PHYSICAL REVIEW E **85**, 016106 (2012)

Analytical and belief-propagation studies of random constraint satisfaction problems with growing domains

Chunyan Zhao,[1,*] Pan Zhang,[2,†] Zhiming Zheng,[1,‡] and Ke Xu[3,§]

[1]*LMIB and School of Mathematics and Systems Science, Beihang University, Beijing 100191, China*

[2]*Politecnico di Torino, C.so Duca degli Abruzzi 24, I-10129 Torino, Italy*

[3]*State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China*

(Received 28 June 2011; published 10 January 2012)

We study solution-space structure and solution-finding algorithms of a representative hard random constraint satisfaction problem with growing domains known as Model RB. Using rigorous methods, we show that solutions are grouped into disconnected clusters before the theoretical satisfiability phase transition point. Using the cavity method, it is shown that the entropy density obtained by belief propagation (BP) on random Model RB instances, which corresponds well to the analytical results, vanishes as the control parameter (constraint tightness) approaches the satisfiability threshold. From an algorithmic point of view, we find that reinforced BP, which performs much better than all existing algorithms, allows us to find solutions efficiently for instances in the regime that is very close to the satisfiability transition. These results also can shed light on the effectiveness of BP reinforcement on problems with a large number of states.

DOI: 10.1103/PhysRevE.85.016106 PACS number(s): 89.75.Fb, 02.50.−r, 64.70.P−, 89.20.Ff

I. INTRODUCTION

Constraint satisfaction problems (CSPs) have played an extremely significant role in the interdiscipline of theoretical computer science, information theory, and statistical physics [1]. They can find applications in many practical problems, including pattern recognition, resource allocation, verification, and more. Roughly speaking, an instance of a random CSP consists of N variables and M constraints. Each variable takes values from its domain D. Each constraint contains a subset of k distinct variables chosen uniformly at random and prohibits these variables from taking some disallowed assignments out of $|D|^k$ possible ones. The instance is satisfiable if there exists a solution, that is, an assignment of the variables that satisfies all the constraints simultaneously. Most of the interesting CSPs belong to the class of nondeterministic polynomial-complete (NP-complete) problems, and finding a solution of such problems appears to require exponential time in the worst case [2]. In ensembles of random CSPs, empirical evidence indicates that there exists a sharp satisfiability threshold separating a phase where instances are almost always satisfiable (SAT) from a phase where instances are almost always unsatisfiable (UNSAT) [3]. Moreover, the hardest instances are found near the SAT-UNSAT transition.

Many of the previously studied problems (such as the random k-SAT, random vertex covers, random graph q-coloring, and so on) are random CSPs with fixed domains [4–7] in which the domain size (number of states) of each variable is independent of the system size. In recent years, studies of random CSPs with growing domains have attracted much interest in computer science and statistical physics [8–15]. Model RB, as a prototypical random CSP with growing domain size, was first proposed in Ref. [8] for overcoming the trivial asymptotic insolubility of the standard CSP Model B [16]. For Model RB, not only has the existence of the satisfiability phase transition been established rigorously, but also the threshold point has been located precisely. For some random CSPs with fixed domains, there is no rigorous proof yet of the existence of this transition [17,18], while methods of statistical physics provide powerful tools to investigate various transitions in the geometrical organization of solutions [19–23]. On the other hand, it has been proved theoretically and experimentally that all instances of Model RB are hard at the threshold [11–13]. Therefore, benchmarks based on Model RB have been used widely in various kinds of algorithm competitions, including SAT (2004,2009,2011), CSP (since 2005), pseudo-Boolean (since 2005), Max-SAT (since 2008), and so on, and the results confirmed the intrinsic hardness of these benchmarks. Finding solutions of a single Model RB instance is very challenging, and attempts have been limited to cases of $N \sim 10^2$ [24]. In Ref. [14], it is shown that message-passing algorithms guided by belief propagation (BP) [25,26] based on the cavity method from statistical physics can construct solutions efficiently for random instances of Model RB in the SAT phase. Recently, reinforced BP has been used as a solver in several NP-complete problems [27–30]. The main idea of reinforced BP, originally proposed in Ref. [31], is to introduce an external set of reinforcement messages enforcing BP equations toward a solution.

