Long Range Frustrations in a Spin Glass Model of the Vertex Cover Problem

Haijun Zhou

Max-Planck-Institute of Colloids and Interfaces, 14424 Potsdam, Germany and
Institute of Theoretical Physics, the Chinese Academy of Sciences, Beijing 100080, China

(Dated: February 2, 2008)

In a spin glass system on a random graph, some vertices have their spins changing among different configurations of a ground–state domain. Long range frustrations may exist among these unfrozen vertices in the sense that certain combinations of spin values for these vertices may never appear in any configuration of this domain. We present a mean field theory to tackle such long range frustrations and apply it to the NP–hard minimum vertex cover (hard–core gas condensation) problem. Our analytical results on the ground–state energy density and on the fraction of frozen vertices are in good agreement with known numerical and mathematical results.

PACS numbers: 75.10.Nr, 89.75.-k, 05.20.-y, 02.10.Ox

The energy landscape of a large spin glass system is very complex. There may exist (exponentially) many ground–state and metastable domains in the configurational space; these domains are mutually separated from each other by (infinitely) high energy barriers. At low temperature, the system may get trapped in one of these configurational domains, and ergodicity is broken. In the cavity field formalism \[1\], of mean field theory of finite temperature, the system may get trapped in one of these macroscopic states (hereafter, a macroscopic state is simply referred to as a ‘state’ and a microscopic configuration as a ‘configuration’). In a given state each vertex \(i\) feels a cavity field \(h_i\) that may be different for different vertices, and the fluctuation of this field among all the states is characterized by a probability distribution \(P_i(h_i)\) that again may be different for different vertices.

The ground–state energy landscape of a spin glass system can be studied by the zero temperature limit of the cavity field theory \[2\]. In this limit and in a given state \(\alpha\), the spin value \(\sigma_i\) of a vertex \(i\) either is positively frozen \((\sigma_i = +1\) in all configurations\) or is negatively frozen \((\sigma_i = -1\) in all configurations\) or is unfrozen \((\sigma_i\) fluctuates over \(\pm1\) among configurations of state \(\alpha\)). A crucial assumption of the cavity field theory \[2\] is that, with probability unity each of the \(2^n\) combinations of spin values for \(n\) randomly chosen unfrozen vertices is realized in configurations of state \(\alpha\). However, this conventional cavity field theory leads to negative values of structural entropy \(\Sigma\) \[2\] when loops of spin–spin interactions become abundant (see, e.g., \[3\]) or even causes certain type of divergence in the population dynamics \[7\]. To overcome these difficulties, a positive re-weighting parameter \(y\) can be introduced and its value be determined self–consistently by requiring \(\Sigma(y) = 0\) \[2\]. This procedure is however not quite satisfactory; in case of the vertex cover problem, it predicts a ground–state energy that is systematically lower than the actual value \[3\].

Here we discuss the possibility of long range correlations among spins of different unfrozen vertices. Both the spins of two unfrozen vertices \(i\) and \(j\) will certainly fluctuate among configurations of a state \(\alpha\). On the other hand, we find that with certain probability \(\sigma_i\) and \(\sigma_j\) may be prohibited to take certain combination of values (e.g. \(\sigma_i = \sigma_j = -1\) in all configurations of state \(\alpha\), even if \(i\) and \(j\) are far apart from each other in terms of shortest path length. To detect such long range frustrations among unfrozen vertices, our idea is to flip the spin of one unfrozen vertex and then check whether this perturbation propagates to other unfrozen ones. This Letter reports our calculations on a spin glass model \[8\] of the NP-hard minimum vertex cover problem \[12\], which is equivalent to the hard–core gas condensation of physics \[13\]. A long range frustration order parameter \(R\) is defined. In this model the quenched randomness comes from the underlying random graph. Work on systems with additional quenched randomness of spin–spin interactions is reported in an accompanying paper \[11\].

