A simple one dimensional glassy Kac model

Andrea Montanari\textsuperscript{1} and Antoine Sinton\textsuperscript{2}

\textsuperscript{1} Departments of Electrical Engineering and Statistics, Stanford University, Stanford, CA 9305, USA
\textsuperscript{2} Laboratoire de Physique Théorique de l’Ecole Normale Supérieure, 24 rue Lhomond, F-75231, Paris Cedex 05, France
E-mail: montanari@stanford.edu and sinton@ens.fr

Received 7 May 2007
Accepted 11 June 2007
Published 1 August 2007

Online at stacks.iop.org/JSTAT/2007/P08004
doi:10.1088/1742-5468/2007/08/P08004

Abstract. We define a new family of random spin models with one-dimensional structure, finite-range multispin interactions, and bounded average degree (number of interactions in which each spin participates). Unfrustrated ground states can be described as solutions of a sparse, band diagonal linear system, thus allowing for efficient numerical analysis.

In the limit of infinite interaction range, we recover the so-called XORSAT (diluted $p$-spin) model that is known to undergo a random first-order phase transition as the average degree is increased. Here we investigate the most important consequences of a large but finite interaction range: (i) fluctuation-induced corrections to thermodynamic quantities; (ii) the need of an inhomogeneous (position-dependent) order parameter; (iii) the emergence of a finite mosaic length scale. In particular, we study the correlation length divergence at the (mean field) glass transition.

Keywords: disordered systems (theory), structural glasses (theory), exact results

ArXiv ePrint: 0705.0054
A simple one dimensional glassy Kac model

1. Introduction

A large class of disordered mean field spin models exhibit a behaviour that is reminiscent of the structural glass transition in fragile glasses [1]–[3]. As temperature is lowered, they undergo a ‘dynamical phase transition’ characterized by a diverging relaxation time at a critical temperature $T_d$. The reason for such a dynamical arrest can in turn be ascribed to ergodicity breaking: below $T_d$ the Boltzmann measure decomposes into an exponential number of pure states. While equilibration is fast within each state, it takes an exponentially large (in the system size) time to equilibrate across states.

Below $T_d$, the system can be meaningfully characterized through its complexity $\Sigma$, which gives the exponential growth rate of the number of pure states (i.e. the number of such states is about $e^{N\Sigma}$, $N$ being the size). The complexity decreases as temperature is further lowered, and vanishes linearly at a second (static) transition temperature $T_s$. This corresponds to an actual thermodynamic phase transition.

A strikingly similar scenario has been found to hold in a large array of random constraint satisfaction [4]–[6] problems of interest in theoretical computer science$^3$. The role of temperature is played here by the number of constraints per variable $\gamma$, while the Boltzmann distribution is replaced by the uniform measure over solutions of the problems. As the constraint density crosses a critical value $\gamma_d$, the set of solutions splits into ‘lumps’ analogous to pure states. Above a second threshold $\gamma_s$ the set of constraints becomes, with high probability, unsatisfiable.

$^3$ In a typical constraint satisfaction problem, one seeks an assignment of values to $L$ discrete variables in such a way to satisfy $M$ constraints.
In the last few years there has been a consistent effort in interpreting disordered mean field models as a genuine mean field theory for the structural glass transition. This is highly non-trivial since in any finite-dimensional model there cannot be coexistence of an exponentially large number of pure states. Imagine trying to select one such state through appropriate boundary conditions on a box of size \( \ell \). This will imply an energetic bias towards the selected state, which is of order \( \sigma \ell^\theta \), where \( 0 \leq \theta \leq d - 1 \) is a surface tension exponent. On the other hand, the entropic advantage of the other states is of order \( \Sigma \ell^d \), because of their number. Therefore, for \( \ell \gtrsim \ell_s \equiv (\beta \sigma / \Sigma)^{1/(d - \theta)} \), pure states are no longer stable.

According to the ‘mosaic state’ scenario, below \( T_d \) a typical configuration of the system can be described as a patchwork \([7]–[9]\). Each patch corresponds to the configuration being close to a particular pure state in a localized region whose length scale is \( \ell_s \). Since \( \Sigma \sim (T - T_s) \) at the static transition, the mosaic length scale diverges as \( \ell_s \sim (T - T_s)^{-\nu} \) with \( \nu = 1/(d - \theta) \).

While the mosaic scenario is appealing, its consistency and implications, as well as its precise meaning, are far from obvious. An important step forward was achieved in \([10]\) where a concrete ‘gedanken experiment’ was introduced to define \( \ell_s \). This length scale was interpreted in \([11]\) as a point-to-set correlation length, and its divergence was rigorously proved to be equivalent to a divergent correlation time. In \([12]\), \( \ell_s \) was actually shown to diverge at \( T_s \) in a class of disordered Kac models with continuous scalar spins.

Unhappily the models considered in \([12]\) can currently be treated only in the Kac limit, and through somewhat formal techniques such as the replica method. As a consequence, many interesting questions (such as the relevance of this limit for realistic interaction ranges, non-perturbative fluctuation effects or a precise definition of states) cannot be addressed in this context. The present paper aims at introducing a new class of models that share some features with the ones treated in \([12]\), while being tractable within alternative approaches (e.g. numerically).

We follow the route of generalizing one of the ensembles of random constraint satisfaction problems mentioned above, and referred to as \( k \)-XORSAT \([13,14]\). We shall require constraints to have finite range with respect to an underlying one-dimensional geometry. Our motivation is twofold: (i) because of its underlying linear structure, the \( k \)-XORSAT is very well understood. In particular a wealth of information regarding pure states and their geometry is accessible through rigorous techniques \([15]–[18]\); (ii) the ensembles of random constraint satisfaction problems studied within the computer science and statistical mechanics communities have so far lacked any geometrical structure (in physics terms, they are mean field models). This is, of course, a poor model of real world instances and it is surely instructive to explore alternative—structured—models.

Constraint satisfaction problems with finite interaction range were already considered in \([19]\), without, however, considering the interaction range as a parameter. Further, the most important questions that we shall consider in this paper were not studied there. Several papers \([20]–[22]\) investigated the behaviour of thermodynamic quantities and local order in Kac spin glasses. In particular, Kac spin glasses with bounded average degree and multispin interactions were introduced and studied in \([24]\). Finally, a one-dimensional Kac spin glass, with a different (continuous) phase transition, was recently studied numerically in \([23]\).
This paper is organized as follows. In section 2 we introduce our Kac-XORSAT model, and its variants, and discuss some of their most basic properties in section 3. We investigate thermodynamic quantities (in particular the ground state entropy) in section 4 and the correlation length divergence in section 5. Finally a discussion of our results is presented in section 6 and several technical details are contained in the appendices.

2. Definition of the model

Let us recall that an instance of the \( k \)-XORSAT problem is defined by an \( M \times L \) matrix binary \( \mathbb{H} \), with row weight \( k \) and a binary vector \( \mathbf{b} \) of length \( M \). Solving such an instance requires determining whether the linear system

\[
\mathbb{H} \mathbf{x} = \mathbf{b} \mod 2,
\]

admits a solution \( \mathbf{x} \in \{0, 1\}^L \). This question is equivalent to asking whether the ground state energy of a certain Ising spin model is zero or not. More precisely, let \( \{i_1(a), \ldots, i_k(a)\} \subseteq [L] \) denote the indices of the non-zero entries in the \( a \)th row of \( \mathbb{H} \), and \( J_a = (-1)^{b_a} \) (here \( a \in [M] \), and \([n] \equiv \{1, \ldots, n\}\)). The relevant spin model is defined by letting the energy of configuration \( \sigma \equiv (\sigma_1, \ldots, \sigma_L) \in \{+1, -1\}^L \) be

\[
E(\sigma) = \sum_{a=1}^{M} (1 - J_a \sigma_{i_1(a)} \cdots \sigma_{i_k(a)}).
\]

In the following we shall refer to a particular XORSAT instance as a ‘formula’ or a ‘sample’.

