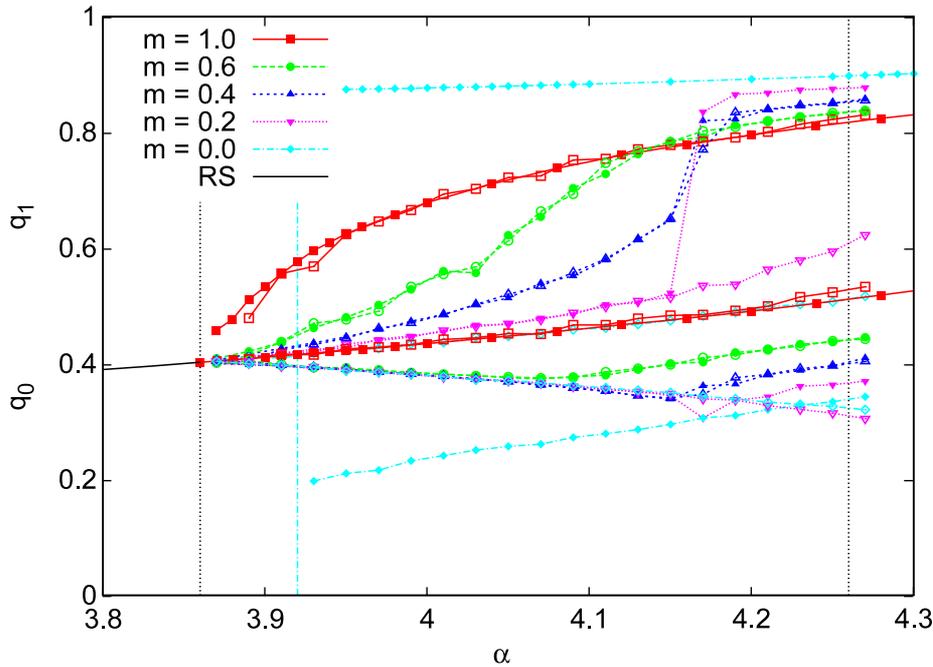
Figure C.1. Intra- and inter-state overlap, q_0 and q_1, for $k = 3$ and some values of the Parisi parameter m. Data below (resp. above) the RS line are for q_0 (resp. q_1). Full (resp. open) symbols refer to data measured while increasing (resp. decreasing) α.

For larger values of α these two solutions thus coexist[16]. A natural conjecture is that two solutions also coexist for $m \neq 0$. The iterative population dynamics algorithm converges to one of them depending on the initialization (more precisely, on the fraction of hard fields in the initial populations).

Our data suggest that the interval of α in which the two solutions coexist shrinks when m grows from 0. For instance, in figure C.1 one clearly see two branches for $m = 0.2$ at high enough values of α, whereas for $m = 0.6$ the two curves obtained by increasing and decreasing α at fixed m are superimposed within numerical precision.

It remains to understand which, if any, of these solutions is the correct one. In principle one should test their stability with respect to higher level of replica symmetry breaking [50]–[52]. However, it is an extremely demanding numerical task that we did not undertake. A simpler consistency argument can be invoked by computing the internal entropy of the pure states. This should be an increasing function of m. We can see in the curves of figure C.2 that this condition is not respected for all the values of α and m (full symbols refer to consistent solutions, while open symbols are for inconsistent ones). For values of α smaller than roughly 4.15 we are not able to find a consistent solution in the whole range of $m \in [0, 1]$ (a consistent solution exists only for m large enough). While for α roughly larger than 4.15 two solutions coexist at small values of m and the consistent

