A Letters Journal Exploring the Frontiers of Physics

September 2014

EPL, **107** (2014) 57011
doi: 10.1209/0295-5075/107/57011

www.epljournal.org

Message passing and Monte Carlo algorithms: Connecting fixed points with metastable states

A. Lage-Castellanos[1], R. Mulet[1] and F. Ricci-Tersenghi[2]

[1] *Department of Theoretical Physics, Physics Faculty, University of Havana - La Habana, CP 10400, Cuba*
[2] *Dipartimento di Fisica, INFN - Sezione di Roma 1 and CNR - IPCF, UOS di Roma, Università La Sapienza P.le A. Moro 5, 00185 Roma, Italy*

received 24 June 2014; accepted in final form 21 August 2014
published online 8 September 2014

PACS 75.10.Nr – Spin-glass and other random models
PACS 05.10.Ln – Monte Carlo methods

Abstract – Mean field-like approximations (including naive mean-field, Bethe and Kikuchi and more general cluster variational methods) are known to stabilize ordered phases at temperatures higher than the thermodynamical transition. For example, in the Edwards-Anderson model in 2 dimensions these approximations predict a spin glass transition at finite T. Here we show that the spin glass solutions of the Cluster Variational Method (CVM) at plaquette level do describe well the actual metastable states of the system. Moreover, we prove that these states can be used to predict non-trivial statistical quantities, like the distribution of the overlap between two replicas. Our results support the idea that message passing algorithms can be helpful to accelerate Monte Carlo simulations in finite-dimensional systems.

Introduction. – Monte Carlo (MC) methods are the most celebrated and used techniques to computationally explore the configuration space of Hamiltonian systems [1]. Unfortunately, in many practical cases, usually at very low temperatures or close to phase transitions, the dynamics becomes very slow and the time needed to average the system diverges with the system size. The situation is specially frustrating when studying problems that are computationally demanding. In these cases it is natural to first try to understand the properties of the phase space that make the problems *hard* in the computational sense and then, with the help of this comprehension, to design efficient algorithms [2–5].

A very promising tool in this direction are message passing algorithms that are derived from an approximated free energy of the specific model of interest [6]. Up to now, most of the attention to this approach has been concentrated on Bethe-like approximations [4,6–8]. However, the applicability of this approximation is usually restricted to systems with very large loops, $\sim \log(N)$, where N is the system size, but is of limited value to study finite-dimensional systems.

A more sophisticated approach is the Cluster Variational Method (CVM) [9–13] that in principle may consistently account for the presence of short loops in the model, providing also a more natural connection with MC methods in finite-dimensional systems.

It is tempting to combine message passing and Monte Carlo techniques to exploit the potentialities of both approaches. In a recent paper, [14] this was done for the first time. In that contribution, the standard Metropolis technique, was *guided* by the marginals estimated by a message passing algorithm defined on a proper tree-like structure. Yet, there is a lot of room for improvement. In particular, to use this, or similar techniques in finite-dimensional systems. But to firmly progress in this direction it remains to understand what the connection (if any) is between the fixed-points solutions of message passing techniques and Monte Carlo simulations in finite-dimensional systems.

This is the main aim of this work. In what follows we present new data supporting that non-paramagnetic fixed points of plaquette-CVM are indeed connected with the configurational space explored by the Metropolis algorithm. These fixed points correspond to actual metastable states of the system and are a useful tool to extract non-trivial information about the dynamics, for example the overlap between Monte Carlo replicas. We will use $\pm J$ Edwards-Anderson (EA) 2D model as proof of concept, but we expect that similar properties hold in other disordered models.