The random \( k \)-XORSAT (rXOR) ensemble is defined by letting \( \mathbb{H} \) be a uniformly random binary matrix (with dimensions \( M \times L \) and row weight \( k \)) and \( \mathbf{b} \) a uniformly random vector in \( \{0, 1\}^L \). It is also useful to consider the unfrustrated random \( k \)-XORSAT ensemble, defined by setting \( \mathbf{b} = 0 \) (the all 0’s vector) deterministically. Such an ensemble exhibits a particularly rich behaviour in the ‘thermodynamic’ limit \( L \to \infty, M \to \infty \) with \( \gamma = M/L \) kept fixed.

The Kac \( k \)-XORSAT (KacXOR) ensembles add to the above features a one-dimensional (or, in linear algebra terms, a band diagonal) structure. One such ensemble is characterized by the parameters introduced so far, namely \( k, L \) and \( \gamma \), plus an ‘interaction range’ \( R \). Unlike in the rXOR ensemble, \( \gamma \) is required to be in \([0, 1]\), although generalizations are not difficult. Further, the interaction range is an integer such that \( 2R + 1 \geq k \). Given such parameters, the matrix \( \mathbb{H} \) is sampled as follows. Rows of \( \mathbb{H} \) are indexed by a subset \( F \) of \([L]\): for each \( a = 1, \ldots, L, a \in F \) independently from the others with probability \( \gamma \). In particular, the number of rows of \( \mathbb{H}, M \), is a binomial random variable

\[
\mathbb{P}\{M\} = \left( \begin{array}{c} L \\ M \end{array} \right) \gamma^M (1 - \gamma)^{L-M}.
\]

As \( L \to \infty \), the number of rows is with high probability\(^5\), close to \( L\gamma \). For each \( a \in F \), the corresponding row in \( \mathbb{H} \) is sampled independently from the others by

\(^4\) The row weight is the number of non-vanishing entries in each row of the matrix.

\(^5\) Following the use in probability theory, we say that something happens with high probability (w.h.p.) if its probability approaches 1 in the thermodynamic limit.
A simple one dimensional glassy Kac model

Figure 1. Factor graph representation of a portion of a KacXOR formula with \( k = 3 \) and \( R = 3 \). Empty circles correspond to variables (columns of \( \mathbb{H} \)) and filled squares to equations (rows of \( \mathbb{H} \)).

Letting the indices of non-zero entries \((i_1(a), \ldots, i_k(a))\) be a uniformly random subset of \( \{a - R, \ldots, a + R\} \) (i.e. each of the \( \binom{2R + 1}{k} \) subsets has the same probability). We shall refer to \( \{a - R, \ldots, a + R\} \) as to the range of equation \( a \).

Finally, we let the entries of \( b \) be indexed by \( F \) as well, and iid random in \( \{0, 1\} \).

A XORSAT formula admits a natural representation as a factor graph \( G \). This is a bipartite graph including one ‘parity check node’ for each row in \( \mathbb{H} \) (i.e. for each equation in the linear system) and one ‘variable node’ for each column (i.e. for each variable in the linear system). A parity check and a variable node are connected by an edge if the corresponding entry of \( \mathbb{H} \) is non-vanishing. An example of such a representation is presented in figure 1.

There is still one point of the above definition to be clarified. When \( a \leq R \) or \( a \geq L - R \), the range for equation \( a \) is not included in the sets of variable indices, and it might be that \( i_l(a) \leq 0 \) or \( i_l(a) > L \). We shall consider two types of boundary conditions. With periodic boundary conditions, variable indices are interpreted modulo \( L \). Therefore, if for some index we have \( i_l(a) > L \), this is identified with \( i_l(a) - L \) while, if \( i_l(a) \leq 0 \), this is identified with \( i_l(a) + L \).

In the case of fixed boundary condition we will let the set potential indices of row of \( \mathbb{H} \) be \( \{-R + 1, \ldots, L + R\} \). Namely \( F \) includes each \( a \) in this set independently with probability \( \gamma \). To define a fixed boundary condition, we shall fix a doubly infinite reference configuration \( \underline{x}^{(0)} = \{x_i^{(0)} : i \in \mathbb{Z}\} \). If in building row \( a \) we get an index \( i_l(a) \not\in [L] \), the corresponding non-zero entry is not included in \( \mathbb{H} \), but rather the value \( x_{i_l(a)}^{(0)} \) is added to \( b_a \). This corresponds to fixing \( x = x^{(0)} \) ‘outside’ \( \{1, \ldots, L\} \). Finally, we shall agree that, whenever considering the unfrustrated ensemble, the reference configuration will be the all 0’s sequence \( \underline{x}^{(0)} = \underline{0} \).

6 Notice that the boundary condition depends on the reference configuration \( \underline{x}^{(0)} \) only through \( x^{(0)}_{-2R+1}, \ldots, x^{(0)}_0 \) and \( x^{(0)}_{L+1}, \ldots, x^{(0)}_{L+2R} \).
3. Frustrated versus unfrustrated ensemble

The most important feature of the rXOR ensemble in the large size limit is that it undergoes a SAT–UNSAT phase transition at a well defined constraint density $\gamma_s(k)$. More precisely, a random XORSAT formula of the type (1) is solvable (SAT) with high probability if $\gamma < \gamma_s(k)$, while it is not solvable (UNSAT) for $\gamma > \gamma_s(k)$ [13]–[16].

It is a convenient feature of XORSAT that this phase transition can be studied by considering uniquely the unfrustrated ensemble. This simplification can be explained through the well-known identity

$$P\{H \vec{x} = \vec{b} \text{ is SAT}\} = E\{2^{L-M}/Z(H)\}, \quad (4)$$

where $Z(H)$ denotes the number of solutions of the homogeneous linear system $H \vec{x} = \vec{0} \mod 2$. The identity holds irrespective of the distribution of $H$ provided the right-hand side of equation (1), i.e. the vector $\vec{b}$, is uniformly random. In order to prove it, it is sufficient to notice that $H \vec{x} = \vec{b}$ is SAT if and only if $\vec{b}$ is in the image of $H$. Let $\ker(H) \equiv \{\vec{x} \text{ s.t. } H \vec{x} = \vec{0} \mod 2\}$ denote the null space of $H$. Since the dimension of the image of $H$ is $\text{rank}(H) = L - \dim \ker(H)$, this happens with probability $2^{L-\dim \ker(H)}/2^M$. On the other hand, $Z(H) = 2^{\dim \ker(H)}$. Equation (4) follows by taking expectation with respect to $H$.

Within the rXOR ensemble, for $\gamma < \gamma_s(k)$, $Z(H)$ is tightly concentrated around $2^{L-M}$, implying $P\{H \vec{x} = \vec{b} \text{ is SAT}\} \approx 1$. Vice versa for $\gamma > \gamma_s(k)$, typically $Z(H) \approx 2^{L\phi(\gamma)}$ (here $\phi(\gamma)$ denotes equality to leading exponential order), with $\phi(\gamma) > 1 - \gamma$ and therefore the formula is SAT with exponentially small probability.

Furthermore, as long as the non-homogeneous solution has at least one solution, its number of solutions is independent of $\vec{b}$, and is given by $Z(H)$. Even more, the set of solutions is an affine space obtained by translating the linear space of solutions of the homogeneous system. In other words, conditional to the problem being solvable (which happens with high probability for $\gamma < \gamma_s(k)$) the frustrated and unfrustrated ensemble are essentially equivalent.

