Cavity approach to sphere packing in Hamming space

A. Ramezanpour  
Physics Department and Center for Computational Sciences,  
Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

R. Zecchina  
Physics Department and Center for Computational Sciences,  
Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy  
Human Genetics Foundation, Torino, via Nizza 52, 10126 Torino, Italy and  
Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy  
(Dated: December 21, 2013)

In this paper we study the hard sphere packing problem in the Hamming space by the cavity method. We show that both the replica symmetric and the replica symmetry breaking approximations give maximum rates of packing that are asymptotically the same as the lower bound of Gilbert and Varshamov. Consistently with known numerical results, the replica symmetric equations also suggest a crystalline solution, where for even diameters the spheres are more likely to be found in one of the subspaces (even or odd) of the Hamming space. These crystalline packings can be generated by a recursive algorithm which finds maximum packings in an ultra-metric space. Finally, we design a message passing algorithm based on the cavity equations to find dense packings of hard spheres. Known maximum packings are reproduced efficiently in non trivial ranges of dimensions and number of spheres.

I. INTRODUCTION

The problem of packing rigid objects, and spheres in particular, is a fundamental problem which appears across disciplines\cite{1,2}. In general, the objects could have arbitrary shapes and the ambient space can be an abstract space $\Lambda$. Given the space and the objects, the main question is that of finding the densest packings.

In coding theory one is interested in finding an optimal representation of $N$ symbols in binary strings of length $n$, that is $\Lambda = \{0,1\}^n$ is the Hamming space of dimension $n$. This optimal coding contains as many as possible symbols and ensures that after transmitting through a noisy channel, which at most flips $\frac{d-1}{2}$ variables, one can recover the original messages. The ratio between the characteristic length of the symbols $l_c \equiv \log_2 N$ and length of the transmitted strings $n$ defines the rate of coding (or packing). The maximum rate of coding is denoted by $R$. Indeed people are interested to know the asymptotic form of $R$ when $n,d \to \infty$ and $\delta \equiv d/n$ remains constant. So far there is a considerable difference between the best lower and upper bounds for $R$\cite{3,10}. This means that for large $n$ the lower bound and upper bounds for the number of symbols differ by a factor of order $2^n$.

Physically, one can consider the set of symbols as a system of identical particles in the Hamming space of dimension $n$, interacting by a hard core potential of range $d$\cite{11,12}. The aim is then to study the physical states of different densities. Clearly for small densities the system is in the liquid phase respecting the translational symmetry. In this case it is easy to calculate, for example, the entropy. At higher densities the liquid entropy becomes incorrect (negative) signaling the onset of other stable phases, either crystalline or glassy\cite{13,14}.

In this study we formulate the packing problem as a constraint satisfaction problem. We consider $N$ variables (the physical particles or the strings of symbols) which take values in $\Lambda$ and for each pair of the variables we consider a constraint that forbids overlapping assignments of the two variables. A packing is thus an assignment of the variables that satisfies all the constraints. This representation differs substantially with the so called lattice gas models where binary variables (representing occupied or empty positions) are defined one each point of the space $\Lambda$.

The cavity method provides analytical and numerical tools which can be extremely useful in solving optimization problems (or constraint satisfaction problems) over random structures\cite{15,16}. In certain cases it is known to provide sampling results which cannot be obtained by Monte Carlo Markov chains in subexponential times [e.g. optimization problems in the one-step replica symmetry breaking (1RSB) phase]. As an optimization tool it often outperforms linear programming methods\cite{20}. The development of such algorithm is, however, by no means obvious due to the choice of the representation of the problem and to the need of writing the cavity equations in an algorithmically efficient form.

As we shall discuss in this paper, insights from the application of the cavity method will turn out to be also useful in the study of the type of packing problems we are interested in.

In this paper we study the cavity equations for the packing problem in the replica symmetric (RS) and in the one-step replica symmetry breaking (1RSB) approximations\cite{19}. These equations are called belief propagation (BP) and survey propagation (SP) equations, respectively\cite{17,21}. In the RS approximation, besides a liquid solution we
find a crystalline phase where with higher probability spheres are found in one of the sublattices (even or odd) of the Hamming space. This phase has already been observed in Monte Carlo simulations of Ref. [14]. Both the liquid and crystalline solutions predict a maximum rate of packing that behaves asymptotically like the best known lower bound [10]. The same result has been obtained in Ref. [12] where the liquid entropy is computed in the hypernetted chain approximation. To discuss the exactness of the cavity free entropy we also consider some interpolation techniques which connect the cavity free entropy to the true entropy of the system [22–24].

In the 1RSB case, we provide an approximate solution of the SP equations to calculate the configurational entropy, defined as the (log of) the number of pure states or clusters in the solution space. This quantity acquires a nonzero value at a clustering transition where the liquid entropy is still positive. The maximum rate of packing that is achievable still coincides with the one obtained in the RS approximation.

Finally we design a message passing algorithm based on the BP equations which allows us to find dense packings in not too large dimensions. These packings are typically hard to find by other simple methods like Monte Carlo based algorithms. Unfortunately the computation time and memory increase exponentially with the space dimension \( n \), making larger dimensions very difficult to explore. To partially overcome this problem we introduce an approximate update rule for the message passing algorithm which is restricted to a subspace and makes the computation more efficient. This improvement, together with the distributive nature of the algorithm, could help to run the algorithm in larger dimensions. We also discuss another iterative algorithm that, given a packing configuration of spheres with diameter \( d \), finds another packing for larger diameter \( d + 1 \) by increasing the space dimension \( n \). This is a polynomial time algorithm generating maximum packings in an ultrametric space. When applied to our problem, we find crystalline packings predicted by the RS cavity equations for even sphere diameters.

The structure of the paper is as follows. In Sec. II we define the problem more precisely and give a summary of known results. In Secs. III and IV we present the BP and SP equations and study their consequences for the hard sphere packing problem. In Sec. V we study the packing algorithms and Sec. VI is devoted to extension to the \( q \)-ary Hamming spaces. Finally the concluding remarks are given in Sec. VII. In the first two appendixes we give the details of calculations for checking the stability of the BP solutions and deriving the SP equations. The interpolation methods and some of their properties are presented in Appendix C.

II. DEFINITIONS AND KNOWN RESULTS

Consider \( N \) hard spheres of diameter \( d \) indexed by \( i = 1, \ldots, N \) and a Hamming space of dimension \( n \). The set of points in this space are denoted by \( \Lambda = \{0, 1\}^n \) with size \( |\Lambda| = 2^n \). We index the points in this space by \( \sigma \). A point can be represented by a binary vector of \( n \) elements \( \sigma \in \{0, 1\} \). The Hamming space can be partitioned into two subspaces: even and odd. Points in the even subspace have an even number of 1’s (even parity) and those in the odd subspace have an odd number of 1’s (odd parity). The Hamming distance \( D(\sigma, \sigma') \) between two points \( \sigma \) and \( \sigma' \) is equal to the number of different elements in the binary representation of the two points. For each point \( \sigma \) we define the set \( V_d(\sigma) \) as

\[
V_d(\sigma) \equiv \{\sigma'|D(\sigma, \sigma') < d\}, \quad V_d \equiv |V_d(\sigma)| = \sum_{l=0}^{d-1} \binom{n}{l}, \tag{1}
\]

The aim is to find a non-overlapping configuration of spheres such that \( D(\sigma_i, \sigma_j) \geq d \), for any two spheres \( i \) and \( j \). The above problem is a constraint satisfaction problem with \( N(N - 1)/2 \) constraints to satisfy. We index these constraints by \( (ij) \). A configuration \( \underline{\sigma} \equiv \{\sigma_i | i = 1, \ldots, N\} \) that satisfies all the constraints is called a solution of the problem. The partition function \( Z \) counts the number of such solutions,

\[
Z = \sum_{\underline{\sigma}} \prod_{i < j} I_{ij}(\sigma_i, \sigma_j), \tag{2}
\]

where \( I_{ij}(\sigma_i, \sigma_j) \) is an indicator function for constraint \( (ij) \); it is 1 if \( D(\sigma_i, \sigma_j) \geq d \) and 0 otherwise. The maximum rate of packing is defined as

\[
R \equiv \lim_{n, d \to \infty} \frac{1}{n} \log_2(N_{\text{max}}), \tag{3}
\]

with \( \delta = d/n = \text{const} \) and \( N_{\text{max}} \) is the maximum number of spheres such that \( Z > 0 \).

Let us here mention some known lower and upper bounds for \( R \). The Gilbert and Varshamov (GV) lower bound [1, 5] states that

\[
N_{\text{max}} \geq \frac{2^n}{V_d}, \tag{4}
\]
resulting in the following lower bound for the maximum rate of packing
\[
R \geq R^{GV} = 1 - H(\delta),
\] (5)

where
\[
H(\delta) \equiv -\delta \log_2 \delta - (1 - \delta) \log_2 (1 - \delta).
\] (6)

Notice that \(1 - H(\delta)\) is also the Shannon rate for a binary symmetric channel with error probability \(\delta\). A better lower bound is obtained with graph theoretical methods \([10]\) and gives
\[
R \geq R^{JV} = 1 - H(\delta) + \log_2 \left[ c n H(\delta) \right],
\] (7)

where \(c\) is a constant. As far as we know, this is the best lower bound reported for the maximum rate of packing. However, it still behaves asymptotically like \(R^{GV}\). The authors in Ref. \([12]\) use the liquid entropy of the system of hard spheres to find a maximum rate of packing that is very close to the above lower bound
\[
R^{PZ} \equiv 1 - H(\delta) + \log_2 \left( \frac{2 \ln 2}{n} H(\delta) \right).
\] (8)

On the other side, we have the Hamming upper bound \([3]\)
\[
N_{max} \leq \frac{2^n}{V^2},
\] (9)

resulting in
\[
R \leq R^{H} \equiv 1 - H \left( \frac{\delta}{2} \right),
\] (10)

The best upper bound for \(R\) is obtained by the linear programming methods \([6]\). An overestimate of this linear programming bound is
\[
R \leq R^{MRRW} \equiv H \left( \frac{1}{2} - \sqrt{\delta(1 - \delta)} \right).
\] (11)

In Fig. [1] we have compared the above bounds to show the large gap between the best lower and upper bounds.
III. REPLICA SYMMETRIC SOLUTIONS: BP EQUATIONS