For the vertex cover model, we show that long range frustration builds up \((R > 0)\) when the mean vertex degree \(c\) of the graph exceeds \(c = e = 2.7183\). Analytical predictions on the ground–state energy density and on the fraction of frozen vertices are both in very good agreement with known numerical and mathematical results. The calculations are carried out through the cavity approach. It remains open whether the same results are achievable by the replica method. Our approach is essentially replica symmetric in the sense that (a) we focus attention on just one of all possible macroscopic states, (b) the statistical property of this state is specified by just three mean field parameters to be defined, \(R\), \(q_+\), and \(q_0\). Competitions among multiple states will be included in the theory in future work.

We first introduce the random graph vertex cover problem. A random graph \(G(N, c)\) has \(N\) vertices; and between any two vertices an edge is present with probability

\[c\]
are randomly picked up from $\alpha$ negatively frozen, and unfrozen vertices in state $\pm$. The total number of affected vertices may also be finite. In their spins as a consequence?

The percolation clusters evoked by two type-I unfrozen probability for this to happen is denoted as $h$ happens, vertex $\sigma$ vertex cover problem consists of finding a vertex cover $\Lambda$ with size $|\Lambda| \leq n_0$, $n_0$ being a prescribed integer. This problem is mapped to a spin glass model with energy functional

$$E[\{\sigma_i\}] = -\sum_{i=1}^{N} \sigma_i + \sum_{(i,j) \in E(G)} (1+\sigma_i)(1+\sigma_j) \cdot (1)$$

$\sigma_i = -1$ if vertex $i \in \Lambda$ (covered) and $\sigma_i = +1$ otherwise.

The ground–state configurations of model (1) correspond to vertex cover patterns with the global minimum size $\tilde{R}$. These configurations may be grouped into different states $\tilde{R}$. Two configurations in the same state are mutually reachable by flipping a finite number of spins in one configuration and then letting the system relax. (According to this definition of states, two configurations of the same state can have a Hamming distance scaling linearly with system size $N$.) Let us focus on one state, say $\alpha$. In state $\alpha$, the spin value of a randomly chosen vertex $i$ may be fixed to $\sigma_i \equiv +1$, or to $\sigma_i \equiv -1$, or fluctuate over $\pm 1$. The fraction of positively frozen, negatively frozen, and unfrrozen vertices in state $\alpha$ is $q_+$, $q_-$, and $q_0$ respectively. (By the way, we notice that in the minimum vertex cover problem, the parameters $(q_+, q_-, q_0)$ are the same for different ground–state states, due to the fact that the energy density is determined by Eq. (1).) The probability that among $k$ vertices that are randomly picked up from $G(N,c)$ are unfrrozen, $k_+$ positively frozen, and $k_-$ negatively frozen is $k!(k_0 k_1 k_2 \ldots) q_0^{k_0} q_+^{k_+} q_-^{k_-}$ (in the large $N$ limit).

Since the spin of an unfrrozen vertex $i$ fluctuates among different configurations of state $\alpha$, the ‘correlation length’ of this fluctuation is an important issue. We ask the following question: If $\sigma_i$ is externally fixed to $\sigma_i = -1$, how many other unfrrozen vertices must eventually fix their spins as a consequence?

For a random graph of size $N \to \infty$, the total number $s$ of affected vertices may scale linearly with $N$. If this happens, vertex $i$ is referred to as type-I unfrrozen. The probability for this to happen is denoted as $R$ (which defines our long range frustration order parameter). The total number of affected vertices may also be finite. In this case, vertex $i$ is type-II unfrrozen. Based on insights gained from studies on random graphs [17], we know that the percolation clusters evoked by two type-I unfrrozen vertices have a nonzero intersection (of size proportional to $N$). Therefore, the spin values of all the type-I unfrrozen vertices must be strongly correlated. If we randomly choose two type-I unfrrozen vertices $i$ and $j$, then with probability one-half their spin values can not be negative simultaneously: if $\sigma_i = -1$, then $\sigma_j$ must be $+1$; if $\sigma_j = -1$, then $\sigma_i$ must be $+1$. On the other hand, two randomly chosen type-II unfrrozen vertices are mutually independent, since each vertex can only influence the spin values of $s \sim O(1)$ other unfrozen vertices while the shortest path length between two randomly chosen vertices of $G(N,c)$ scales as $\ln N$ and becomes divergent when $N \to \infty$ [17]. Denote $f(s)$ as the probability that a randomly chosen unfrozen vertex $i$, when flipped to $\sigma_i = -1$, will eventually fix the spin values of $s$ unfrrozen vertices with $s$ being finite and therefore $\lim_{N \to \infty} s/N = 0$.

