Coding Theory and Markov-Ansätze for Spin-Glasses

Pál Ruján

Fachbereich 8 Physik and Institut für Chemie und Biologie des Meeres, Carl von Ossietzky Universität, Postfach 2503, D-26111 Oldenburg, Germany

(March 4, 1997)

Viewing spin-glass models as decoding Hamiltonians of convolution error-correcting codes leads to simple but not trivial results. I show that the decoding problem maps into doubly quenched spin-glass models, generalizing Sourlas' results. The internal energy and a subclass of correlation functions is calculated analytically at specific temperatures, generalizing Nishimori's gauge invariant solutions. Shannon's channel coding theorem provides accurate estimates for the extent of the ordered phase(s) in spin-glass models.

PACS numbers: 02.50.+s 02.70.+p 05.50+q 64.60cn 75.10hk 89.90+h

Consider a system where the Ising spin variables $\mathbf{s} = (s_1, s_2, \ldots, s_N)$, $s_i = \pm 1$, relax to equilibrium much faster than the lattice defects. On an intermediate time-scale the spins are already in thermal equilibrium, while the other dynamic variables, represented by the effective couplings $\{J\}$, are frozen. Hence, a physical observable A should be calculated as [1]

$$[\langle A(\mathbf{s}, \{J\})\rangle] = \int \mathcal{D}\{J\} p(\{J\}) \sum_{\{\mathbf{s}\}} [p(\mathbf{s}|\{J\}) A(\mathbf{s}, \{J\})] \tag{1}$$

where the coupling constants follow the nonequilibrium distribution $p(\{J\})$. As usual, $\langle\ldots\rangle$ denotes a thermal and $[\ldots]$ a 'quenched' average $\int \mathcal{D}\{J\} p(\{J\})\ldots$.

Eq. (1) makes explicit that thermal averages over the spin variables are conditional on the actual realization of the random couplings. In the Edwards-Anderson model, for example, the conditional probability $p(\mathbf{s}|\{J\})$ is defined as

$$p(\mathbf{s}|\{J\}) = \frac{\mathrm{e}^{-\beta E(\mathbf{s}|\{J\})}}{Z(\{J\})} \tag{2}$$

It is the presence of the normalization constant, the partition function

$$Z(\{J\}) = \sum_{\mathbf{s}} \mathrm{e}^{-\beta E(\mathbf{s}|\{J\})} \tag{3}$$

which makes the *quenched averages* Eq. (1) technically difficult. The energy functional is

$$E(\mathbf{s}|\{J\}) = -\sum_{\langle i,j\rangle} J_{ij} s_i s_j \tag{4}$$

where $\langle i,j\rangle$ runs over the edges of a d-dimensional lattice. The couplings J_{ij} are independently and uniformly sampled from a binary

$$p(J_\alpha) = p\delta(J_\alpha + 1) + (1-p)\delta(J_\alpha - 1) \tag{5}$$

or a Gaussian distribution. More precisely, $p = p(\{J\}; \vec{\lambda})$, where $\vec{\lambda}$ denotes the control *parameters* of the distribution.

The main idea of this Letter is to avoid the quenched averages (1) by using the Bayes identity

$$p(\mathbf{s}, \{J\}) = p(\mathbf{s}|\{J\}) p(\{J\}) = p(\{J\}|\mathbf{s}) p(\mathbf{s}) \tag{6}$$

and expressing the averages (1) alternatively as

$$\langle [A(\mathbf{s}, \{J\})]\rangle = \sum_{\{\mathbf{s}\}} \left[\int \mathcal{D}\{J\} p(\mathbf{s}) p(\{J\}|\mathbf{s}) A(\mathbf{s}, \{J\}) \right] \tag{7}$$

At first glance this transformation does not make sense because $p(\mathbf{s})$ and $p(\{J\}|\mathbf{s})$ do not have a direct (measurable) physical interpretation. However, as first shown by Sourlas [2–4] - see also [5] - under certain conditions Eq. (7) can be reinterpreted in terms of convolution error-correcting parity codes. A physicist should at this stage consider both $p(\mathbf{s})$ and $p(\{J\}|\mathbf{s})$ as *Ansätze* satisfying the conditions

$$p(\{J\}|\mathbf{s}) = \prod_{\alpha=1}^{M} p(J_\alpha|\gamma_\alpha) \sim \mathrm{e}^{\sum_\alpha E_\alpha(J_\alpha|\prod_{i\in\alpha} s_i)} = \mathrm{e}^{-\tilde{\beta}E(\mathbf{s}|\{\mathbf{J}\})} \tag{8}$$

$$p(\{J\}) = \sum_{\mathbf{s}} p(\{J\}|\mathbf{s})p(\mathbf{s}) \stackrel{!}{=} \prod_\alpha p(J_\alpha) \tag{9}$$

The distribution of bonds (9) should replicate Eq. (5). The resolution of the constraints (8) - and in particular of (9) - is in principle as difficult as the problem of quenched averages itself. However, as explained below, *special* solutions can be constructed using the analogy to coding theory.