The basic objects in the cavity method are the cavity marginals or messages passed along the edges of the interaction graph \[15, 17, 18\]. In our problem the messages are \(2^n\)-component vectors with elements \(\in \{0, 1\}\). There are two kinds of messages: (i) The cavity bias \(u_{i \rightarrow j}\) represents the warning that variable \(i\) sends to \(j\); \(u_{i \rightarrow j} = 1(0)\) says that point \(\sigma \in \Lambda\) is (not) forbidden by \(i\) for variable \(j\). (ii) The cavity field \(u_{i \rightarrow j}\) is sum of warnings that \(i\) receives in absence of \(j\): \(h_{i \rightarrow j}^\ast\) gives the number unsatisfied constraints, in the absence of \(j\), if variable \(i\) takes state \(\sigma\). The cavity biases are

\[
u_{i \rightarrow j} \in \{e_{\sigma} | \sigma \in \Lambda\}, \quad e_\sigma' = \begin{cases} 1, & \text{if } \sigma' \in V_d(\sigma); \\ 0, & \text{otherwise}. \end{cases} \tag{12}\]

In the RS framework we assume that all the packings or solutions belong to the same cluster of solutions in the configuration space. Suppose the interaction graph is a tree and \(e_\ast\) is a solution of the problem in the absence of variable \(j\). Then we represent the cavity bias \(u_{i \rightarrow j}\) by \(e_{\sigma_i^*}\). Notice that according to our definitions a cavity solution uniquely determines the cavity message \(u_{i \rightarrow j}\). The histogram of cavity biases among the solutions is given by

\[
Q_{i \rightarrow j}(u) = \sum_\sigma \eta_{i \rightarrow j}^\sigma \delta_{u,e_\sigma}. \tag{13}
\]

Assuming a tree interaction graph one can write equations governing the cavity probabilities \(\eta_{i \rightarrow j}^\sigma\):

\[
\eta_{i \rightarrow j}^\sigma \propto \prod_{k \in V(i) \setminus j} \left( \sum_{\sigma_k} I_{ik}(\sigma_i, \sigma_k) \eta_{k \rightarrow i}^{\sigma_k} \right), \tag{14}
\]

where \(V(i)\) denotes the set of variables interacting with variable \(i\). The above equations are called BP equations \[21, 22\].

In the Bethe approximation the entropy density of the system is written as

\[
s = \frac{1}{N} \left[ \sum_i \Delta s_i - \sum_{i<j} \Delta s_{ij} \right], \tag{15}\]

where \(\Delta s_i\) and \(\Delta s_{ij}\) are the entropy shifts by adding variable \(i\) and interaction \((ij)\), respectively. For these quantities we have

\[
\exp(\Delta s_i) = \prod_{\sigma \in V(i)} \left( \sum_{\sigma'} I_{ij}(\sigma, \sigma') \eta_{j \rightarrow i}^{\sigma'}(1) \right) = Z_i, \tag{16}
\]

\[
\exp(\Delta s_{ij}) = \sum_{\sigma, \sigma'} I_{ij}(\sigma, \sigma') \eta_{i \rightarrow j}^{\sigma} \eta_{j \rightarrow i}^{\sigma'} = Z_{ij}. \tag{17}
\]

In words, \(Z_i\) is the probability that variable \(i\) can occupy at least one point in the Hamming space and \(Z_{ij}\) is the probability that interaction \((ij)\) is satisfied.

Using the above equations for hard spheres we obtain

\[
\eta_{i \rightarrow j}^{\sigma} = \frac{\prod_{k \in V(i) \setminus j} \left( 1 - \sum_{\sigma' \in V_d(\sigma)} \eta_{k \rightarrow i}^{\sigma'} \right)}{\sum_{\sigma'} \prod_{k \in V(i) \setminus j} \left( 1 - \sum_{\sigma'' \in V_d(\sigma')} \eta_{k \rightarrow i}^{\sigma''} \right)}, \tag{18}
\]

and

\[
Z_i = \sum_{\sigma} \prod_{j \in V(i)} \left( 1 - \sum_{\sigma' \in V_d(\sigma)} \eta_{j \rightarrow i}^{\sigma'} \right), \quad Z_{ij} = 1 - \sum_{\sigma, \sigma': D(\sigma, \sigma') < d} \eta_{i \rightarrow j}^{\sigma} \eta_{j \rightarrow i}^{\sigma'}. \tag{19}\]

Now we can try different solutions to the BP equations.
FIG. 2: Comparing $N^{BPL}_{max}$ (BP) with some lower bounds (LB) from [27]. Up to $n = 15$ all the lower bounds are exact.

1. Liquid solution

The liquid solution is obtained by taking $\eta_{i \rightarrow j} = \eta = \frac{1}{2^n}$ for any $i$ and $j$. Evaluating $Z_i$ and $Z_{ij}$ at the liquid solution we can write the BP entropy

$$s = \ln(2^n) + \frac{N - 1}{2} \ln(1 - v_d),$$

where $v_d = \frac{\sqrt{\bar{v}}}{2}$. This entropy vanishes at

$$N^{BPL}_{max} = 1 - \frac{2 \ln(2^n)}{\ln(1 - v_d)}.$$  (21)

For large $n$ it gives

$$N^{BPL}_{max} v_d \approx (2 \ln 2)n.$$  (22)

In Fig. 2 we compare this quantity with some known exact results and lower bounds in small dimensions. The maximum rate of packing predicted by the liquid solution of the BP equation is

$$R^{BPL} \approx 1 - H(\delta) + \frac{\log_2[(2 \ln 2)n]}{n}.$$  (23)

where we have used

$$v_d \approx 2^{-n[1-H(\delta)]}.$$  (24)

We check the stability of the liquid solution in Appendix A; we find no continuous glass transition as long as

$$2 \ln 2 < \frac{1}{4v_d},$$  (25)

which is the case for $\delta < \frac{1}{2}$ as $v_d$ is exponentially small in this region.

2. Crystalline solution

In some situations the spheres may prefer one of the sublattices of the Hamming space to the other [12]. Indeed for even values of $d$, a sphere at the origin forbids more points from the odd sublattice than the even one. For high
densities the neighboring spheres will be at distance $d$ from the origin, i.e., they occupy points that again belong to the even sublattice. The above observation suggests that we could have ordered states in which nearly all the spheres are in the even or odd sector of the Hamming space. Here we will consider this situation by assigning different messages $\eta^e$ and $\eta^o$ to points in the even and odd sublattices, respectively. Again we use the fact that all the points in one sublattice are equivalent. Now the BP equations are

$$
\eta^e = \frac{(1 - V^e_d \eta^e - V^o_d \eta^o)^{N-2}}{2^{n-1}[(1 - V^e_d \eta^e - V^o_d \eta^o)^{N-2} + (1 - V^o_d \eta^o - V^e_d \eta^e)^{N-2}]} \quad \text{(26)}
$$

$$
\eta^o = \frac{(1 - V^o_d \eta^o - V^e_d \eta^e)^{N-2}}{2^{n-1}[(1 - V^e_d \eta^e - V^o_d \eta^o)^{N-2} + (1 - V^o_d \eta^o - V^e_d \eta^e)^{N-2}]} \quad \text{(27)}
$$

where $V^e_d$ and $V^o_d$ are given by

$$
V^e_d = \sum_{l=0}^{d-1} \frac{1 + (-1)^l}{2} \binom{n}{l}, \quad V^o_d = \sum_{l=0}^{d-1} \frac{1 - (-1)^l}{2} \binom{n}{l}. \quad \text{(28)}
$$

Notice that $\eta^e > 0$ and $\eta^o = 0$ cannot be a solution of the above equations unless for $d = n$. Indeed to have such a solution we need $\eta^e = \frac{1}{2^{n-1}} = \frac{1}{V^e_d}$, which holds only when $d = n$. We will look for other solutions where $\eta^e > 0, \eta^o > 0$. In this case we have

$$
\eta^e (1 - V^o_d \eta^o - V^e_d \eta^e)^{N-2} = \eta^o (1 - V^e_d \eta^e - V^o_d \eta^o)^{N-2}. \quad \text{(29)}
$$

Let us introduce the rescaled variables $v^e_d \equiv V^e_d / 2^{n-1}$, $v^o_d \equiv V^o_d / 2^{n-1}$, $p^e \equiv 2^{n-1} \eta^e$ and $p^o \equiv 2^{n-1} \eta^o$. Using the normalization condition $p^e + p^o = 1$, we obtain an equation for $p^e$,

$$
p^e [1 - v^e_d p^e - v^o_d (1 - p^e)]^{N-2} = (1 - p^e) [1 - v^o_d p^o - v^e_d (1 - p^e)]^{N-2}. \quad \text{(30)}
$$

For large $n$ and $d$ a nontrivial solution appears at $N^{BPC}_0$; see Fig. 3. At this point, the two sides of Eq. (29) have the same slope at $p^e = 1/2$. Thus we find

$$
N^{BPC}_0 = 2 + \frac{2 - v_d}{v^e_d - v^o_d}. \quad \text{(31)}
$$

The entropy of the crystalline phase is obtained given $Z_i$ and $Z_{ij}$,

$$
Z_i = 2^{n-1}[(1 - v^e_d p^e - v^o_d p^o)^{N-1} + (1 - v^o_d p^o - v^e_d p^e)^{N-1}], \quad \text{(32)}
$$

$$
Z_{ij} = 1 - p^e (v^e_d p^e + v^o_d p^o) - p^o (v^e_d p^e + v^e_d p^o). \quad \text{(33)}
$$
FIG. 4: Comparing the liquid and crystal entropies predicted by the BP equations.

To get a simple expression for the entropy we do the following approximation: \( p^e = 1 \) and \( p^o = 0 \). Notice that when \( d \) is even, \( v^e_d > v^o_d \) and from Eq. \( 29 \) we see that for large \( N \), \( p^o \) should be much smaller than \( p^e \). In this approximation the entropy reads

\[
s \approx \ln(2^{n-1}) + \frac{N - 1}{2} \ln(1 - v^c_d).
\]

(33)

This entropy vanishes at

\[
N_{\text{max}}^{\text{BPC}} = 1 - \frac{2 \ln(2^{n-1})}{\ln(1 - v^c_d)}.
\]

(34)

For large \( n \) the leading term is

\[
N_{\text{max}}^{\text{BPC}} v^c_d \approx (2 \ln 2)n.
\]

(35)

For \( n, d \to \infty \) the ratio \( v_d/v^c_d \) approaches to a constant and we recover asymptotically the bound provided with the liquid solution. In Fig. 4 we have compared the liquid and crystal entropies for a given \( n \) and even \( d \). Later in Sec. VIB we will introduce an iterative algorithm constructing the above crystalline packings.

