Statistical Physics of Irregular Low-Density Parity-Check Codes

[Statistical Physics of Irregular Codes]

Renato Vicente † §, David Saad † and Yoshyiuki Kabashima ‡

† The Neural Computing Research Group, Aston University, Birmingham B4 7ET, UK

‡ Department of Comptational Intelligence and System Science, Tokyo Institute of Technology, Yokohama 2268502, Japan

Abstract. Low-density parity-check codes with irregular constructions have been recently shown to outperform the most advanced error-correcting codes to date. In this paper we apply methods of statistical physics to study the typical properties of simple irregular codes. We use the replica method to find a phase transition which coincides with Shannon's coding bound when appropriate parameters are chosen. The decoding by belief propagation is also studied using statistical physics arguments; the theoretical solutions obtained are in good agreement with simulations. We compare the performance of irregular with that of regular codes and discuss the factors that contribute to the improvement in performance.

PACS numbers: 89.70+c 89.90+n 05.50+q

1. Introduction

Error-correction mechanisms are essential for preventing loss of information in transmissions through noisy environments. They are of increasing technological importance with applications ranging from high capacity storage media to satellite communication. The surprising fact that error-free communication is possible if the information is encoded to include a minimum amount of redundancy was discovered by Shannon in 1948 [1]. Shannon proved that a message encoded at rates R (message information content/code-word length) up to the channel capacity C can be decoded with vanishing average error probability $P_E \to 0$ as the length of the message increases $M \to \infty$. This theorem was then progressively refined by Gallager and others (see [2]

§ E-mail: vicenter@aston.ac.uk

and references therein) to say that the average over messages and codes of the error probability is bounded by

$$P_E < \mathrm{e}^{-ME(R)}, \tag{1}$$

where $E(R)$ is the error exponent that is greater than zero for rates up to the channel capacity C.

These proofs were presented in a non-constructive form by assuming encoding processes by ensembles of unstructured random codes and impractical decoding methods like maximum likelihood or typical set decoding [3]. No encoding-decoding scheme that is practical and attains the coding bound has been found to date.

The most successful code in use to date is the Turbo code [4]. However, the current performance record is owned by an irregular low-density parity check code (LDPC), more specifically an irregular Gallager code [5] †. This code was first proposed by Gallager in 1962 [6, 7], and were all but forgotten soon after due to technical limitations of the time. Recently a variation of the original proposal by Gallager named MN code has been proposed by MacKay and Neal [8, 9]; they showed that this code has good performance, what attracted renewed interest to LDPCs. Since then LDPCs have been reconsidered in a variety of architectures [10, 11]. Some of which reported close to optimal performance [12, 13].

Representing a message by a binary vector $\boldsymbol{\xi} \in \{0,1\}^N$, the LDPC encoding process consists of producing the binary vector $\boldsymbol{t} \in \{0,1\}^M$ defined by $\boldsymbol{t} = \boldsymbol{G^T s}$ (mod 2), where all operations are performed in the field $\{0,1\}$ and are indicated by (mod 2) and $\boldsymbol{G^T}$ is a $M \times N$ generator matrix. The transmission is then corrupted by noise, that we assume to be a binary vector $\boldsymbol{\zeta} \in \{0,1\}^M$, and the received vector takes the form $\boldsymbol{r} = \boldsymbol{G^T \xi} + \boldsymbol{\zeta}$ (mod 2). The decoding process is performed by applying a suitable parity-check matrix to the received message to produce the *syndrome* vector $\boldsymbol{z} = \boldsymbol{Ar}$ (mod 2). The parity-check matrix $\boldsymbol{A}$ defines the code structure and can be represented by a bipartite undirected graph with check and bit nodes. This gives rise to the classification of LDPCs to regular (those forming regular graphs) and irregular codes.

The parity-check matrix for Gallager codes is a concatenation $\boldsymbol{A} = [\boldsymbol{C_1} \mid \boldsymbol{C_2}]$ of two very sparse matrices, with $\boldsymbol{C_2}$ (of dimensionality $(M-N) \times (M-N)$) being invertible and the rectangular matrix $\boldsymbol{C_1}$ of dimensionality $(M-N) \times N$. The generator matrix of a Gallager code is $\boldsymbol{G} = [\boldsymbol{I} \mid \boldsymbol{C_2^{-1} C_1}]$ (mod 2), where $\boldsymbol{I}$ is the $N \times N$ identity matrix, implying that $\boldsymbol{AG^T}$ (mod 2) $= 0$ and that the message itself is set as the first N bits in the transmission. The syndrome vector is then $\boldsymbol{z} = \boldsymbol{Ar} = \boldsymbol{A\zeta}$ (mod 2) from which the noise can be estimated and subtracted from the received message. For a MN code the generator matrix has the form $\boldsymbol{G^T} = \boldsymbol{C_n^{-1} C_s}$ (mod 2), where $\boldsymbol{C_n}$ is an $M \times M$ invertible matrix and $\boldsymbol{C_s}$ is $M \times N$. The matrix applied by the decoder is given by $\boldsymbol{C_n}$ producing

† See *http://www331.jpl.nasa.gov/public/JPLtcodes.html* for JPL's " imperfectness" contest.

$\boldsymbol{z} = \boldsymbol{C_n r} = \boldsymbol{C_s \xi} + \boldsymbol{C_n \zeta}$ (mod 2), from which the most probable message vector can be predicted.

Although Gallager and MN codes can be analysed by the same methods of information theory [9], they represent different physical systems with different properties. In this paper we will restrict the analysis to irregular MN codes, the analysis of Gallager codes will appear elsewhere.

Statistical physics has first been applied to the analysis of error-correcting codes in the seminal work of Sourlas [14] which has been recently extended to the case of finite code rates [15, 16]. Similar methods have been recently applied to the case of Turbo codes [19] and regular MN codes [17, 18], providing a detailed description of the system's phases and capabilities for various parameter choices. Here we analyse irregular MN codes using the standard replica calculation to find a free energy that is a measure of the likelihood of typical solutions to the decoding problem, given an ensemble of code matrices $\boldsymbol{C_s}$ and $\boldsymbol{C_n}$ (*code construction*), channel and message models (*noise level* and *message bias*).

We show that three types of solutions emerge depending on the parameters provided: successful errorless decoding (number of incorrect bits less than $\mathcal{O}(N)$), imperfect decoding (number of incorrect bits of order N) and complete failure (number of correct bits less than $\mathcal{O}(N)$). We also show, as in [17, 18], that the line separating errorless and complete failure phases can coincide with the coding limit; this fact itself is not particularly surprising as the statistical physics analysis relies on the same kind of arguments used in the original coding bounds, using averages over ensembles of codes and maximum likelihood decoding. The main difference here is that the matrices in the ensemble have some structure.