---

[16] Let us signal a peculiarity of the $m = 0$ 'soft' solution. It is easy to realize from equations (60) and (62) that the average of the distributions $P(h)$ and $Q(u)$ in this solution verify the RS equations. As a consequence its intra-overlap q_1 coincides with the RS overlap, its complexity vanishes and its internal entropy equals the RS one.
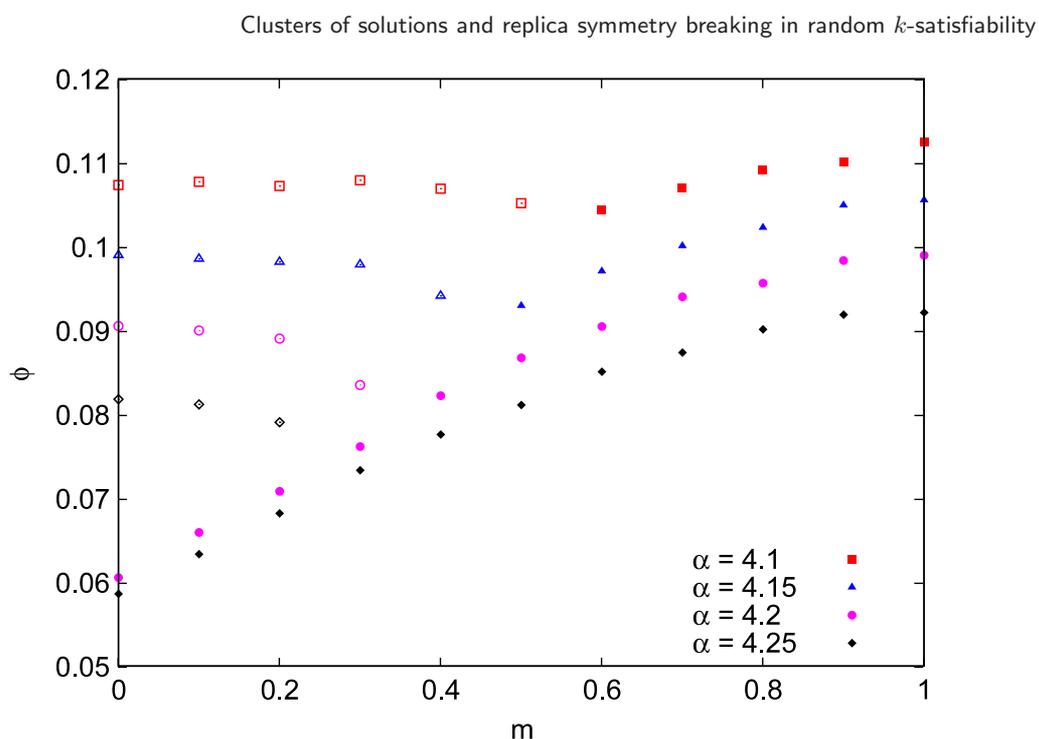
Figure C.2. The internal entropy should be a non-decreasing function of m if the solution is consistent. Filled (resp. empty) symbols refer to solutions with $\partial_m \phi > 0$ (resp. $\partial_m \phi < 0$), for $k = 3$.
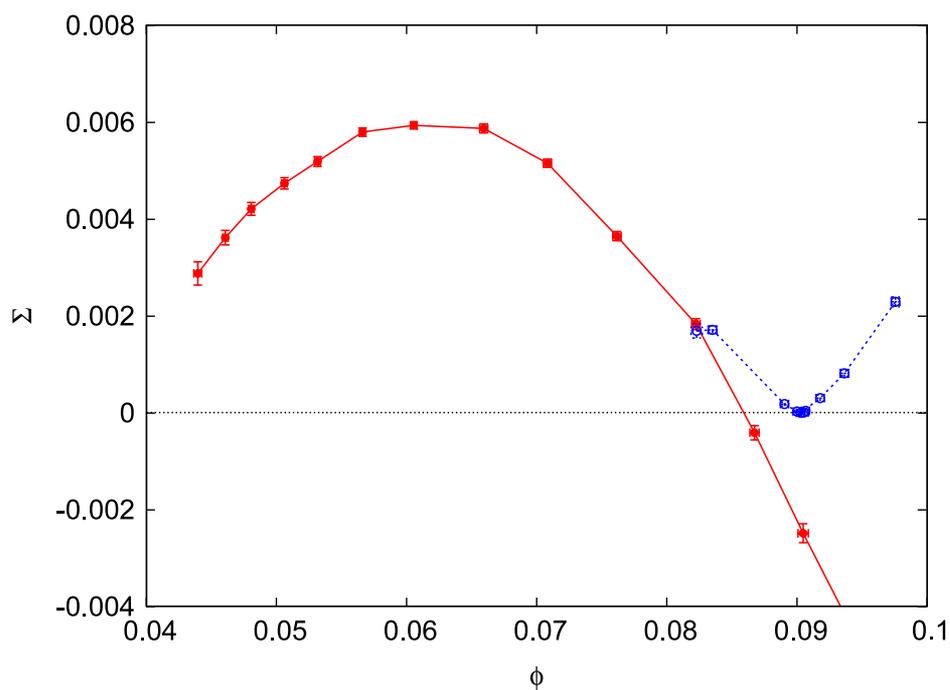