IV. ONE-STEP RSB SOLUTION: SP EQUATIONS

In the 1RSB framework we assume that there are an exponentially large number \( N_c \sim e^{\Sigma} \) of clusters of solutions. A cluster of solutions consists of packing solutions in the configuration space that are connected to each other by paths of finite Hamming distances in the thermodynamic limit. In each cluster we could have frozen and unfrozen variables. Suppose the interaction graph is a tree and consider all cavity solutions (in absence of variable \( j \)) that belong to a given cluster; in a tree graph, fixing the boundary variables is equivalent to fixing the cluster of solutions. It may happen that in all the solutions of a cluster, variable \( i \) takes only one state, say \( \sigma \). Then the survey bias \( U_{i \to j} \) will be \( e_\sigma \) and we say variable \( i \) is a completely frozen variable in that cluster. We could have partially frozen variables that are frozen on a subset \( V^f \) of the Hamming space. In this case we represent the survey bias by \( U_{i \to j} = \sum_{\sigma \in V^f} e_\sigma \) ignoring the degeneracy of each state. In general we could write \( U_{i \to j} = \sum_{\sigma \in V_{i \to j}} w_{i \to j}^\sigma e_\sigma \) where \( w_{i \to j}^\sigma > 0 \) is the number of times that variable \( i \) appears in state \( \sigma \). A variable that takes all possible values is called an unfrozen variable and its survey bias is represented by \( 0 \).

The survey biases defined above depend on the cluster of solutions and change from one cluster to another. For an edge \((ij)\) of the interaction graph, we define the histogram of the survey biases among the clusters

\[
Q_{i \to j}(U) = \eta^0_{i \to j} \delta_{U, 0} + \sum_{m=1}^{2^{\Sigma} - 1} \sum_{\sigma_1 < \cdots < \sigma_m} \eta^\sigma_1 \cdots \sigma_m \delta_{U, \sum_{\sigma \in \{\sigma_1, \ldots, \sigma_m\}}} e_\sigma.
\]

(36)
Notice that compared with the RS case here we have the extra option $0$. The survey fields are determined by the survey biases:

$$\mathcal{H}_{i \rightarrow j}^\sigma = \sum_{k \in V(i) \setminus j} \min_{\sigma' \in V_i^k} \{1 - I_{ik}(\sigma, \sigma')\}. \quad (37)$$

In words, $\mathcal{H}_{i \rightarrow j}^\sigma$ is the minimum number of unsatisfied constraints, in the absence of $j$, if sphere $i$ takes position $\sigma$. Here we need only to know if $\mathcal{H}_{i \rightarrow j}^\sigma$ is zero or greater than zero. So we change the field’s definition to

$$\mathcal{H}_{i \rightarrow j}^\sigma = \min \left\{ 1, \sum_{k \in V(i) \setminus j} \min_{\sigma' \in V_i^k} \{1 - I_{ik}(\sigma, \sigma')\} \right\}. \quad (38)$$

The aim is to write an equation for $\eta_{i \rightarrow j}^{\sigma_1, \ldots, \sigma_m}$. The value of $\eta_{i \rightarrow j}^0$ is determined by the normalization condition. Variable $i$ sends the survey bias $\mathcal{U}_{i \rightarrow j} = e_{\sigma_1} + \cdots + e_{\sigma_m}$ to variable $j$ if its survey field is

$$\mathcal{H}_{i \rightarrow j}^\sigma = \begin{cases} 0, & \text{if } \sigma \in \{\sigma_1, \ldots, \sigma_m\}; \\ 1, & \text{otherwise}, \end{cases} \quad (39)$$

that is if the only choices for variable $i$ are the states in $V^f = \{\sigma_1, \ldots, \sigma_m\}$. We write this probability as

$$1 - \text{Prob} \left( \bigcup_{\sigma \in V_f} \mathcal{H}_{i \rightarrow j}^\sigma = 1 \text{ OR } \bigcup_{\sigma \in \Lambda \setminus V_f} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right), \quad (40)$$

where $\bigcup_{\sigma} \mathcal{H}_{i \rightarrow j}^\sigma = 0$ is the event that at least one of the $\mathcal{H}_{i \rightarrow j}^\sigma$ is zero. We have also the condition that variable $i$ does not receive contradictory biases. It means that the cavity field $\mathcal{H}_{i \rightarrow j}$ should not be equal to $1 \equiv (1, \ldots, 1)$. This probability is given by $\text{Prob}(\bigcup_{\sigma} \mathcal{H}_{i \rightarrow j}^\sigma = 0)$; so we obtain

$$\eta_{i \rightarrow j}^{\sigma_1, \ldots, \sigma_m} = \frac{1 - \text{Prob} \left( \bigcup_{\sigma \in V_f} \mathcal{H}_{i \rightarrow j}^\sigma = 1 \text{ OR } \bigcup_{\sigma \in \Lambda \setminus V_f} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right)}{\text{Prob}(\bigcup_{\sigma} \mathcal{H}_{i \rightarrow j}^\sigma = 0)}. \quad (41)$$

One can use the standard tools in the probability theory to find the above probabilities in terms of $\eta$’s. The result is the so called SP equation and is derived with more details in Appendix B. Here we simplify the analysis by assuming that all partially frozen surveys are zero, that is, only completely frozen and unfrozen surveys are taken into account. This would be a reasonable approximation when $N$ approaches $N_{\text{max}}$. We also expect to obtain a larger complexity within this approximation. In this attempt the SP equation reads

$$\eta_{i \rightarrow j}^{\sigma} = \frac{\sum_{m=0}^{2^n-1} (-1)^m \sum_{\sigma_1 < \cdots < \sigma_m \in \Lambda} \prod_{k \in V(i) \setminus j} \left(1 - \sum_{\sigma' \in V_k(\sigma, \sigma_1, \ldots, \sigma_m)} \frac{\eta_{k \rightarrow i}^{\sigma'}}{\sum_{m=1}^{2^n} (-1)^{m+1} \sum_{\sigma_1 < \cdots < \sigma_m} \prod_{k \in V(i) \setminus j} \left(1 - \sum_{\sigma' \in V_k(\sigma_1, \ldots, \sigma_m)} \eta_{k \rightarrow i}^{\sigma'}\right)\right)}{\sum_{m=0}^{2^n-1} (-1)^m \sum_{\sigma_1 < \cdots < \sigma_m \in \Lambda} \prod_{k \in V(i) \setminus j} \left(1 - \sum_{\sigma' \in V_k(\sigma_1, \ldots, \sigma_m)} \eta_{k \rightarrow i}^{\sigma'}\right)}, \quad (42)$$

where $V_i(\sigma_1, \ldots, \sigma_m)$ is the union volume of $V_d(\sigma_1)$, $V_d(\sigma_2)$, \ldots and $V_d(\sigma_m)$.

In our problem all the edges of the interaction graph are equivalent. Moreover, due to the translational symmetry, the surveys $\eta_{i \rightarrow j}^{\sigma}$ do not depend on $\sigma$. Considering these simplifications, the SP equation can be rewritten in a more compact form as

$$\eta = \frac{1}{2^n} \sum_{V=V_d} G(V)(1-V\eta)^{N-2}, \quad (43)$$

where

$$G(V) = \sum_{m=1}^{2^n} (-1)^{m+1} g_m(V), \quad G'(V) = \sum_{m=1}^{2^n} (-1)^{m+1} mg_m(V). \quad (44)$$

Here $g_m(V)$ is the total number of configurations that $m$ distinct points can take in the Hamming space such that the union volume $V_i$ is equal to $V$. More precisely we have

$$g_m(V) = \sum_{\sigma_1 < \cdots < \sigma_m} \delta_{V,V_i(\sigma_1, \ldots, \sigma_m)} \quad (45)$$
To obtain more explicit results we consider a naive approximation where $V_{ij}(\sigma_1, \ldots, \sigma_m) \approx mV_d$. This is a good approximation for small $d/n$. We also approximate $(1 - mV_d\eta)^N$ by $e^{-(N-2)mV_d\eta}$ to compensate the error made by overestimating the union volume $V_{ij}(\sigma_1, \ldots, \sigma_m)$ for large $m$. Then the SP equation reads

$$\eta = \frac{1}{2^n} \sum_{m=1}^{2^n} m(-1)^{m-1} \left( \frac{2^n}{m} \right) e^{-m(N-2)V_d\eta}.$$ (46)

Summing over $m$ we get

$$\eta = \frac{e^{-(N-2)V_d\eta}(1 - e^{-(N-2)V_d\eta})^{2^n-1}}{1 - (1 - e^{-(N-2)V_d\eta})^{2^n}}.$$ (47)

Let us see where the above equation admits a nontrivial solution $\eta > 0$. To this end we will consider the following scalings: $(N - 2)V_d/2^n = cn$ and $2^n\eta = 1 - g(n,c)$. For $n \to \infty$ we expect to have $c \to \text{const}$ and $g \to 0$. Replacing these in the above equation and expanding for small $2^n e^{-cn(1-g)}$ we obtain

$$g \simeq 2^n e^{-cn(1-g)}.$$ (48)

To have a solution for $g$ we need $c > \ln 2 + c'(\ln n)/n$ with

$$(n \ln 2 + c' \ln n)e^{-c' \ln n} = 1.$$ (49)

Increasing $N$, the frozen variables appear for the first time at $N^{SP}_c$, where

$$c' \simeq 1 + O\left(\frac{1}{\ln n}\right), \quad c \simeq \ln 2 + \frac{\ln n}{n} + O\left(\frac{1}{n}\right),$$ (50)

and therefore

$$N^{SP}_c v_d \simeq n \ln 2 + \ln n + \text{const.}$$ (51)

This gives the clustering transition, when the solution space splits into an exponentially large number of clusters.

3. Complexity

We can compute the complexity in the Bethe approximation as

$$\Sigma = \frac{1}{N} \left[ \sum_i \Delta \Sigma_i - \sum_{i<j} \Delta \Sigma_{ij} \right].$$ (52)

As before, $\Delta \Sigma_i$ and $\Delta \Sigma_{ij}$ are the complexity shifts by adding variable $i$ and interaction $(ij)$. The complexity is zero for small densities of the particles where we expect to have a single cluster of solutions. As we increase $N$, we encounter the clustering transition at $N_c$, where $\Sigma$ becomes nonzero. The complexity is a decreasing function of $N$ and finally vanishes at $N_{\text{max}}$. Here we shall focus on the most numerous clusters, which are not necessarily the relevant ones. More accurate results can be obtained by a large deviation study of the problem, which involves computing the complexity of different clusters [26].