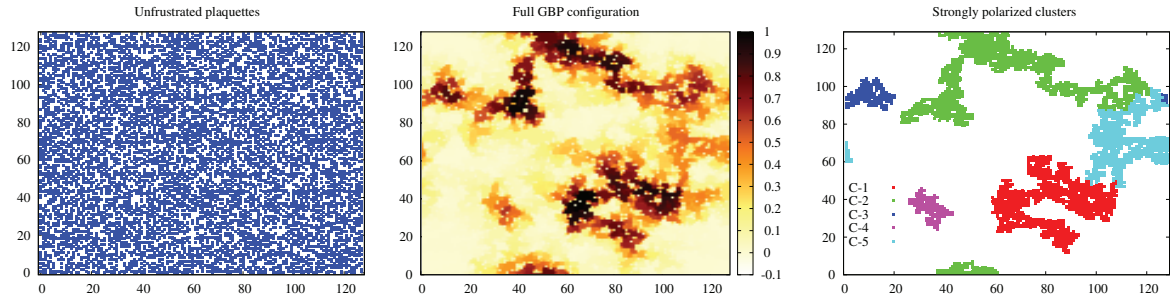

Fig. 1: (Colour on-line) For a 128×128 EA 2D system, the pattern of unfrustrated plaquettes (left panel, blue squares) gives neither an obvious hint for the appearance of strongly magnetized regions (middle panel) nor the distinction of clusters (right panel). In the clusters where the magnetization appears, there is a slightly higher concentration of non-frustrated plaquettes (53% *vs.* 49% in the whole system).

Cluster variational method in Edwards-Anderson 2D. – The celebrated Edwards-Anderson model in statistical mechanics [15] is defined by a set $\sigma = \{s_1 \dots s_N\}$ of N Ising spins $s_i = \pm 1$ placed at the nodes of a square lattice (in our case in two dimensions), and random interactions $J_{i,j} = \pm 1$ at the edges, with a Hamiltonian

$$\mathcal{H}(\sigma) = -\sum_{\langle i,j \rangle} J_{i,j} s_i s_j \qquad (1)$$

where $\langle i, j \rangle$ runs over all couples of neighboring spins.

The direct computation of the partition function Z, or any marginal probability distribution like $p(s_i, s_j) = \sum_{\sigma \setminus s_i, s_j} P(\sigma)$, is unattainable in general, and therefore approximations are required. Among all of them, we will explore the CVM, a technique that includes mean-field and Bethe approximations [16] as particular cases, and was first derived by Kikuchi [17] for the homogeneous system, and later extended to disordered models. In its modern presentation [6,12], it consists in replacing the exact (Boltzmann-Gibbs) distribution $P(\sigma)$, by a reduced set of its (approximated) marginals $\{b_R(\sigma_R)\}$ over certain degrees of freedom grouped in regions. With this reduction, the approximated free energy can be minimized in a numerically treatable manner. The consistency between the probability distributions of regions that share one or more degrees of freedom, is forced by Lagrange multipliers [6]. The latter are connected by self-consistent relations, that are solved by an iterative procedure, the so-called Generalized Belief Propagation (GBP). In what follows we use this approximation with the square plaquettes of the 2D lattice as the largest set of marginals considered. We skip the details and point the reader to [9] where the precise form of these equations for the plaquette-CVM in EA 2D can be found. Other approaches, similar in spirit, have been followed in references [10,18–20].

Solutions of GBP. – When running GBP for the plaquette-CVM approximation in EA 2D we find a paramagnetic solution at high temperature, as expected. However, above $\beta_c \simeq 0.79$ (below $T_c \simeq 1.26$) GBP finds, not one, but many fixed points with non-zero local magnetizations. Suggesting then, a transition from a paramagnetic to a spin glass phase [10,21]. On the other hand the use of a provably convergent method called Double Loop [22] showed that [11], while having the same set of fixed-point solutions as GBP, at low temperatures, in order to keep converging the algorithm is set back to the paramagnetic solution. Moreover, at still lower temperatures (above $\beta_{\rm conv} \simeq 1.2$) [9], GBP stops converging. It is this region of intermediate temperatures, where GBP finds many non-trivial solutions that will concentrate our attention on here.

Already in references [21] and [10] it was noted that the non-paramagnetic solutions have inhomogeneous magnetizations, not only in their sign as expected in a disordered system, but also in their spatial distribution: connected clusters of magnetized spins are surrounded by a sea of unmagnetized ones (see fig. 1).

Relation to Monte Carlo. – Though GBP solutions are not thermodynamic states, we will show that Monte Carlo dynamics remain most of the time near the GBP solutions in the range of temperatures ($\beta_c - \beta_{\rm conv}$), as schematized in fig. 2.