The statistical physics approach can be regarded as complementary to that of information theory; it enables one to attain a more complete picture by analysing the decoding problem in the infinite message limit and by looking at global properties of the free energy. It allows for a transparent analysis of the possible performance of different codes characterised by different choices of construction parameters, and has already resulted in new practical high performance codes [13].

In this framework, Bayes-optimal decoding generally corresponds to finding the global minimum of a TAP free energy [20, 21] which is very costly if the landscape has multiple local minima. A practical decoding algorithm that has been used in LDPCs is the scheme known as belief propagation, broadly used by the Bayesian inference community [26, 27]. Belief propagation is equivalent to solving iteratively a set of coupled equations for finding extrema (local or global) of the TAP free energy [16, 18, 28]. This method is very sensitive to the presence of local minima and can be easily trapped in sub-optimal solutions.

In this paper we study the dependence of the free energy surface on the noise level and the message bias; this allows us to study the solutions which exist in each one of the cases and to detect the emergence of suboptimal solutions that will interfere in the practical decoding dynamics.

This paper is organised as follows: Section 2 presents irregular MN codes, while the statistical physics analysis is outlined in Section 3; the relations between the belief propagation approach and statistical physics are discussed in Section 4 and employed to examine the decoding performance in Sections 5 and 6. Concluding remarks are given in Section 7.

2. Irregular MN codes

Although the best irregular LDPCs found so far are defined in q-ary alphabets [30], we will restrict the current analysis to the binary alphabet $\{0,1\}$.

We suppose that the binary messages $\boldsymbol{S}$ comprise independent bits sampled from the prior distribution $P(S) = (1-p)\,\delta(S) + p\,\delta(S-1)$, where $\delta(S)$ stands for the Dirac's delta distribution. We also assume a simple memoryless Binary Symmetric Channel (BSC) with binary vectors $\boldsymbol{\tau}$ having independent components sampled from a similar prior distribution of the form $P(\tau) = (1-f)\,\delta(\tau) + f\,\delta(\tau-1)$. From now on we will reserve the symbols $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ for the actual message and noise, using $\boldsymbol{S}$ and $\boldsymbol{\tau}$ for denoting random variables in the message and noise models.

The goal is then to find the Bayes-optimal estimate $\widehat{S}_j = \mathrm{argmax}_{S_j} \mathrm{Tr}_{S_{i\neq j},\boldsymbol{\tau}} P(\boldsymbol{S},\boldsymbol{\tau} \mid \boldsymbol{z})$; the matrices $\boldsymbol{C_n}$ and $\boldsymbol{C_s}$ are also given, but were omitted for brevity.

One can use Bayes formula to incorporate the prior knowledge on message and noise and write the adequate posterior probability:

$$P(\boldsymbol{S},\boldsymbol{\tau} \mid \boldsymbol{z}) = \frac{1}{Z}\,\chi\left\{\boldsymbol{C_s S} + \boldsymbol{C_n \tau} = \boldsymbol{z} \ (\mathrm{mod}\ 2)\right\} P(\boldsymbol{S})P(\boldsymbol{\tau}), \qquad (2)$$

where the indicator function is $\chi\{A\} = 1$ if A is true and 0 otherwise.

The matrices are chosen at random in such a way that $\boldsymbol{C_n}$ is invertible over the field $\{0,1\}$ and a row m in $\boldsymbol{C_s}$ and $\boldsymbol{C_n}$ contains K_m and L_m non-zero elements respectively. In the same way, each column j of $\boldsymbol{C_s}$ contains C_j non-zero elements and each column l of $\boldsymbol{C_n}$ contains D_l non-zero elements.

Parity-checks for the signal and noise bits are specified by the matrices $\boldsymbol{C_s}$ and $\boldsymbol{C_n}$ respectively. The system can be mapped onto a bipartite graph represented by $(\boldsymbol{C_s} \mid \boldsymbol{C_n})$ (adjacency matrix in the graph theory jargon), to say, each one of the M rows lists the bit nodes connected to a check node and each one of the $N+M$ columns lists the checks conveying information about the particular bit node. Therefore, the sets $\{K_m\}_{m=1}^{M}$ and $\{L_n\}_{n=1}^{M}$ give the order of check nodes, $\{C_j\}_{j=1}^{N}$ and $\{D_l\}_{l=1}^{M}$ the order of bit nodes. Clearly this sets must obey the relations:

$$\sum_{j=1}^{N} C_j = \sum_{m=1}^{M} K_m \qquad \sum_{l=1}^{M} D_l = \sum_{m=1}^{M} L_m, \tag{3}$$

standing for the number of edges in the signal and noise graphs respectively.

The information rate of the code is given by $R = H_2(p)\ M/N$, where $H_2(p) = -p\log_2(p) - (1-p)\log_2(1-p)$ is the binary entropy of the source.

Alternatively one can write $R = H_2(p)\overline{K}/\overline{C}$, where:

$$\overline{K} = \frac{1}{M}\sum_{m=1}^{M} K_m \qquad \overline{C} = \frac{1}{N}\sum_{j=1}^{N} C_j \tag{4}$$

To simplify the calculations we change, as in the original work by Sourlas [14], the representation of the variables, replacing the field $\{0,1\}$ by $\{\pm 1\}$ and modulo 2 sums by products. Moreover, we restrict our analysis to the case of irregular bit nodes (sets $\{C_j\}_{j=1}^{N}$ and $\{D_l\}_{l=1}^{M}$) and regular check nodes (fixed K and L). The case with regular bit nodes and irregular check nodes is the basis for high performance codes studied in [13].

3. Equilibrium theory

To assess the performance of irregular MN codes we compute, using standard techniques, the free energy of the system $f = -\lim_{N\to\infty}\frac{1}{N}\langle \ln Z\rangle$ where Z is the normalisation in (2). The average $\langle ...\rangle$ is performed over the matrices $\boldsymbol{C_n}$ and $\boldsymbol{C_s}$, the messages $\boldsymbol{\xi}$ and the noise $\boldsymbol{\zeta}$ and will provide information about the *typical* performance of these codes.