We calculate the parameters $q_0$, $q_+$, $q_-$ by the cavity field method [1, 5]: First a random graph $G(N,c)$ is generated; then a new vertex $i$ is connected to a set $V_i$ of $k$ randomly chosen vertices in $G(N,c)$, $k$ following the distribution $P_c(k)$: the unfrrozen/frozen probabilities $\{q_0(i), q_+(i), q_-(i)\}$ of vertex $i$ in the enlarged graph (denoted as $G'$) is then calculated. We assume the following convergence condition: $\lim_{N \to \infty} \{q_0(i), q_+(i), q_-(i)\} = \{q_0, q_+, q_-\}$. This enables us to write down a set of self-consistent equations in the large $N$ limit.

If the new vertex $i$ is positively frozen, then none of the vertices in $V_i$ are positively frozen in graph $G$. Furthermore, there are two possible situations: (i) no vertices in $V_i$ is type-I unfrrozen in $G$; or (ii) some of the vertices in $V_i$ are type-I unfrrozen. In case (ii), all these type-I unfrrozen vertices will take spin value $-1$ simultaneously in some configurations of state $\alpha$, so that vertex $i$ will have $\sigma_i \equiv +1$ as it is added into the system. With this analysis, we get a self-consistent equation for $q_+$:

$$q_+ = \sum_{k=1}^{\infty} P_c(k) \sum_{l=1}^{k} C_k^l (q_0 R)^l (q_0 (1 - R) + q_-)^{k-l} 2^{1-l}$$

$$+ \sum_{k=0}^{\infty} P_c(k) (q_0 (1 - R) + q_-)^k$$

$$= 2 e^{-c_{q_+} - (1/2)c_{q_0} R - e^{-c_{q_+} - c_{q_0} R}} , \quad (2)$$

where $C_k^l = k!/(l!(k-l)!)$. Equations (2), (3), (4) and (5) are derived elsewhere [13].

If the new vertex $i$ is unfrrozen, there are also two possibilities concerning the spin values of vertices in $V_i$: (iii) none of them is positively frozen in $G$; or (iv) one of them is positively frozen in $G$. To ensure vertex $i$ will be unfrrozen, in situation (iii) two or more of the vertices in $V_i$ must be type-I unfrrozen in $G$, among which one is in conflict with all the others; and in situation (iv) some of the vertices in $V_i$ may be type-I unfrrozen in $G$, but they must be capable of taking spin value $-1$ simultaneously. Therefore, we get a self-consistent equation for $q_0 [13]$: \begin{align}
q_0 &= \left( 2c_{q_+} + c_{q_0} R \right) e^{-c_{q_+} - (1/2)c_{q_0} R} \\
&\quad - \left( c_{q_+} + c_{q_0} R + (c_{q_0} R)^2/4 \right) e^{-c_{q_+} - c_{q_0} R} . \quad (3)
\end{align}

If the spin of an unfrrozen vertex $i$ is flipped to $\sigma_i = -1$, it may not affect any other vertices ($s = 0$), provided
its local environment is described by the above mentioned situation (iii). This happens with probability $p_1 = 1 - cq_0^2 / q_0$. With probability $1 - p_1$, the unfrozen vertex $i$ encounters a local environment of type (iv), that is, one of its nearest neighbors vertex $j$ is positively frozen in graph $G$. This vertex $j$ must face the local environment of type (i) in graph $G$ if vertex $i$ is type-II unfrozen. (If vertex $j$ has the local environment of type (ii), flipping the spin value of vertex $i$ to $\sigma_i = -1$ would cause a percolation cluster of size proportionally to $N$.) With these preparations, we obtain the following self-consistent equation for the distribution $f(s)$:

$$f(s) = p_1 \delta_s^0 + (1 - p_1) \sum_{l=0}^1 P_c(l) \prod_{m} f(s_m) \delta_{s-1}^{l_1+l_2+\ldots+s_l}.$$  