One of the many possible systematic approaches (although heuristic and non-exhaustive) to locate *all* GBP solutions is divided in two steps: i) to identify the clusters of connected and strongly magnetized spins, then ii) to explore all possible combinations of orientations for those clusters. Locating the clusters starts from a given non-paramagnetic solution of GBP equations (the reference GBP state) at the desired temperature. Then it iterates the following procedure, starting from $c = 1$:

1) *Leader spin*: Take the most magnetized spin that does not belong to a cluster already defined, call its magnetization m_c.

2) *Grow cluster*: Starting with the neighbors of the leader spin, add to cluster c all nearby spins that have a magnetization with modulus greater than $\theta|m_c|$, then iterate over the neighbors of the newly added spins, and so on. Stop when the neighborhood of the cluster has magnetizations smaller than $\theta|m_c|$.

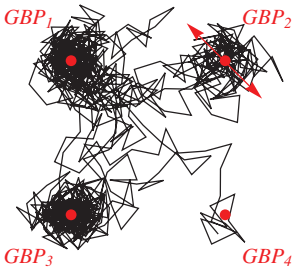

Fig. 2: (Colour on-line) Schematic representation of the Monte Carlo dynamics in the configurational space. Most of the time MC dynamics is in the vicinity of one GBP solution.

The parameter θ is an arbitrary threshold. We used here $\theta = 0.8$, but other values (higher) produced equivalent results.

3) If cluster c is not in touch with any previously defined cluster, then go to 1 with $c = c + 1$, else stop.

The result of the procedure for a particular instance is depicted in the rightmost panel of fig. 1. Cluster 5 is in touch with clusters 1 and 2, and therefore is the last cluster to be considered. We studied more than a dozen of single instances and all provided results consistent with the data shown in this work.

Once the clusters are identified, we use the message passing program starting from the given solution, and seek convergence after reverting the sign of all messages pointing to the spins in a given cluster. This is tantamount to reverting all magnetizations in the given cluster, while keeping the others in their original state. If we have found n clusters, and all of them can flip independently, we can try convergence to 2^n solutions. Our final set of GBP states will be created out of all different GBP solutions found by this procedure. In fig. 3 we show three different solutions obtained in this way.

One case example. Take, for instance, the sample of fig. 1. From the possible $2^5 = 32$ different GBP initial conditions (including the trivial symmetry of the system), we found only 24 solutions at $T = 0.75$. To illustrate the connection between these states and Monte Carlo dynamics, we run a MC simulation with Metropolis updating rule, of the $N = 128 \times 128$ system and averaged the local magnetization in a time window of 1000 MC steps:

$$m_i^{MC}(K = \lfloor t/1000 \rfloor) = 1/1000 \sum_{t=1+1000K}^{1000(K+1)} s_i(t), \qquad (2)$$

where K is a mesoscopic Monte Carlo time. Then we project this quantity over the GBP local magnetization of each state α, defining the quantity

$$q_{\alpha,MC}(K) = \frac{1}{N} \sum_i m_i^\alpha m_i^{MC}(K). \qquad (3)$$

In fig. 4 we show this projection as a function of MC mesoscopic time K. It can be seen how the projection

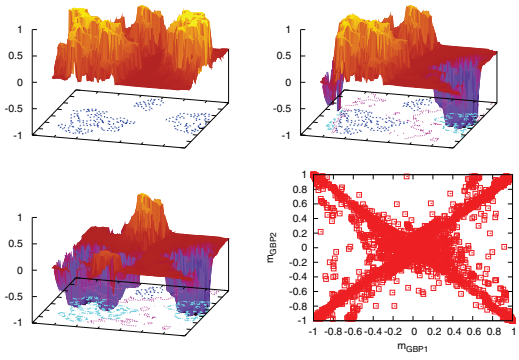

Fig. 3: (Colour on-line) Different GBP solutions obtained for the same $N = 64 \times 64$ EA 2D system. The first 3 panels show the magnetization of each spin at positions (x, y) in the lattice. To help the eye in recognizing the three different clusters of spins, the first GBP state is used as a reference. The z-axis is the projection of each site magnetization onto the direction of the first solution found $m^{\beta,1}(x, y) = m^\beta(x, y)\text{Sign}(m^1(x, y))$. Three different mostly independent clusters can be seen in the contour surfaces of the top-left plot, and 3D plots show how they can switch directions from one GBP solution to the other. In the bottom-right we plot m_i^1 *vs.* m_i^2 for each spin in the system in two different GBP solutions.

over the GBP states is non-trivial, growing in absolute value and in time persistence as temperature goes down, and how the system switches from one state to the other, remaining most of the time near one of these states.