In the ± 1 representation, the syndrome vector $\boldsymbol{z} = \boldsymbol{C_n r} = \boldsymbol{C_s \xi} + \boldsymbol{C_n \zeta} \pmod 2$ becomes $\mathcal{J}_{\mu\sigma} = \prod_{j\in\mu}\xi_j \prod_{l\in\sigma}\zeta_l$, where $\mu = \langle i_1, \cdots, i_K\rangle$ and $\sigma = \langle l_1, \cdots, l_L\rangle$ are sets of indices corresponding to the non-zero elements in one of the M rows of $\boldsymbol{C_s}$ and $\boldsymbol{C_n}$ respectively.

The prior distribution over the message bits $S_j \in \{\pm 1\}$ becomes $P(S_j) = (1-p)\,\delta(S_j - 1) + p\,\delta(S_j + 1)$, while for the noise bits $\tau_l \in \{\pm 1\}$ one has $P(\tau_l) = (1-f)\,\delta(\tau_l - 1) + f\,\delta(\tau_l + 1)$.

The code construction is specified by the tensor $\mathcal{A}_{\mu\sigma} \in \{0,1\}$ that determines the set of indices $\mu\sigma$ which correspond to non-zero elements in a particular row of the matrix $(\boldsymbol{C_s} \mid \boldsymbol{C_n})$. To cope with non-invertible $\boldsymbol{C_n}$ matrices one can start by considering an ensemble with uniformly generated $M\times M$ matrices. The non-invertible instances can then be made invertible by eliminating a $\epsilon \sim \mathcal{O}(1)$ number of rows and columns, resulting in an ensemble of $(M-\epsilon)\times(M-\epsilon)$ invertible $\boldsymbol{C_n}$ matrices and $(M-\epsilon)\times(N-\epsilon)$ $\boldsymbol{C_s}$ matrices. As we are interested in the thermodynamical limit we can neglect $\mathcal{O}(1)$ differences and compute the averages in the original space of $M\times M$ matrices. The averages are then performed over an ensemble of codes generated as follows:

(i) sets of numbers $\{C_j\}_{j=1}^{N}$ and $\{D_l\}_{l=1}^{M}$ are sampled independently from distributions $\mathcal{P}_C$ and $\mathcal{P}_D$ respectively;

(ii) tensors $\mathcal{A}_{\mu\sigma}$ are generated such that $\sum_{\mu\sigma}\mathcal{A}_{\mu\sigma} = M$, $\sum_{\{\mu:j\in\mu\}}\mathcal{A}_{\mu\sigma} = C_j$ and $\sum_{\{\sigma:l\in\sigma\}}\mathcal{A}_{\mu\sigma} = D_l$, where $\{\mu : j \in \mu\}$ stands for all sets of indices that contain j.

The indicator χ in (2) can be replaced by a more tractable function that is $E(\boldsymbol{S},\boldsymbol{\tau};\mathcal{A}) = 1$, if the dynamical variables $\boldsymbol{S}$ and $\boldsymbol{\tau}$ satisfy $\mathcal{J}_{\mu\sigma} = \prod_{j\in\mu} S_j \prod_{l\in\sigma}\tau_l$ and $E(\boldsymbol{S},\boldsymbol{\tau};\mathcal{A}) = 0$ otherwise. This function has the form:

$$E(\boldsymbol{S},\boldsymbol{\tau};\mathcal{A}) = \lim_{\beta\to\infty}\exp\left\{-\beta\sum_{\mu\sigma}\mathcal{A}_{\mu\sigma}\,\left[\mathcal{J}_{\mu\sigma}\prod_{j\in\mu}S_j\prod_{l\in\sigma}\tau_l - 1\right]\right\}. \tag{5}$$

The priors over message and noise take the form of external fields in the statistical physics framework and can be written in an exponential form with the normalisation incorporated in the partition function Z:

$$P(\boldsymbol{S},\boldsymbol{\tau}) \sim \exp\left(F_s\sum_{j=1}^{N}S_j \;+\; F_n\sum_{l=1}^{M}\tau_l\right), \tag{6}$$

the fields are then $F_s = \mathrm{atanh}(1-2p)$ and $F_n = \mathrm{atanh}(1-2f)$.

As in [17, 18], the partition function becomes:

$$Z = \lim_{\beta\to\infty}\mathrm{Tr}_{\boldsymbol{S},\boldsymbol{\tau}}\exp\left[\beta\left(\sum_{\mu\sigma}\mathcal{A}_{\mu\sigma}\left(\mathcal{J}_{\mu\sigma}\prod_{j\in\mu}S_j\prod_{l\in\sigma}\tau_l - 1\right) + \frac{F_s}{\beta}\sum_{j=1}^{N}S_j + \frac{F_\tau}{\beta}\sum_{l=1}^{M}\tau_l\right)\right]. \tag{7}$$

Performing the gauge transformation $S_j \mapsto \xi_j S_j$ and $\tau_l \mapsto \zeta_l\tau_l$ one obtains:

$$\mathcal{H} = -\sum_{\mu\sigma}\mathcal{A}_{\mu\sigma}\left(\prod_{j\in\mu}S_j\prod_{l\in\sigma}\tau_l - 1\right) - \frac{F_s}{\beta}\sum_{j=1}^{N}\xi_j S_j - \frac{F_\tau}{\beta}\sum_{l=1}^{M}\zeta_l\tau_l. \tag{8}$$

The resulting Hamiltonian represents a multi-spin ferromagnet in a random field, the disorder is transformed as $\mathcal{J}_{\mu\sigma} \mapsto 1$ under the gauge transformation, and therefore, is trivial and there is no frustration in the system. The different phases that will appear are then due to competition between the local fields and ferromagnetic interactions. Due to the structure of (5) all the thermodynamics is obtained in the *zero temperature* limit unlike the Sourlas' code case where optimal decoding must be carried out at finite temperatures [15, 16, 22, 23, 24, 25].

