Long Range Frustration in Finite–Connectivity Spin Glasses: Application to the random K-satisfiability problem

Haijun Zhou

Max-Planck-Institute of Colloids and Interfaces,
14424 Potsdam, Germany

November 3, 2004

Abstract

A long range frustration index R is introduced to the finite connectivity Viana–Bray $\pm J$ spin glass model as a new order parameter. This order parameter is then applied to the random K-satisfiability (K-SAT) problem to understand its satisfiability transition and to evaluate its global minimum energy. Associated with a jump in R from zero to a finite value, SAT-UNSAT transition in random 3-SAT occurs when the clauses-to-variables ratio α approaches $\alpha_c(3) = 4.1897$. This transition in random 2-SAT occurs at $\alpha_c(2) = 1$, with R remaining to be zero as long as $\alpha < 4.459$. An accumulation of long range frustration in random 3-SAT of $\alpha \geq \alpha_c(3)$ may explain why it is NP-complete; its absence in random 2-SAT of $\alpha_c(2) \leq \alpha < 4.459$ suggests that, the maximum of satisfied clauses in such a system may be determined with times scaling polynomially with system sizes. The zero–temperature phase–diagram of the Viana–Bray $\pm J$ spin glass model is found to be identical to that of the random 2-SAT.

PACS: 89.75.-k, 75.10.Nr, 02.10.Ox

Spin glass statistical physics has been applied to hard combinatorial optimization problems for many years. This interaction between statistical physics and computer science has improved our understanding of NP-completeness and computation complexity [1]; on the other hand, it also stimulates further development of spin glass theories. The present work represents another step along this direction. We introduce a new order parameter to a spin glass model of the random K-satisfiability (K-SAT) problem as well as the $\pm J$ Viana–Bray model [2] and presents new insights on the ground states of such systems.

The random K-SAT is at the root of computation complexity [1]. A K-SAT formula involves N Boolean variables and $M = \alpha N$ constraints or clauses, each of which is a disjunction of K randomly chosen variables or their negations. A K-SAT formula is satisfiable (SAT) if there exists an assignment of the Boolean variables such that all clauses are satisfied; otherwise it is unsatisfiable (UNSAT). For example, the 2-SAT formula $(x_1 \vee \bar{x_2}) \wedge (x_1 \vee x_2) \wedge (\bar{x_1} \vee x_3) \wedge (x_2 \vee \bar{x_3})$ with $N = 3$ and $M = 4$ is SAT through $x_1 = x_2 = x_3 = \texttt{TRUE}$. It was observed [3, 4, 5] that, the probability p of a randomly constructed K-SAT formula being satisfiable drops from $p \simeq 1$ to $p \simeq 0$ over a small range of the clauses-to-variables ratio α around certain α_c (for $K = 3$, $\alpha_c \sim 4.2$ [4, 5, 6]). Furthermore, at the vicinity of α_c, an exponential slowing down in a search algorithm for random K-SAT with $K \geq 3$ was observed [3]. Such threshold phenomena are reminiscent of what is usually observed in a physical system around its phase transition or critical point. Concepts of spin glass physics, such as replica symmetry breaking and proliferation of metastable states, were applied to random K-SAT in some recent contributions [7, 8, 9, 10, 11]. Our understanding of the satisfiability transition, however, is still far from being complete.

In this paper, with the introduction of a long range frustration index R, random K-SAT and the zero temperature phase diagram of the $\pm J$ Viana–Bray model [2] is studied from the viewpoint of long range frustration. We find that at the satisfiability transition point of random 3-SAT $\alpha_c(3) = 4.189724$, there is a jump in R from zero to $R = 0.2605$, while the fraction of unfrozen vertices q_0 drops from unity to $q_0 = 0.5270$. This differs qualitatively from what happens in random 2-SAT. In the later case, R remains to be zero at the two sides of the transition point $\alpha_c(2) = 1$; it becomes positive only when $a \geq 4.459$. We suggest that the NP-completeness of random 3-SAT at $\alpha \geq \alpha_c(3)$ may be due to an accumulation of long range frustration. If this conjecture is correct, then there must be efficient algorithms to determine the maximal number of satisfied clauses for random 2-SAT in the parameter range of $1 \leq \alpha < 4.459$. The zero temperature phase diagram of the $\pm J$ Viana–Bray model is found to be identical to that of the random 2-SAT. Long range frustration in the vertex–cover problem was also studied in an earlier paper [12], where we were able to achieve an apparently exact expression for the minimal vertex–cover size.