Now consider the same problem from the point of view of a computer scientist. Messages (binary strings $\mathbf{s}$) are sampled from a source distribution, $p(\mathbf{s})$, and encoded using the Hamiltonian

$$H_{enc} = -\sum_{\alpha=1}^{M} \gamma_\alpha \prod_{i\in\alpha} \sigma_i \tag{10}$$

where

$$\gamma_\alpha = \prod_{i\in\alpha} s_i = \pm 1 \tag{11}$$

α denotes one of the $M > N$ spin products (parity-checks) used by the code. The redundancy of this code is measured by the *rate*, $R = \frac{N}{M} < 1$. In the absence of noise, the original message $\mathbf{s}$ is stored (and retrieved) as the unique ground state of the encoding Hamiltonian H_{enc} (10), $\mathbf{s} = \arg\min_\sigma H_{enc}(\sigma)$.

The (binary) couplings γ_α are called *codewords* and are sent through a noisy transmission channel. The channel accepts binary inputs and might deliver continuous or discrete outputs. The channel noise is modeled as a Markov process, each variable γ_α is distorted independently and identically according to the probability $p(J_\alpha|\gamma_\alpha)$. In general, the channel properties will define the spin-groups of the inputs, outputs, and the stochastic maping between them. Iff the channel is described by an ergodic Markov process, one can use the parametrization

$$p(J_\alpha|\gamma_\alpha) \sim \mathrm{e}^{\sum_{\alpha=1}^{M} E_\alpha(J_\alpha|\prod_{i\in\alpha}\sigma_i)} \tag{12}$$

Hence, Eq. (6) leads to the decoding Hamiltonian

$$-\beta H_{dec} = \sum_{\alpha=1}^{M} E_\alpha(J_\alpha|\prod_{i\in\alpha}\sigma_i) + E_{prior} \tag{13}$$

where $E_{prior} = \ln p(\mathbf{s})$ reflects the decoder's expectation about the source of original messages. The optimal procedure for decoding H_{dec} has been discussed in [5].

Shannon's channel coding theorem proves the *existence* of codes which in the thermodynamic limit reconstruct the original messages with vanishing average error as long as the code rate R is smaller than the channel capacity $R \le C$ [6]. For these *optimal codes* one can prove the strong inverse channel-theorem: once $R > C$ the average decoding error is maximal (no retrieval is possible) [7]. In physical terms, define the order parameter as the overlap between the inferred and the original message. For optimal codes the order parameter jumps in the thermodynamic limit from one to zero at the noise level corresponding to the channel capacity. Unfortunately, none of the known practicable codes does saturate the Shannon-bound [7]. Therefore: 1) for some *given* code the average decoding error might be nonvanishing already at rates well below the channel capacity, $R < C$ and 2) a positive overlap with the original message might persist even for $R > C$. In fact, most convolutional codes correspond to spin chains with short range interactions, where no phase transition can occur.

The main points of this Letter are illustrated below using the binary asymmetric and the binary erasure channel. The generalization to other spin-groups, spin-interactions, and noise distributions is straightforward. For both channels the encoding Hamiltonians are given by Eq. (10), where the γ's correspond to products of nearest-neighbor Ising spins.

The asymmetric binary channel has binary inputs γ_α and outputs J_α and is defined as

$$p(J_\alpha|\gamma_\alpha) = q_1\delta_{J_\alpha,1}\delta_{\gamma_\alpha,1} + p_1\delta_{J_\alpha,-1}\delta_{\gamma_\alpha,1} + p_2\delta_{J_\alpha,1}\delta_{\gamma_\alpha,-1} + q_2\delta_{J_\alpha,-1}\delta_{\gamma_\alpha,-1} \tag{14}$$

whith $q_{1,2} + p_{1,2} = 1$. The decoding Hamiltonian can be calculated from Eq. (12) as

$$-\beta H_{dec} = \sum_\alpha [a + b\prod_{i\in\alpha}\sigma_i + cJ_\alpha + dJ_\alpha\prod_{i\in\alpha}\sigma_i] \tag{15}$$

with $a = \frac{1}{4}\ln(q_1p_1q_2p_2)$, $b = \frac{1}{4}\ln(q_1p_2/q_2p_1)$, $c = \frac{1}{4}\ln(q_1p_1/q_2p_2)$, and $d = \frac{1}{4}\ln(q_1q_2/p_1p_2)$. The decoding Hamiltonian has fixed parameters, determining the 'temperature' of the model.