An important novelty within the KacXOR ensemble is that the linear system (1) is always UNSAT with high probability if we let $L \to \infty$ with $\gamma, R$ fixed. More precisely, we expect that

$$P\{H \vec{x} = \vec{b} \text{ is SAT}\} \approx e^{-L\Lambda(\gamma, R)}, \quad (5)$$

for some $\Lambda(\gamma, R)$ strictly positive and non-decreasing in $\gamma$. This phenomenon was already observed in [19] for a related model. The basic reason for this behaviour is that small subsets of nearby rows of $H$ have a fair chance of being linearly dependent. If this is the case, the corresponding linear subsystem is unsolvable with finite probability. In the large $L$ limit, the expected number of such subsystems is of order $L$, and the probability that none is present is exponentially small, thus leading to the above behaviour.

It is not difficult to prove the above statement, and indeed to prove lower bounds on the rate $\Lambda(\gamma, R)$ by combinatorial techniques. The basic idea is to select a particular type of substructure that leads to unsatisfiability and estimate the probability that no such substructure is present. The simple such substructure is obtained when two lines of $H$
coincide but the corresponding entries of $h$ do not. Using the Janson inequality (see [25] for an introduction to this tool), one can prove that

$$\Lambda(\gamma, R) \geq K_1 \gamma^2 - K_2 \gamma^3 (1 - K_0 \gamma^2)^{-1},$$

where

$$K_0 \equiv \left(\frac{2R + 1 - 1}{k}\right) \left(\frac{2R + 1}{k}\right)^{-2}, \quad K_1 \equiv \frac{1}{2(2R + 1)^2} \sum_{p=1}^{2R+1-k} \left(\frac{2R + 1 - p}{k}\right),$$

$$K_2 \equiv \frac{3\gamma^3}{8(2R + 1)^3} \sum_{p=2}^{2R+1-k} (p - 1) \left(\frac{2R + 1 - p}{k}\right).$$

Such a lower bound is easily seen to be strictly positive for $\gamma$ small enough. We refer to appendix A for a derivation of this formula.

Notice that the lower bound in equation (6) vanishes as $1/R^k$ when $R \to \infty$. We expect the same behaviour to hold for the actual exponent as long as $\gamma$ is below the (mean field) satisfiability threshold $\gamma_s(k)$. Explicitly

$$\Lambda(\gamma, R) = \Lambda_1(\gamma)/R^k + O(1/R^{k+1}),$$

where $\Lambda_1(\gamma) \uparrow +\infty$ as $\gamma \uparrow \gamma_s(k)$.

In the following we shall avoid dealing with the above phenomenon by focusing directly on the unfrustrated ensemble: this will enable us to use efficient linear algebra techniques for numerical computations. There are several justifications for doing this:

(1) The two ensembles become equivalent in the Kac limit, which is our main concern here.

(2) We are interested in the long distance properties of the model, rather than in the effect of small substructures. We think that the two decouple for large $R$.

(3) Even if the frustrated ensemble is, with high probability, unsatisfiable, one can always consider ‘almost satisfying’ configurations. Equivalently, one can study the Boltzmann measure for the energy (2) at a small non-vanishing temperature $T$. We expect the effect of small frustrated substructures on the thermodynamics to be small, and indeed vanishing as $R^{-k}$ for large $R$.

In this perspective, we shall introduce an improved ensemble which reduces the effect of small substructures, while keeping the large $R$ behaviour unchanged. This is particularly convenient in numerical simulations.

Ideally, one would like to consider a uniform ensemble conditioned on some class of substructures being absent. In practice it can be exceedingly difficult to sample matrices $H$ from such a conditional ensemble. We shall define the improved ensemble by the following sequential procedure. First, generate the random set $F \subseteq [L]$ by letting $i \in F$ independently for each $i = 1, \ldots, L$ with probability $\gamma$. The set $F$ will index lines of $H$ as above. Then we choose a uniformly random ordering $(i(1), \ldots, i(M))$ of the elements of $F$, and generate the corresponding lines of $H$ following such an order. For each $t = 1, \ldots, M$ we try to generate the line of $H$ indexed by $i(t)$ by drawing its $k$ non-zero elements uniformly at random in $\{i(t) - R, \ldots, i(t) + R\}$. If the newly generated line has $k - 1$ or $k$ non-vanishing entries in common with one of the previously generated lines...
A simple one dimensional glassy Kac model

Figure 2. Ground state entropy density in the thermodynamic limit $\phi_R(\gamma) = \lim_{L \to \infty} \phi_{L,R}(\gamma)$, cf equation (10), for the standard and improved ensembles. Here $k = 3$ and $\gamma = 0.4$. The horizontal line marks the $R \to \infty$ limit $\phi(\gamma) = 0.6$.

\{i(1), \ldots, i(t - 1)\}, we reject it and re-sample it. We repeat the trial–rejection step for at most 100 times. If no valid line is generate within this round, the whole system generated so far is rejected and the procedure is re-initiated from scratch.

We shall refer to the first ensemble introduced above as the *standard*, whenever it will be necessary to distinguish it from the *improved* ensemble. In figure 2 we compare the $R \to \infty$ behaviour of the ground state entropy for the two ensembles. Although the limits clearly coincide, the improved ensemble is close to it even for $R = 5$.

4. Ground state entropy

The simplest thermodynamic quantity that is relevant for the study of the unfrustrated KacXOR problem is the ground state entropy, i.e. the logarithm of the number of solutions of the linear system. Let us denote by $Z(\mathbb{H})$ the number of solutions of the linear system (1) for a random binary matrix $\mathbb{H}$. Then the average entropy density is defined as

$$\phi_{L,R}(\gamma) = \frac{1}{L} \mathbb{E} \log_2 Z(\mathbb{H}).$$

(10)

In order to compare analytical predictions and numerical data it will be convenient to define the ‘subtracted’ entropy density \( \hat{\phi}_{L,R}(\gamma) \equiv \phi_{L,R}(\gamma) - \phi_{\text{naive}}(\gamma) \), where $\phi_{\text{naive}}(\gamma) = 1 - \gamma$. Notice that $\phi_{\text{naive}}(\gamma)$ is the naive prediction that would be obtained by assuming the lines of $\mathbb{H}$ to be linearly independent.

Given a matrix $\mathbb{H}$, the corresponding number of solutions takes the form of a partition function

$$Z(\mathbb{H}) = \sum_{x} \prod_{a=1}^{L} \psi_a(x_{a-R}, \ldots, x_{a+R}),$$

(11)
A simple one dimensional glassy Kac model

where (denoting by $\oplus$ the sum modulo 2)

$$\psi_a(x_{a-R}, \ldots, x_{a+R}) = \begin{cases} 1 & \text{if } a \in F, \\ 0 & \text{if } a \notin F. \end{cases}$$

Due to the finite interaction range $R$, $Z = Z(\mathbb{H})$ can be computed through a transfer matrix algorithm which recursively computes left and right partition functions, respectively $Z_{\rightarrow}$ and $Z_{\leftarrow}$. These are indexed by $\vec{z} = (z_1, \ldots, z_{2R}) \in \{0, 1\}^{2R}$ and defined as

$$Z_{\rightarrow}(\vec{z}) \equiv \sum_{x_1, \ldots, x_i} \prod_{a=1}^{i-R} \psi_a(x_{a-R}, \ldots, a+R),$$

$$Z_{\leftarrow}(\vec{z}) \equiv \sum_{x_1, \ldots, x_i} \prod_{a=i+R}^{L} \psi_a(x_{a-R}, \ldots, a+R),$$

where we used the notation $x_{j+2R} = (x_{j+1}, \ldots, x_{j+2R})$. A recursion naturally follows:

$$Z_{\rightarrow}(z_1, \ldots, z_{2R}) = \sum_{z_0 \in \{0, 1\}} \psi_{i-R+1}(z_0, z_1, \ldots, z_{2R}) Z_{\rightarrow}(z_0, \ldots, z_{2R-1}),$$

which together with the analogous recursion for $Z_{\leftarrow}$ gives

$$\phi_R(\gamma) \equiv \lim_{L \to \infty} \phi_{L,R}(\gamma).$$

Next, we let $R \to \infty$. In appendix C we will show that $\phi_R(\gamma)$ can be expanded for large $R$ as follows:

$$\phi_R(\gamma) = \phi^{(0)}(\gamma) + \frac{1}{2R+1} \phi^{(1)}(\gamma) + O \left( \frac{1}{R^2} \right).$$

