Landscape statistics of the binary perceptron

J.F. Fontanari, R. Köberle

To cite this version:


HAL Id: jpa-00212455
https://hal.archives-ouvertes.fr/jpa-00212455
Submitted on 1 Jan 1990

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Landscape statistics of the binary perceptron

J. F. Fontanari and R. Köberle

Instituto de Física e Química de São Carlos, Universidade de São Paulo, 13560 São Carlos SP, Brasil

(Reçu le 20 décembre 1989, accepté sous forme définitive le 13 mars 1990)

Abstract. — The landscape of the binary perceptron is studied by Simulated Annealing, exhaustive search and performing random walks on the landscape. We find that the number of local minima increases exponentially with the number of bonds, becoming deeper in the vicinity of a global minimum, but more and more shallow as we move away from it. The random walker detects a simple dependence on the size of the mapping, the architecture introducing a nontrivial dependence on the number of steps.

1. Introduction.

The single-layer perceptron has been object of intense theoretical research since Gardner’s seminal paper [1]. The tools for this area of research have been developed in the analysis of the infinite range spin-glasses [2]. Remarkable outcomes of these studies were the rederivation of the maximal storage capacity of the network for the random input/output mapping [1] and the calculation of the distribution of relaxation times for the optimal stability perceptron learning algorithm of Krauth and Mezard [3]. Most of the results derived in the 1960’s [4, 5] and also recently [1, 3, 6-9] concerned perceptrons with real weights, though Gardner’s approach can easily be extended to the case where the weights are constrained to take on only a limited number of values. Clearly, knowing how the discretisation of the weights affects the performance of the perceptron is important from the technological viewpoint because some sort of discretisation is always made when implementing the network in hardware. There is also a biological motivation for considering discrete weights, since the fine tuning of the synaptic efficacies is not a plausible assumption for highly robust systems as biological brains.

In this paper we consider an extreme case of discretisation where the weights are constrained to take the values ±1 and we refer to this model as the binary perceptron [6, 10]. The network consists of an input layer composed of \( N \) binary units \( \{ \xi_i = \pm 1, i = 1, ..., N \} \) each one connected to a binary output unit \( S = \pm 1 \) through the weights \( \{ J_i = \pm 1, i = 1, ..., N \} \). Once the values of the units in the input layer (input pattern) are known, the value of the output unit is determined by the equation

\[
S = \text{sgn} \left( \sum_{i=1}^{N} J_i \xi_i \right).
\]  

(1.1)
The perceptron's task consists of learning a mapping between \( p \) input patterns and \( p \) output states. Such mapping can be emulated by the perceptron if there exists a vector \( \mathbf{J} = (J_1, J_2, \ldots, J_N) \) such that the \( p \) inequalities

\[
S^\mu \sum_i J_i \xi_i^\mu > 0 \quad \mu = 1, \ldots, p
\]

are simultaneously satisfied.

Despite the apparent simplicity of the binary perceptron its analysis has been much harder to carry out than for the perceptron with real weights. For instance, consider the computation of the maximal storage capacity \( \alpha_c \) defined as the ratio between the maximal number of input patterns which can be correctly mapped into their respective outputs and the size of the input field \( N \). For the real weights perceptron the replica symmetric ansatz is exact and, for the random input/output mapping, gives \( \alpha_c = 2 \) [1] while it gives a value larger than 1 [6], the Information theoretic bound, for the binary perceptron. Only recently the issue of the maximal storage capacity of the binary perceptron was settled through extensive numerical simulations [11] and analytical calculations using Parisi's ansatz for breaking the replica symmetry [12]. Its value, \( \approx 0.83 \), indicates that the discretisation of the weights does not cause a serious damage to the perceptron performance. Although this result asserts that for \( \alpha \approx 0.83 \) there exist solution vectors \( \mathbf{J} \) which satisfy the inequalities (1.2), there is no known algorithm for finding them, in contrast to the case of real weights where the perceptron algorithm is guaranteed to converge in a finite time to one of the solutions [4].