Usual spin-glass models are recovered when sending repeatedly the same *reference* message, *ie.* a ferromagnetic state. A physicist on the decoder side does not know this reference message and sees the output couplings as a *sample* of the random distribution $p(\{J\})$. The overlap between the reference and the decoded message is in this case the average magnetization. Hence, the ferromagnetic phase defines the region where the code works.

Applying the identity $\delta_{\sigma,\sigma'} = \frac{1}{2}(1+\sigma\sigma')$, $s, s" = \pm 1$, and summing first over J_α one obtains from (7)

$$[\langle J_\alpha\rangle] = p_2 - p_1 + (1 - p_1 - p_2)\langle\gamma_\alpha\rangle \tag{16}$$

and

$$[\langle J_\alpha\gamma_\alpha\rangle] = 1 - p_1 - p_2 + (p_2 - p_1)\langle\gamma_\alpha\rangle \tag{17}$$

If one sends always the same reference message $\mathbf{r}$, $\gamma_\alpha = \prod_{i\in\alpha} r_i$.

Let $\mathbf{r} = (1, 1, \ldots, 1)$ and $p_1 = p_2 = p$ (symmetric channel). Then $a = 1/2\ln(qp)$, $b = c = 0$, and $d = 1/2\ln(q/p) = \beta_N$. Both the encoding and decoding Hamiltonians are invariant under the gauge transformation

$$\begin{aligned} J_\alpha &\rightarrow J_\alpha\prod_{i\in\alpha}\tau_i \\ \sigma_i &\rightarrow \sigma_i\tau_i \end{aligned} \tag{18}$$

where $\tau_i = \pm 1$ are local gauge variables. Since **we** know $\mathbf{r}$, we can evaluate (but not the decoding physicist!) the internal energy directly as the energy of the reference message *after* transmission [5], $\langle E\rangle = H_{dec}(\mathbf{r}|\{J\}) = -M(1-2p)$, or from Eq. (17). Hence, we recover the result of Nishimori [8,9] for all 'gauge invariant' correlation depending only on combinations of $J_\alpha\gamma_\alpha$-type products. However, I did not use the gauge symmetry (18) when deriving Eqs. (16)-(17), valid for the general asymmetric channel.

For the Edwards-Anderson model (4) the code rate equals $R = 2/z$, where z is the coordination number. Numerical evidence [5,10] supports the idea that the *error-correcting capability* [11] of the binary symmetric code will be lost around the critical noise level p_S satisfying

$$R = \frac{2}{z} \simeq C = 1 + \frac{1}{2}\left[p_1\log_2 p_1 + q_1\log_2 q_1 + p_2\log_2 p_2 + q_2\log_2 q_2\right] \stackrel{p_1=p_2}{=} 1 + p_S\log_2 p_S + (1-p_S)\log_2(1-p_S) \tag{19}$$

The minimal average error per bit and therefore the 'strongest' ferromagnetic signature can be extracted on the Nishimori line, $\beta = \beta_N$ [5,12,13]. Since Shannon's theorem uses maximum likelihood, corresponding to $T = 0$ decoding [2], the estimate (19) is valid for all temperatures $T \leq T_N$. For example, the ferromagnetic phase in the EA-model with binary distributed couplings would extend only over $p \leq p_c \approx 0.11003$ for a square and $p \leq p_c \approx 0.17395$ for a simple cubic lattice, which are very close to numerical results obtained for these models. Note that $p_c = 1 - p_S$. The ferromagnetic phase is bounded at p_c by $k_BT/J \leq k_BT_N/J = 1/\beta_N(p_S)$. Since the Nishimori variety must pass through a tricritical point where the paramagnetic, the ferromagnetic, and the spin-glass phase meet, $\beta_N(p_S)$ provides also an estimate of the spin-glass transition. The physical interpretation of these results is very simple: the model (4) is used as a code for the ferromagnetic state. As long as the redundance of the code outweights the quenched disorder the average magnetization does not vanish.