The leading term gives the mean field limit and coincides with the ground state entropy density within the rXOR ensemble [15, 16]. It can be expressed in the form

$$\phi^{(0)}(\gamma; \varphi) = \max_{\varphi \in [0,1]} \phi^{(0)}(\gamma; \varphi),$$

where

$$\phi^{(0)}(\gamma; \varphi) = -\gamma(1 - \varphi^k) + k \gamma \varphi^{k-1}(1 - \varphi) + e^{-k \gamma \varphi^{k-1}}.$$
A simple one dimensional glassy Kac model

Figure 3. Subtracted entropy density $\hat{\phi}_{L,R}(\gamma) = \phi_{L,R}(\gamma) - 1 + \gamma$ for various values of $L$ and $R = 25$ constant. We also plot the result of an $L \to \infty$ extrapolation and the analytical mean field prediction $\phi^{(0)}(\gamma) - 1 + \gamma$ (continuous line).

It is easy to show that the max is achieved for a value of the order parameter $\varphi$ that satisfies the equation $\varphi = 1 - \exp\{-k\varphi^{k-1}\}$. For $\gamma < \gamma_s(k)$, the maximum is at $\varphi = 0$, yielding $\phi^{(0)}(\gamma) = 1 - \gamma$. In other words, the rank of $H$ is smaller than the maximum possible value by a fraction of order $1/R$. For $\gamma \geq \gamma_s(k)$, the maximum is at $\varphi = \varphi^* > 0$ strictly, and $\phi^{(0)}(\gamma) > 1 - \gamma$: the rank of $H$ remains strictly smaller than its maximum possible value even as $R \to \infty$. For instance we have $\gamma_s(k) \approx 0.917935$ for $k = 3$.

The first-order contribution $\phi^{(1)}(\gamma)$ is related to fluctuation around the saddle point in an appropriate path integral representation of equation (11). Its expression is given in appendix C.

In figure 3 we plot the numerical estimates for the subtracted entropy density $\hat{\phi}_{L,R}(\gamma)$, as obtained with our transfer matrix algorithm for $R = 25$ and several system sizes. Data points are the result of averaging over 1000 realizations of $H$ with $k = 3$. The same statistics and value of $k$ will be used in the other numerical experiments below: we shall omit mentioning it again. Further, unless otherwise stated, we will keep using the improved ensemble. We also show the result of an $1/L$ extrapolation to $L = \infty$. The control of the thermodynamic limit is quite good (although corrections at moderate values of $L$ are large). It is clear that the $L = \infty$ extrapolation is not compatible with the mean field prediction, and that $1/R$ corrections must be taken into account.

Figure 4 shows the result of such an $L \to \infty$ extrapolation for several values of $R$. The data seem to approach the mean field prediction $\phi^{(0)}(\gamma) - 1 + \gamma$ as $R \to \infty$, although the approach is rather slow.

In order to better study the large-$R$ limit, for four values of $\gamma$ we computed the ground state entropy for a wide range of $R$. The result is compared in figure 5 with the asymptotic expression (17). In this case we used the standard ensemble which presents larger $1/R$ corrections (computing the first-order correction $\phi^{(1)}$ within the improved
A simple one dimensional glassy Kac model

**Figure 4.** Subtracted entropy density in the thermodynamic limit: \( \hat{\phi}_R(\gamma) = \phi_R(\gamma) - 1 + \gamma \) for various values of \( R \), together with the mean field prediction \( \phi(0)(\gamma) - 1 + \gamma \) (continuous line).

**Figure 5.** Ground state entropy density \( \phi_R(\gamma) \), extrapolated to the thermodynamic limit, versus the inverse interaction range \( 1/(2R + 1) \) for various values of \( \gamma \). Straight lines correspond to the analytic prediction, cf equation (17).

The second limit we shall consider is \( L, R \to \infty \) with \( \ell \equiv L/R \) kept fixed. We thus define

\[
\phi^*_\ell(\gamma) \equiv \lim_{R \to \infty} \phi_{R\ell,R}(\gamma).
\]  

This is the mean field limit for a system of finite size. The limit can be computed exactly by maximizing an appropriate action functional over a position-dependent order parameter.
More precisely we have \( \phi^*_t(\gamma) = \max_\varphi A[\varphi] \), where \( \varphi : [0, \ell] \to \mathbb{R} \) is the order parameter and

\[
A[\varphi] = \frac{1}{\ell} \int_0^\ell \left\{ \gamma - k \gamma \varphi(z)^{k-1} + (k-1)\gamma \varphi(z) - \exp \left[ -\frac{k \gamma}{2} \int_{-1}^1 \varphi(z + u)^{k-1} du \right] \right\} \, dz. \tag{20}
\]

By differentiating with respect to \( \varphi \), we obtain the mean field equation

\[
\varphi(z) = 1 - \frac{1}{2} \int_{-1}^{z+1} \exp \left\{ -\frac{k \gamma}{2} \int_{-1}^{u+1} \varphi(z + u + v)^{k-1} dv \right\} \, du. \tag{21}
\]

We refer to appendix C for a derivation of these formulae and limit ourselves to discuss their consequences here.

In the case of a homogeneous order parameter \( \varphi(z) = \varphi \) independent of \( z \), equation (21) is satisfied if \( \varphi \) is a solution of the standard mean field equation, \( \varphi = 1 - \exp \{-k \gamma \varphi^{k-1}\} \). The action (20) then reduces to the mean field free energy equation (18).

In the general case the order parameter \( \varphi(z) \) has a simple interpretation. Consider the linear system \( \mathbb{H} x = \mathbf{0} \mod 2 \) and let \( i \in \{1, \ldots, L\} \). Then, one of the following must happen: either \( x_i = 0 \) in all of the solutions; or \( x_i = 0 \) in half of the solutions and \( x_i = 1 \) in the other half. We shall call \( x_i \) (or, sometimes, \( i \)) a frozen variable in the first case, and a free variable in the second one. Given \( z \in [0, \ell] \), the number of frozen variables with \( i \in [Rz, R(z + dz)] \) in a typical random linear system from our ensemble is about \( R \varphi(z) \, dz \). Equivalently, the probability for \( x_i, i = [Rz] \), to be frozen converges to \( \varphi(z) \).

We shall come back to this interpretation in the next section, while using it here to derive the appropriate boundary conditions for equation (21). If the linear system is defined with periodic boundary conditions, then we have to use periodic boundary conditions in equation (21) as well, namely \( \varphi(z + \ell) = \varphi(z) \). If, on the other hand, we adopt fixed boundary conditions with respect to the reference solution \( \underline{x}^{(0)} = \mathbf{0} \), then we have to impose \( \varphi(z) = 1 \) for \( z \leq 0 \) and \( z \geq \ell \) in equation (21). As a consequence, the homogeneous solution is no longer relevant in this case.

Once the boundary conditions have been established, equation (21) can be solved numerically by iteration (after discretizing \( z \) on a sufficiently fine mesh). In the regime in which multiple fixed points exist, the relevant one is obtained by maximizing the action.

The result of such a computation is compared in figure 6 with the outcome of numerical simulations. The agreement is good already at moderate interaction ranges. The main effect of a finite size is a decrease in the number of solutions due to the fact that variables close to the boundary are more highly constrained (and thus more likely to be frozen). This effect is accurately reproduced by the analytical calculation.