In this paper, we study the solution space of Model RB analytically and algorithmically. Furthermore, we compare experimental results coming from the statistical-physics methods to the rigorous asymptotic predictions. It seems that the BP entropy density normalized by $\ln N$ is in good agreement with the rigorous results. We also find that reinforced BP can serve as an efficient algorithm to solve random Model RB instances in the hard-SAT region where the solution pairs at the Hamming distance around $N/2$ begin to disappear.

*zhaocy@ss.buaa.edu.cn

†pan.zhang@polito.it

‡zzheng@pku.edu.cn

§Corresponding author: kexu@nldse.buaa.edu.cn

II. QUENCHED AND ANNEALED RESULTS OF MODEL RB

Let us first briefly review Model RB. An instance of Model RB contains N variables $\vec{\sigma} = (\sigma_1,\sigma_2,\ldots,\sigma_N)$ and $M = rN\ln N$ ($r > 0$ is a constant) constraints. Each σ_i takes values from its domain $D = \{1,2,\ldots,d_N\}$, where $d_N = N^\alpha$ ($\alpha > 0$ is a constant) grows polynomially with N. Each constraint a involves k ($k \geqslant 2$) different variables $(\sigma_a^1,\sigma_a^2,\ldots,\sigma_a^k)$ chosen uniformly at random, and we define $Q_a \subset D^k$ as the set of disallowed assignments for these variables. For each a, we randomly pick $|Q_a| = pd_N^k$ disallowed assignments without repetition out of d_N^k possible ones where $p \in (0,1)$ measures the tightness of the constraint. Constraint a is satisfiable by the assignment of $\vec{\sigma}$ if and only if $(\sigma_a^1,\sigma_a^2,\ldots,\sigma_a^k) \notin Q_a$. Solving the instance amounts to finding an assignment of the variables that satisfies all the constraints or proving that no such assignment exists.

A random instance of Model RB admits a natural factor graph representation [32]. The mean connectivity of a variable is $r\ln N$, therefore, the factor graph is dense with loops. The typical length of a loop scales as $\ln N/\ln\ln N$, so the factor graph is locally treelike when N is sufficiently large. Let $\psi_{\sigma_i}^{a\to i(t)}$, $\chi_{\sigma_j}^{j\to a(t)}$ denote the messages that are passed at time t along all the edges in the factor graph. More precisely, $\psi_{\sigma_i}^{a\to i}$ is the probability that constraint a being satisfied if variable i takes value σ_i, and $\chi_{\sigma_j}^{j\to a}$ is the probability that j takes value σ_j in the absence of a. Using the cavity method [19], BP equations can be written as

$$\psi_{\sigma_i}^{a\to i(t)} = \frac{1}{Z^{a\to i}} \sum_{\sigma_j,\, j\in\partial a\setminus i} \phi_a(\vec{\sigma}_j,\sigma_i) \prod_{j\in\partial a\setminus i} \chi_{\sigma_j}^{j\to a(t)}, \quad (1)$$

$$\chi_{\sigma_j}^{j\to a(t+1)} = \frac{1}{Z^{j\to a}} \prod_{b\in\partial j\setminus a} \psi_{\sigma_j}^{b\to j(t)}, \quad (2)$$

where $\partial a\setminus i$ denotes the set of variables connected to constraint a except i, $Z^{a\to i}$, and $Z^{j\to a}$ are normalization constants, the function $\phi_a(\vec{\sigma}_a)$ is equal to 0 if $(\sigma_a^1,\sigma_a^2,\ldots,\sigma_a^k) \in Q_a$, and is equal to 1 otherwise.