In Eq. (4), $\delta$ is the Kronecker symbol; and $c' = cq_0(1 - R)$ is the mean number of type-II unfrozen vertices adjacent to a positively frozen vertex. Since $R = 1 - \sum_{s=0}^\infty f(s)$, we establish that the long range frustration order parameter $R$ is determined by the following equation:

$$R = \left(\frac{cq_0^2}{q_0} \right) \left(1 - e^{-cq_0 R (1 - R)}\right).$$  

A positive $R$ signifies the appearance of a percolation cluster of unfrozen vertices whose spin values are strongly correlated.

Figure 1 shows the value of $R$ as a function of mean vertex degree $c$. $R \equiv 0$ when $c \leq e$; this is consistent with Ref. 10 that, a minimal vertex cover pattern can be found by a polynomial leaf-removal algorithm. When $c > e$, a finite fraction of the unfrozen vertices are long-rangely frustrated; the leaf-removal algorithm outputs a looped subgraph 19. At mean vertex degree $c \approx 40$, the order parameter $R$ reaches a maximal value; then it gradually decays as $c$ is further increased.

The fraction of vertices that are covered in a minimal vertex cover is

$$X_{\text{min}} = 1 - q_+ - q_0/2.$$  

At large $c$ values, Eq. (3) is in agreement with a rigorous asymptotic expression given by Frieze 20; at low values of $c$, it is in agreement with the exact enumeration results of Weigt and Hartmann 13. These excellent agreements are quite encouraging, in view of the fact that all previous efforts failed 4, 13, 14. It has already been established that when $c > e$ the replica symmetric solution of vertex cover problem becomes unstable 13, 14; but earlier replica symmetry breaking solutions either resulted in negative structural entropy or predicted a minimal vertex cover size noticeably lower than the actual value 4.

So far we have focused on only one ground-state state of the vertex cover problem. When $c > e$ it is believed that there are many such states (replica symmetry breaking). This is consistent with our observation that, when $c > e$ the fraction of frozen vertices ($= q_+ + q_-$, dashed lines in Fig. 2) in one state is much higher than the actual fraction of frozen vertices estimated numerically (symbols in Fig. 2). This is easy to understand: A frozen vertex in one state may be unfrozen or be frozen to the opposite spin value in another state. At the moment we are unable to construct a theory to include the competitions among different states. As a first attempt, we make the following conjectures: (a) if a vertex is positively frozen in one state, it is positively frozen in all states; and (b) a vertex is negatively frozen in all states only if it is adjacent to two or more positively frozen vertices. Then an expression on the fraction of frozen vertices is obtained 18:

$$\Gamma = q_+ + 1 - e^{-cq_+} - cq_+ e^{-cq_+}.$$  

The agreement of Eq. (7) with the numerical data of Ref. 14 is quite good (Fig. 2). This is an issue to be understand more deeply.
FIG. 3: Fraction of frozen vertices in all macroscopic states (Eq. (7), solid lines) and its comparison with the numerical results of Ref. [14] (symbols) and the fraction of frozen vertices (Eq. (7), solid lines) and its comparison with the numerical results of Ref. [14] (symbols) and the fraction of frozen vertices in one macroscopic state (dashed lines).