5. Point-to-set correlation function

As we have seen in the previous section, the thermodynamic behaviour of the KacXOR ensemble at finite \( R \) carries several traces of the mean field limit. Here we want to investigate some structural features of the uniform measure over solutions of the linear system:

\[
\mu(x) = \frac{1}{\mathcal{Z}} \mathbb{I} \left( \mathbb{H} x = \mathbf{0} \right) \equiv \frac{1}{\mathcal{Z}} \prod_a \psi_a(x_{a-R}, \ldots, x_{a+R}). \tag{22}
\]
Figure 6. Subtracted entropy density \( \hat{\phi}_{L,R}(\gamma) = \phi_{L,R}(\gamma) - 1 + \gamma \) as a function of \( \gamma \) for several at \( \ell = L/R = 50 \) fixed. The continuous line corresponds to the analytical prediction \( \phi^*_L(\gamma) - 1 + \gamma \) in the \( R \to \infty \) limit.

In particular, we want to understand whether the mean field ergodicity breaking transition shows up in the long range correlations of this measure, as predicted within the mosaic state scenario. Since the relevant length scale is predicted to diverge, within this scenario, as \( \gamma \uparrow \gamma_s \), we will focus on the \( \gamma < \gamma_s \).

It is expected that the long range order emerging at a glass transition cannot be probed through ordinary point-to-point correlation functions, and that point-to-set correlation functions have to be used instead \cite{11}. These can be defined through the following ‘experiment’ \cite{10} (we refer here to the one-dimensional case we are studying). Consider a large sample \( L \gg R \), let \( i \) be a node in its bulk: \( i \gg R, L - i \gg R \), and \( x^* \) a ‘reference’ configuration sampled from the measure \( \mu(\cdot) \). Then fix some \( 1 \leq \tilde{L} \ll L \) and consider a second configuration that is forced to coincide with \( x^* \) on sites \( j \) with \( |j - i| > \tilde{L} \), and free otherwise, and compute the probability that \( x_i \neq x^*_i \). The expectation of this probability with respect to \( x^* \) and the sample realization yields the desired point-to-set correlation.

In order to express formally the same definition, denote \( \mu_A(\cdot) \) the marginal distribution of \( \underline{x}_A \) for a subset \( A \) of the variable nodes, and \( x^*_B \) distributed according to \( (22) \). Analogously, let \( \mu_{A|B}(\cdot|\underline{x}_B) \) be the conditional distribution of \( \underline{x}_A \) given the assignment \( \underline{x}_B \) of the variables in \( B \). Finally, we let \( \mathcal{B}(i, \tilde{L}) = \{ j : |i - j| \leq \tilde{L} \} \) be the box of size \( 2\tilde{L} + 1 \) around \( i \), and \( \overline{\mathcal{B}}(i, \tilde{L}) \) its complement. We then define the point-to-set correlation function as

\[
\tilde{G}(\tilde{L}, R, \gamma) \equiv \mathbb{E}\{1 - 2\mu_i[\overline{\mathcal{B}}(i, \tilde{L})](x_i \neq x^*_i | \underline{x}_B^{\mathcal{B}(i, \tilde{L})})\}.
\]

Here the thermodynamic limit \( L \to \infty \) is assumed to be taken at the outset, \( \mathbb{E} \) denotes expectation both with respect to the matrix \( \underline{H} \) and the reference configuration \( x^* \) (distributed according to \( \mu \)), and the redefinition \( 1 - 2(\cdot \cdot \cdot) \) is for future convenience.

The linear structure of our problem implies two simplifications. First, the conditional probability appearing in equation (23) is indeed independent of \( x^* \) (that can be ‘gauged
A simple one dimensional glassy Kac model

away'). Therefore we can fix $x^* = 0$ and eliminate the expectation over the reference configuration $x^*$. The resulting conditional measure is just the distribution of a system with fixed boundary conditions $0$ as discussed in section 4. This implies a second simplification (already noticed above). The conditional probability $\mu_{i=0}(x_i \neq 0 | x_i^* = 0)$ can take the value 1/2 (if $x_i$ is ‘free’) or 0 (if it is ‘frozen’). We thus get

$$\tilde{G}(\tilde{L}, R, \gamma) = \mathbb{P}_L \{ x_{\tilde{L}+1} \text{ is free} \}.$$ 

Here $\mathbb{P}_L$ denotes probability with respect to a matrix $H$ with $2\tilde{L} + 1$ columns and fixed 0 boundary conditions. In fact it is interesting to generalize the above definition and consider the correlation between any point inside the box of size $2\tilde{L}$ and its boundary

$$\tilde{G}(n; \tilde{L}, R, \gamma) \equiv \mathbb{P}_L \{ x_{\tilde{L}+n} \text{ is free} \}.$$ 

The original definition is recovered for $n = 0$.

We expect $\tilde{G}(n; \tilde{L}, R, \gamma)$ to be close to 1 when $n$ approaches the boundaries of the box (i.e. $n \approx \tilde{L}$ or $n \approx -\tilde{L}$) and to decrease in the interior. If the box is large enough, it will approach its thermodynamic value, independent of the boundary condition, near the centre (for $n \approx 0$). In figure 7 we show the outcomes of a numerical calculation of $\tilde{G}(n; \tilde{L}, R, \gamma)$ for several values of its parameters.

We are particularly interested in the mean field limit. This is obtained by defining

$$G(z; \ell, \gamma) \equiv \lim_{R \to \infty} \tilde{G}(n = Rz; \tilde{L} = R\ell, R, \gamma),$$

that is, by measuring lengths in terms of the interaction range and letting $R \to \infty$. In agreement with the interpretation of section 4, we expect $G(z; \ell, \gamma) = \varphi(z)$, where $\varphi(z)$ solves equation (21) with boundary condition $\varphi(z) = 1$ for $z \leq -\ell$, and for $z \geq \ell$. The comparison with numerical data in figure 7 is satisfactory although the convergence to the $R \to \infty$ limit gets slower and slower as $\gamma_s(3) \approx 0.917935$ is approached.

The point-to-set correlation function can be used to define a correlation length, namely the smallest box size such that the correlation is below a pre-established constant $\varepsilon$. Here we will choose $\varepsilon = 1/2$. In formulae

$$\ell_s(\gamma, R) = \min \{ \ell : \tilde{G}(\tilde{L} = R\ell, R, \gamma) \leq 1/2 \}.$$ 

An analytical prediction in the $R \to \infty$ limit can be obtained by solving equation (21) with boundary conditions $\varphi(z) = 1$ for $z \not\in [-\ell, \ell]$. The resulting length can be shown to diverge at $\gamma_s$ as $\ell_s(\gamma, R = \infty) \sim (\gamma_s - \gamma)^{-1}$, in agreement with the mosaic picture (indeed $\Sigma(\gamma) \sim (\gamma_s - \gamma)$ close to the transition).

In figure 8 we compare this prediction with the estimates from numerical simulations at finite $R$. These two are clearly consistent, although the convergence is rather slow in the critical regime.

---

8 Any strictly positive constant below the Edwards–Anderson parameter (in this case given by the largest solution of $\varphi = 1 - \exp\{-k\gamma\varphi^{k-1}\}$) should provide an equivalent definition.
Figure 7. Correlation $\bar{G}(n; \tilde{L}, R, \gamma)$ between the boundary of a box of size $2\tilde{L} + 1 = 2\ell R + 1$ and a point in its interior (at distance $n = zR$ from the centre). In the right frames: blow-up of the region near the boundary. The continuous line (partially hidden by data points) corresponds to the analytic prediction obtained by solving equation (21).
6. Discussion

We defined a simple ensemble of constraint satisfaction problems (more precisely, an ensemble of linear problems over integers modulo 2), with one-dimensional Kac structure. The model is exactly soluble for infinite interaction range $R \to \infty$ and exhibits in this limit a glassy phase with an exponential number of pure states and a SAT–UNSAT transition.