Furthermore, if GBP states are the metastable states, then the time T_α that the system is near any GBP solution α, should be proportional to the exponential of its free energy F_α that we can estimate by GBP,

$$T_\alpha \propto w_\alpha = \frac{\exp(-\beta F_\alpha)}{\sum_{\alpha'=1}^n \exp(-\beta F_{\alpha'})}. \qquad (4)$$

In fig. 5 this is shown to be the case for a system of $N = 64 \times 64$ spins, at three different temperatures. We define the system to be near state α at time t if its overlap is the highest:

$$\forall_\gamma \; q_{\gamma,MC}(t) \leq q_{\alpha,MC}(t). \qquad (5)$$

An *experimental* frequency of each state is computed as the amount of Monte Carlo time T_α the system stays in the vicinity of GBP state α, divided by the total Monte Carlo time of the experiment. This frequency is very well predicted by w_α.

Monte Carlo replicas overlap. – So far we have shown that GBP locates metastable states in the Monte Carlo dynamics of the EA in 2D. Therefore, the Boltzmann measure can be approximated as the linear combination of each GBP state measure

$$P(\mathbf{s}) = \sum_\alpha w_\alpha P_\alpha(\mathbf{s}) \mathbb{I}(\mathbf{s} \in \alpha). \qquad (6)$$

Next we use this fact to predict some non-trivial MC quantities using only GBP fixed points.

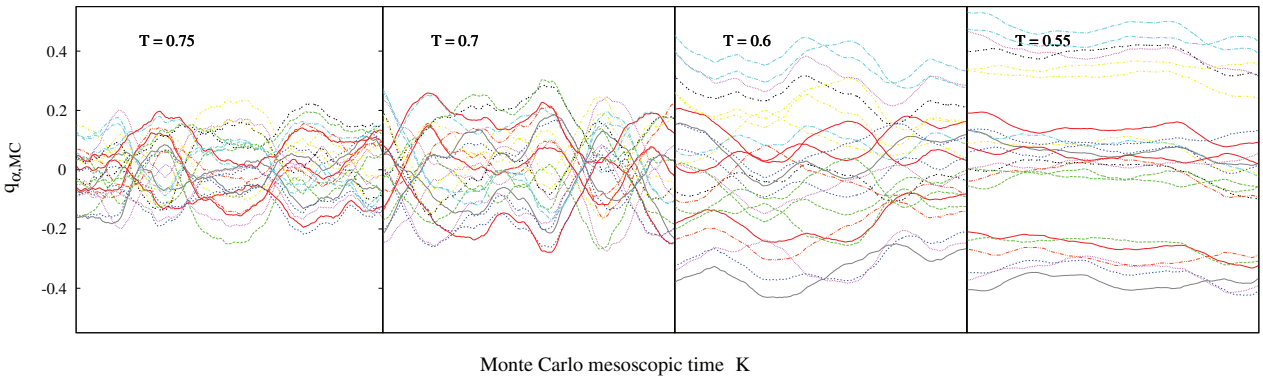

Fig. 4: (Colour on-line) Overlap $q_{\alpha,MC}$ between Monte Carlo magnetizations over a short time window (1000 MCS, see eq. (2)) and the 24 GBP predicted magnetizations, as a function of Monte Carlo mesoscopic time $K = \lfloor t/1000 \rfloor$. Each line corresponds to the projection on a different GBP solution. The set of GBP solutions is computed at $T = 0.75$, and projected over MC dynamics going from $K = 1$ to $K = 10^3$ at four different temperatures. The data is smoothed with a nearest-neighbor smoother, to average out high-frequency Monte Carlo noise.