Lacking a specific algorithm for the binary perceptron we need to resort to heuristic methods to find the solutions of a given mapping [13]. The process of finding these solutions can be viewed as a search in the weight space for the global minima of the energy function

\[
E = p - \sum_{\mu = 1}^p \Theta \left( S^\mu \sum_i J_i \xi_i^\mu \right)
\]

which represents the number of misassociated patterns \( 0 \leq E \leq p \) and \( \Theta(x) = 1 \) for \( x \geq 0 \) and 0 otherwise. Moreover by restricting the moves in the weight space to one bond flip one can think of equation (1.3) as defining an \( N \)-dimensional landscape which may have many valleys of low energy (local minima) and zero-energy (solutions). It's clear that the performance of the heuristic algorithms in trying to minimize \( E \) will depend upon several properties of the energy landscape like how mountainous it is, the number of minima and its energy distribution, the distance between minima and so on.

The goal of this paper is to study the statistical structure of the energy landscape as a function of \( N \) characterizing the network architecture and \( p \) characterizing the mapping structure. Since we want to guarantee that at least one solution exists, we consider a mapping where the \( p \) input patterns are chosen randomly with equal probability while the outputs are given by

\[
S^\mu = \text{sgn} \left( \sum_{i = 1}^N K_i \xi_i^\mu \right) \quad \mu = 1, \ldots, p
\]

where \( K_i = \pm 1 \) at random [10]. Of course, we can choose \( K = 1 \) without loss of generality because the \( \xi_i^\mu \)'s are random. It is interesting that there exists a maximal number of input patterns \( p_{\text{max}} \approx 1.35 N \) above which the only solution to the problem is \( K \) [10]. We do not pursue this issue in this paper.

The paper is organized as follows. In section 2 we study the distribution of minima of the energy landscape. These minima are obtained by an exhaustive search procedure for small \( N \).
and by a greedy (gradient descend) algorithm for larger $N$. In section 3 we study the distribution of global minima (solutions) obtained by Simulated Annealing [14]. In section 4 we study a random walk on the energy landscape to obtain information about the landscape’s correlation. The details of this calculation are given in an appendix. Finally, in section 5 we summarize our results and present some concluding remarks.

2. Distribution of minima.

In this section we address the issue of the distribution of minima, disregarding whether local or global, of the energy landscape defined by equation (1.3). The algorithm we use to generate the minima is the following. Starting from an initial guess vector $J_0$ with a given energy, one computes the energy of its $N$ neighbors (vectors differing by 1 bond from $J_0$). The vector with lower energy is kept and the process is repeated until it reaches a vector which has lower energy than any of its neighbors. There are several questions of interest whose answers may shed some light on the structure of the landscape:

1. What is the average energy of the minima and how does it depend on the distance to the solution vector $1$?
2. What is the average distance between minima?
3. What is the average number of steps the above algorithm takes to reach a minimum?
4. How does the number of minima depend on $N$ and $p$?

To investigate these questions we have performed simulations for networks and mappings of different sizes. Typically our results are averages over 20 different mappings and 200 initial guess vectors. In the presentation of our results we introduce the parameter $\alpha$ defined as the ratio between the size of the mapping and the size of the network, $\alpha = p/N$.

Table I shows the average and the width of the distribution of energy for several values of $N$ and $\alpha$ in the case where the initial guess vector components are chosen randomly with equal probability. Both quantities, the average and the width, are divided by $p$ to facilitate the comparisons between the data. These results indicate that the average energy increases linearly with $p$ while the width increases with $\approx p^{1/2}$. When $N$ increases the average energy approaches $p/2$, the energy of a random configuration. This result resembles Kauffman’s complexity catastrophe [15] since by increasing the complexity of the network (measured by $N$) one makes the local minima poorer to such a point that they are not better than a randomly chosen configuration. In figure 1 we present the energy distribution of the minima obtained by exhaustive search (all minima were found) for $N = 15$ and several values of $\alpha$. Notice that as $\alpha$ increases the average energy tends to a well defined limit, which depends on $N$, in agreement with the results of table I.

Table I. — *Average and width (between parenthesis) of the distribution of energy of the minima in the case where the components of $J_0$ are ±1 with equal probability.*

<table>
<thead>
<tr>
<th>$\alpha/N$</th>
<th>51</th>
<th>301</th>
<th>501</th>
<th>1 001</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.37(0.23)</td>
<td>0.43(0.09)</td>
<td>0.45(0.07)</td>
<td>0.46(0.05)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.32(0.11)</td>
<td>0.42(0.05)</td>
<td>0.43(0.04)</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.33(0.09)</td>
<td>0.43(0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>0.35(0.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1. — Energy distribution of the minima obtained by exhaustive search for $N = 15$ and $\alpha = 1.2$ (Δ), 2.0 (●) and 3.0 (□). For each curve $P(E)$ is divided by $P_{\text{max}}(E) < 1$.