The free energy $f(p,f,\alpha,\mathcal{P}_C,\mathcal{P}_D) = -\lim_{N\to\infty}\frac{1}{N}\langle\ln Z\rangle_{\mathcal{A},\boldsymbol{\xi},\boldsymbol{\zeta}}$ can be determined using the replica method along the same lines as reported in [15, 16, 17], but for the irregular case it also depends on the probability distributions $\mathcal{P}_C$ and $\mathcal{P}_D$ used to generate the ensemble of codes. The auxiliary variables $q_{\alpha_1\cdots\alpha_m} = N^{-1}\sum_j Z_j S_j^{\alpha_1}\cdots S_j^{\alpha_m}$

and $r_{\alpha_1\cdots\alpha_m} = M^{-1}\sum_l Y_l \tau_l^{\alpha_1}\cdots\tau_l^{\alpha_m}$, and their conjugates $\widehat{q}_{\alpha_1\cdots\alpha_m}$ and $\widehat{r}_{\alpha_1\cdots\alpha_m}$, emerge from the calculation. The replica symmetry assumption is enforced by using the ansätze:

$$q_{\alpha_1\cdots\alpha_m} = \int dx\ \pi(x)\ x^m \qquad \widehat{q}_{\alpha_1\cdots\alpha_m} = \int d\widehat{x}\ \widehat{\pi}(\widehat{x})\ \widehat{x}^m \tag{9}$$

and

$$r_{\alpha_1\cdots\alpha_m} = \int dy \rho(y)\ y^m \qquad \widehat{r}_{\alpha_1\cdots\alpha_m} = \int d\widehat{y}\ \widehat{\rho}(\widehat{y})\ \widehat{y}^m. \tag{10}$$

The expression for the free energy then follows:

$$\begin{aligned} f(p,f,\alpha,\mathcal{P}_C,\mathcal{P}_D) = \mathrm{Extr}_{\{\widehat{\pi},\pi,\widehat{\rho},\rho\}}\Bigg\{ & \alpha\ \ \ln 2 \\ & -\alpha \int \left[\prod_{j=1}^{K} dx_j \pi(x_j)\right] \left[\prod_{l=1}^{L} dy_l \rho(y_l)\right]\ \ln\left(1+\prod_{j=1}^{K} x_j \prod_{l=1}^{L} y_l\right) \\ & +\ \overline{C}\int dx\ \pi(x)\ d\widehat{x}\ \widehat{\pi}(\widehat{x})\ \ln\left(1+x\widehat{x}\right) +\ \alpha\ \overline{L}\int dy\ \rho(y)\ d\widehat{y}\ \widehat{\rho}(\widehat{y})\ \ln\left(1+y\widehat{y}\right) \\ & -\sum_C\ \mathcal{P}_C(C)\ \int \left[\prod_{j=1}^{C} d\widehat{x}_j\ \widehat{\pi}(\widehat{x}_j)\right] \left\langle\ \ln\left[e^{\xi F_s}\prod_{j=1}^{C}(1+\widehat{x}_j) + e^{-\xi F_s}\prod_{j=1}^{C}(1-\widehat{x}_j)\right]\right\rangle_{\xi} \\ & -\ \alpha\sum_D\ \mathcal{P}_D(D)\ \int \left[\prod_{l=1}^{D} d\widehat{y}_l\ \widehat{\rho}(\widehat{y}_l)\right] \left\langle\ \ln\left[e^{\tau F_\tau}\prod_{l=1}^{D}(1+\widehat{y}_l) + e^{-\tau F_\tau}\prod_{l=1}^{D}(1-\widehat{y}_l)\right]\right\rangle_{\tau}\ \Bigg\}, \end{aligned} \tag{11}$$

where $\alpha = M/N = \overline{C}/\overline{K}$.

The system's states are obtained by the extremization above, resulting in the saddle-point equations :

$$\begin{aligned} \widehat{\pi}(\widehat{x}) &= \int \prod_{j=1}^{K-1} dx_j\ \pi(x_j) \prod_{l=1}^{L} dy_l\ \rho(y_l)\ \delta\left[\widehat{x}\ -\ \prod_{j=1}^{K-1} x_j \prod_{l=1}^{L} y_l\right], \\ \widehat{\rho}(\widehat{y}) &= \int \prod_{j=1}^{K} dx_j\ \pi(x_j) \prod_{l=1}^{L-1} dy_l\ \rho(y_l)\ \delta\left[\widehat{y}\ -\ \prod_{j=1}^{K} x_j \prod_{l=1}^{L-1} y_l\right], \\ \pi(x) &= \sum_C \frac{C}{\overline{C}}\mathcal{P}_C(C) \int \prod_{j=1}^{C-1} d\widehat{x}_j\ \widehat{\pi}(\widehat{x}_j) \left\langle \delta\left[x\ -\ \tanh\left(F_s\xi + \sum_{l=1}^{C-1} \mathrm{atanh}\ (\widehat{x}_l)\right)\right]\right\rangle_{\xi}, \\ \rho(y) &= \sum_D \frac{D}{\overline{D}}\mathcal{P}_D(D) \int \prod_{l=1}^{D-1} d\widehat{y}_l\ \widehat{\rho}(\widehat{y}_l) \left\langle \delta\left[y\ -\ \tanh\left(F_\tau\zeta + \sum_{l=1}^{D-1} \mathrm{atanh}\ (\widehat{y}_l)\right)\right]\right\rangle_{\zeta}. \end{aligned} \tag{12}$$

The exact meaning of the fields π, $\widehat{\pi}$, ρ and $\widehat{\rho}$ were presented in [16, 28] and will be further discussed in the next section.

Due to (5) the estimate for the message is $\widehat{\boldsymbol{S}} = \mathrm{sgn}(\langle \boldsymbol{S}\rangle_{\beta\to\infty})$, where the average is thermal with Hamiltonian (8) in the zero temperature limit. The decoding performance can be measured by

$$m = \frac{1}{N}\left\langle \sum_{i=1}^{N} \widehat{S}_i \xi_i \right\rangle_{\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{A}} = \int dh\ \phi(h)\ \mathrm{sgn}(h), \tag{13}$$

where, as in [18]

$$\phi(h) = \sum_C \mathcal{P}_C(C) \int \prod_{j=1}^{C} d\widehat{x}_j \ \widehat{\pi}(\widehat{x}_j) \left\langle \delta \left[h \ - \ \tanh \left(F_s \xi + \sum_{l=1}^{C} \mathrm{atanh} \ (\widehat{x}_l) \right) \right] \right\rangle_{\xi} . \tag{14}$$

Solutions can be found easily for the case where $F_s = 0$ (unbiased messages) and the code constructions are generated by distributions $P_D(D)$ and $P_C(C)$ that vanish for $0 \leq C, D < 2$ (codes with at least two checks per bit). For $K, L > 2$ one finds just two types of solutions: a ferromagnetic state with magnetization $m = 1$,

$$\begin{aligned} \pi(x) &= \delta[x - 1] \qquad & \widehat{\pi}(\widehat{x}) &= \delta[\widehat{x} - 1] \\ \rho(x) &= \delta[y - 1] \qquad & \widehat{\rho}(\widehat{y}) &= \delta[\widehat{y} - 1], \end{aligned} \tag{15}$$

and a paramagnetic state with $m = 0$,

$$\begin{aligned} \pi(x) &= \delta[x] \qquad & \widehat{\pi}(\widehat{x}) &= \delta[\widehat{x}] \\ \rho(x) &= \langle \delta[y - \tanh(\zeta F_\tau)] \rangle_\zeta \qquad & \widehat{\rho}(\widehat{y}) &= \delta[\widehat{y}]. \end{aligned} \tag{16}$$

For other parameter choices, suboptimal ferromagnetic states with $0 < m < 1$ can also be found by solving the saddle-point equations (12) numerically.