Mean field theory (as interpreted within the mosaic picture) seems to describe the behaviour of the system at moderately large $R$. Indeed we were able to get quantitative predictions by taking into account the principal modifications of naive mean field theory, namely a position-dependent order parameter, and $1/R$ corrections. In particular we checked for the first time the divergence of the mosaic length scale in a concrete model, by comparing the result of a controlled approximation (large $R$ limit) with exact numerical calculations.

We think the KacXOR model can be a useful playground for many ideas developed in the physics of glasses. Among several interesting research directions, one might consider:
(i) studying the frustrated ensemble (corresponding to an inhomogeneous linear system);
(ii) introducing a non-vanishing temperature and studying the corresponding Boltzmann distribution;
(iii) studying the behaviour of Glauber dynamics, and in particular the relation between relaxation time and mosaic length scale.

On a different theme, ensembles of random constraint satisfaction problems have been recurrently used to test heuristic algorithms [13]. Such tests have limited scope because in practical applications instances are often structured. It might be insightful therefore to consider ensembles with some tunable ‘structure parameter’, such as the interaction range $R$ in the present model.
Appendix A. Counting small substructures

Consider the random linear system $\mathbb{H}x = b$ defined in section 2. If two lines $i, j \in F$ in $\mathbb{H}$ are equal, while the corresponding entries in $b$ (namely $b_i$ and $b_j$) are different, then the system has no solution. We call such a pair $(i, j)$ a ‘bad pair,’ and will write $B_{ij} = 1$ if $(i, j)$ is bad, and $B_{ij} = 0$ otherwise. Therefore

$$\mathbb{P}\{\mathbb{H}x = b \text{ is SAT}\} \leq \mathbb{P}\{\cap_{i,j}[B_{ij} = 0]\},$$

where the intersection ranges over $i, j$ such that $i < j \leq i + 2R + 1 - k$. Let $B = \sum_{(ij)} B_{ij}$ be the number of bad pairs. In order to bound the right-hand side above, we use Janson’s inequality [25], which implies

$$\mathbb{P}\{\mathbb{H}x = b \text{ is SAT}\} \leq \exp\{-E[B] + \Delta / 2(1 - \epsilon)\}.$$  

Here

$$\epsilon = \sup_{(ij)} E[B_{ij}], \quad \Delta = \sum_{(ij) \sim (lm)} E[B_{ij}B_{lm}],$$

where the sum over $(ij) \sim (lm)$ runs over all the couples of distinct pairs $(ij)$ and $(lm)$ such that $B_{ij}$ and $B_{lm}$ are not independent.

It is easy to realize that both $E[B]$ and $\Delta$ are of order $\Theta(L)$ since they are sums of $\Theta(L)$ positive terms. Since we are only interested in the coefficient of the order $L$ term, we shall always consider pairs $(ij)$ in the bulk. Then we have

$$E[B_{ij}] = \frac{\gamma^2}{2^{(2R+1)/k}} \left(\frac{2R+1 - |i-j|}{k}\right).$$

The factor $\gamma^2$ has to be included for having $i, j \in F$ (the two equations must present), $(2R+1)/k$ is the probability that the two lines in $\mathbb{H}$ coincide and $1/2$ is the probability that $b_i \neq b_j$.

Since the above expression is maximized for $|i-j| = 1$, we have $\epsilon = K_0\gamma^2$, with $K_0$ as in equation (7). Further, by summing over $i, j$ we obtain $E[B] = K_1\gamma^3L + O(1)$ for $L \to \infty$.

As for the term $\Delta$, the only non-vanishing contribution comes from the case in which there are three distinct indices among $\{i, j, l, m\}$. If we denote by $\mathbb{H}_i$ the line indexed by $n$ in $\mathbb{H}$, we get

$$\Delta = \frac{3}{4} \sum_{i<j<l} \mathbb{P}\{i, j, l \in F \text{ and } \mathbb{H}_i = \mathbb{H}_j = \mathbb{H}_l\}.$$  

The factor 3 counts the number of different couples of pairs in $\{i, j, l\}$ and $1/4$ is the probability that the corresponding entries in $b$ are different. By computing the above probability and summing over $i, j, l$ we get $\Delta = K_2\gamma^3L + O(1)$, thus proving equation (6).
Appendix B. Polynomial transfer matrix algorithm

Consider the constrained partition function (13) and the corresponding transfer matrix recursion (15). In this appendix we shall consider only left-to-right iterations and drop the arrow $\rightarrow$ in subscripts. We shall further set $n = 2R$ and use the vector notation $x^+_{j+n} = (x_{j+1}, \ldots, x_{j+n})$.

The constrained partition function $Z_i(\vec{z})$ is just the number of solutions in an inhomogeneous linear system, obtained by retaining the lines of $\mathbb{H}$ with index in $\{1, \ldots, i - R\}$ (and the corresponding equations), and adding the $n$ equations $x_{i-n+1} = z_1, \ldots, x_i = z_n$. As a consequence, for all the choices of $\vec{z}$ such that this linear system has a solution, it has the same number of solutions as the corresponding homogeneous system. Further, the number of solutions of the homogeneous system is a power of 2 (because it is the size of a linear space over $\mathbb{F}_2$). Finally, the vectors $\vec{z}$ for which a solution exists form a linear space. Therefore, there exists a binary matrix $A_i$ and an integer $\Phi_i$ such that

$$Z_i(\vec{z}) = \begin{cases} 2^{\Phi_i} & \text{if } A_i \vec{z} = \vec{0}, \\ 0 & \text{otherwise.} \end{cases} \tag{B.1}$$

The matrix $A_i$ can always be chosen as an $n \times n$ matrix by eventually eliminating linearly dependent lines.

We therefore reduced the memory requirements from $\Theta(2^n)$ to $\Theta(n^2)$. We have now to show that the $A_i$ and $\Phi_i$ can be computed recursively in polynomial time as well. Consider the recursion (15) and let $a_i = (a_{i,1}, \ldots, a_{i,n+1})$ be the binary vector defined as follows. If $i - R + 1 \not\in F$ (the new equation added in the recursion is not present), then $a_i \equiv 0$. Otherwise, $a_{i,j} \equiv H_{i-R+1,i-2R+j}$ ($a_i$ encodes the newly added line of $\mathbb{H}$, properly shifted). Then define the $(n+1) \times (n+1)$ matrix $B_i$ as follows:

$$B_i = \begin{pmatrix} A_i & 0 \\ \vdots & \iddots \\ a_{i,1} & \ldots & a_{i,n+1} \end{pmatrix}. \tag{B.2}$$

Denote by $b_i$ the first column of $B_i$ (i.e. a column vector) and by $\tilde{B}_i$ the $(n+1) \times n$ matrix formed by its last $n$ columns. By using equation (B.1) the recursion (15) can be written as

$$Z_{i+1}(\vec{z}) = 2^{\Phi_i} \sum_{z_0 \in \{0,1\}} I(b_i z_0 + \tilde{B}_i \vec{z} = 0). \tag{B.3}$$

Let us now consider two cases:

- If $b_i = 0$, then we get immediately the form (B.1) for $Z_{i+1}$, by letting $\Phi_{i+1} = \Phi_i + 1$, and $A_{i+1}$ the matrix obtained by eliminating linear dependences among rows of $\tilde{B}_i$.

- If $b_i \neq 0$, then there exists at least one vector $\hat{b}_i$ of dimension $(n+1)$ such that $\hat{b}_i^T b_i = 1 \mod 2$. The only non-vanishing term in the sum (B.3) is therefore obtained for $z_0 = -\hat{b}_i^T B_i \vec{z} \mod 2$. Substituting this value of $z_0$, we obtain that $Z_{i+1}$ can again be written in the form (B.1). The new matrix $A_{i+1}$ is obtained by eliminating linearly dependent rows from $(1 - b_i \hat{b}_i^T) \tilde{B}_i$, while the number of solutions is updated by $\Phi_{i+1} = \Phi_i$. 