Figure C.3. The entropic complexity $\Sigma(\phi)$ for $k = 3$ and $\alpha = 4.2$. The two different branches correspond to the consistent (full line) and inconsistent solution (dashed line).

one is the one with more hard fields. We also notice that this inconsistency is accompanied by the decreasing of the inter-overlap q_0 with α: in other words, we empirically find that the quantities $\partial_m \phi$ and $\partial_\alpha q_0$ always have the same sign. This observation makes it easier to locate in figure C.1 consistent solutions (those with q_0 increasing with α). In order to make connection with previous studies where consistent and inconsistent solutions were found [9, 15, 34] we plot in figure C.3 the entropic complexity curve for $\alpha = 4.2$: the full (resp. dashed) curve corresponds to the consistent (resp. inconsistent) branch.

References

[1] Garey M R and Johnson D S, 1983 *Computers and Intractability, A Guide to the Theory of NP-Completeness* (San Francisco, CA: Freeman)
[2] Mitchell D, Selman B and Levesque H, 1992 *Proc. 10th Nat. Conf. Artif. Intell.* p 459
[3] Friedgut E, 1999 *J. Am. Math. Soc.* **12** 1017
[4] Franco J, 2001 *Theor. Comput. Sci.* **265** 147
[5] Dubois O, 2001 *Theor. Comput. Sci.* **265** 187
[6] Achlioptas D and Peres Y, 2004 *J. Am. Math. Soc.* **17** 947
[7] Fernandez de la Vega W, 2001 *Theor. Comput. Sci.* **265** 131
[8] Monasson R and Zecchina R, 1997 *Phys. Rev.* E **56** 1357
[9] Biroli G, Monasson R and Weigt M, 2000 *Eur. Phys. J.* B **14** 551
[10] Mézard M and Zecchina R, 2002 *Phys. Rev.* E **66** 056126
[11] Mézard M, Parisi G and Zecchina R, 2002 *Science* **297** 812
[12] Mertens S, Mézard M and Zecchina R, 2006 *Random Struct. Algorithms* **28** 340
[13] Mézard M, Parisi G and Virasoro M A, 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[14] Krzakala F, Montanari A, Ricci-Tersenghi F, Semerjian G and Zdeborová L, 2007 *Proc. Nat. Acad. Sci.* **104** 10318
[15] Zdeborová L and Krzakala F, 2007 *Phys. Rev.* E **76** 031131
[16] Zhou H, 2008 *Preprint* 0801.0205
[17] Dall'Asta L, Ramezanpour A and Zecchina R, 2008 *Preprint* 0801.2890
[18] Talagrand M, 2003 *Spin Glasses: A Challenge for Mathematicians* (Berlin: Springer)
[19] Kschischang F, Frey B J and Loeliger H-A, 2001 *IEEE Trans. Inf. Theory* **47** 498
[20] Georgii H O, 1988 *Gibbs Measures and Phase Transitions* (Berlin: De Gruyter)
[21] Parisi G, 2003 *Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter* (*Les Houches Lecture Notes, Session LXXVII*) ed J-L Barrat *et al* (Berlin: Springer) p 271
[22] Cugliandolo L F, 2003 *Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter* (*Les Houches Lecture Notes, Session LXXVII*) ed J-L Barrat *et al* (Berlin: Springer) p 367
[23] Monasson R, 1995 *Phys. Rev. Lett.* **75** 2847
[24] Ruelle D, 1987 *Commun. Math. Phys.* **108** 225
[25] Mézard M and Parisi G, 2001 *Eur. Phys. J.* B **20** 217
[26] Mézard M and Montanari A, 2006 *J. Stat. Phys.* **124** 1317
[27] Biroli G and Bouchaud J P, 2004 *J. Chem. Phys.* **121** 7347
[28] Martinelli F, Sinclair A and Weitz D, 2004 *Commun. Math. Phys.* **250** 301
[29] Cavagna A, Grigera T S and Verrocchio P, 2007 *Phys. Rev. Lett.* **98** 187801
[30] Franz S and Parisi G, 1995 *J. Physique* **5** 1401
[31] Berger E N, Kenyon C, Mossel E and Peres Y, 2005 *Probab. Theory Relat. Fields* **131** 311
[32] Montanari A and Semerjian G, 2006 *J. Stat. Phys.* **125** 23
[33] Mézard M, Palassini M and Rivoire O, 2005 *Phys. Rev. Lett.* **95** 200202
[34] Mézard M and Parisi G, 2003 *J. Stat. Phys.* **111** 1
[35] Daudé H, Mézard M, Mora T and Zecchina R, 2008 *Theor. Comput. Sci.* **393** 260
[36] Achlioptas D and Ricci-Tersenghi F, 2006 *Proc. 38th ACM Symp. on Theory of Computing* p 130
[37] Mora T and Zdeborova L, 2007 *Preprint* 0710.3804
[38] Mézard M, Ricci-Tersenghi F and Zecchina R, 2003 *J. Stat. Phys.* **111** 505
[39] Cocco S, Dubois O, Mandler J and Monasson R, 2003 *Phys. Rev. Lett.* **90** 047205
[40] Janson S, Luczak T and Rucinski A, 2000 *Random Graphs* (New York: Wiley)
[41] Mézard M and Montanari A, 2008 *Information, Physics, Computation: Probabilistic Approaches* available at http://www.stanford.edu/~montanar/BOOK/book.html
[42] Franz S and Leone M, 2003 *J. Stat. Phys.* **111** 535