To compute the BP entropy density of Model RB, we first define ΔS_i, ΔS_a, and $\Delta S_{(i,a)}$, respectively, as the entropy shift after the addition of a variable i, a constraint a, and an edge (i,a). After Eqs. (1) and (2) converge to a fixed point, the entropy density (the logarithm of the number of solutions divided by N) can be estimated by the following equation [19]:

$$s = \frac{1}{N}\sum_i \Delta S_i + \frac{1}{N}\sum_a \Delta S_a - \frac{1}{N}\sum_{(i,a)} \Delta S_{(i,a)}, \quad (3)$$

where

$$\Delta S_i = \ln\left(\sum_{\sigma_i}\prod_{a\in\partial i}\psi_{\sigma_i}^{a\to i}\right), \quad (4)$$

$$\Delta S_a = \ln\left[\sum_{\sigma_i, i\in\partial a}\phi_a(\vec{\sigma}_i)\prod_{i\in\partial a}\chi_{\sigma_i}^{i\to a}\right], \quad (5)$$

$$\Delta S_{(i,a)} = \ln\left(\sum_{\sigma_i}\psi_{\sigma_i}^{a\to i}\chi_{\sigma_i}^{i\to a}\right). \quad (6)$$

TABLE I. Given $\alpha = 0.8$ and $r = 3$, the domain size d_N and the number of constraints M for different N are shown in this table. The theoretical satisfiability threshold is 0.234.

N	α	d_N	r	M	p_s
100	0.8	40	3	1382	0.234
150	0.8	55	3	2255	0.234
200	0.8	69	3	3179	0.234
300	0.8	96	3	5133	0.234
400	0.8	121	3	7190	0.234

Since Model RB is NP-complete in the cases of $k \geqslant 2$, for the sake of simplicity, we use binary Model RB ($k = 2$) as the tested problem. In particular, we choose $\alpha = 0.8$ and $r = 3$ to generate random instances of size $N \in [100,400]$ (on random instances with other groups of parameters, we also obtain similar results). The corresponding parameters for different N are shown in Table I.

In Fig. 1, the BP entropy densities for different system sizes, scale as $O(\ln N)$, decrease monotonically with constraint tightness p, and vanish at $p_s \simeq 0.234$, which is the theoretical satisfiability threshold. In fact, in Ref. [8], it was shown that, under the conditions of $\alpha > 1/k$ and $ke^{-\alpha/r} \geqslant 1$, as $N \to \infty$, we precisely can locate $p_s = 1 - e^{-\alpha/r}$, below which the probability of a random instance of Model RB being satisfiable tends to 1 and above which the probability tends to 0. Let $\langle Z\rangle$ denote the expectation of the solution number, as reported in Ref. [8], $\langle Z\rangle$ [$=N^{\alpha N}(1-p)^{rN\ln N}$] computed on a single instance can roughly match the typical number of solutions due to $\langle Z\rangle^2 = \langle Z^2\rangle$. The inset in Fig. 1 shows that BP entropy density normalized by $\ln N$ (dashed blue line) coincides very well with the annealed entropy $\ln\langle Z\rangle$ divided by $N\ln N$ (solid red line). This suggests that the replica symmetry (RS)