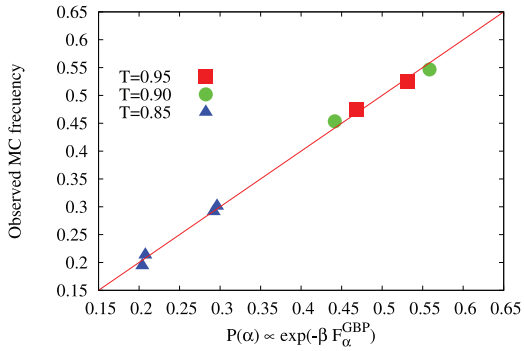

Fig. 5: (Colour on-line) At temperatures $T = 0.95, 0.9, 0.85$ GBP finds two, two and three independent clusters, and therefore there are 2×2^1, 2×2^1 and 2×2^2 GBP solutions, respectively. The first factor 2 corresponds to the natural symmetry $s_i \to -s_i$. Since symmetric solutions are equivalent in all senses, they will be taken as one solution. The observed time fraction during which the Monte Carlo dynamics stays in the vicinity of each GBP solution is plotted against its predicted value from the GBP free energy in this particular 2D EA instance (see eq. (4)).

One key parameter in disordered systems is the overlap between two different replicas of the system:

$$q = \frac{1}{N} \sum_i^N s_i^1 s_i^2. \tag{7}$$

The probability distribution of the overlap $P(q)$ provides information on the structure of states.

If each replica of the system stays close to one of the GBP states (for a time that can be estimated from the GBP free energy), we should be able to reproduce the statistics of the overlap q from GBP data alone. We will consider that the random variable q is given by a two-steps stochastic process: the first one is the choice of the GBP states where the replicas are (see scheme in fig. 2), the second considers the stochastic fluctuation in the given states. Therefore, the distribution of q is a weighted sum of the probabilities of the random variables $q_{\alpha\beta}$, where α and β are states indices,

$$P(q) = \sum_{\alpha,\beta} w_\alpha w_\beta P_{\alpha\beta}(q), \tag{8}$$

where

$$P_{\alpha\beta}(q) = \sum_{\mathbf{s}^1 \mathbf{s}^2} P_\alpha(\mathbf{s}^1) P_\beta(\mathbf{s}^2) \delta\left(q - \frac{1}{N} \sum_i s_i^1 s_i^2\right) \tag{9}$$

is the distribution of the overlap between two replicas when they are in states α and β, and $w_\alpha w_\beta$ is the probability of such a situation.

The expected value of $q_{\alpha\beta}$ is readily given in terms of averages in the GBP states

$$q_{\alpha\beta} = \left\langle \frac{1}{N} \sum_i s_i^\alpha s_i^\beta \right\rangle = \frac{1}{N} \sum_i m_i^\alpha m_i^\beta. \tag{10}$$

On the other hand, the computation of the variance is harder and requires the estimation of the correlations as we will show in the following.

Estimating the variance. The variance of the overlap between states α and β is given by

$$\sigma_{\alpha\beta}^2 = \left\langle \left(\frac{1}{N} \sum_i s_i^\alpha s_i^\beta - q_{\alpha\beta} \right)^2 \right\rangle = \left\langle \left(\frac{1}{N} \sum_i s_i^\alpha s_i^\beta \right)^2 \right\rangle - q_{\alpha\beta}^2. \tag{11}$$

The first term on the second line can be written in terms of connected correlations in a state, $C_{ij}^\alpha = \langle s_i^\alpha s_j^\alpha \rangle - m_i^\alpha m_j^\alpha$,

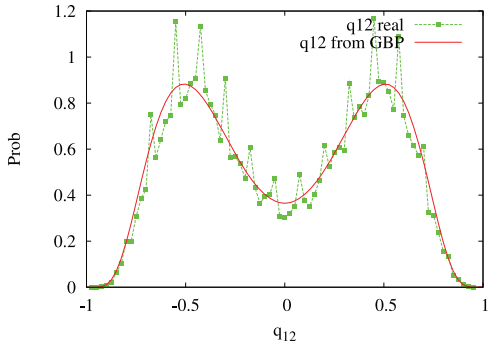

Fig. 6: (Colour on-line) Symbols: distribution of the overlap q_{12} between two independent Monte Carlo simulations of a 2D EA system ($N = 64 \times 64$) at temperature $T = 0.65$. Solid line: distribution (8) obtained from GBP states at the same temperature.