To compute the complexity we need the two quantities $\Delta \Sigma_i$ and $\Delta \Sigma_{ij}$. Again for the sake of simplicity we only work with the completely frozen surveys. In this approximation we get

$$e^{\Delta \Sigma_i} = \sum_{m=1}^{2^n} (-1)^{m+1} \prod_{\sigma_1, \ldots, \sigma_m \in V(i)} \left( 1 - \sum_{\sigma' \in V(i,\sigma_1, \ldots, \sigma_m)} \eta_{ij}^{\sigma' \rightarrow i} \right) \equiv Z_i,$$ (53)

and the link contribution is

$$e^{\Delta \Sigma_{ij}} = 1 - \sum_{\sigma, \sigma': D(\sigma, \sigma') < d} \eta_{ij}^{\sigma \rightarrow j} \eta_{ij}^{\sigma' \rightarrow i} \equiv Z_{ij}.$$ (54)
FIG. 5: Comparing the RS entropy with the 1RSB complexity in the naive approximation. Here $n = 25$ and $d = 2$.

Using our notation in the previous subsection we can write a more compact form of the complexity for uniform surveys

$$\Sigma = \ln \left[ \sum_G G(V)(1 - V\eta)^{N-1} \right] - \frac{N-1}{2} \ln [1 - 2^n V_d \eta^2].$$  \hspace{1cm} (55)

In the naive approximation, where we approximate $V_\cup$ by $m V_d$, we find

$$\Sigma = \ln [1 - (1 - e^{-(N-1) V_d \eta}) 2^n] - \frac{N-1}{2} \ln [1 - 2^n V_d \eta^2].$$  \hspace{1cm} (56)

The complexity vanishes when

$$1 - (1 - e^{-(N-1) V_d \eta}) 2^n = (1 - 2^n V_d \eta^2)^{\frac{N-1}{2}}.$$  \hspace{1cm} (57)

Using again the scalings $(N - 1)V_d/2^n = c n$ and $2^n \eta = 1 - g(c,n)$, the above equation can be rewritten as

$$1 - e^{-x} = e^{-\frac{1}{2} c n (1-g)^2}, \hspace{1cm} x \equiv e^{n [\ln 2 - c(1-g)]}.$$  \hspace{1cm} (58)

Expanding the left side for the exponentially small $x$ we obtain

$$2 \ln 2 = c(1 - g^2).$$  \hspace{1cm} (59)

Notice that $g$ approaches to zero for large $n$ as $1/n$, so the above equation suggests

$$N_{SP} v_d = (2 \ln 2) n + O(1/n).$$  \hspace{1cm} (60)

Using Eqs. 47 and 50 we can find the complexity in the naive approximation. In Fig. 5 we have compared this complexity with the BP entropy for $n = 25$ and $d = 2$, where the naive approximation is expected to work. We see the jump in the complexity that happens at the clustering transition $N_{SP}$. Moreover, the complexity is always smaller than the RS entropy, as it should be; it is the sum $\Sigma + s_{\text{cluster}}$ that gives the total entropy of the system. Here $s_{\text{cluster}}$ is the internal entropy of the clusters.

Let us summarize the main approximations we used in the above calculations: First, the SP equation ignores the soft part of the messages and works only with the hard fields. We know that this could give an underestimate of the clustering transition but works well in predicting the satisfiability-unsatisfiability (SAT-UNSAT) transition in some constraint satisfaction problems. Next, we neglected the partially frozen states and considered only the completely frozen and unfrozen parts to write a simpler expression for the SP equation. This approximation seems reasonable as for maximum packings the spheres should be strongly localized by their neighborhood. And finally, we resorted to another approximation by replacing the union volume of $m$ distinct spheres with $m V_d$. One can improve on this by using an annealed approximation of $V_\cup$,

$$V_\cup(\sigma_1, \ldots, \sigma_m) \simeq 2^n [1 - (v_d)^m],$$  \hspace{1cm} (61)
where \((1 - v_d)^m\) is the probability that a point in the Hamming space is outside the \(m\) spheres centered at \(\sigma_1, \ldots, \sigma_m\). Further, we can approximate \(g_m(V)\) by

\[
g_m(V) \simeq \left( \frac{2^n}{m} \right) \left( \frac{2^n}{V} \right) \left[ 1 - (1 - v_d)^m \right] V \left[ 1 - (1 - v_d)^m \right]^{2^n - V}.
\]

The annealed volume is approximately given by \(mV_d\) when \(m \ll \frac{2^n}{V_d}\). Moreover, the completely frozen approximation we used, suggests that the main contribution in the SP equation comes from small values of \(m\), see Appendix B. Therefore, we do not expect to observe exponentially large deviations in \(N_{SP}^c\) and \(N_{SP}^{\max}\) beyond the naive approximation.

V. SOME ALGORITHMS FOR THE PACKING PROBLEM

A. Hard spheres in an ultrametric space

Exact solutions are always useful in that we obtain some insights about more complex problems [28, 29]. As a simple example which can be treated exactly, we consider the packing problem in an ultrametric space.

An ultrametric space is a metric space where for any three points \(i_1, i_2,\) and \(i_3\) we have \(D(i_1, i_2) \leq \max\{D(i_1, i_3), D(i_2, i_3)\}\). Consider a rooted binary tree \(T_n\) with \(n\) generations. The number of points at generation \(l\) is \(2^l\). The space of points is given by the \(2^n\) leaves of this tree. The distance \(D(i_1, i_2)\) between two points \(i_1\) and \(i_2\) is defined as the number of generations that one should go up in the tree to find the first common ancestor.

The problem is to find a packing of \(N\) hard spheres such that for any two spheres \(D(i_1, i_2) \geq d\). The packing problem is trivial for \(d = 1\), so we can start from this simple case and try to find packings for larger \(d\). Let us start with a binary tree of \(n_1\) generations and \(2^{n_1}\) leaves, \(T_{n_1}\). We put one sphere at each point of this ultrametric space to obtain a packing of \(N = 2^{n_1}\) spheres with diameter \(d = 1\); see Fig. 6. This is the densest configuration of hard spheres for the above parameters. The idea is to increase \(d\) by 1 and add the minimum number of necessary generations to find a new packing with the new diameter for the spheres. Indeed in an ultrametric space we need only one additional generation to remove all the overlaps in the previous configuration. After adding the new generation we are free to put each sphere in one of its two descendants. One can continue the above process to find a valid packing for an arbitrary value of \(d\). It turns out that these are the densest configurations of the spheres with the parameters \(n\) and \(d\); the packing indeed partitions the whole space into regions occupied by the spheres. Therefore, given \(n\) and \(d\), the number of packings and maximum number of spheres read

\[
Z = N! \left( \frac{2^{n-d}}{N} \right) (2^d)^N, \quad N_{\text{max}} = 2^{n-d}.
\]

The maximum rate of packing will be

\[
R^{UM} = \frac{\log_2 N_{\text{max}}}{n} = 1 - \delta.
\]

It is interesting that this is exactly the Shannon rate for a binary erasure channel with error probability \(\delta\).
The above example is one of the exactly solvable packing problems that one can compare the exact $N_{\text{max}}$ with the BP one computed at the liquid solution:

$$N_{\text{max}}^{BPL} = 1 - \frac{(2 \ln 2)n}{\ln(1 - 2^{d-n})} \to (2 \ln 2)n 2^{n-d}.$$  \hspace{1cm} (65)

Here $N_{\text{max}}^{BPL} > N_{\text{max}}$ but we find asymptotically $R^{BPL} = 1 - \delta$.

B. Hard spheres in the Hamming space

Here we can use the same strategy as above to find packings of hard spheres in the Hamming space. The points of an $n$-dimensional Hamming space can be represented by the leaves of a binary tree $T_n$. A configuration of $N$ spheres, $\sigma \equiv \{\sigma_i \in \Lambda | i = 1, \ldots, N\}$, is a packing of hard spheres with diameter $d$ if it satisfies the following set of constraints

$$C(d) \equiv \{ D(\sigma_i, \sigma_j) \geq d | i \neq j \},$$  \hspace{1cm} (66)

where $D(\sigma_i, \sigma_j)$ is the Hamming distance of points $\sigma_i$ and $\sigma_j$. We define the energy $E[\sigma] = \sum_{i<j} [1 - I_{ij}(\sigma_i, \sigma_j)]$ as the number of unsatisfied constraints. The conflict graph $G(\sigma)$ represents a graph of $N$ nodes where edge $(ij)$ is present if the corresponding constraint is not satisfied. To find a packing we do the following steps:

- We start with $N = 2^n$ spheres occupying all the leaves of $T_n$. This is the densest configuration of spheres that satisfies $C(1)$. We use $\sigma^1$ to represent this configuration of spheres.

- For $t = 2, \ldots, d$, we start with $(\sigma^{t-1}, T_{n_{t-1}})$ and find $(\sigma^t, T_{n_t})$ such that all the constraints in $C(t)$ are satisfied. Set $n = n_{t-1}$ and $\Delta n_t = 0$, then

  - If $E > 0$, add a new generation to $T_n$, that is $n = n + 1$ and $\Delta n_t = \Delta n_t + 1$.

  - For each sphere find the best value of $\sigma_i(n) \in \{0, 1\}$ and call the new configuration $\sigma^{t-1, \Delta n_t}$ and $n_t = n$.

Notice that every time we add a new generation we have to solve an optimization problem to find the best configuration, $\bar{\sigma} \equiv \{\sigma_i(n) | i = 1, \ldots, N\}$. Given $t$ and $\Delta n_t$, we need to find the ground state of the following energy function

$$E[\bar{\sigma}] = \sum_{(ij) \in G(\sigma^{t-1, \Delta n_t})} \delta_{\bar{s}_i, \bar{s}_j},$$  \hspace{1cm} (67)
The aim is to assign different values to the new components $s_i$ and $s_j$ when edge $(ij)$ belongs to the conflict graph. In our numerical simulations we use simulated annealing to solve the above optimization problem.