The paramagnetic and ferromagnetic free energies can be easily computed by inserting (15) and (16) in (11) to give $f_{\mathrm{para}} = \alpha \ \ln 2 - \alpha \ \ln (\ 2 \cosh F_\tau)$ and $f_{\mathrm{ferro}} = - \ (1 - 2f) \ F_\tau$ respectively. One can instantly obtain a phase transition occurring at the critical code rate for the BSC $R_c = 1 - H_2(f)$, that is valid for every code construction under the restrictions $K, L > 2$, $C_j > 1$ and $D_l > 1$. This is the same phase transition as the one described in [17]. The critical code rate saturates the channel capacity and therefore Shannon's coding limit.

It is important to stress that the coding bound can *only* be attained in the case of unbiased messages. For biased messages ($F_s \neq 0$) the paramagnetic state (16) is not a solution for the saddle-point equations (12) and the thermodynamical transition can only be obtained numerically and must be bellow the Shannon's bound as can be shown by a simple upper bound proposed in [9].

The upper bound is based on the fact that each bit of the syndrome vector $\boldsymbol{z} = \boldsymbol{C_n r} = \boldsymbol{C_s \xi} + \boldsymbol{C_n \zeta}$ (mod 2) is a sum (or product, depending on the representation adopted) of K message bits with bias p with L noise bits with flip rate f. The probability of $z_i = +1$ is $p_z^+(K, L) = 1/2 \ (1 + (1 - 2p)^K (1 - 2f)^L)$. The maximum information content in the syndrome vector is then $MH_2(p_z^+)$. For the decoding process one has $MH_2(p_z^+) \geq NH_2(p) + MH_2(f)$, resulting in the bound $R \leq H_2(p_z^+) - H_2(f)$. Shannon's bound is recovered for unbiased patterns $p_z^+ = 1/2$, while for biased patterns the attainable rates must be bellow Shannon's bound as $H_2(p_z^+) < 1$.

The main question that remains to be addressed is the accessibility of the various states by a practical decoding algorithm. In particular, we will focus on the belief propagation decoding process. In this practical scenario the energy landscape may be dominated by the basin of attraction of paramagnetic or suboptimal ferromagnetic states even when the ferromagnetic state is the global minimum, degrading the practical performance of the code.

4. Statistical physics and belief propagation

The decoding problem focuses on finding a Bayes-optimal estimate (also known as *marginal posterior maximiser*, MPM) $\widehat{\boldsymbol{S}}$ for the original message, given the code structure, the syndrome vector $\mathcal{J}$ and prior probabilities p and f.

The Bayes-optimal estimator is defined as an estimator that minimises the posterior average of some determined loss function. Using the overlap between message and estimate as a loss function, the Bayes-optimal estimator that emerges is of the form $\widehat{S}_j = \mathrm{sgn}\langle S_j \rangle_{P(S_j|\mathcal{J})}$ [25]. The task of computing this estimator is usually very difficult as no simple form is known for the posterior $P(S_j \mid \mathcal{J})$ and an exponential number of operations is required.

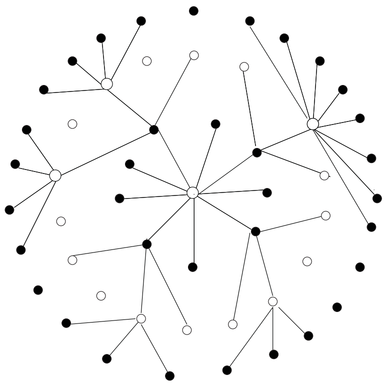

Figure 1. Tanner graph representing the neighbourhood of a bit node in an irregular MN code. Black circles represent checks and white circles represent bits.

The problem can be solved in practical time scales by applying the belief propagation (BP)[26] framework. In this framework, an approximation for the marginal posterior probabilities $P(S_j \mid \mathcal{J})$ can be computed iteratively in linear time. For that, a graphical representation (belief network) for dependencies between check nodes (or evidence nodes) and signal nodes can be constructed. By identifying proper substructures in the belief network one can write a closed set of equations whose solutions provide the approximation to the posterior probabilities. These substructures can be uniquely identified with conditional distributions. For LDPCs these probability distributions are:

$$q_{\mu j}^{(S)} = P(S_j = S \mid \{\mathcal{J}_{\nu\sigma\in\mathcal{M}_s(j)\setminus\mu}\}) \qquad \widehat{q}_{\mu j}^{(S)} = P(\mathcal{J}_{\mu\sigma} \mid S_j = S, \{\mathcal{J}_{\nu\kappa\neq\mu\sigma}\}) \tag{17}$$

$$r_{\sigma l}^{(\tau)} = P(\tau_l = \tau \mid \{\mathcal{J}_{\mu\kappa\in\mathcal{M}_n(l)\setminus\sigma}\}) \qquad \widehat{r}_{\sigma l}^{(\tau)} = P(\mathcal{J}_{\mu\sigma} \mid \tau_l = \tau, \{\mathcal{J}_{\nu\kappa\neq\mu\sigma}\}), \tag{18}$$

where $\mathcal{M}_s(j) \setminus \mu$ $(\mathcal{M}_n(l) \setminus \sigma)$ denote the set of checks connected to the signal bit j (noise bit l) excluding the check containing the bits in μ (noise bits in σ). Using Bayes' theorem, the posterior probabilities $\mathcal{P}(S_j \mid \mathcal{J})$ can then be written in terms of $\widehat{q}_{\mu j}^{(S)}$ and a priori distributions $P_0(S)$ [28].