We begin with the random K-SAT problem. We follow Ref. [11] and represent a random K-SAT formula by an equivalent spin–glass system of N variable nodes and $M = \alpha N$ function nodes. Each variable node i represents a Boolean variable; it has a binary spin state $\sigma_i = \pm 1$; it is connected to k function nodes, with k obeying the Poisson distribution $f_{K\alpha}(k) = e^{-K\alpha}(K\alpha)^k/k!$ for N sufficiently large. Each function node a represents a a clause (constraint); it is linked to K randomly chosen variable nodes $i_1, \ldots, i_K$; it has energy E_a that is either 0 (clause satisfied) or 1 (clause violated):

$$E_a = \prod_{r=1}^{K} \frac{1 - J_a^r \sigma_{i_r}}{2}, \tag{1}$$

where J_a^r is the edge strength between a and variable node i_r: $J_a^r = -1$ or 1 depending on whether Boolean variable i_r in clause a is negated or not. The edge (a, i_r) is violated (satisfied) if $J_a^r \sigma_{i_r} = -1$ $(= 1)$. In a given random K-SAT formula, J_a^r is a quenched random variable with value equally distributed over ± 1. The total energy of the system for each of the 2^N possible spin configurations is expressed as $E[\{\sigma_i\}] = \sum_{a=1}^{M} E_a$. As we are interested in the SAT-UNSAT transition, only those configurations which have the global minimum energy need to be considered.

In the thermodynamic limit of $N \to \infty$, there can exist an exponential number of ground–energy configurations for a given random K-SAT formula. Following Ref. [13], these configurations are grouped into different (macroscopic) states. A state of the system contains (exponentially) many configurations. These configurations all have the same minimum energy; any two configurations are mutually reachable by first flipping a finite number of spins in one configuration and then *letting the system relax* [12]. We focus on a randomly chosen state β. In this state, the spin values of a total number of $q_+ N$ variable nodes are fixed (frozen) to $\sigma = +1$ in all configurations, and those of another $q_- N$ variable nodes are frozen to $\sigma = -1$, while each of the remaining $q_0 N$ variable nodes has spin value $\sigma = +1$ in some (but not all) configurations and $\sigma = -1$ in the other configurations [12]. Obviously, $q_+ = q_- = (1 - q_0)/2$.

We now address the possibility of long range frustration among these $q_0 N$ unfrozen variable nodes [12]. Question: if we fix the spin value of such a variable node i to $\sigma_i = \sigma_i^*$ with σ_i^* being equally distributed over ± 1, how many unfrozen variable nodes will eventually have their spin values be fixed as a consequence?

The total number s (≥ 1) of affected variable nodes may reach infinity as $N \to \infty$. If this happens, variable node i is referred to as type-I unfrozen and σ_i^* is referred to as its canalizing value. This happens with probability R, with R being our long range frustration index. The infinite percolation clusters evoked by two randomly chosen type-I unfrozen variable nodes i and j must have an infinite intersection, therefore *with probability one-half* their canalizing values σ_i^* and σ_j^* will be in conflict [12]: if $\sigma_i = \sigma_i^*$, then $\sigma_j = -\sigma_j^*$; if $\sigma_j = \sigma_j^*$, then $\sigma_i = -\sigma_i^*$. On the other hand, if s is finite with probability distribution $f(s)$, then variable node i is type-II unfrozen. Obviously, $R = 1 - \sum_{s=1}^{\infty} f(s)$. We find that $f(s)$ is determined by the following self-consistent equation

$$f(s) = f_{\lambda_0}(0)\delta_s^1 + \sum_{k=1}^{\infty} f_{\lambda_0}(k) \sum_{\{s_l\}} \delta_{s_1 + \ldots s_k}^{s-1} \prod_{l=1}^{k} f(s_l), \tag{2}$$

where $\lambda_0 = (3/2)\alpha q_0 (1 - q_0)(1 - R)$ for random 3-SAT and $\lambda_0 = \alpha q_0 (1 - R)$ for random 2-SAT; $f_{\lambda_0}(k)$ is again the Poisson distribution with mean λ_0; δ is the Kronecker symbol. The derivation of Eq. (2) is quite tedious and will be given elsewhere [14]. Here we only mention that, when the spin of a type-II unfrozen variable node i is fixed to $\sigma_i = \sigma_i^*$, on average λ_0 nearest neighboring unfrozen variable nodes will also have their spins fixed (two variable nodes are referred to as nearest neighbors if they share a common function node). From Eq. (2) we realize that R is determined by