We observe that for even $t$ the change in the number of generations $\Delta n_t$ is 1; i.e., the conflict graph is a bipartite graph connecting only spheres in different sublattices, odd and even. Consequently, for even $d$ all spheres are in the same sublattice of the Hamming space, as expected from the crystalline phase of Sec. V. In Fig. 7 we have compared the packings constructed by the above algorithm with the known lower bounds. We also checked the fraction of \textit{locally frozen} spheres in these packings. A locally frozen sphere is one that, fixing the other spheres, cannot be locally displaced without violating some of the constraints. Figure 8 displays this quantity as a function of $d$. We see that for even $d$ a considerable fraction of the spheres are frozen. This indicates that finding such packings is very difficult by a random sequential adding algorithm [30] where spheres are added randomly one by one.

C. An algorithm based on the BP equations

Let us again write the general BP equations for the packing problem:

$$\eta_{i\rightarrow j}(\sigma_i) \propto \prod_{k \in V(i)\setminus j} \left( \sum_{\sigma_k} I_{ik}(\sigma_i, \sigma_k) \eta_{k\rightarrow i}(\sigma_k) \right).$$

To solve these equations one starts with random initial values for the cavity messages and updates the messages iteratively to reach a fixed point. At the fixed point we have the local marginals given by

$$\eta_i(\sigma_i) \propto \prod_{j \in V(i)} \left( \sum_{\sigma_j} I_{ij}(\sigma_i, \sigma_j) \eta_{j\rightarrow i}(\sigma_j) \right).$$

One can find a packing configuration by decimation, fixing the spheres one by one according to the above local marginals; after each decimation one has to run again the BP equations to obtained the new marginals. Here we take another approach, by using the local marginals to converge the equations to a polarized fixed point defining a packing configuration. We found it useful here to add also some additional random potential $w_i(\sigma_i)$ to break the high level of symmetry present in the problem. The modified BP equations read

$$\eta_{i\rightarrow j}(\sigma_i) \propto [\eta_i(\sigma_i)]^{r} e^{\beta w_i(\sigma_i)} \prod_{k \in V(i)\setminus j} \left( \sum_{\sigma_k} I_{ik}(\sigma_i, \sigma_k) \eta_{k\rightarrow i}(\sigma_k) \right),$$

$$\eta_i(\sigma_i) \propto [\eta_i(\sigma_i)]^{r} e^{\beta w_i(\sigma_i)} \prod_{j \in V(i)} \left( \sum_{\sigma_j} I_{ij}(\sigma_i, \sigma_j) \eta_{j\rightarrow i}(\sigma_j) \right).$$
where cavity messages are biased toward their local marginals. These are the reinforced BP (rBP) equations and the positive parameter \( r \) is called the reinforcement parameter. The next step is to write the above rBP equations in the limit \( \beta \to \infty \), assuming the messages scale like \( \eta_{i \to j}(\sigma_i) \propto e^{\beta m_{i \to j}(\sigma_i)} \). The resulting reinforced max-sum equations are

\[
m_{i \to j}(\sigma_i) = w_i(\sigma_i) + r m_i(\sigma_i) + \max_{k \in V(i) \setminus j} \sigma_k : I_{ik}(\sigma_i, \sigma_k) = 1 m_{k \to i}(\sigma_k),
\]

\[
m_i(\sigma_i) = w_i(\sigma_i) + r m_i(\sigma_i) + \max_{j \in V(i)} \sigma_j : I_{ij}(\sigma_i, \sigma_j) = 1 m_{j \to i}(\sigma_j).
\]

We solve these equations by iteration starting from random initial messages and zero reinforcement. The reinforcement parameter \( r \) is increased gradually as \( r(t + 1) = r(t) + \delta r \). At the end, the states that maximize the \( m_i(\sigma_i) \) define a packing configuration.

The time complexity of this algorithm grows as \( [N(2^n - V_d)]^2 \). We checked the algorithm by finding some packings given in [27]: for instance, maximum packings with parameters \((n = 10, d = 4, N = 40)\), \((n = 11, d = 3, N = 144)\), \((n = 11, d = 5, N = 24)\), \((n = 15, d = 7, N = 32)\). The main difficulty with large dimensions and number of spheres is the computation time and memory. One way to reduce both time and memory is to work with restricted search spaces. For example, we may take an initially random search space \( \Lambda_i \) of size \( S = 2^{n_0} \) for each sphere. Given these search spaces we run the reinforced max-sum equations to obtain the cavity and local messages. Then for each sphere we replace a part of its search space with new states. The above steps are repeated to find better search spaces and maybe a packing.

More precisely, the algorithm starts by initially randomly selected search spaces \( \Lambda_i = \{\sigma_i^{(1)}, \ldots, \sigma_i^{(S)}\} \). Then we do the following steps:

- Run the reinforced max-sum equations: The messages are updated for sufficiently large number of iterations \( T \) using the restricted search spaces \( \Lambda_i \). For example, we use \( T = 100 \) if we increase the reinforcement parameter by \( \delta r = 10^{-2} \).
- Update the search spaces: Compute the local messages \( \{m_i(\sigma) | \sigma \in \Lambda_i\} \) and sort them in a decreasing order to put the good states at the beginning of the list. Replace a number of states at the end of this list with other states close to the best one. In our simulations we replace half of the search spaces.

In this way we could, for example, find the maximum packings for \((n = 10, d = 4, N = 40)\) and \((n = 11, d = 3, N = 144)\) with a search space of dimension \( n_0 = 3 \) and \( n_0 = 5 \), respectively. This also allowed us to find some packings in larger dimensions, for instance \((n = 53, d = 29, N = 10)\), \((n = 54, d = 29, N = 12)\). These are the best known packings (maybe maximum) with those parameters. And in some cases we find denser packings with larger sphere diameters: \((n = 48, d = 32, N = 4)\), \((n = 51, d = 30, N = 6)\).

VI. HARD SPHERE PACKING IN THE \( q \)-ARY HAMMING SPACE

The above results can easily be extended to the \( q \)-ary Hamming spaces \( \Lambda = \{0, \ldots, q-1\}^n \). There are \( q^n \) points in this space represented by vectors of \( n \) elements in \( \{0, \ldots, q-1\} \). As before, the Hamming distance \( D(\sigma, \sigma') \) gives the number of different elements in two vectors \( \sigma \) and \( \sigma' \). Therefore the number of points at distance less than \( d \) from a point is given by

\[
V_d = |V_d(\sigma)| = \sum_{l=0}^{d-1} \binom{n}{l} (q-1)^l.
\]

In the asymptotic limit \( n, d \to \infty \) and \( q, \delta = d/n \) finite, we get \( v_d = \frac{V_d}{q^n} \simeq e^{-n[H_q(\delta)]} \) where

\[
H_q(\delta) = \delta \log_q(q-1) - \delta \log_q(\delta) - (1-\delta) \log_q(1-\delta), \quad 0 \leq \delta \leq 1 - \frac{1}{q}.
\]

In this section we shall only discuss the results obtained within the replica symmetric approximation. The BP equations remain the same as in Eq. 18 so we write directly the BP entropy computed at the liquid solution, i.e., \( \eta_{i \to j} = \frac{1}{q^{n_i}} \):

\[
s = N \frac{1}{2} \log(1-v_d) + \ln(q^n).
\]
This entropy vanishes at

\[ N_{\text{max}}^{BP} = 1 - \frac{2 \ln(q^n)}{\ln(1 - v_d)}, \]  

which gives a rate of packing that is not asymptotically different from the Gilbert-Varshamov (GV) one,

\[ R^{BP} \simeq 1 - H_q(\delta) + \frac{\log_q(n)}{n}. \]  

But, we know that for \( q \)-ary alphabets and \( q \) large enough, the rate of packing could asymptotically exceed the GV bound \([32]\). These packings, known as algebraic-geometry codes in coding theory, result in the following lower bound when \( q \) is a square:

\[ R \geq R^{TV} = 1 - \frac{1 - \frac{1}{\sqrt{q} - 1}}{1 - \frac{1}{\sqrt{q} - 1}}, \quad 0 \leq \delta \leq 1 - \frac{1}{\sqrt{q} - 1}. \]  

And if \( q \geq 49 \) there exists always an interval in which \( R^{TV} > R^{GV} \). We believe that these are some crystalline or ordered solutions that only happen for large values of \( q \). The fact that at \( \delta = 0 \) the rate is smaller than 1 indicates that the above packings are restricted to a subspace of \( \Lambda \) which is exponentially smaller than \( q^n \). Moreover, comparing \( R^{TV} \) with the rate of packing in an ultra-metric space suggests that this subspace is effectively ultrametric; see Sec. V.A. More precisely, an ultra-metric subspace of size \( \simeq q^{n(1 - \frac{1}{\sqrt{q} - 1})} \) where each sphere occupies \( \simeq q^d \) points would result to the same rate of packing as \( R^{TV} \). We remind that in ultrametric spaces both the BP and GV bounds are asymptotically exact. Notice that also for \( q \rightarrow \infty \) one obtains \( R^{BP} = R^{GV} = 1 - \delta \).

Indeed the number of points that their Hamming distance from a given point is equal to their ultrametric distance is \( U_n = \frac{1}{q-1}[(q-1)n - 1] \), which is exponentially large as long as \( q \geq 3 \). This is an upper bound for the size of an ultrametric subspace. Moreover, given diameter \( d \), we need only a subspace that is ultra-metric up to distances less than \( d \). Let us define \( \mu \) as the probability that a point belongs to an ultrametric subspace. A naive estimation of \( \mu \) can be obtained by

\[ \mu \simeq (1 - \mu)^{V_d - U_d}. \]  

Note that \( U_d \) is exponentially smaller than \( V_d \), thus asymptotically \( \mu \simeq q^{-nH_q(\delta)} \). In this case, each sphere will occupy only \( O(1) \) points of the subspace, so we recover the GV rate of packing. The above argument says the gap between the typical and maximal ultra-metric subspaces is very large, assuming the TV bound comes from packings in an optimal ultrametric subspace.

VII. CONCLUSION

In summary, we have written the BP and SP equations to study the hard sphere packing problem in the Hamming space. Within these approximations we obtained a maximum rate of packing that is asymptotically the same as the lower bound of Gilbert and Varshamov. In the RS approximation we also found a crystalline phase where for even values of \( d \) the spheres prefer to be in one sector of the Hamming space. The BP solutions were stable with respect to continuous glass transitions as long as \( \delta < 1/2 \). This suggests that phase transitions, if any, would be of discontinuous type.