The Gibbs weight appearing in Equation (7), as observed in [22, 28], is proportional to $P(\mathcal{J} \mid \boldsymbol{S})P_0(\boldsymbol{S})$ and can be used to write update formulas for the distributions. Introducing $m_{\mu j}^s = q_{\mu j}^{(+1)} - q_{\mu j}^{(-1)}$ and $m_{\nu l}^n = r_{\nu l}^{(+1)} - r_{\nu l}^{(-1)}$, following the steps described in [28] one can find the following set of equations:

$$m_{\mu l}^s = \tanh\left[\sum_{\nu\in\mathcal{M}_s(l)\setminus\mu} \mathrm{atanh}(\widehat{m}_{\nu l}^s) + F_s\right] \qquad \widehat{m}_{\mu j}^s = \mathcal{J}_\mu \prod_{i\in\mathcal{L}_s(\mu)\setminus j} m_{\mu i}^s \prod_{l\in\mathcal{L}_n(\mu)} m_{\mu l}^n, \tag{19}$$

$$m_{\sigma l}^n = \tanh\left[\sum_{\nu\in\mathcal{M}_n(l)\setminus\sigma} \mathrm{atanh}(\widehat{m}_{\nu l}^n) + F_n\right] \qquad \widehat{m}_{\mu j}^n = \mathcal{J}_\mu \prod_{i\in\mathcal{L}_s(\mu)} m_{\mu i}^s \prod_{l\in\mathcal{L}_n(\mu)\setminus j} m_{\mu l}^n, \tag{20}$$

where the set of signal bits (noise bits) in a check μ (σ) is represented by $\mathcal{L}_s(\mu)$ $(\mathcal{L}_n(\mu))$. The notation $\mathcal{L}_s(\mu) \setminus l$ indicates all bits in check μ excluding bit l, Greek letters run from 1 to M and Latin letters run from 1 to N.

The estimate for the message is $\widehat{S}_j = \mathrm{sgn}(m_j^s)$, where m_j^s is computed as:

$$m_j^s = \tanh\left[\sum_{\nu\in\mathcal{M}_s(j)} \mathrm{atanh}(\widehat{m}_{\nu j}^s) + F_s\right] \tag{21}$$

The BP decoding dynamics consists of updating Equations (19) and (20) until a certain halting criteria is reached, and then computing the estimate for the message using equation (21). The initial conditions are set to reflect the prior knowledge about the message $m_{\mu j}^s(0) = 1 - 2p$ and noise $m_{\sigma l}^n(0) = 1 - 2f$.

The BP algorithm is known to provide the *exact* posterior when the Tanner graph (see [29] and references therein) associated to the system has a tree architecture. A Tanner graph is a bipartite graph where checks are represented by black circles, bits are represented by white circles and an edge connects bits to their related checks.

When very sparse matrices are used, the probability for a loop in the related graph in a finite number of generations decays as γ/N, where $\gamma \sim \mathcal{O}(1)$ [31]. For finite systems one can expect that a limited neighbourhood of node has a tree structure. When applying the thermodynamical limit $N \to \infty$, the topology actually converges to a tree and BP equations become exact. In Figure 1 we show a Tanner graph representing the neighbourhood of a bit node in a large irregular MN code.

Equations (19) and (20) can also be obtained by looking for extrema of the TAP free-energy [18]:

$$
\begin{aligned}
f_{\mathrm{TAP}}(\boldsymbol{m},\widehat{\boldsymbol{m}}) \; &= \frac{M}{N}\ln 2 + \frac{1}{N}\sum_{\mu=1}^{M}\sum_{i\in\mathcal{L}_s(\mu)} \ln\left(1+m^s_{\mu i}\widehat{m}^s_{\mu i}\right) + \frac{1}{N}\sum_{\mu=1}^{M}\sum_{j\in\mathcal{L}_n(\mu)} \ln\left(1+m^n_{\mu j}\widehat{m}^n_{\mu j}\right) \\
&\quad - \frac{1}{N}\sum_{\mu=1}^{M}\ln\left(1+\mathcal{J}_\mu \prod_{i\in\mathcal{L}_s(\mu)} m^s_{\mu i} \prod_{j\in\mathcal{L}_n(\mu)} m^n_{\mu j}\right) \\
&\quad - \frac{1}{N}\sum_{i=1}^{N}\ln\left[e^{F_s}\prod_{\mu\in\mathcal{M}_s(i)}\left(1+\widehat{m}^s_{\mu i}\right) + e^{-F_s}\prod_{\mu\in\mathcal{M}_s(i)}\left(1-\widehat{m}^s_{\mu i}\right)\right] \\
&\quad - \frac{1}{N}\sum_{j=1}^{M}\ln\left[e^{F_n}\prod_{\mu\in\mathcal{M}_n(j)}\left(1+\widehat{m}^n_{\mu j}\right) + e^{-F_n}\prod_{\mu\in\mathcal{M}_n(j)}\left(1-\widehat{m}^n_{\mu j}\right)\right] \; . \qquad (22)
\end{aligned}
$$

Observe that the TAP free energy described above is not equivalent to the variational mean-field free energy introduced in [10, 32]. Here no essential correlations except those related to the presence of loops are disregarded.

The meaning of the fields introduced in the previous section can be understood by first applying the gauge transformations $m^s_{\mu j} \mapsto \xi_j m^s_{\mu j}$, $\widehat{m^s}_{\mu j} \mapsto \xi_j \widehat{m^s}_{\mu j}$, $m^n_{\sigma l} \mapsto \zeta_l m^n_{\sigma l}$ and $\widehat{m^n}_{\sigma l} \mapsto \zeta_l \widehat{m^n}_{\sigma l}$ to the TAP free energy and introducing new variables $x \equiv m^s_{\mu j}$, $\widehat{x} \equiv \widehat{m^s}_{\mu j}$, $y \equiv m^n_{\sigma l}$ and $\widehat{y} = \widehat{m^n}_{\sigma l}$. If $x, \widehat{x}, y$ and $\widehat{y}$ are interpreted as random variables generated by the probability distributions $\pi, \widehat{\pi}, \rho$ and $\widehat{\rho}$ respectively, one recovers the replica symmetric free energy (11) (see also [16]).

From the statistical physics point of view, belief propagation is one of many possible ways to find minima of the TAP free energy, representing simple iterative fixed point maps. The ferromagnetic state, corresponding to perfect decoding is the global minimum up to Shannon's limit in the case of unbiased messages (or very close to it in the case of biased messages). However, this equations are very sensitive to the presence of local minima in the landscape and the convergence to the global minimum is only expected if the initial conditions are set up within the basin of attraction of the ferromagnetic state, which requires prior knowledge about the message sent what is not the case in practical applications.

In the next sections we will try to address how the free energy landscape changes with the parameters.