by

$$\left\langle \left(\frac{1}{N} \sum_i s_i^\alpha s_i^\beta \right)^2 \right\rangle = \frac{1}{N^2} \sum_{ij} (C_{ij}^\alpha + m_i^\alpha m_j^\alpha)(C_{ij}^\beta + m_i^\beta m_j^\beta) = \frac{1}{N^2} \sum_{ij} \left(C_{ij}^\alpha C_{ij}^\beta + m_i^\alpha m_j^\alpha C_{ij}^\beta + m_i^\beta m_j^\beta C_{ij}^\alpha \right) + q_{\alpha\beta}^2. \tag{12}$$

From this we finally get

$$\sigma_{\alpha\beta}^2 = \frac{1}{N^2} \sum_{ij} \left(C_{ij}^\alpha C_{ij}^\beta + m_i^\alpha m_j^\alpha C_{ij}^\beta + m_i^\beta m_j^\beta C_{ij}^\alpha \right). \tag{13}$$

Connected correlations between spins C_{ij}^α can be approximated in two ways: with a generalized susceptibility propagation algorithm, or using fluctuation-dissipation relations within GBP approximation. The generalized susceptibility algorithm, though somehow intuitive, to the best of our knowledge has not been developed so far. Instead, we can obtain the connected correlations C_{ij}^α in an *experimental* way from GBP, by introducing a small external field over the spins (one at a time) and using the fluctuation dissipation relation $C_{ij}^\alpha = \frac{\partial m_i^\alpha}{\partial h_j}$.

This procedure is a little bit more cumbersome. We need to run GBP, and within every solution found, compute C_{ij}^α for every pair of spins in the system. This calculation requires the introduction of a small field δh_i over spin i, then running GBP some more steps until convergence, and then computing $C_{ij}^\alpha \simeq \delta m_j^\alpha / \delta h_i$. Note that every time we put the probe field δh_i we get, after convergence, an estimate for N correlations. Fortunately, for estimating $\sigma_{\alpha\beta}^2$ the correlations are averaged over all site pairs, and thus it is enough to sample a random, and large enough, subset of the correlations. Therefore, we have selected 50 random spins in the system, and run GBP with the external field on each of them to get $50 \times N$ estimates of C_{ij}^α.

Given $q_{\alpha\beta}$ and $\sigma_{\alpha\beta}^2$, we would like to approximate $P_{\alpha\beta}(q)$ by a suitable function with average $q_{\alpha\beta}$ and variance $\sigma_{\alpha\beta}^2$. Unfortunately, a simple Gaussian ansatz is deemed to fail because q is bounded in $[-1, 1]$. We alleviate this problem by assuming normal fluctuations for the unbounded variable $h \equiv \text{arctanh}(q)$, which has been proved effective in previous works [23].

In fig. 6 we show the results of our analysis. The figure compares Monte Carlo measurements for $P(q)$ in a system of $N = 64 \times 64$ spins at $T = 0.65$ with function (8) showing a remarkable coincidence between the two.

Conclusions. – We have used the plaquette-CVM approximation to the free energy, and the corresponding Generalized Belief Propagation algorithm to study the intermediate temperature regime of the 2D Edwards-Anderson model. We have shown that the spin glass solutions obtained in the temperature range $\beta \in [0.79, 1.2]$ give very useful information about the dynamics of the actual finite-size system. Indeed, the Monte Carlo dynamic stays near the GBP solutions a fraction of time proportional to the statistical weigth predicted by the plaquette-CVM approximation. Moreover, the overlap distribution $P(q)$ can be well approximated from the GBP fixed-point solutions.

In our opinion this is a very promising result which may pave the way towards a better use of CVM approximations to improve the numerical study of complex and disordered systems. For example, one can think of speeding up a Monte Carlo simulation by proposing a cluster flipping move, using the clusters found by the GBP algorithm.

Furthermore, it could be interesting to study whether some specific features of the aging dynamics, as, *e.g.*, the strong timescale sepration in the flipping times at low temperatures [24,25], can be extracted from the GBP fixed-point solutions.