We have also introduced two new algorithms. First a message passing algorithm based on the BP equations which finds dense packings of hard spheres in finite dimensions. An approximate scheme is used to reduce the time and memory complexity of the algorithm, which can still be improved and hopefully lead to new packing results. An almost identical algorithm can be used in continuous spaces. As a proof of concept we used it to find some known local dense packings in two-dimensional Euclidean space.

Second, we introduced an iterative algorithm to find packings of hard spheres starting from small diameters and dimensions. For even diameters the algorithm generates packings with all spheres in one subspace of the Hamming space, as expected from the crystalline solution of the BP equations.

There are still some points that need more effort to be clarified. It is of extreme relevance to establish the relation between the exact maximum rate of packing and the one provided by the BP equations. This means, for instance, that we need to study some interpolating functions between the exact and the BP entropy. Preliminary results are discussed in Appendix C. We expect the BP bound to be asymptotically exact in the binary Hamming space.

Finally, it would be nice if one could obtain the algebraic-geometry lower bound for the maximum rate of packing in the \( q \)-ary Hamming spaces with some physical arguments. These are probably crystalline solutions that cannot be
captured within the RSB formalism as the RS entropy is expected to be larger than that of the glassy solutions in the RSB phase.

As a concluding remark, we should mention that the 1RSB study presented here is not complete; the SP equations only consider the hard or frozen part of the cavity messages. A more complete study of the 1RSB approximation also takes the soft part of the messages into account and asks for the stability of these solutions \[33, 35\].

Acknowledgments

We would like to thank C. Baldassi, A. Braunstein, H. Cohn, S. Franz, S. Torquato, and F. Zamponi for their help and useful discussions. AR and RZ acknowledge the ERC grant OPTINF 267915.

Appendix A: Stability of the liquid solution

The liquid solution of the BP equations would be stable as long as the ferromagnetic (linear) and spin glass (nonlinear) susceptibilities are finite \[33, 35\]. These conditions can be expressed in terms of the maximum eigenvalue of the response matrix \(M\):

\[
M_{\sigma,\sigma'} \equiv \frac{\partial \eta_{\sigma \rightarrow j}}{\partial \eta_{\sigma' \rightarrow i}} = \begin{cases} -\frac{1}{2n}, & \text{if } \sigma' \in V_d(\sigma); \\ V_d, & \text{otherwise.} \end{cases}
\]

that has been evaluated at the liquid solution. Then the stability conditions read

\[
N^{BPL}_1|\lambda_{max}| = 1, \quad N^{BPL}_2|\lambda_{max}|^2 = 1,
\]

where \(\lambda_{max}\) is the maximum eigenvalue (in absolute value) of \(M\).

More precisely, for \(N < N^{BPL}_1\) and \(N < N^{BPL}_2\) we would have no continuous phase transition to an ordered and spin glass phase, respectively. Notice that these conditions do not exclude discontinuous phase transitions.

The response matrix \(M\) is symmetric with real eigenvalues and orthogonal eigenvectors. Using the translational symmetry by even vectors we write the eigenvectors as

\[
u^\sigma_{\lambda} = e^{i\pi k.\sigma},
\]

where \(\sigma\) is the binary vector representing a point in the Hamming space and \(k\) is a binary wave vector. The eigenvalues are obtained by plugging the above expression in the eigenvalue equation

\[
\lambda \nu^\sigma_{\lambda} = -\frac{1}{2n} \sum_{\sigma' \in V_d(\sigma)} \nu^\sigma_{\lambda} + \frac{V_d}{2n(2^n - V_d)} \sum_{\sigma' \notin V_d(\sigma)} \nu^\sigma_{\lambda}.
\]

Then for the eigenvalues we obtain

\[
\lambda = -\frac{1}{2n} \sum_{\sigma' \in V_d(\sigma)} e^{i\pi k. (\sigma' - \sigma)} + \frac{V_d}{2n(2^n - V_d)} \sum_{\sigma' \notin V_d(\sigma)} e^{i\pi k. (\sigma' - \sigma)}.
\]

The above expression is independent of \(\sigma\) and for simplicity we choose \(\sigma = (0, 0, \ldots, 0)\). Consider the maximum wave vector \(k = (1, 1, \ldots, 1)\). The corresponding eigenvalue is

\[
\lambda(1) = -\frac{1}{2n} \sum_{l=0}^{d-1} (-1)^l \binom{n}{l} + \frac{V_d}{2n(2^n - V_d)} \sum_{l=d}^{n} (-1)^l \binom{n}{l}.
\]

Obviously, the maximum eigenvalue is bounded by \(2V_d\). Let us approximate the maximum eigenvalue by the largest contribution at \(l = n/2\),

\[
|\lambda_{max}| \approx \frac{V_d}{2n(2^n - V_d)} \left(\frac{n}{2}\right).
\]
Notice that for large $n$ this is still asymptotically equivalent to the trivial bound $2v_d$. Using the Sterling approximation for large $n$ we get

$$|\lambda_{max}| \simeq \frac{2}{\sqrt{n}} \frac{V_d}{(2^n - V_d)}$$

which according to Eq. (A2) gives

$$N_1^{BP} v_d \simeq \frac{\sqrt{n}}{2}(1 - v_d), \quad N_2^{BP} v_d \simeq \frac{n}{4v_d} (1 - v_d)^2.$$  \hspace{1cm} (A9)

**Appendix B: Survey propagation equation**

Let us start from Eq. (41) and find a more suitable form of the SP equation,

$$\eta_{i \rightarrow j}^{\sigma_1, \ldots, \sigma_m} = \frac{1 - \text{Prob}\left(\bigcup_{\sigma \in V_f} H_{i \rightarrow j}^{\sigma} = 1 \text{ OR } \bigcup_{\sigma \in \Lambda \setminus V_f} H_{i \rightarrow j}^{\sigma} = 0\right)}{\text{Prob}\left(\bigcup_{\sigma} H_{i \rightarrow j}^{\sigma} = 0\right)},$$ \hspace{1cm} (B1)

where $V_f = \{\sigma_1, \ldots, \sigma_m\}$. Using the inclusion-exclusion theorem we write

$$\text{Prob}\left(\bigcup_{\sigma} H_{i \rightarrow j}^{\sigma} = 0\right) = \sum_{\sigma_1} \text{Prob}\left(\bigcup H_{i \rightarrow j}^{\sigma_1} = 0\right) - \sum_{\sigma_1 < \sigma_2} \text{Prob}\left(\bigcap_{\alpha = \sigma_1, \sigma_2} H_{i \rightarrow j}^{\alpha} = 0\right) + \sum_{\sigma_1 < \sigma_2 < \sigma_3} \text{Prob}\left(\bigcap_{\alpha = \sigma_1, \sigma_2, \sigma_3} H_{i \rightarrow j}^{\alpha} = 0\right) - \cdots + (-1)^{2^n - 1} \text{Prob}\left(\bigcap H_{i \rightarrow j}^{\sigma} = 0\right).$$ \hspace{1cm} (B2)

Here $\text{Prob}\left(\bigcap_{\sigma = \sigma_1, \ldots, \sigma_m} H_{i \rightarrow j}^{\sigma} = 0\right)$ is the probability of having $H_{i \rightarrow j}^{\sigma_1} = 0, H_{i \rightarrow j}^{\sigma_2} = 0, \ldots, H_{i \rightarrow j}^{\sigma_m} = 0$.

The probability in the numerator can be rewritten in the same way as

$$\text{Prob}\left(\bigcup_{\sigma \in V_f} H_{i \rightarrow j}^{\sigma} = 1 \text{ OR } \bigcup_{\sigma \in \Lambda \setminus V_f} H_{i \rightarrow j}^{\sigma} = 0\right) = \left\{ \text{Prob}\left(\bigcup_{\sigma \in V_f} H_{i \rightarrow j}^{\sigma} = 1\right) + \sum_{\sigma_1} \text{Prob}\left(\bigcup_{\sigma \in \Lambda \setminus V_f} H_{i \rightarrow j}^{\sigma_1} = 0\right) \right\} - \left\{ \sum_{\sigma_1 \in \Lambda \setminus V_f} \text{Prob}\left(\bigcup_{\sigma \in V_f} H_{i \rightarrow j}^{\sigma_1} = 1 \text{ AND } H_{i \rightarrow j}^{\sigma_1} = 0\right) \right\} + \cdots.$$ \hspace{1cm} (B3)

In the above equation we have probabilities like

$$\text{Prob}\left(\bigcap_{\sigma \in V_f} H_{i \rightarrow j}^{\sigma} = 1 \text{ AND } \bigcap_{\sigma \in \Lambda \setminus V_f} H_{i \rightarrow j}^{\sigma} = 0\right),$$ \hspace{1cm} (B4)

which, using the normalization condition, can be rewritten as

$$\text{Prob}\left(\bigcap_{\sigma \in V_f} H_{i \rightarrow j}^{\sigma} = 0\right) - \text{Prob}\left(\bigcap_{\sigma \in \Lambda \setminus V_f} H_{i \rightarrow j}^{\sigma} = 0 \text{ AND } \bigcap_{\sigma \in V_f} H_{i \rightarrow j}^{\sigma} = 0\right).$$ \hspace{1cm} (B5)

Plugging these into our expression for the numerator, Eq. (B3) we find

$$\text{Prob}\left(\bigcup_{\sigma \in V_f} H_{i \rightarrow j}^{\sigma} = 1 \text{ OR } \bigcup_{\sigma \in \Lambda \setminus V_f} H_{i \rightarrow j}^{\sigma} = 0\right) = \left\{ \text{Prob}\left(\bigcup_{\sigma \in V_f} H_{i \rightarrow j}^{\sigma} = 1\right) + \sum_{\sigma_1} \text{Prob}\left(\bigcup_{\sigma \in \Lambda \setminus V_f} H_{i \rightarrow j}^{\sigma_1} = 0\right) \right\} - \left\{ \sum_{\sigma_1 \in \Lambda \setminus V_f} \text{Prob}\left(\bigcup_{\sigma \in V_f} H_{i \rightarrow j}^{\sigma_1} = 1 \text{ AND } H_{i \rightarrow j}^{\sigma_1} = 0\right) \right\} + \cdots.$$ \hspace{1cm} (B6)
Notice that the last term in the first bracket is canceled with the first term in the second bracket. This cancellation indeed happens for any two subsequent brackets. Simplifying the above expression we find

\[
1 - \text{Prob} \left( \bigcup_{\sigma \in V^f} \mathcal{H}_{i \rightarrow j}^\sigma = 1 \text{ OR } \bigcup_{\sigma \in \Lambda \setminus V^f} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right) = \text{Prob} \left( \bigcap_{\sigma \in V^f} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right) - \sum_{\sigma_1 \in \Lambda \setminus V^f} \text{Prob} \left( \bigcap_{\sigma \in V^f} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right) \text{ AND } \mathcal{H}_{i \rightarrow j}^\sigma = 0 \]

\[+ \sum_{\sigma_1 < \sigma_2 \in \Lambda \setminus V^f} \text{Prob} \left( \bigcap_{\sigma \in V^f} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right) \text{ AND } \bigcap_{\sigma = \sigma_1, \sigma_2} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right) - \cdots + (-1)^{2^d - |V^f|} \text{Prob} \left( \bigcap_{\sigma} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right). \quad (B7)
\]