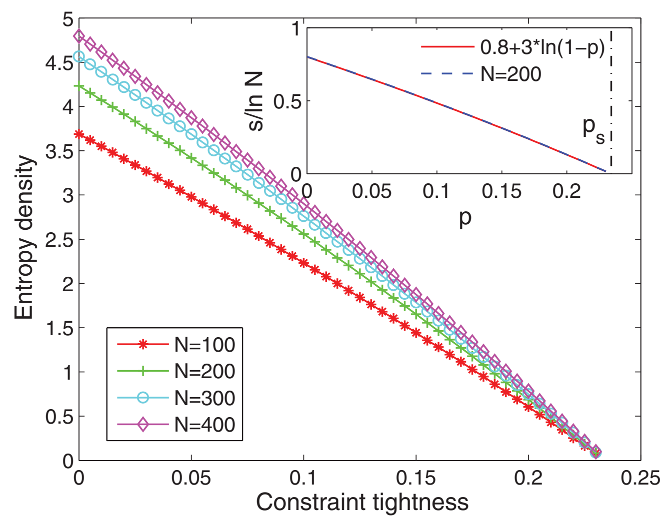

FIG. 1. (Color online) BP entropy densities averaged over 100 random binary Model RB instances with $\alpha = 0.8$ and $r = 3$ as a function of p for $N \in [100,400]$. Inset, BP entropy density divided by $\ln N$ is plotted against p for $N = 200$ (dashed blue line). It is independent of N and vanishes as p crosses p_s. The curve of $s/\ln N$ meets with that of $\ln\langle Z\rangle/(N\ln N)$ for $\alpha = 0.8$ and $r = 3$ (solid red line) as p is varied.

solution should always be stable locally, thus, the condensation transition that is observed in the random k-SAT problem [20] is absent in Model RB. It is possible that, before the SAT-UNSAT transition, one-step replica symmetric breaking (1RSB) with Parisi parameter $m = 1$ has a nontrivial solution and a positive complexity. We cannot study the 1RSB solution for Model RB by population dynamics in the thermodynamic limit or even large systems because the number of states grows rapidly with system size, and without simplification at $m = 1$, to study 1RSB equations on a single instance, one has to use populations on each edge of the graph to represent 1RSB messages, which are computationally heavy and imprecise. So we leave the 1RSB part to our future paper.

III. BP REINFORCED ALGORITHM

We now show the reinforced BP algorithm, which is proved to be a very efficient solver for random binary Model RB instances in the hard region. The idea is to add an external field (with probability $1 - \gamma_t$) into Eq. (2) enforcing the equations toward a solution. We modify Eq. (2) to the following equations:

$$\chi_{\sigma_j}^{j\to a(t+1)} = \frac{1}{Z^{j\to a}}\mu_{\sigma_j}^{j(t)} \prod_{b\in\partial j\backslash a} \psi_{\sigma_j}^{b\to j(t)}, \tag{7}$$

$$\mu_{\sigma_j}^{j(t+1)} = \frac{1}{Z^j}\prod_{b\in\partial j} \psi_{\sigma_j}^{b\to j(t)}. \tag{8}$$

In the updating process, we use $\gamma_t = (1+t)^{-\nu}$, where ν is a parameter that needs to be optimized. Note that, if we choose $\nu = 0$, reinforced BP gives back the original BP. At each iteration, we check if the configuration $\vec{\sigma} = \{\max_{\sigma_i} \mu_{\sigma_i}^i\}$ is a solution. If it is not, we proceed at maximum T (=10^3) times or the iteration converges (precision $\epsilon = 10^{-6}$).