To summarize, we have studied long range frustrations among unfrozen vertices in a macroscopic state of a spin glass system. We found that, with certain probability, the fluctuations of the spin values of two or more distinctly separated unfrozen vertices are highly correlated. A long range frustration order parameter \( R \) was calculated to quantify this strong correlation. When applying our method to the NP-hard minimum vertex cover (hard-core gas condensation) problem, the analytical predictions concerning the ground-state energy density and the fraction of frozen vertices are in good agreement with known numerical and rigorous results. The basic idea behind this paper is also applicable to other spin glass systems.

We emphasize that the appearance of many macroscopic states in the energy landscape of a spin glass system does not necessarily mean the existence of long range frustrations among unfrozen vertices in a single macroscopic state. As an counter-example, in the maximum matching problem, there is no long range frustrations \( (R \equiv 0) \) but there exist an exponential number of macroscopic states. It is interesting to notice that the maximum matching problem can be solved by polynomial algorithms. It appears that, the proliferation of macroscopic states is not the real reason of the computational complexity in finding a ground-state configuration for a disordered system. As another example, there are many macroscopic states in a typical random 3-satisfiability formula when \( 3.921 < \alpha < 4.267 \) (here \( \alpha \) is the clauses-to-variables ratio); but the survey propagation algorithm is able to find a solution efficiently.

On the other hand, we believe the existence of long range frustrations among unfrozen vertices will make it intrinsically difficult for a search algorithm to find a ground-state configuration. Because of these long range effects, it is difficult (a) to determine whether a vertex is frozen or unfrozen in a macroscopic state and, (b) to trace the percolation cluster associated with a given unfrozen vertex. Recently, some NP-hard combinatorial optimization problem in computer science were studied by zero temperature cavity field method. We hope the present work, besides improving our understanding of finite connectivity spin glasses, will stimulate further efforts in finding more efficient algorithms. We are presently implementing the physical picture of this paper into an algorithm for the vertex cover problem.

H.Z. is grateful to R. Lipowsky, Z.-C. Ou-Yang, and L. Yu for their support, to A. K. Hartmann and M. Weigt for sharing their numerical results, and to L. Yu and R. Lipowsky for helpful discussions.

[1] M. Mézard and G. Parisi, Eur. Phys. J. B 20, 217 (2001).
[2] L. Viana and A. J. Bray, J. Phys. C 18, 3037 (1985).
[3] I. Kanter and H. Sompolinsky, Phys. Rev. Lett. 58, 164 (1987).
[4] R. Monasson, J. Phys. A 31, 513 (1998).
[5] M. Mézard and G. Parisi, J. Stat. Phys. 111, 1 (2003).
[6] M. Mézard, G. Parisi, and R. Zecchina, Science 297, 812 (2002).
[7] M. Mézard and R. Zecchina, Phys. Rev. E 66, 056126 (2002).
[8] R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. Lett. 89, 268701 (2002).
[9] H. Zhou, Eur. Phys. J. B 32, 265 (2003).
[10] O. Rivoire, J. Stat. Phys. 117, 453 (2004).
[11] H. Zhou, New J. Phys. 7, 123 (2005).
[12] M. Garey and D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness (Freeman, San Francisco, 1979).
[13] M. Weigt and A. K. Hartmann, Phys. Rev. Lett. 84, 6118 (2000).
[14] M. Weigt and A. K. Hartmann, Phys. Rev. E 63, 056127 (2001).
[15] A. K. Hartmann and M. Weigt, J. Phys. A 36, 11069 (2003).
[16] G. Biroli and M. Mézard, Phys. Rev. Lett. 88, 025501 (2002).
[17] B. Bollobás, Random Graphs (Academic Press, London, 1985).
[18] See EPAPS Document No. E-PRLTAO-94-030524 for technical details of the calculation. A direct link to this document may be found in the online article’s HTML reference section. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html) or ftp.aip.org in the directory /epaps/. See the EPAPS homepage for more information.
[19] M. Bauer and O. Golinelli, Eur. Phys. J. B 24, 339 (2001).
[20] A. M. Frieze, Discrete Math. 81, 171 (1990).
[21] H. Zhou and Z.-C. Ou-Yang, Maximum matching on random graphs, e-print: cond-mat/0309348 (2003).