In practice we found more convenient to reduce \( B_i \) in upper triangular form by Gaussian elimination before computing \( A_{i+1} \) and \( \Phi_{i+1} \) as just described.

The initialization of the above recursion depends on the choice of boundary conditions. When using fixed boundary conditions with reference solution \( \bar{x}^{(0)} = 0 \), we set \( A_0 = 1 \) and \( \Phi_0 = 0 \).

It is clear that the above procedure can be implemented in a time that is polynomial in the interaction range. Indeed the most complex operation to be performed consists in eliminating linearly dependent lines from the matrix \( \tilde{B}_i \), or \((1 - b_i \hat{b}_i^T) \tilde{B}_i \). This can be done via Gaussian elimination in time \( O(R^3) \). The total complexity is therefore \( O(LR^3) \).

### Appendix C. Analytical calculations

#### C.1. Replicas

In order to compute the ground state entropy and the point-to-set correlation function, we shall follow the replica approach, see [26]. Each site \( i \in \{1, \ldots, L\} \) thus carries \( n \) binary variables \( \bar{x}_i = (x_1^i, \ldots, x_n^i) \) corresponding to the \( n \) replicas.

Let us consider first a particularly simple instance consisting of a single equation labelled by \( i \in F \) and \( 2R + 1 \) variables on sites \( j \in \{i - R, \ldots, i + R\} \). Denote by \( \bar{c}_i(\bar{x}) \) the fraction of nodes \( j \) such that \( \bar{x}_j = \bar{x} \).

\[
\bar{c}_i(\bar{x}) = \frac{1}{2R+1} \sum_{j=i-R}^{i+R} \mathbb{I}(\bar{x}_j = \bar{x}). \tag{C.1}
\]

The probability that a randomly sampled equation at \( i \) (with range \( \{1 - R, \ldots, i + R\} \)) is satisfied by all of the \( n \) replicas, is a function of \( \bar{c}_i \), call it \( F_{k,R}(\bar{c}_i) \). For large \( R \) it is easy to show that

\[
F_{k,R}(\bar{c}) = k \left( \bar{c} \right) + \frac{1}{2R+1} \left[ \frac{k}{2} \left( \mathbb{J}_k(\bar{c}) - \mathbb{J}_{k-2}(\bar{c}) \right) + O(R^{-2}) \right], \tag{C.2}
\]

where

\[
\mathbb{J}_1(\bar{c}) = \prod_{\bar{x}_1, \ldots, \bar{x}_1} \mathbb{I}(x_1^n = 0) \bar{c}(\bar{x}_1) \cdots \bar{c}(\bar{x}_1). \tag{C.3}
\]

Consider now the full linear system and the partition function (11). We shall implicitly assume periodic boundary conditions in order to lighten the notations. Fixed boundary conditions can be recovered by properly constraining the expressions that we will derive. It follows from the above that

\[
\mathbb{E}\{ Z^n \} = \sum_{(\bar{x}_i)} \prod_{i=1}^L \left[ 1 - \gamma + \gamma F_{k,R}(\bar{c}_i) \right]. \tag{C.4}
\]

Next we introduce two variables \( \bar{\lambda}_i(\bar{x}), c_i(\bar{x}) \) indexed by \( \bar{x} \in \{0, 1\}^n \) for each \( i \in \{1, \ldots, L\} \), using the identity

\[
1 = \int d\bar{c}_i(\bar{x}) \int_{-i\infty}^{+i\infty} \frac{d\bar{\lambda}_i(\bar{x})}{2\pi i} \exp \{-\bar{\lambda}_i(\bar{x})(c_i(\bar{x}) - \bar{c}_i(\bar{x}))\}. \tag{C.5}
\]
This allows us to perform the sum over \( \vec{x} \) in equation (C.4). If we expand the resulting expression for large \( R \) we get, after some lengthy but straightforward calculations,

\[
E\{Z^n\} = \int dc_i(\vec{x}) \int_{-i\infty}^{i\infty} d\lambda_i(\vec{x}) \exp \{ -(2R + 1) S_0[c, \lambda] - S_1[c] + O(1/R) \},
\]

where

\[
S_0[c, \lambda] = \frac{1}{2R + 1} \sum_{i=1}^{L} \left\{ -\log[1 - \gamma + \lambda \mathcal{J}_k(c_i)] + \sum_{\vec{x}} \lambda_i(\vec{x}) c_i(\vec{x}) - \log \left[ \sum_{\vec{x}} \exp \left( \sum_{j \in D(i)} \frac{\lambda_j(\vec{x})}{2R + 1} \right) \right] \right\},
\]

\[
S_1[c] = \frac{1}{2R + 1} \sum_{i=1}^{L} \left( k \gamma \right) \frac{\mathcal{J}_{k-2}(c_i) - \mathcal{J}_k(c_i)}{1 - \gamma + \gamma \mathcal{J}_k(c_i)},
\]

and we introduced the notation \( D(i) \equiv \{ j : |i - j| \leq R \} \).

### C.2. Mean field limit

In the \( R \to \infty \) limit, the integral (C.6) is dominated by the saddle points of \( S_0[c, \lambda] \). We neglect for the moment the correction given by \( S_1[c] \), and look for a saddle point of the type

\[
c_i(\vec{x}) = \varphi_i \delta_{\vec{x}, \vec{x}_0} + \frac{1}{2^n} (1 - \varphi_i), \quad \lambda_i(\vec{x}) = \omega_i \delta_{\vec{x}, \vec{x}_0} + \frac{1}{2^n} \omega_0,
\]

where \( \vec{x}_0 \equiv (0, 0, \ldots, 0) \) and \( \delta_{\vec{x}, \vec{y}} \) is the \( n \)-dimensional Kronecker delta function. There are several reasons for this ansatz: (i) the algebra of functions of the form \( f(\vec{x}) = f_0 + f_1 \delta_{\vec{x}, \vec{x}_0} \) is closed; (ii) this ansatz is known to give the correct thermodynamic behaviour for the rXOR ensemble (i.e. in the mean field limit); (iii) although it is replica symmetric, it yields the correct one-step replica symmetry breaking physics (it is a peculiarity of XORSAT that replica symmetry can be explicitly broken).

By substituting in equation (C.7) and letting \( n \to 0 \), we get \( S_0[c, \lambda] = A_0[\varphi, \omega] n \log 2 + O(n^2) \) where

\[
A_0[\varphi, \omega] = \frac{1}{2R + 1} \sum_{i=1}^{L} \left\{ \gamma(1 - \varphi_i^k) - \omega_i (1 - \varphi_i) - \exp \left( - \sum_{j \in D(i)} \frac{\omega_j}{2R + 1} \right) \right\}.
\]

By differentiating with respect to \( \varphi_i \) and \( \omega_i \) we get the saddle point equations

\[
\varphi_i = 1 - \frac{1}{2R + 1} \sum_{j \in D(i)} \exp \left( - \sum_{l \in D(j)} \frac{\omega_l}{2R + 1} \right), \quad \omega_i = k \gamma \varphi_i^{k-1}.
\]

The second of these equations can be used to eliminate \( \omega_i \) from the action.

If we finally assume that \( \varphi_i \) only depends on \( i \) on a scale of order \( R \), we can set (with an abuse of notation) \( \varphi_i = \varphi(i/R) \) and let \( R \to \infty \) with \( L = \ell R \), thus getting equations (19)–(21).

doi:10.1088/1742-5468/2007/08/P08004
C.3. 1/R corrections

In computing the 1/R corrections we shall assume the system to be homogeneous. For instance, we can think of imposing periodic boundary conditions, or letting $L \to \infty$ at the outset. As a consequence, in the leading order calculation we have $\varphi_i = \phi$ independent of $i$ and thus $A_0[\varphi, \omega] = L \phi^{(0)}(\gamma; \varphi)$ with $\phi^{(0)}(\gamma; \varphi)$ as in equation (18). Hereafter $\varphi$ will denote a solution of the mean field equation $\varphi = 1 - \exp\{-k\gamma \varphi^{k-1}\}$, and we let $\omega = k\gamma \varphi^{k-1}$.