* * *

We acknowledge discussions with Tommaso Rizzo in the initial part of this work. This research has received financial support from the Italian Research Ministry through the FIRB project No. RBFR086NN1. We also acknowledge the support from the ERC-AdG grant OPTINF No. 267915.

REFERENCES

[1] Newman M. E. J. and Barkema J. T., *Monte Carlo Methods in Statistical Physics* (Oxford University Press) 2001.
[2] Mézard M. and Montanari A., *Information, Physics, and Computation*, Oxford Graduate Texts (Oxford University Press, Oxford) 2009.
[3] Mézard Marc and Parisi G., *J. Stat. Phys.*, **111** (2003) 1.

[4] Mézard Marc, Parisi G. and Zecchina R., *Science*, **297** (2002) 812.

[5] Mulet R., Pagnani A., Weigt M. and Zecchina R., *Phys. Rev. Lett.*, **89** (2002) 268701.

[6] Yedidia J., Freeman W. T. and Weiss Y., *IEEE Trans. Inf. Theory*, **51** (2005) 2282.

[7] Braunstein A., Mulet R., Pagnani A., Weigt M. and Zecchina R., *Phys. Rev. E*, **68** (2003) 036702.

[8] Mézard Marc and Zecchina Riccardo, *Phys. Rev. E*, **66** (2002) 056126.

[9] Dominguez Eduardo, Lage-Castellanos Alejandro, Mulet Roberto, Ricci-Tersenghi Federico and Rizzo Tommaso, *J. Stat. Mech.: Theory Exp.* (2011) P12007.

[10] Zhou Haijun and Xiao Jing-Qing, *J. Phys. A: Math. Theor.*, **44** (2011) 425001.

[11] Lage-Castellanos Alejandro, Mulet Roberto, Ricci-Tersenghi Federico and Rizzo Tommaso, *Phys. Rev. E*, **84** (2011) 046706.

[12] Pelizzola Alessandro, *J. Phys. A*, **38** (2005) R309.

[13] Rizzo Tommaso, Lage-Castellanos Alejandro, Mulet Roberto and Ricci-Tersenghi Federico, *J. Stat. Phys.*, **139** (2010) 375.

[14] Krzakala Florent and Decelle Aurélien, *Phys. Rev. B*, **89** (2014) 214421.

[15] Edwards S. F. and Anderson P. W., *J. Phys. F: Met. Phys.*, **5** (1975) 965.

[16] Bethe H. A., *Proc. R. Soc. A*, **150** (1935) 552.

[17] Kikuchi R., *Phys. Rev.*, **81** (1951) 988.

[18] Wang Chuang and Zhou Haijun, *J. Stat. Phys.*, **148** (2013) 513.

[19] Zhou Haijun, Wang Chuang, Xiao Jing-Qing, and Bi Zedong, *J. Stat. Mech.: Theory Exp.* (2011) L12001.

[20] Teodorescu Razvan, Chertkov Michael and Chernyak Vladimir Y., *J. Stat. Mech.* (2008) P05003.

[21] Lage-Castellanos Alejandro, Mulet Roberto, Ricci-Tersenghi Federico and Rizzo Tommaso, *J. Phys. A: Math. Theor.*, **46** (2013) 135001.

[22] Heskes T., Albers K. and Kappen B., *Approximate inference and constrained optimization*, in *Proceedings of the Nineteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-03)* (Morgan Kaufmann, San Francisco, Cal.) 2003, pp. 313–320.

[23] Baños R. A., Cruz A., Fernandez L. A., Gil-Narvion J. M., Gordillo-Guerrero A., Guidetti M., Iñiguez D., Maiorano A., Mantovani F., Marinari E., Martin-Mayor V., Monforte-Garcia J., Muñoz Sudupe A., Navarro D., Parisi G., Perez-Gaviro S., Ricci-Tersenghi F., Ruiz-Lorenzo J. J., Schifano S. F., Seoane B., Tarancón A., Tripiccione R. and Yllanes D., *Phys. Rev. B*, **84** (2011) 174209.

[24] Ricci-Tersenghi F. and Zecchina R., *Phys. Rev. E*, **62** (2000) R7567.

[25] Romá F., Bustingorry S. and Gleiser P. M., *Phys. Rev. Lett.*, **96** (2006) 167205.