In table II we present the results of a similar experiment except that now the initial guess vector components are chosen equal to $-1$ with probability $1/4$ so that $J_0$ differs in average by $N/4$ bonds from the solution vector $1$. These results indicate that the average energy of the local minima decreases as they come closer to a global minimum. Supporting this conjecture we present in figure 2 the energy distribution of the minima obtained by exhaustive search for $N = 15$ and $\alpha = 2$, where we have separated the minima in different samples according to their Hamming distance (normalized to 1 throughout this paper) to the solution vector $1$. This rather curious property of the binary perceptron energy landscape, reminiscent of a « Massif Central » since there is a high concentration of very low energy minima around the global minima, seems to be a characteristic of some classes of landscapes [15]. In figures 1 and 2 we have divided $P(E)$ by its maximum value $P_{\text{max}}(E) < 1$ for each curve so that the data for different $\alpha$ and $d_1$ can be presented in the same scale.

Table II. — Average and width (between parenthesis) of the distribution of energy of the minima in the case where the components of $J_0$ are $-1$ with probability $1/4$.

<table>
<thead>
<tr>
<th>$\alpha/N$</th>
<th>51</th>
<th>301</th>
<th>501</th>
<th>1 001</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.16(0.16)</td>
<td>0.23(0.08)</td>
<td>0.24(0.06)</td>
<td>0.26(0.05)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.14(0.08)</td>
<td>0.22(0.04)</td>
<td>0.24(0.03)</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.14(0.07)</td>
<td>0.23(0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>0.18(0.06)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now we turn to the distribution of the Hamming distances $d_H$ between minima. We have found that the average of this distribution, $\bar{d}_H$, is extremely insensitive to $p$ and $N$; already for $N = 51$ it coincides with its value for $N = 1 001$. The width, however, though insensitive to $p$...
Fig. 2. — Energy distribution of the minima obtained by exhaustive search for $N = 15$ and $\alpha = 2.0$ for $d_1 \leq 1$ (●), $d_1 = 1/5$ (□), $d_1 \leq 1/3$ (△) and $d_1 \geq 4/5$ (+). For each curve $P(E)$ is divided by $P_{\text{max}}(E) < 1$.

decreases with $N^{-1/2}$. The same comments are also valid for the distribution of distances between the minima and the solution vector 1, $d_1$. However, $d_{H}$ and $d_1$ depend on the choice of the initial guess vector $J_0$. Table III shows this dependence for $N = 1001$ and $\alpha = 0.1$ with $J_0$ chosen inside a Hamming sphere centered on 1 and of radius equal to 0.1, 0.25 and 0.5. Notice how the minima become closer to each other as they approach the solution 1 thus corroborating the « Massif Central » conjecture.

Table III. — Average Hamming distance between minima ($d_{H}$) and between the minima and the solution 1 ($d_1$) for $J_0$ chosen inside a Hamming sphere centered on 1 of radius 0.1, 0.25 and 0.5. The numbers between parenthesis are the width of the distribution.

<table>
<thead>
<tr>
<th>Radius</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{H}$</td>
<td>0.092(0.004)</td>
<td>0.20(0.01)</td>
<td>0.50(0.01)</td>
</tr>
<tr>
<td>$d_1$</td>
<td>0.17(0.01)</td>
<td>0.32(0.01)</td>
<td>0.50(0.01)</td>
</tr>
</tbody>
</table>

In figure 3 we show the number of steps the greedy algorithm takes to reach a minimum as a function of $\alpha$. This quantity scales nicely with $N^{1/2}$ (we have tested this scaling for $N = 15$ up to 1 001), increases linearly for small $\alpha$ and seems to saturate for large $\alpha$. This behavior might be a consequence of a decreasing in the number of minima as $\alpha$ increases, since it is clear that as the minima become rarer more time will be needed to find them. To check it we present in figure 4 the ratio between the number of minima and the total number of vectors against $\alpha$ for $N = 7, 9$ and 15. The minima were obtained by exhaustive search. For
Fig. 3. — Number of steps to reach a minimum (scaled by $N^{-1/2}$) as a function of $\alpha$ for $N = 101$ ($\bullet$) and 201 ($\Delta$).