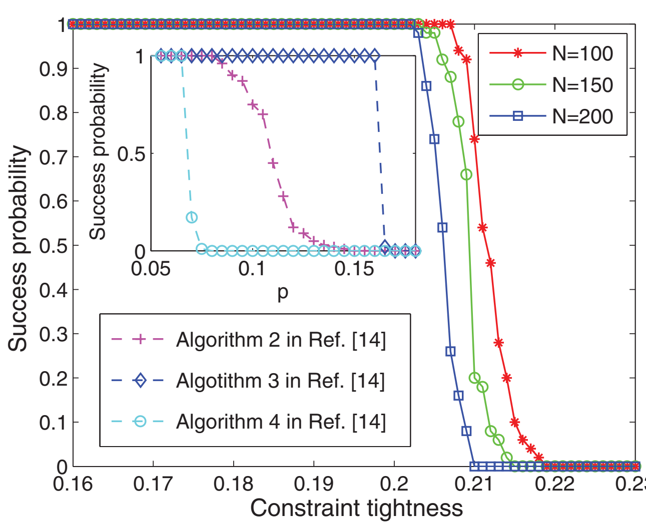

FIG. 2. (Color online) Success probability (averaged over 50 random binary Model RB instances) of the reinforced BP algorithm with the optimal parameter $\nu = 0.10$ as a function of p with $\alpha = 0.8$ and $r = 3$ for $N = 100,150,200$. Inset, the performance of message-passing algorithms guided by BP in Ref. [14] averaged over 100 random binary Model RB instances with $\alpha = 0.8$ and $r = 3$ for $N = 100$.

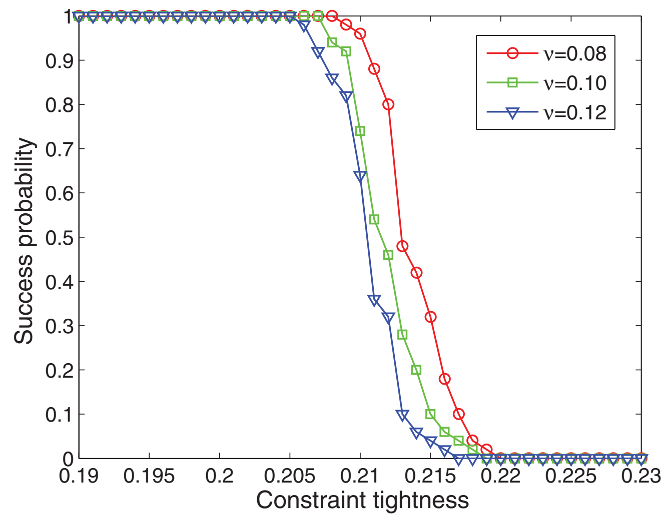

FIG. 3. (Color online) Success probability (over 50 random binary Model RB instances) of the reinforced BP algorithm as a function of p. The curves are plotted for $\alpha = 0.8$, $r = 3$, $N = 100$, and $\nu = 0.08,0.10,0.12$.

In Fig. 2, we plot the fraction of successful runs over 50 random binary Model RB instances with $\alpha = 0.8$ and $r = 3$ for $N \in [100,200]$. We choose the optimal parameter $\nu = 0.10$, and the performance of the algorithm depends on the optimization of ν (see Fig. 3). It is observed that the reinforced BP algorithm can find solutions efficiently when constraint tightness $p < 0.206$ ($p_s \simeq 0.234$). However, it fails at the point $p = 0.210$. We would like to note that, although the reinforced BP algorithm cannot escape from the extreme hard-SAT regime, it works much better than message-passing algorithms guided by BP and other state-of-the-art algorithms (see the inset of Fig. 2) [14]. This indicates that BP reinforcement can be extended to other classes of problems in which variables involve a large number of states.

IV. THE SOLUTION-SPACE STRUCTURE OF MODEL RB

Now, we analyze the geometrical structure of the solution space of Model RB in order to shed light on the performance of the reinforced BP algorithm. Let $d_{\vec{\sigma},\vec{\tau}} \equiv Nx$ ($x \in [0,1]$) be the Hamming distance (number of distinct values) between a solution pair $(\vec{\sigma},\vec{\tau})$ [33], and let $Z(x,p)$ be the number of solution pairs at fixed distance Nx, then, the expectation of $Z(x,p)$ can be written as (similar to the derivation of $\langle Z^2\rangle$ in Ref. [8])

$$\langle Z(x,p)\rangle = N^{\alpha N}\binom{N}{Nx}(N^\alpha - 1)^{Nx}\{(1-p)^2 + p(1-p) \times [(1-x)^k + g(x)]\}^{rN\ln N}\left[1 + O\left(\frac{1}{N}\right)\right], \tag{9}$$