[43] Panchenko D and Talagrand M, 2004 *Probab. Theory Relat. Fields* **130** 319
[44] Montanari A and Shah D, 2007 *Proc. XVIII Symp. Discr. Algorithms (New Orleans)*
[45] Abou-Chacra R, Thouless D J and Anderson P W, 1973 *J. Phys. C: Solid State Phys.* **6** 1734
[46] Semerjian G, 2008 *J. Stat. Phys.* **130** 251
[47] Montanari A and Semerjian G, 2006 *J. Stat. Phys.* **124** 103
[48] Pagnani A, Parisi G and Ratieville M, 2003 *Phys. Rev.* E **68** 046706
[49] Castellani T, Krzakala F and Ricci-Tersenghi F, 2005 *Eur. Phys. J.* B **47** 99
[50] Montanari A, Parisi G and Ricci-Tersenghi F, 2004 *J. Phys. A: Math. Gen.* **37** 2073
[51] Montanari A and Ricci-Tersenghi F, 2003 *Eur. Phys. J.* B **33** 339
[52] Krzakala F and Zdeborová L, 2008 *Eur. Phys. Lett.* **81** 57005
[53] Seitz S, Alava M and Orponen P, 2005 *J. Stat. Mech.* P06006
[54] Alava M, Ardelius J, Aurell E, Kaski P, Krishnamurthy S, Orponen P and Seitz S, 2007 *Preprint* 0711.4902
[55] Krzakala F and Kurchan J, 2007 *Phys. Rev.* E **76** 021122
[56] Maneva E, Mossel E and Wainwright M J, 2007 *J. Am. Chem. Soc.* **54** 1
[57] Pretti M, 2005 *J. Stat. Mech.* P11008
[58] Montanari A, Ricci-Tersenghi F and Semerjian G, 2007 *Proc. 45th Allerton Conf. on Comm. Control and Computing*
[59] Ricci-Tersenghi F and Semerjian G, 2008 in preparation