Fig. 4. — Ratio between the number of minima and the total number of vectors as a function of $\alpha$ for $N = 7$ ($\triangle$), 9 ($\square$) and 15 ($\bullet$).

fixed $N$, its number decreases as $\alpha^{-x}$ and for $N = 15$ one finds $x \approx 0.73$. Notice the linear decreasing of the number of minima for small $\alpha$ and the saturation for larger $\alpha$. One could be lead to think that finding the global minimum becomes easier as $\alpha \to \infty$ since there are fewer traps (local minima) on the way, however, the possibility the landscape takes a golf course shape as $\alpha$ increases cannot be dismissed. Indeed, this seems to be the case: for $N = 15$ we have measured the energy of the lowest energy local minima, obtained by exhaustive search, for $\alpha = 33$ and 66. The results were 0.17 and 0.29 respectively. Thus the
energy gap between the global minimum and the first excitation increases with $\alpha$. We have also observed that the energy gaps between local minima are much smaller than the ones stated above so that a golf course scenario seems to be appropriate to describe the landscape for very large $\alpha$. From figure 4 one can also see that the number of minima increases exponentially with $N$.

3. Distribution of global minima.

In this section we focus on the global minima of the energy landscape. We have used Simulated Annealing (SA) [14] to obtain in average 50 different solutions per mapping. The average and the width of the distribution of $d_H$ are computed and the results averaged over $\approx 100$ mappings. However, there are some inherent problems in this method. First, for $p$ near $p_{\text{max}} \approx 1.35 N$ there are very few solutions to a mapping so one should find practically all of them in order to have a significative statistics. Since finding solutions in this region has proven to be an exceedingly difficult task even for a powerful heuristic as SA we restrict our analysis to the region $\alpha \approx 0.7$. Second, the solutions found by SA might not be a representative sample of the space of solutions since they could possess some particular properties which made them accessible to SA. However, avoiding to collect biased solutions for our statistics was the main reason we have chosen SA among other heuristic algorithms: the stochasticity of SA makes it very improbable to be attracted to some special class of minima.

Now we turn to the results of the simulations. Keeping $\alpha = 0.1$ fixed we measure the mean and the width of the distribution of $d_H$. The results, presented in table IV for several values of $N$, show that the mean is practically insensitive to $N$ while the width becomes narrower as $N$ increases. The decreasing width is fitted by the curve $0.54 N^{-0.52}$, indicating then that in the thermodynamic limit the solution space, at least for small $\alpha$, has a very simple structure more similar to a ferromagnet whose ground states are related to each other by the symmetry of the model than to an infinite range spin-glass where there are no symmetry operators leading from one ground state to the others. These results also show that the replica symmetric ansatz should give a good description of the binary perceptron's solution space for small $\alpha$, in agreement with the results of [12] though the problem considered there was the random input/output mapping.

Table IV. — Average and width of the distribution of Hamming distances between global minima for $\alpha = 0.1$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>120</th>
<th>240</th>
<th>300</th>
<th>500</th>
<th>1 000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average distance</td>
<td>0.473</td>
<td>0.472</td>
<td>0.471</td>
<td>0.471</td>
<td>0.470</td>
</tr>
<tr>
<td>Width</td>
<td>0.045</td>
<td>0.032</td>
<td>0.028</td>
<td>0.022</td>
<td>0.015</td>
</tr>
</tbody>
</table>

One of the key points to the calculation of $\alpha_c$ for the binary perceptron was to realize that the solutions do not become arbitrarily close to each other as $\alpha \to \alpha_c$. In table V we show the dependence on $\alpha$ for $N = 120$ of the mean and the width of the distribution of $d_H$. The solutions become closer to each other as $\alpha$ increases and the width decreases very slowly with $\alpha$. As observed in the beginning of this section the sharp reduction in the number of solutions prevents us of further increasing $\alpha$ and then verifying whether $d_H$ goes to zero or not when $p$ approaches $p_{\text{max}}$. 
4. Random walks on the energy landscape.