Now we should write a more explicit relation for \( \text{Prob} \left( \bigcap_{\sigma = \sigma_1, \ldots, \sigma_m} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right) \). First note that

\[
\text{Prob} \left( \bigcap_{\sigma = \sigma_1, \ldots, \sigma_m} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right) = \prod_{k \in V \setminus \{i\} \cup j} \left( 1 - \sum_{m' = 1}^{2^n} \sum_{\sigma_1' < \cdots < \sigma_{m'}'} I[\sigma_1' \in V \setminus \{i\} \cup j] \eta_{k \rightarrow i}^{\sigma_1' \cdots \sigma_{m'}'} \right), \quad (B8)
\]

where \( V \setminus \{i\} \cup j \) is the intersection of sets \( V \setminus \{i\} \cup j \). This is the probability that no variable in \( V \setminus \{i\} \cup j \) sends a survey that forbids state \( \sigma_1 \) for variable \( i \). Here \( I(C) \) is an indicator function for condition \( C \). And in general

\[
\text{Prob} \left( \bigcap_{\sigma = \sigma_1, \ldots, \sigma_m} \mathcal{H}_{i \rightarrow j}^\sigma = 0 \right) = \prod_{k \in V \setminus \{i\} \cup j} \left( 1 - \sum_{m' = 1}^{2^n} \sum_{\sigma_1' < \cdots < \sigma_{m'}'} \eta_{k \rightarrow i}^{\sigma_1' \cdots \sigma_{m'}'} \right), \quad (B9)
\]

where

\[
\sum_{\sigma_1' < \cdots < \sigma_{m'}'} \eta_{k \rightarrow i}^{\sigma_1' \cdots \sigma_{m'}'} = \sum I(\{\sigma_1, \ldots, \sigma_m\} \cap V \setminus \{i\} \cup j \neq \emptyset). \quad (B10)
\]

Using the above probabilities in Eqs. (B2) and (B7) we obtain the SP equation,

\[
\eta_{i \rightarrow j}^{\sigma_1, \ldots, \sigma_m} = \frac{\sum_{m' = 0}^{2^n - |V^f|} (-1)^{m'} \sum_{\sigma_1' < \cdots < \sigma_{m'}'} \prod_{k \in V \setminus \{i\} \cup j} \left( 1 - \gamma_{k \rightarrow i}(\sigma_1', \ldots, \sigma_{m'}', \sigma_1, \ldots, \sigma_m) \right)}{\sum_{m' = 1}^{2^n} (-1)^{m'+1} \sum_{\sigma_1' < \cdots < \sigma_{m'}'} \prod_{k \in V \setminus \{i\} \cup j} \left( 1 - \gamma_{k \rightarrow i}(\sigma_1', \ldots, \sigma_{m'}') \right)}, \quad (B11)
\]

with

\[
\gamma_{k \rightarrow i}(\sigma_1', \ldots, \sigma_{m'}') \equiv \sum_{m'' = 1}^{2^n} \sum_{\sigma_1'' < \cdots < \sigma_{m''}} \eta_{k \rightarrow i}^{\sigma_1'' \cdots \sigma_{m''}}. \quad (B12)
\]

Appendix C: Interpolating between the BP and exact entropies

In any approximation it is important to know how well the method approximates the correct behavior. Considering the BP approximation, we know that it is exact at least on tree interaction graphs. On loopy graphs, the BP approximation may overestimate or underestimate the correct entropy. This in general depends on the nature of the interactions and configuration space.

1. A static interpolation

The liquid solution of the BP equations is indeed equivalent to treating all the constraints independently and replacing a check \( I_{ij}(\sigma_i, \sigma_j) \) with its average value when both \( \sigma_i \) and \( \sigma_j \) are free. To model this solution, on each
interaction edge \((ij)\) we introduce auxiliary variables \((s_{ij}, s_{ji})\) besides the original variables. The new variables will serve as independent copies of variables \((\sigma_i, \sigma_j)\) on edge \((ij)\). Clearly if we want to recover the exact partition function we should force all the copies to be the same as their originals. Summing all together we write the following interpolating partition function:

\[
Z(\beta) = \sum_{\mathcal{Z}} \prod_{i < j} \left( \frac{1}{1 + e^{-\beta}} \right)^{2} \sum_{s_{ij}, s_{ji}} \left( I_{ij}(s_{ij}, s_{ji}) e^{-\beta[D(\sigma_i, s_{ij}) + D(\sigma_j, s_{ji})]} \right) ,
\]

with an inverse temperature \(\beta\) scaling as \(1/N\). Then we have \(Z_{BPL}^L = Z(0)\) and \(Z = Z(\infty)\). The above interpolating function can be defined for any constraint satisfaction problem that admits a uniform liquid solution.

Notice that without the normalization factor \(1/(1 + e^{-\beta})^{2m}\) we would obtain a replicated partition function \(Z_r(\beta)\) that is a convex and decreasing function of \(\beta\). In fact \(Z_r(\beta) \geq Z_r(\infty) = Z\), providing a simple upper bound for the true partition function for any \(\beta\). To get a good upper bound one needs to solve the replicated problem for the largest possible \(\beta\). Adding the normalization factor destroys the upper bound property but it is necessary if we want to recover \(Z_{BPL}^L\) at \(\beta = 0\).

Let us define variables \(X_i\) as

\[
X_i = \sum_{j \neq i} D(\sigma_i, s_{ij}).
\]

Taking the \(m\)th derivative of \(\ln(Z)\) with respect to \(\beta\) we obtain

\[
\frac{\partial^m \ln Z}{\partial(-\beta)^m} = \frac{\partial^m \ln \hat{Z}}{\partial(-\beta)^m} - \frac{\partial^m \ln \hat{Z}_0}{\partial(-\beta)^m},
\]

with

\[
\hat{Z} = \sum_{\mathcal{Z}} e^{-\beta \sum_i X_i} \prod_{i < j} I_{ij}(s_{ij}, s_{ji}), \quad \hat{Z}_0 = \sum_{\mathcal{Z}} e^{-\beta \sum_i X_i}.
\]

One can easily check that at \(\beta = 0\) and for \(m < 6\)

\[
\frac{\partial^m \ln \hat{Z}}{\partial(-\beta)^m} = \frac{\partial^m \ln \hat{Z}_0}{\partial(-\beta)^m}.
\]

Indeed, it is at \(m = 6\) that for the first time we encounter closed loops connecting three spheres in a diagrammatic expansion of the partition function. So for \(m < 6\) we have

\[
\frac{\partial^m \ln Z}{\partial(-\beta)^m}\big|_{\beta=0} = 0,
\]

which means the interpolating partition function is nearly flat at \(\beta = 0\). This is also true for \(\beta = \infty\) where all distances \(D(\sigma_i, s_{ij})\) should be zero.

For \(m = 6\) we have

\[
\frac{\partial^6 \ln Z}{\partial(-\beta)^6}\big|_{\beta=0} = \sum_{i_1 < i_2 < i_3} \frac{\partial^6 \ln Z_3}{\partial(-\beta)^6}\big|_{\beta=0},
\]

where \(Z_3\) is the interpolating partition function of three spheres:

\[
Z_3 = \frac{2^n}{(1 + e^{-\beta})^6} \sum_{l_1, l_2, l_3} S_l Q_{l_1,l_2}(l_3) \left( \sum_{r_1 \geq \delta = 1} R_{r_1,r_2}(r_3; l_4) e^{-\beta r_1 - \beta r_2} \right) ,
\]

Here \(S_l\) is the number of points at distance \(l\) from a given point and \(Q_{\ell_1,\ell_2}(l)\) is the number of points at distance \(l_1\) from \(\sigma_1\) and distance \(l_2\) from \(\sigma_2\) when \(D(\sigma_1, \sigma_2) = l\). More precisely, we have

\[
Q_{\ell_1,\ell_2}(l) = \left( \frac{n - l}{L_1 + L_2 - l} \right) \left( \frac{L_1 + L_2 - l}{l + l_1 - l_2} \right).
\]
And finally $R_{r_1,r_2}(r_3;l)$ is the number of pairs $(s_1,s_2)$ that satisfy the following conditions: $D(s_1,\sigma_1) = r_1$, $D(s_2,\sigma_2) = r_2$, and $D(s_1,s_2) = r_3$ given $D(\sigma_1,\sigma_2) = l$. This number can be written as

$$R_{r_1,r_2}(r_3;l) = \sum_r Q_{r_1,r}(l)Q_{r_2,r_3}(r).$$  \hfill (C10)

In Fig. 9 we have plotted the typical behavior of $Z_3(\beta)$ for some value of $n$ and $d$. Despite the fact that $Z_3(\beta)$ is decreasing with $\beta$ we find that the first nonzero term in the Taylor expansion is positive, signaling the nonperturbative nature of $Z_3(\beta)$ with respect to $\beta$.

### 2. A dynamic interpolation

Here we use an idea previously introduced in Refs. [23, 24] to replace the interactions in the original problem with the effective ones coming from the BP approximation.