$$R = 1 - \exp(-\lambda_0 R). \tag{3}$$

Now we determine the value of the other order parameter q_0. For this purpose, we first construct k new function nodes $a_1, a_2, \ldots, a_k$, with k following the Poisson distribution $f_{K\alpha}(k)$. Each of these function nodes, say a_l, is connected to $K - 1$ randomly chosen variable nodes with strengths equally distributed over ± 1. Before connecting to a_l, the spins of these $K - 1$ variable nodes may already be frozen to specific values that violate all the edges linking to function node a_l. The mean number of such new function nodes is λ_1, with $\lambda_1 = (3/4)\alpha(1-q_0)^2$ for random 3-SAT and $\lambda_1 = \alpha(1-q_0)$ for random 2-SAT. Another important possibility is that some or all of the $K-1$ neighboring variable nodes of a_l were originally unfrozen while the others were already frozen to spin values that violate the corresponding edges to a_l. In this case, a_l is satisfiable by flipping one of its originally unfrozen neighbors. However, a careful analysis leads to the observation that, even in this case, an infinite number of originally unfrozen variable nodes may need to be fixed *in all configurations of state* β if a_l is to be satisfied. The mean number of new function nodes with this property is $2\lambda_2$, with $\lambda_2 = (3/2)\alpha q_0 R(1-q_0+q_0 R/2)$ for random 3-SAT and $\lambda_2 = \alpha q_0 R$ for random 2-SAT. Two such new function nodes, although both are satisfiable, might not be *simultaneously* satisfiable due to long range frustration.

After these k new function nodes are added, we then add a new variable node 0 to the system and connect it to these function nodes with edge strengths equally distributed over ± 1. Suppose we fix the spin value of variable node 0 to $\sigma_0 = +1$ (or $\sigma_0 = -1$), then m of its neighboring function nodes will be violated. This number obeys the following probability distribution $P_{\rm v}(m)$:

$$P_{\rm v}(m) = \sum_{n=0}^{m} f_{\lambda_1/2}(n) P_{\rm f}(m-n), \tag{4}$$

where $P_{\rm f}(n)$ is the probability that, n neighboring function nodes of variable node 0, although all are separately satisfiable, are violated due to long range frustration. The expression of $P_{\rm f}(n)$ is

$$P_{\rm f}(n) = f_{\lambda_2}(2n) C_{2n}^{n} 2^{-2n} + \sum_{n'=2n+1}^{\infty} f_{\lambda_2}(n') C_{n'}^{n} 2^{1-n'}, \tag{5}$$

where $C_{n'}^{n} = n'!/[n!(n'-n)!]$. A detailed derivation of Eqs. (4) and (5) will be given in a later paper [14].

As $N \to \infty$, the fraction of unfrozen variable nodes q_0 equals to the probability for variable node 0 to be unfrozen, namely

$$q_0 = \sum_{m=0}^{\infty} P_{\rm v}(m)^2; \tag{6}$$

and the energy increase due to the addition of variable node 0 is

$$\epsilon_1(\alpha) = \sum_{m=0}^{\infty} \sum_{m'=0}^{\infty} P_{\rm v}(m) P_{\rm v}(m') \min(m, m'). \tag{7}$$

After addition of variable node 0, the system has $N + 1$ variable nodes and, on average, $\alpha N + K\alpha$ function nodes, with clauses-to-variables ratio $\alpha' = \alpha + (K-1)\alpha/(N+1)$. Suppose the energy density ϵ of the system is a function only of α. Then we get from $(N + 1)\epsilon(\alpha') = N\epsilon(\alpha) + \epsilon_1$ the expression

$$\epsilon(\alpha) = \frac{1}{(K-1)\alpha^{1/(K-1)}} \int_0^{\alpha} \tilde{\alpha}^{\frac{2-K}{K-1}} \epsilon_1(\tilde{\alpha}) \mathrm{d}\tilde{\alpha}. \tag{8}$$

In the case of random vertex–cover, the energy density derived through this way is a lower bound when there exists long range frustration. When $R > 0$, an upper bound for the energy density can also be derived. For the enlarged system to have clauses-to-variables ratio α, on average $(K - 1)\alpha$ function nodes must be removed [11]. The energetic contribution of these function nodes must also be removed. This energy comes from two parts: (i) the energy sum of these individual function nodes, $\Delta E = (K-1)\alpha[(1-q_0)/2]^K$ and, (ii) additional energy $\Delta E'$ caused by long range frustration among these function nodes. Therefore,

$$\epsilon(\alpha) = \epsilon_1 - \Delta E - \Delta E'. \tag{9}$$

If we set $\Delta E' = 0$, Eq. (9) gives an upper bound.