5. Error-correction: regular vs. irregular codes

Irregularity improves the practical performance of a MN code. We now illustrate this for the simplest possible irregular constructions with a probability distribution describing

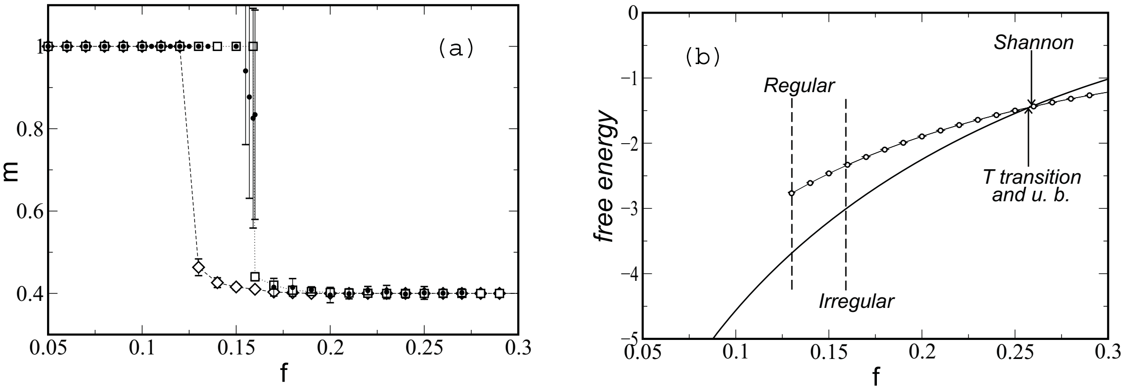

Figure 2. (a) Magnetization as a function of the noise level f for codes with $K = L = 3$ and $\overline{C} = 15$ with message bias $p = 0.3$. Analytical RS solutions for the regular code are denoted as $\diamond$ and for the irregular code; with $C_o = 4$ and $C_e = 30$ denoted as $\square$. Results are averages over 10 runs of the TAP/BP algorithm in an irregular code of size $N = 6000$ starting from fixed initial conditions (see the text) ; they are plotted as $\bullet$ in the rightmost curve for comparison. TAP/BP results for the regular case agree with the theoretical solutions and have been omitted to avoid overloading the figure. (b) Free energies for the ferromagnetic state (full line) and for the failure state (line with $\circ$). The transitions observed in (a) are indicated by the dashed lines. Arrows indicate the thermodynamical (T) transition, the upper bound (u.b.) of Section 3 and Shannon's limit.

connectivities of the signal matrix $\boldsymbol{C_s}$ chosen to be:

$$\mathcal{P}_C(C) = (1-\theta)\ \delta(C - C_o)\ +\ \theta\ \delta(C - C_e). \tag{23}$$

The mean connectivity is $\overline{C} = (1-\theta)\ C_o\ +\ \theta\ C_e$ and $C_o < \overline{C} < C_e$; bits in a group with connectivity C_o will be refered as *ordinary* bits and bits in a group with connectivity C_e as *elite* bits. The noise matrix $\boldsymbol{C_n}$ is chosen to be regular.

To gain some insight on the effect of irregularity on solving the TAP/BP equations (19) and (20) we performed several runs starting from the fixed initial conditions $m^s_{\mu j}(0) = 1 - 2p$ and $m^n_{\sigma l}(0) = 1 - 2f$ as prescribed in the last section. For comparison we also iterated the saddle-point equations (12) obtained in the replica symmetric (RS) theory, setting the initial conditions to be $\pi_0(x) = (1-p)\ \delta(x - m^s_{\mu j}(0))\ +\ p\ \delta(x + m^s_{\mu j}(0))$ and $\rho_0(y) = (1-f)\ \delta(y - m^n_{\sigma l}(0))\ +\ f\ \delta(y + m^n_{\sigma l}(0))$, as suggested from the interpretation of the fields $\pi(x)$ and $\rho(y)$ in the last section.

In Figure 2 (a) we show a typical curve for the magnetization as a function of the noise level. The RS theory agrees very well with TAP/BP decoding results. The addition of irregularity improves the performance considerably. In Figure 2 (b) we show the free energies of the two emerging states. The free energy for the ferromagnetic state with magnetization $m = 1$ is shown as a full line, the failure state (in Figure 2 (a) with magnetization $m = 0.4$) is shown as a line marked with $\circ$. The transitions seen in Figure 2 (a) are denoted by dashed lines. It is clear that they are far below the thermodynamical (T) transition, indicating that the system becomes trapped in suboptimal states for noise levels f between the observed transitions and the thermodynamical transition. The thermodynamical transition coincides with the upper bound (u.b.) in Section 3 and is very close to, but below, Shannon's limit which is shown for comparison. Similar behaviour has already been observed in regular MN codes with $K = 1$ in [18].

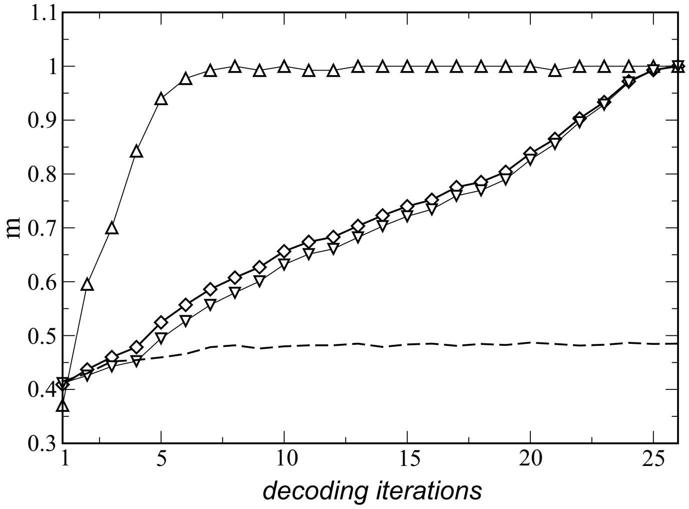

Figure 3. Magnetization monitored during the TAP/BP decoding process as a function of the number of iterations for $N = 4000$. Elite nodes magnetization is represented by $\triangle$. Ordinary nodes magnetization is represented by $\bigtriangledown$. The overall magnetization is represented by $\diamond$. The long dashed line shows the dynamics of the regular code. The constructions employed have parameters $K = L = 3$, $\overline{C} = 6$, $C_e = 20$ and $C_o = 5$. The noise level is $f = 0.065$ and the message bias is $p = 0.3$.

It is instructive to look how the magnetization of elite (m_e) and ordinary (m_o) nodes evolve throughout the iterative decoding process. In Figure 3 we show this dynamics for a regular and an irregular code at a noise level where the irregular code converges to the ferromagnetic state while the regular code fails (long-dashed lines). One can see that the magnetization of ordinary nodes follow that of the regular code in the first iterations, elite nodes are then corrected quickly achieving high magnetization values. These highly reliable nodes then lead the correction of ordinary nodes (around the fifth iteration), producing successful decoding. From the decoding dynamics point of view irregular MN codes can be qualitatively regarded as a mixture of low and highly connected regular codes where elite nodes can tolerate higher noise levels while ordinary nodes allow for higher code rates.