In this section we study random walks on the energy landscape of the binary perceptron. The quantity we focus on is the average variation of energy after $s$ random steps. Each step consists of flipping one randomly chosen bond. Clearly on one hand, in a very correlated landscape a random walker would need many steps to reach a position, i.e. a vector $J$, whose energy differs significantly from the initial one. On the other hand, in a very uncorrelated landscape our random walker would need only a few steps to reach such a position. Thus studying the variation of energy in a random walk we expect to obtain information about the landscape’s correlation.

The walker starts from a random initial position, i.e. from a vector $J$ whose components $\pm 1$ are chosen randomly with equal probability. We first calculate the probability distribution for an energy change $\Delta E$ due to flipping $m$ bonds of $J$ and then express it in terms of the number of steps $s$ taken by the random walker.

We will be mainly interested in the first two moments of $P_{\Delta E}$. In the following we only present and discuss the results, giving the details of the calculations in the appendix. We have found

\begin{equation}
\langle \Delta E \rangle = 0 \tag{4.1a}
\end{equation}

\begin{equation}
\langle (\Delta E)^2 \rangle = pP_m \tag{4.1b}
\end{equation}

Clearly the first moment is zero since a bond flip can increase or decrease the energy of a random configuration with equal probability. $P_m$ in the equation for the second moment is given by

\begin{equation}
P_m = m \sqrt{\frac{2}{\pi N}} \times \begin{cases} 2^{-m} C_{m/2}^{m/2}, & m \text{ even} \\ 2^{-m-1} C_{(m-1)/2}^{(m-1)/2}, & m \text{ odd} \end{cases}
\end{equation}

for $m \ll N$ and odd $N \gg 1$. From this equation one can see that $P_m = P_{m+1}$ for odd $m$.

In the limit $m = \beta N$ we can write $P_m \rightarrow P(\beta)$ where

\begin{equation}
P(\beta) = \frac{2}{\pi^{1/2}} \int_0^\infty dz \, e^{-z^2} \text{erf}(z(\beta^{-1} - 1)^{-1/2}). \tag{4.3}
\end{equation}

Before comparing these results with our simulations $m$ has to be related to the number of steps taken in our random walk. We thus ask for the probability $P_{s,m}$ of obtaining $m$ flipped bonds in $s$ random steps. This is a well known problem proposed by Ehrenfest [16], which can be solved writing a recursion relation for $P_{s,m}$:

\begin{equation}
P_{s,m} = \left( 1 - \frac{m - 1}{N} \right) P_{s-1,m-1} + \frac{m + 1}{N} P_{s-1,m+1} \tag{4.4}
\end{equation}
The mean number of flipped bonds after $s$ steps is:

$$\langle m \rangle / N = \frac{1}{2} \left( 1 - e^{-2s/N} \right).$$  \hfill (4.5)

The stationary solution for the number of flipped bonds exhibits a sharp Gaussian peak of width $N^{-1/2}$ around the mean, so that we can replace the number of flipped bonds by its mean value. In this way we obtain the dependence of $P_{\Delta E}$ on the number of steps $s$. We compare $\langle (\Delta E)^2 \rangle$, computed by replacing $\beta$ in equation (4.3) by the r.h.s. of equation (4.5) and substituting the result in equation (4.1b), with our simulations in figure 5. A very good convergence with $N$ to the theoretical prediction can be seen.

Some comments regarding the interpretation of $\langle (\Delta E)^2 \rangle$ are in order. It is a product of two factors, the linear $p$-behavior being displayed in one of them reflects the dependence on the mapping via a sum of $p$ random independent variables. The nontrivial dependence on $s/N$ is determined by the perceptron's architecture.

![Graph showing second moment scaled by $p$ of distribution for $\Delta E$ in terms of number of steps $N$ for $N = 201$ (□) and 401 (●). The lower curve is the theoretical prediction.](image)

5. Conclusion.

The aim of this study is to understand general features of the binary perceptron landscape, which would eventually guide our intuition in the construction of algorithms searching for global minima. Our results are restricted to mapping problems which have at least one solution [10]. The general problem of finding the global minima of the binary perceptron energy is NP-complete [17] since it is equivalent to integer programming although a typical problem may be tractable [18]. Next we summarize our main results. We find that the number of local minima increases exponentially with $N$, but far from a global minimum they become ever shallower as they proliferate. In the vicinity of a global minimum the average energy of the local minima tends to decrease, indicating the existence of a « Massif Central » à la Kauffman [15]. For fixed $N$ the number of minima decreases as $\alpha^{-x}$ with $x > 0$. However,
finding the global minimum does not become easier since the landscape seems to take a golf
course shape. An analysis of the distribution of distances between global minima obtained by
Simulated Annealing for small $\alpha$ indicates that it tends to a delta function in the
thermodynamic limit.