The values of the long range frustration index R and the fraction of unfrozen variable nodes q_0 are shown in Fig. 1 for random 3-SAT and in Fig. 2 for random 2-SAT. For random 3-SAT, when $\alpha < \alpha_c(3)$, all the variable nodes are unfrozen ($q_0 = 1$) and not frustrated ($R = 0$). The system is in the SAT phase, with zero energy density (Fig. 3). This solution of $q_0 = 1$ and $R = 0$ is locally stable in all values of α. At $\alpha = \alpha_c(3)$, another stable solution appears: the long range frustration index R suddenly jumps to $R = 0.2605$ and q_0 drops to $q_0 = 0.5270$, suggesting that there is a sudden accumulation of long range frustration among the unfrozen variables. When $\alpha > \alpha_c(3)$, the system is in the UNSAT phase with positive energy density (Fig. 3).

For random 2-SAT, SAT-UNSAT transition occurs at $\alpha_c(2) = 1$. When $\alpha \geq \alpha_c(2)$, the fraction of unfrozen variable nodes gradually decreases from unity and the energy density gradually increases from zero; however, the long range frustration index R remains to be zero for $\alpha < 4.459$, suggesting the absence of long range frustration among the unfrozen variables. Long range frustration only builds up when $\alpha \geq 4.459$ (inset of Fig. 2).

A discontinuity of the order parameter q_0 in random 3-SAT at the SAT-UNSAT transition point was first noticed in [9]. This was suggested to be a

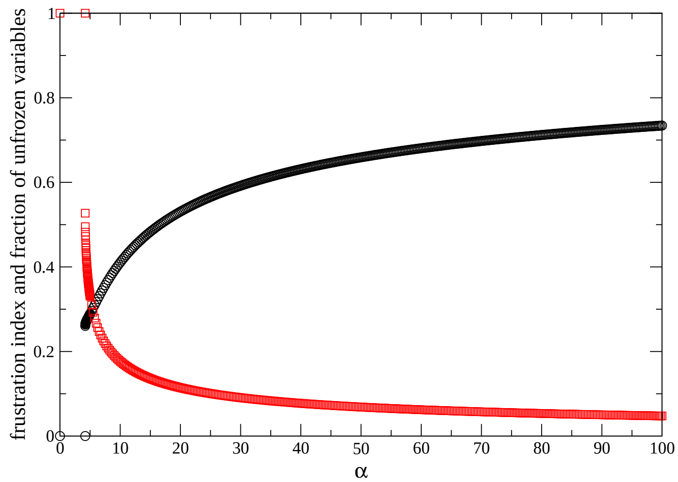

Figure 1: Long range frustration index R (circles) and fraction of unfrozen variables q_0 (squares) as a function of clause-to-variable ratio α for random 3-SAT.

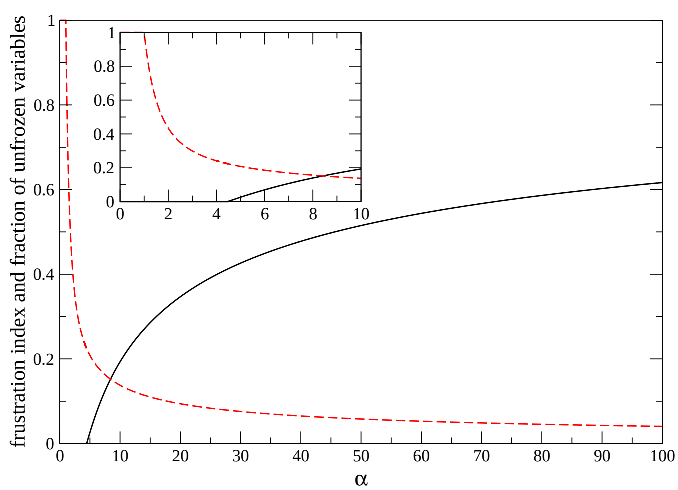

Figure 2: R (solid line) and q_0 (dashed line) as a function of α for random 2-SAT.