What kind of model is generated by the usual coding protocol, sending all messages with the appropriate weight? Consider first sending the ferromagnetic state, $\gamma_\alpha = 1$, $\forall\alpha$. Each received set of couplings, $\{J\}$, can be considered as an independent sample of the binary distribution Eq. (5) with $p = p_1$. If the reference is the antiferromagnetic state, $\gamma_\alpha = -1$, $\forall\alpha$, the recipient will detect a binary distribution but whith $p = q_2$. In general, when sampling messages from $p(\mathbf{s})$, the decoder will compute for each message $\mathbf{s}$ a binary distribution Eq. (5) whose parameter p equals

$$p = \frac{1}{2}(1-\epsilon)(p_1 - q_2) + q_2 \tag{20}$$

where $\epsilon = (1/M)\sum_{\langle i,j\rangle}\langle s_i s_j\rangle$ and the thermal average is over $p(\mathbf{s})$. Therefore, error-correcting codes generate spin-glass systems where the parameters of the nonequilibrium distribution $p(\{J\};\vec{\lambda})$ are themselves *random variables* determined by the source distribution, $p(\mathbf{s})$. One could call such models 'doubly quenched' spin-glasses.

Our second example is the binary erasure channel. In this case the channel has three outputs, $J_\alpha = -1, 0, 1$,

$$p(J_\alpha|\gamma_\alpha) = q\delta(J_\alpha - \gamma_\alpha) + p\delta(J_\alpha) + r\delta(J_\alpha + \gamma_\alpha), \tag{21}$$

and $p + q + r = 1$. This Ansatz leads to the diluted spin-glass Hamiltonian

$$-\beta H_{dec} = a + b\sum_\alpha J_\alpha^2 + \beta_N \sum_\alpha J_\alpha \prod_{i\in\alpha} s_i \tag{22}$$

where $a = \ln(p)$, $b = 1/2\ln(q/p) - a$ and the generalized Nishimori temperature is $\beta_N = 1/2\ \ln(q/r)$. The internal energy per bond ($\epsilon = -(1-p-r)$) and correlation functions involving $J_\alpha \prod_{i\in\alpha} s_i$ products can be trivially evaluated from Eq. (7). It is elementary to compute the corresponding channel capacity, which leads to the estimate

$$R = \frac{2}{z} \simeq C = q + r + q\log_2 q + r\log_2 r - (q+r)\log_2(q+r) \to 1 - p \text{ as } r \to 0 \tag{23}$$

In the limit $r \to 0$ the model describes a randomly diluted spin system. The generalized Nishimori line moves to $T = 0$ ($\beta_N \to \infty$). Therefore, generalized Nishimori solutions exist only for random systems with competing interactions (frustrated systems). In the $r \to 0$ limit Eq. (23) provides an estimate of the bond-percolation threshold ($p_c = 1 - p_S$) as $p_c \approx 2/z$.

In summary, this Letter shows that 1) the Nishimori-solutions can be generalized to spin-glass systems without gauge-invariance by using the Bayes-identity and appropriate Ansätze for the source and channel distributions, 2) casting the statistical inferrence problem in terms of a decoding Hamiltonian for convolution error-correcting codes leads in general to inhomogeneous spin-glass models whose bond-distribution is parametrized in terms of random variables, and 3) the channel theorem provides accurate estimates for the extent of homogeneous ordered phases generated by given reference messages.

The analogy between convolution error-correcting codes and spin-glass models put forward by N. Sourlas has been useful for both field. Statistical physics led to new nonabelian codes [3], finite temperature decoding [5], and analytical methods for computing the average error [10]. This Letter illustrates the gains on the statistical mechanics side. However, I believe this analogy runs deeper: it hints both at how Nature might implement error-correcting mechanisms through common physical interactions and at how *we* could use statistical physics for designing powerful information processing systems. But this is another story.

---

[1] S-K. Ma: *Statistical Mechanics, World Scientific, 1985*
[2] N. Sourlas *Nature* **339** (1989) 693
[3] N. Sourlas in L. Garrido (Ed.) *Statistical Mechanics of Neural Networks*, Lectures Notes in Physics **368**, pp. 317-330, Springer Verlag, 1991
[4] N. Sourlas *Spin-Glass Models and Error-Correcting Codes*, LPTENS 93/4 preprint, submitted to IEEE Trans. on Inf. Theory
[5] P. Ruján *Phys. Rev. Lett.* **70** (1993) 2968
[6] C. E. Shannon *Bell Syst. Tech. J.* **27** (1948) 379 and 623
[7] R. J. McEliece *The Theory of Information and Coding* (Encyclopedia of Mathematics and its Applications), Addison-Wesley, 1977
[8] H. Nishimori *J. Phys. C* **13** (1980) 4071
[9] H. Nishimori *Progr. Theor. Phys.* **66** (1981) 1169
[10] C. Dress, E. Amic, and J. M. Luck *J. Phys. A* **28** (1995) 135
[11] Defined as the average error of encoded - average error of original messages.
[12] H. Nishimori *J. Phys. Soc. Japan* **62** (1993) 2973
[13] N. Sourlas *Europhys. Lett.* **25** (1994) 159