with $g(x) = -k(k-1)x(1-x)^{k-1}/(2N)$. Let $f(x,p) = \lim_{N\to\infty} \ln\langle Z(x,p)\rangle/(N\ \ln\ N)$, and we have

$$f(x,p) = \alpha(1+x) + r\ \ln[(1-p)^2 + p(1-p)(1-x)^k]. \tag{10}$$

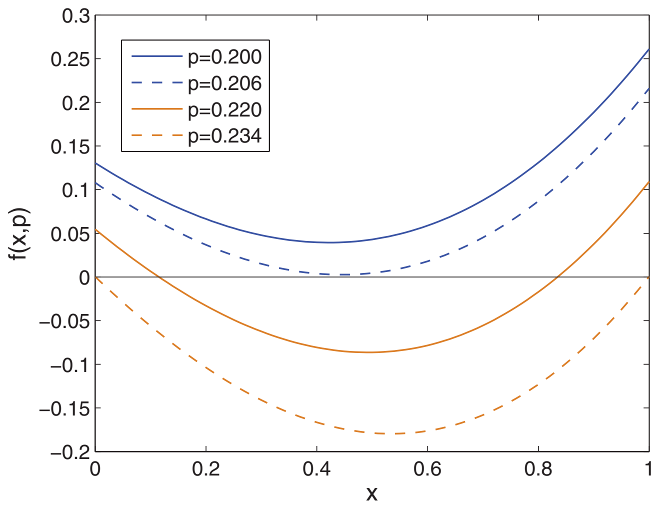

FIG. 4. (Color online) The curves of $f(x,p)$ are plotted against x with $\alpha = 0.8$, $r = 3$, and $k = 2$ for $p = 0.200$ [the solid blue (upper) line], 0.206 [the dashed blue (upper) line], 0.220 [the solid orange (lower) line], and 0.234 [the dashed orange (lower) line]. It is shown that $f(x,p) > 0$ if $p < 0.206$, which means that the solution pairs at distance $x \in [0,1]$ always exist. It is obvious that there is no solution if $p > 0.234$. There exists no solution pair at the distance around 0.5 if $p > 0.206$.

For binary Model RB with $\alpha = 0.8$ and $r = 3$, Fig. 4 shows that $f(x,p) > 0$ for $p < 0.206$, which suggests that the solution pairs at the distance $x \in [0,1]$ always exist. Thus, the solution space of Model RB is a connected cluster when p is below 0.206. Beyond the point 0.206, solution pairs at the distance around 0.5 tend to disappear, which indicates that the set of solutions begins to split into many disconnected clusters. Note that the reinforced BP algorithm efficiently can find a satisfying solution when $p < 0.206$ for random instances of the binary Model RB with $\alpha = 0.8$ and $r = 3$ (see Fig. 2). Therefore, it appears that the algorithm is trapped before the clustering transition point p_c ($\geqslant 0.206$). Here, we suppose that Model RB has the property of $\langle Z^2(x,p)\rangle = \langle Z(x,p)\rangle^2$ induced by $\langle Z^2\rangle = \langle Z\rangle^2$. So, we use the expectation of $Z(x,p)$ instead of $Z(x,p)$.

V. CONCLUSION

To conclude, we have studied Model RB, a random CSP with growing domains, through BP and analytical methods, respectively. It seems that both of the results coincide well with each other. We plan to investigate the 1RSB solution of Model RB in our future paper to describe the solution-space organization in more detail and to exploit the differences between random CSPs with growing domains and with fixed domains. We hope that these studies can provide new insights into the understanding of the hardness in NP-complete problems.

ACKNOWLEDGMENTS

C. Zhao and K. Xu wish to thank Haijun Zhou, Haiping Huang, and Zhiwen Fang for their helpful discussions, and P. Zhang would like to thank Abolfazl Ramezanpour and Riccardo Zecchina for their helpful comments and suggestions. We also thank KITPC (Beijing) for their hospitality. This work has been supported by the International Cooperative Program (Grant No. 2010DFR00700) and the NSFC (Grant No. 60973033).