possible reason why random 3-SAT is NP-complete at around $\alpha_c(3)$, since q_0 only changes gradually at the SAT-UNSAT transition of random 2-SAT. We suggest that a more deeper reason may be the accumulation of long range frustration in random 3-SAT and its absence in random 2-SAT. Long range frustration causes the spin values of distantly separated variable nodes to be strongly correlated. A local search algorithm is unable to recognize such long range frustration effects; at around $\alpha_c(3)$, a finite fraction of all those possible 2^N configurations might need to be surveyed to draw a concrete conclusion on the satisfiability of a 3-SAT formula. To test whether long range frustration is the true reason for NP-completeness, we suggest the following experiment. Generally speaking, to determine the maximal number of satisfied clauses for random 2-SAT (Max-2-SAT) is NP-complete for $\alpha \geq 1$ [15]. However, as shown in this work, when $1 \leq \alpha < 4.459$, there is no long range frustration in this system. If it is necessary to have long range frustration in order to be NP-complete, then there must a polynomial–time algorithm for random Max-2-SAT. This can be checked by testing known or newly designed algorithms, such as survey propagation [16], to random 2-SAT formulas with known minimal energies.

The SAT-UNSAT transition point for random 2-SAT is predicted to be $\alpha_c(2) = 1$. This is in agreement with known rigorous results [15]. For random 3-SAT, we predict $\alpha_c(3) = 4.189724$. This value is located between the rigorously known lower and upper bounds [6]. It is lower than the mean–field value of 4.267 reported in Ref. [11]. This is unsurprising, since long range frustration of the type discussed here was neglected in [11]. Kirkpatrick and Selman [4] reported $\alpha_c = 4.17 \pm 0.05$ by using computer simulation and finite size scaling. Another value of $\alpha_c \simeq 4.258$ was reported in Ref. [5]. The estimate of [4] may be more plausible. In the random vertex–cover problem, the results obtained by similar finite–size scaling method [17] were in full agreement with theoretical predictions [12].

We now briefly discuss the zero temperature phase diagram of the $\pm J$ Viana–Bray model [2]. The model is characterized by the Hamitonian

$$E[\{\sigma_i\} = \sum_{(ij)} J_{ij}\sigma_i\sigma_j, \qquad (10)$$

where the summation is over all the edges (ij) of a Poisson random graph of mean vertex degree c; the edge strength $J_{ij} = \pm 1$, with equal probability; and $\sigma_i = \pm 1$ is the spin value on vertex i. The zero temperature phase diagram

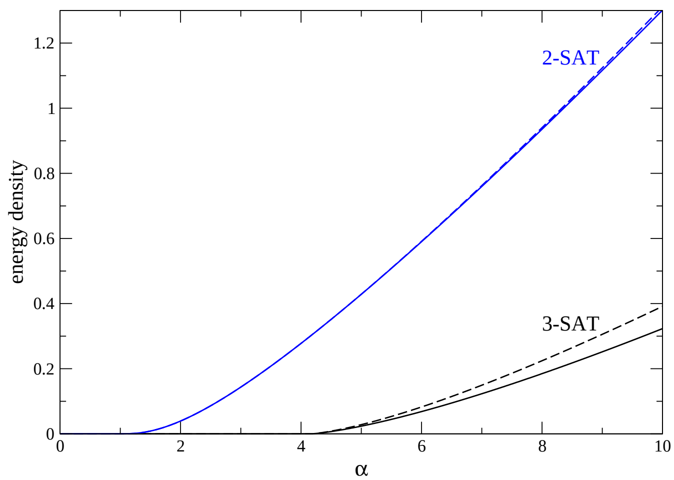

Figure 3: The energy density lower bound (solid lines) and upper bound (dashed lines) for random 3-SAT and 2-SAT as a function of α.

of this model was first reported in Ref. [18]. It was found that, when $c < 1$, the system is in the paramagnetic phase with no frozen vertices, i.e., $q_0 = 1$; when $c \geq 1$, some vertices become to be frozen to positive or negative spin values, with q_0 gradually decreasing from unity.

The Hamitonian Eq. (10) is similar to that of the random 2-SAT and can be studied by the same method mentioned in the previous part of this paper [14]. Actually, we find that the zero temperature phase diagram of Eq. (10) is identical to Fig. 2 of the random 2-SAT: the transition between the paramagnetic phase and the spin glass phase occurs at $c = 1$, confirming the result of Ref. [18]; in the spin glass phase, long range frustration only builds up when the mean vertex degree $c \geq 4.459$. We also noticed that when the long range frustration index is set to $R = 0$, the order parameter q_0 in Eq. (6) has the same expression as Eq. (15) of Ref. [18].