Let us label the edges of an interaction graph by $t = 1, \ldots, M$. Define $E_0$ as the empty set and $E_t = \{e_1, \ldots, e_t\}$ as the set of the first $t$ edges. Now we introduce the following sequence of partition functions

$$Z_t = \sum_\Sigma \prod_{e \in E_t} I_e(\sigma_i,\sigma_j).$$  \hfill (C11)

Clearly, the original partition function is given by $Z = Z_M$ and $Z^{BP_L} = Z_0\prod_{t=1,M} \langle I_t(\sigma_i,\sigma_j) \rangle_0$ where the average is taken with respect to the uniform and independent distribution of $\sigma_i$ and $\sigma_j$. Suppose we add edge $e_{t+1}$ connecting nodes $(i_{t+1},j_{t+1})$. Then

$$Z_{t+1} = Z_t \langle I_{t+1} \rangle_0 (1 + \Delta_{t+1}), \quad \Delta_{t+1} = \frac{\langle I_{t+1} \rangle_l - \langle I_{t+1} \rangle_0}{\langle I_{t+1} \rangle_0},$$  \hfill (C12)

where $\langle I_{t+1} \rangle_l = P_l(I_{t+1})$ is the probability of satisfying constraint $I_{t+1}$ when the interaction set is given by $E_t$. And $\langle I_{t+1} \rangle_0 = P_0(I_{t+1})$ is the same probability when the interaction set is $E_0$. Moreover, $\Delta_{t+1} = 0$ when the interaction graph $E_{t+1}$ is a tree. We rewrite the final entropy as

$$\log Z = \log Z^{BP_L} + \sum_{t=1}^M \log(1 + \Delta_t).$$  \hfill (C13)

For hard spheres $\Delta_t$ may have different signs depending on the interaction graph. For example, suppose the interaction graph is a chain of size $L$ and we add the interaction between the end points. The probability of finding
the end points at distance $r$ can be obtained in a recursive way as

$$P_L(r) = \sum_{r'} P_{L-1}(r') \left( \frac{1}{2^n - V_d} \sum_{r'' \geq d} Q_{r,r''}(r') \right),$$

(C14)

using our expression for $Q_{l_1,l_2}(l)$ in Eq. [C9]. In this case, having $\Delta L$ is enough to know if the BP entropy is an over- or underestimation. Figure 10 displays this quantity for different values of $L$. We observe that $\Delta L$ is always negative for small $d$, but for large $d$ and small $L$ its sign alternates as expected for antiferromagnetic interactions.

Looking again at the BP entropy shows that in general the correction in Eq. [C13] is relevant only if divided by $N$, it is still exponential in dimension $n$. The corrections are irrelevant for instance if $\Delta t$’s behave as independent random numbers with a symmetric distribution and $\Delta t = O(1)$. This is what we expect to happen in high dimensions.

It is instructive here to compare the above hard sphere problem with a system of attractive interactions where the constraints $I_{ij}(\sigma_i, \sigma_j)$ are satisfied if $D(\sigma_i, \sigma_j) \leq d$. Figure 11 displays $\Delta L$ in a chain of attractive interactions; $\Delta L$ is positive for small $L$ but approaches an exponentially small negative value for large $L$ and $d$. In this case, we expect to have asymptotically $\Delta t \geq 0$. This is similar to the Griffiths inequalities [36] for ferromagnetic systems. Actually, using the Fortuin-Kasteleyn-Ginibre (FKG) inequalities [37] one can prove the above statement for soft attractive interactions, where $I_{ij}(\sigma_i, \sigma_j) = e^{-\beta D(\sigma_i, \sigma_j)}$ and $\beta \geq 0$.

Consider the following measure on configurations $\sigma \in \Lambda^N$ where $\Lambda = \{0, 1\}^n$,

$$\mu(\sigma) \propto \prod_{e \in \mathcal{E}} I_e(\sigma_{i_e}, \sigma_{j_e}).$$

(C15)

This measure satisfies the conditions of the FKG inequality; $\Lambda^N$ is a partially ordered set and $\mu(\sigma)$ is convex, that is given two configurations $\sigma$ and $\sigma'$

$$\mu(\sigma^\text{max})\mu(\sigma^\text{min}) \geq \mu(\sigma)\mu(\sigma'),$$

(C16)

where $\sigma^\text{max} = \max(\sigma, \sigma')$ and $\sigma^\text{min} = \min(\sigma, \sigma')$. These are bitwise max and min functions. One can use induction to show that $D(\sigma_i^\text{max}, \sigma_j^\text{max}) + D(\sigma_i^\text{min}, \sigma_j^\text{min}) \leq D(\sigma_i, \sigma_j) + D(\sigma'_i, \sigma'_j)$ which proves the convexity of $\mu(\sigma)$ for the soft attractive interactions.

Having these conditions the FKG inequality says that for decreasing (increasing) functions $f(\sigma)$ and $g(\sigma)$

$$\langle fg \rangle - \langle f \rangle \langle g \rangle \geq 0,$$

(C17)

where the averages are taken with respect to the measure $\mu(\sigma)$. A function $f(\sigma)$ is decreasing if $f(\sigma) \geq f(\sigma')$ for any $\sigma' \geq \sigma$. The latter is a bit-wise inequality.
FIG. 11: Absolute value of $\Delta_L$ in a chain of attractive interactions. Filled and empty symbols show positive and negative values, respectively.

Now consider decreasing functions $f = I[D(\sigma_i, 0) \leq l_i]$ and $g = I[D(\sigma_j, 0) \leq l_j]$ where 0 is the zero member of $\Lambda$. According to the FKG inequality we have

$$\langle I[D(\sigma_i, 0) \leq 0] I[D(\sigma_j, 0) \leq l] \rangle \geq \langle I[D(\sigma_i, 0) \leq 0] \rangle \langle I[D(\sigma_j, 0) \leq l] \rangle,$$

which means $P_i[D(\sigma_i, \sigma_j) \leq l] \geq P_0[D(\sigma_i, \sigma_j) \leq l]$ for any $i, j$, and $l$. In other words, it is more likely to find two spheres closer to each other compared to the uniform measure. As a result, for the soft attractive interactions BP provides a lower bound for the log-partition function.

Actually for the soft interactions $I_{ij}(\sigma_i, \sigma_j) = e^{\pm \beta D(\sigma_i, \sigma_j)}$ the partition function reads

$$Z = \sum_{\mathbf{\sigma}} e^{\pm \beta \sum_{i<j} D(\sigma_i, \sigma_j)} = \left[ \sum_{M=0}^{N} \binom{N}{M} e^{\pm \beta M (1 - M/N)} \right]^n.$$  

One can exactly solve the problem in the thermodynamic limit and compare the results with the BP predictions to confirm the above statement for the attractive interactions. In the case of repulsive interactions the BP and exact results asymptotically coincide.

[1] S. Torquato, and F. H. Stillinger, Reviews of Modern Physics 82, 2633 (2010).
[2] G. Parisi, and F. Zamponi, Reviews of Modern Physics 82, 789 (2010).
[3] R. W. Hamming, Bell Syst. Tech. Jnl. 29, 147 (1950).
[4] E. N. Gilbert, Bell Syst. Tech. Jnl. 31, 504 (1952).
[5] R. R. Varshamov, Dokl. Akad. Nauk SSSR 117, 739 (1957).
[6] R. J. McEliece, E. R. Rodemich, H. Rumsey, and L. R. Welch, IEEE Trans. Inform. Theory 23, 157 (1977).
[7] P. Delsarte and V. I. Levenshtein, IEEE Trans. Inform. Theory, 44(6), 2477 (1998).
[8] A. Barg and D. B. Jaffe, in "Codes and Association Schemes" (A. Barg and S. Litsyn, Eds.), Amer. Math. Soc., Providence,(2001).
[9] A. Samorodnitsky, Journal of Combinatorial Theory, Series A 96, 261 (2001).
[10] T. Jiang and A. Vardy, IEEE Trans. Inform. Theory 50, 1655 (2004).
[11] A. Procaccia and B. Scoppola, J. Stat. Phys. 96, 907 (1999).
[12] G. Parisi and F. Zamponi, J. Stat. Phys. 123, 1145 (2006).
[13] G. Parisi and F. Zamponi, J. Chem. Phys. 123, 144501 (2005).
[14] G. Parisi and F. Zamponi, J. Stat. Mech. (2006), P03017.
[15] M. M´ezard and G. Parisi, Eur. Phys. J. B 20, 217 (2001).
[16] M. M´ezard, G. Parisi, and R. Zecchina, Science 297, 812 (2002).
[17] M. M´ezard and R. Zecchina, Phys. Rev. E 66, 056126 (2002).
[18] M. Mézard and G. Parisi, J. Stat. Phys. 111 (1-2), 1 (2003).
[19] M. Mezard, G. Parisi, and M. A. Virasoro, Spin-Glass Theory and Beyond, vol 9 of Lecture Notes in Physics (World Scientific, Singapore, 1987).
[20] M. Bailly-Bechet, C. Borgs, A. Braunstein, J. Chayes, A. Dagkessamanskaia, J-M Franos, and R. Zecchina, Proc. Natl. Acad. Sci. USA 108, 882 (2011).
[21] A. Braunstein, M. Mezard, and R. Zecchina, Random Structures and Algorithms 27, 201 (2005).
[22] F. Guerra, Commun. Math. Phys. 233, 1 (2003).
[23] S. Franz and M. Leone, J. Stat. Phys. 111(3-4), 535 (2003).
[24] S. Franz, M. Leone, and F. L. Toninelli, J. Phys. A: Math. Gen. 36, 10967 (2003).
[25] F. R. Kschischang, B. J. Frey, and H. -A. Loeliger, IEEE Trans. Infor. Theory 47, 498 (2001)
[26] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborova, Proc. Natl. Acad. Sci. USA 104(25), 10318 (2007).
[27] M. J. A. Sloan, Amer. Math. Monthly 84, 82 (1977). See also the Table of Nonlinear Binary Codes at "http://www2.research.att.com/ njas/codes/And/”.
[28] S. Torquato and F. H. Stillinger, Phys. Rev. E 73, 031106 (2006).
[29] S. Torquato and F. H. Stillinger, Experimental Mathematics 15, 307 (2006).
[30] S. Torquato, Random Heterogeneous Material: Microstructure and Macroscopic Properties, Springer-Verlag, New York (2002).
[31] A. Braunstein and R. Zecchina, Phys. Rev. Lett. 96, 030201 (2006).
[32] M. A. Tsfasman and S. G. Vladut, Algebraic-Geometric Codes, Kluwer, Dordrecht (1991).
[33] A. Montanari and F. Ricci-Tersenghi, Eur. Phys. J. B 33, 339 (2003).
[34] O. Rivoire, G. Biroli, O. C. Martin, and M. Mézard, Eur. Phys. J. B 37, 55 (2004).
[35] A. Montanari, G. Parisi, and F. Ricci-Tersenghi, J. Phys. A: Math. Gen. 37, 2073 (2004).
[36] R. B. Griffiths, J. Math. Phys. 8, 478-481 (1967).
[37] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, Commun. Math. Phys. 22, 89 (1971).