There are two contributions to order $1/R$. The first one comes from the correction to the action and is easy to compute. Substituting our ansatz in equation (C.8) and proceeding as in the previous section we get $S_1[c] = A_1[\varphi] n \log 2 + O(n^2)$, where

$$A_1[\varphi] = \frac{L}{2R + 1} \gamma \left( k \right) \varphi^{k-1}(1 - \varphi^2).$$

(C.12)

The second term comes from Gaussian fluctuations around the saddle point. Let $c_i^*(\vec{x})$, $\lambda_i^*(\vec{x})$ denote the saddle point (C.9) and define

$$c_i(\vec{x}) = c_i^*(\vec{x}) + \nu_i(\vec{x}), \quad \lambda_i(\vec{x}) = \lambda_i^*(\vec{x}) + \xi_i(\vec{x}).$$

(C.13)

By expanding $S_0[c, \lambda]$ to second order around its saddle point, we get

$$S_0[c, \lambda] = S_0[c^*, \lambda^*] + \frac{1}{2(2R + 1)} \sum_{ij} \sum_{\vec{x}, \vec{y}} \{ A(\vec{x}, \vec{y}) \delta_{ij} \nu_i(\vec{x}) \nu_j(\vec{y})
+ \delta_{\vec{x}, \vec{y}} \delta_{ij} \xi_i(\vec{y}) \xi_j(\vec{y}) \} + B(\vec{x}, \vec{y}) \kappa_R(i - j) \xi_i(\vec{x}) \xi_j(\vec{y}) \},$$

(C.14)

where

$$\kappa_R(i) = \begin{cases} (2R + 1 - |i|)(2R + 1)^2 & \text{for } |i| \leq 2R + 1, \\
0 & \text{otherwise}. \end{cases}$$

(C.15)

The coefficients appearing in equation (C.14) have the form

$$A(\vec{x}, \vec{y}) = A_1 + A_2 \delta_{\vec{x}, \vec{y}} + A_3 \delta_{\vec{x}, \vec{x}_0} + \delta_{\vec{y}, \vec{x}_0} + A_4 \delta_{\vec{x}, \vec{x}_0} \delta_{\vec{y}, \vec{x}_0},$$

(C.16)

$$B(\vec{x}, \vec{y}) = B_1 + B_2 \delta_{\vec{x}, \vec{y}} + B_3 \delta_{\vec{x}, \vec{x}_0} + \delta_{\vec{y}, \vec{x}_0} + B_4 \delta_{\vec{x}, \vec{x}_0} \delta_{\vec{y}, \vec{x}_0},$$

(C.17)

where (defining $z = 1 - \gamma(1 - 2^{-n})(1 - \varphi^k)$)

$$A_1 = -\frac{1}{z} k(k - 1) \gamma (1 - \varphi^{k-2}) \frac{1}{z^2} k^2 \gamma^2 \frac{1}{2n} (1 - \varphi^{k-1})^2;$$

(C.18)

$$A_2 = -\frac{1}{z} k(k - 1) \gamma \varphi^{k-2};$$

(C.19)

$$A_3 = \frac{1}{z^2} k^2 \gamma^2 \frac{1}{2n} (1 - \varphi^{k-1}) \varphi^{k-1};$$

(C.20)

$$A_4 = \frac{1}{z^2} k^2 \gamma^2 \varphi^2(k-1);$$

(C.21)

and (defining $\Lambda_0 = (e^\omega - 1 + 2^n)^{-1}$ and $\Lambda = 1 - 2^n \Lambda_0$)

$$B_1 = -2^n \Lambda_0^2; \quad B_2 = \Lambda_0;$$

(C.22)

$$B_3 = -\Lambda \Lambda_0; \quad B_4 = \Lambda(1 - \Lambda).$$

(C.23)
The quadratic form in equation (C.14) can be diagonalized both in position space (by Fourier transform) and in replica space (all the eigenvectors have the form $\zeta(\vec{x}) = \zeta_0 \delta_{\vec{x} \vec{x}_0} + \zeta_1$). One can therefore perform the Gaussian integral, and let $n \to 0$. Putting this contribution together with the action correction, cf equation (C.12), we finally get the entropy correction
\[
\phi^{(1)}(\gamma) = -\left( \frac{k}{2} \right) \gamma \varphi^{k-2}(1 - \varphi^2) - \int_{-\infty}^{+\infty} \left\{ \log(1 - a w(q)) + \frac{b w(q)}{1 - a w(q)} \right\} \frac{dq}{4\pi},
\]
where
\[
a = k(k - 1)\gamma \varphi^{k-2} e^{-\omega},
\]
\[
b = -k(k - 1)\gamma \varphi^{k-2} e^{-\omega}(1 - e^{-\omega}) + k^2 \gamma^2 \varphi^{2(k-2)} e^{-\omega}.
\]

References

[1] Kirkpatrick T R and Wolynes P G, 1987 Phys. Rev. B 36 8552
[2] Kirkpatrick T R, Thirumalai D and Wolynes P G, 1989 Phys. Rev. A 40 1045
[3] Bouchaud J-P, Cugliandolo L F, Kurchan J and Mézard M, Out of equilibrium dynamics in spin-glasses and other glassy system, 1997 Spin Glasses and Random Fields ed A P Young (Singapore: World Scientific)
[4] Biroli G, Monasson R and Weigt M, 2000 Eur. Phys. J. B 14 551
[5] Mézard M, Parisi G and Zecchina R, 2002 Science 297 812
[6] Krzakala F, Montanari A, Ricci-Tersenghi F, Semerjian G and Zdeborova L, 2007 Proc. Natl Acad. Sci. 104 10318
[7] Wolynes P G, 1997 J. Res. NIST 102 187
[8] Xia X and Wolynes P G, 2000 Proc. Natl Acad. Sci. 97 2990
[9] Xia X and Wolynes P G, 2001 Phys. Rev. Lett. 86 5526
[10] Bouchaud J-P and Biroli G, 2004 J. Chem. Phys. 121 7347
[11] Montanari A and Semerjian G, 2006 J. Stat. Phys. 125 23
[12] Franz S and Montanari A, 2007 J. Phys. A: Math. Theor. 40 F251
[13] Creignou N and Daudé H, 1999 Discrete Appl. Math. 96/97 41
[14] Ricci-Tersenghi F, Weigt M and Zecchina R, 2001 Phys. Rev. E 63 026702
[15] Mézard M, Ricci-Tersenghi F and Zecchina R, 2003 J. Stat. Phys. 111 505
[16] Cocco S, Dubois O, Mandjel J and Monasson R, 2003 Phys. Rev. Lett. 90 047205
[17] Montanari A and Semerjian G, 2005 Phys. Rev. Lett. 94 247201
[18] Montanari A and Semerjian G, 2006 J. Stat. Phys. 124 103
[19] Schwarz J M and Middleton A A, 2004 Phys. Rev. E 70 035103(R)
[20] Franz S and Toninelli F L, 2004 Phys. Rev. Lett. 92 030602
[21] Franz S and Toninelli F L, 2004 J. Phys. A: Math. Gen. 37 7433
[22] Franz S and Toninelli F L, 2005 J. Stat. Mech. P01008
[23] Franz S and Parisi G, 2006 Europhys. Lett. 75 385
[24] Franz S and Toninelli F L, 2004 Int. J. Mod. Phys. 18 275
[25] Alon N and Spencer J, 1992 The Probabilistic Method (New York: Wiley)
[26] Monasson R, 1998 J. Phys. A: Math. Gen. 31 513

doi:10.1088/1742-5468/2007/08/P08004