[1] M. Mézard and A. Montanari, *Information, Physics and Computation* (Oxford University Press, New York, 2009).
[2] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, San Francisco, 1979).
[3] O. Dubois, R. Monasson, B. Selman, and R. Zecchina, Theor. Comput. Sci. **265**, 1 (2001).
[4] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).
[5] M. Weigt and H. Zhou, Phys. Rev. E **74**, 046110 (2006).
[6] L. Zdeborová and F. Krząkała, Phys. Rev. E **76**, 031131 (2007).
[7] L. Dall'Asta, P. Pin, and A. Ramezanpour, Phys. Rev. E **80**, 061136 (2009).
[8] K. Xu and W. Li, J. Artif. Intell. Res. **12**, 93 (2000).
[9] B. M. Smith, Theor. Comput. Sci. **265**, 265 (2001).
[10] A. Frieze and M. Molloy, Random Struct. Algorithms **28**, 323 (2006).
[11] K. Xu and W. Li, Theor. Comput. Sci. **355**, 291 (2006).
[12] K. Xu, F. Boussemart, F. Hemery, and C. Lecoutre, Artif. Intell. **171**, 514 (2007).
[13] T. Liu, X. Lin, C. Wang, K. Su, and K. Xu, in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence* (AAAI, Palo Alto, CA, 2011), pp. 611–616.
[14] C. Zhao, H. Zhou, Z. Zheng, and K. Xu, J. Stat. Mech.: Theory Exp. (2011) P02019.
[15] C. Zhao and Z. Zheng, Inform. Process. Lett. **111**, 985 (2011).
[16] B. M. Smith and M. E. Dyer, Artif. Intell. **81**, 155 (1996).
[17] D. Achlioptas, A. Naor, and Y. Peres, Nature (London) **435**, 03602 (2005).
[18] S. Mertens, M. Mézard, and R. Zecchina, Random Struct. Algorithms **28**, 340 (2006).
[19] M. Mézard, M. Palassini, and O. Rivoire, Phys. Rev. Lett. **95**, 200202 (2005).
[20] F. Krząkała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, Proc. Natl. Acad. Sci. USA **104**, 10318 (2007).
[21] J. Ardelius and L. Zdeborová, Phys. Rev. E **78**, 040101(R) (2008).
[22] H. Zhou and H. Ma, Phys. Rev. E **80**, 066108 (2009).
[23] P. Zhang, Y. Zeng, and H. Zhou, Phys. Rev. E **80**, 021122 (2009).
[24] S. Cai, K. Su, and A. Sattar, Artif. Intell. **175**, 1672 (2011).
[25] A. Braunstein, M. Mézard, and R. Zecchina, Random Struct. Algorithms **27**, 201 (2005).
[26] F. Ricci-Tersenghi and G. Semerjian, J. Stat. Mech.: Theory Exp. (2009) P09001.
[27] A. Braunstein and R. Zecchina, Phys. Rev. Lett. **96**, 030201 (2006).

[28] A. Braunstein, F. Kayhan, G. Montorsi, and R. Zecchina, in *IEEE International Symposium on Information Theory, 2007* (IEEE, Piscataway, NJ, 2007), pp. 1891–1895.
[29] L. Zdeborová and M. Mézard, Phys. Rev. Lett. **101**, 078702 (2008).
[30] L. Dall'Asta, A. Ramezanpour, and R. Zecchina, Phys. Rev. E **77**, 031118 (2008).
[31] J. Chavas, C. Furtlehner, M. Mézard, and R. Zecchina, J. Stat. Mech.: Theory Exp. (2005) P11016.
[32] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, IEEE Trans. Inform. Theory **47**, 498 (2001).
[33] M. Mézard, T. Mora, and R. Zecchina, Phys. Rev. Lett. **94**, 197205 (2005).