In summary, we have investigated the possibility of long range frustration in the random K-SAT problem as well as in the finite–connectivity $\pm J$ spin glass model. We found that SAT-UNSAT occurs at $\alpha_c = 4.1897$ and $\alpha_c = 1$ for random 3-SAT and 2-SAT, respectively. The SAT-UNSAT transition is associated with a discontinuity in the order parameters in the 3-SAT systems, namely with a jump in the long range frustration index R and a drop in the fraction of unfrozen variables q_0. The SAT-UNSAT transition in random 2-SAT is associated only with a gradual increase in one order parameter q_0 while R remains to be zero. We suggested that the occurrence of long range frustration may be the real reason why at $\alpha \sim 4.2$, search algorithms for random 3-SAT usually needs exponential computation times to reach a conclusion. As a test of this proposal, one can check whether the random MAX-2-SAT in the range of $1 \leq \alpha < 4.459$ is computationally easy or not. The zero temperature phase diagram of $\pm J$ Viana–Bray model is identical to that of the random 2-SAT.

If long range frustration of the type discussed in this article and in Ref. [12] is really at the heart of computational complexity of NP-complete combinatorial optimization problems, then algorithms must be explored to try to trace such frustration effects. This is anticipated to be a major challenges to theoretical computer scientists.

The method used in the present work may also be applicable to other finite–connectivity spin glass models [2]. As in Ref. [12], we have not yet addressed the issue of multiple macroscopic states. We expect that an appropriate combination of long range frustration and the cavity method at the first order replica symmetry breaking level [19, 13] will offer the statistical physics community a powerful tool to advance our understanding of spin glass statistical physics.

The author acknowledges the kind support of Professor Reinhard Lipowsky. He benefits from an stimulating discussion with Professor Lu Yu at the initial stage of this exploration. He also receives momentum from Dr. Jing Han and Mr. Ming Li.

References

[1] Garey, M & Johnson, D. S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness.* (Freeman, San Francisco).

[2] Viana, L & Bray, A. J. (1985) *J. Phys. C: Solid State Phys.* **18**, 3037–3051.

[3] Cheeseman, P, Kanefsky, B, & Taylor, W. (1991) *Where the really hard problems are.* in Proc. 12th Int. Conf. on Artif. Intelligence. eds. Reiter, R & Mylopoulos, J. (Morgan Kaufmann, San Mateo, CA), pp. 331–337.

[4] Kirkpatrick, S & Selman, B. (1994) *Science* **264**, 1297–1301.

[5] Crawford, J. M & Auton, L. D. (1996) *Artif. Intell.* **81**, 31–57.

[6] Achlioptas, D. (2001) *Theor. Comput. Sci.* **265**, 159–185.

[7] Monasson, R & Zecchina, R. (1996) *Phys. Rev. Lett.* **76**, 3881–3885.

[8] Monasson, R & Zecchina, R. (1997) *Phys. Rev. E* **56**, 1357–1361.

[9] Monasson, R, Zecchina, R, Kirkpatrick, S, Selman, B, & Troyansky, L. (1999) *Nature* **400**, 133–137.

[10] Mézard, M, Parisi, G, & Zecchina, R. (2002) *Science* **297**, 812–815.

[11] Mézard, M & Zecchina, R. (2002) *Phys. Rev. E* **66**, 056126.

[12] Zhou, H. (2004) Long range frustration in a spin glass model of the vertex–cover problem (submitted to Phys. Rev. Lett. on September 01, 2004).

[13] Mézard, M & Parisi, G. (2003) *J. Stat. Phys.* **111**, 1–34.

[14] Zhou, H. (2004) (in preparation).

[15] Fernandez de la Vega, W. (2001) *Theor. Comput. Sci.* **265**, 131–146.

[16] Braunstein, A, Mézard, M, & Zecchina, R. (2002) Survey propagation: an algorithm for satisfiability (e-print: cs.CC/0212002).

[17] Weigt, M & Hartmann, A. K. (2000) *Phys. Rev. Lett.* **84**, 6118–6121.

[18] Kanter, I & Sompolinsky, H. (1987) *Phys. Rev. Lett.* **58**, 164–167.

[19] Mézard, M & Parisi, G. (2001) *Eur. Phys. J. B* **20**, 217–233.