We sample the average features of the energy landscape by a random walker starting at a
random position. We focus on the probability distribution for a change of $\Delta E$ as a function of
the number of steps $s$ taken by the random walker. The first moment is zero whereas the
variance has a trivial dependence on $p$ but a very complex dependence on $s/N$ reflecting the
perceptron architecture.

Somewhat surprisingly, most of the features we studied were found to depend trivially on $p$
(it gives the energy scale), indicating that it is the underlying network architecture which
controls the landscape structure. It should be interesting to see how the landscape features
studied in this paper change when mappings with more complex statistics are considered.

**Acknowledgements.**

JFF is indebted to R. Meir for illuminating discussions on perceptrons and for aid with the
simulations in section 3. JFF also thanks P. Moscato for discussions on Kauffman's work and
J. J. Hopfield for the hospitality at Division of Chemistry (Caltech), where this work was
started. The research of RK was partly supported by CNPq and JFF was partly supported by a
CNPq fellowship during his stay at Caltech.

**Appendix.**

In this appendix we present the details of the calculation of $P_{\Delta E}$, the probability distribution
for an energy change $\Delta E$ due to flipping $m$ bonds of a random vector $J$.

Initially we note that the $p$ fields

$$h^\mu = \sum_{i=1}^{N} J_i \eta^\mu_i \quad \mu = 1, \ldots, p$$

(A1)

where $\eta^\mu_i = S^\mu \xi^\mu_i$ are independent random variables since the $\eta^\mu_i$'s are chosen independently
for different $\mu$'s. Of course, in order to $1$ be a solution of the mapping the $\eta^\mu_i$'s must satisfy

$$\sum_{i=1}^{N} \eta^\mu_i > 0$$

(A2)

for each $\mu$.

A change in energy occurs, whenever some of the fields change sign. Suppose this happens
for $n$ fields, out of which $n_+$ change from plus to minus, whereas $n_-$ do the opposite, so that
$n = n_+ + n_-$ and $\Delta E = n - 2 n_-$. The probability for the occurrence of this energy change is

$$P_{\Delta E} = \sum_{n=0}^{p} P_n C_{n_-}^n 2^{-n}$$

(A3)

with the notation $C_k^n = n! / k! (n-k)!$ and $n_- = (n - \Delta E) / 2$. Since the $p$ fields $h^\mu$ are
independent, $P_n$ is given by

$$P_n = (P_m)^n (1 - P_m)^{p-n}$$

(A4)

where $P_m$ is the probability for some field $h^\mu$ to change sign under a simultaneous flip of $m$
bonds of $J$. In order to compute $P_m$ let us drop the pattern indexes and divide the field
$h^\mu$ into a sum of two terms

$$h = \bar{z} + z$$

(A5)
where \( \tilde{z} \) contains the \( m \) components of \( J \) to be flipped. The probability for \( \sum_{i=1}^{m} J_i \eta_i \) to equal \( z \) is

\[
P_{\tilde{z}} = C_k^m 2^{-m}
\]

with \( k = (m - \tilde{z})/2 \), where we have used the fact that the presence of the \( J_i \)'s turns \( \tilde{z} \) into a sum of independent random variables despite the constraint (A2). Similarly the distribution for \( z \) is given by

\[
P_{z} = C_k^{N-m} 2^{-(N-m)}
\]

with \( k = (m - z)/2 \).

\( \tilde{P}_m \) will be the probability for \( |\tilde{z}| \leq |z| \). Restricting ourselves for simplicity to odd \( N \gg 1 \) we get equation (4.2) for \( m \ll N \). The limit \( m = \beta N \), equation (4.3), is obtained by replacing the binomial distributions of equations (A6) and (A7) by Gaussians.

Thus inserting equation (4.2) or (4.3) into equations (A4) and (A3) we obtain \( \tilde{P}_{\Delta E} \).

References