6. The spinodal point

In the last section we gained some insight on how irregularity affects the practical performance of codes. The dynamical decoding process shown in Figure 3 only provides a qualitative explanation and does not seem to allow some simple analysis.

A possible alternative is to relate the observation that the system gets trapped in suboptimal states (Figure 2) to global properties of the free energy. The TAP/BP algorithm can be regarded as an iterative solution of fixed point equations for the TAP free energy (22), which is sensitive to the presence of local minima in the system. One can expect convergence to the global minimum of the free energy from all initial conditions when there is a single minimum or when the landscape is dominated by the basin of attraction of this minimum when random initial conditions are used.

To analyse this point we rerun the decoding experiments starting from initial conditions $m^s_{\mu j}(0)$ and $m^n_{\sigma l}(0)$ that are random perturbations of the ferromagnetic solution :

$$m^s_{\mu j}(0) = (1-\rho_s)\ \delta(m^s_{\mu j}(0) - \xi_j)\ +\ \rho_s\ \delta(m^s_{\mu j}(0) + \xi_j), \tag{24}$$

and

$$m^n_{\sigma l}(0) = (1-\rho_n)\ \delta(m^n_{\sigma l}(0) - \tau_l)\ +\ \rho_n\ \delta(m^n_{\sigma l}(0) + \tau_l), \tag{25}$$

where for convenience we choose $0 \leq \rho_s = \rho_n = \rho \leq 0.5$.

We performed TAP/BP decoding several times for different values of ρ and noise level f. For $\rho \leq 0.026$ we observed that the system converges to the ferromagnetic state for *all* constructions, message biases p and noise levels f examined. It implies that this state is always stable. The convergence occurs for any ρ for noise levels below the transition observed in practice.

These observations suggest that the ferromagnetic basin of attraction dominates the landscape up to some noise level f_s. The fact that no other solution is ever observed in this region suggests that f_s is the noise level where suboptimal solutions actually appear, namely, it is the noise level that corresponds to the *spinodal point* of the system. This behaviour have already been observed for regular MN codes with $K=1$ or $K=L=2$ [17, 18].

In [17, 18] we have also shown that MN codes can be divided into three categories with different equilibrium properties: (i) $K \geq 3$ or $L \geq 3$, (ii) $K > 1$, $K = L = 2$ and (iii) general L, $K = 1$. In the next two subsections we will discuss these groups separately.

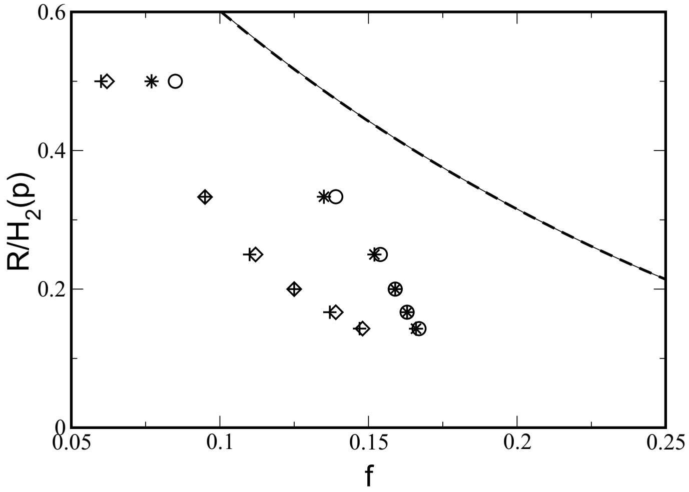

Figure 4. Spinodal point noise level f_s for regular and irregular codes. In both constructions parameters are set as $K = L = 3$. Irregular codes with $C_o = 4$ and $C_e = 30$ are used. TAP/BP decoding is carried out with $N = 5000$ and a maximum of 500 iterations; they are denoted by $+$ (regular) and $*$ (irregular). Numerical solutions for the RS saddle-point equations are denoted by $\diamond$ (regular) and $\circ$ (irregular). Shannon's limit is represented by a full line and the upper bound in Section 3 is represented by a dashed line. The symbols are chosen to be larger than the actual error bars.

6.1. Biased coding: $K \geq 3$ or $L \geq 3$

To show how irregularity affects codes with this choice of parameters we choose $K, L = 3$, $C_o = 4$, $C_e = 30$ and biased messages with $p = 0.3$. These choices are arbitrary but can illustrate what happens with the practical decoding performance. In Figure 4 we show the transition from the decoding phase to the failure phase as a function of the noise level f for several rates R in both regular and irregular codes. Practical decoding ($\diamond$ and $\circ$) results are obtained for systems of size $N = 5000$ with a maximum number of iterations set to 500. Random initial conditions are chosen and the whole process repeated 20 times. The practical transition point is found when the number of failures equals the number of successes.

These experiments were compared with theoretical values for f_s obtained by solving the RS saddle-point equations (12) (represented as $+$ and $*$ in Figure 4) and finding the noise level for which a second solution appears. For comparison the coding limit is represented in the same figure by a full line.

As the constructions used are chosen arbitrarily one can expect that these transitions

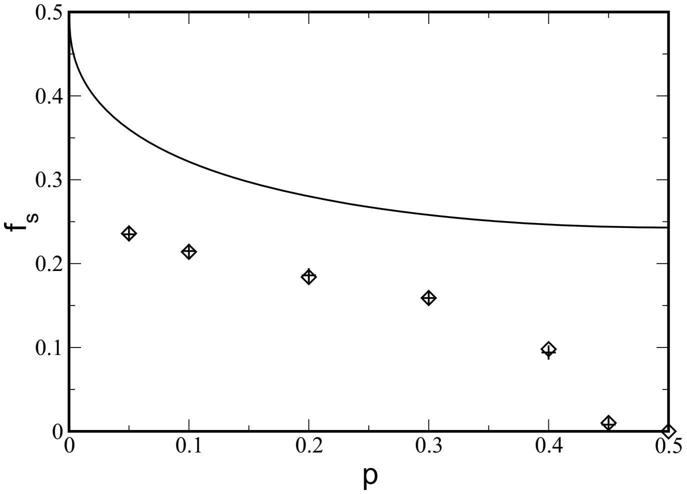

Figure 5. Spinodal point f_s for irregular codes as a function of the message bias p. The construction is parametrised by $K = L = 3$, $C_o = 4$ and $C_e = 30$ with $\overline{C} = 15$. TAP/BP decoding is carried out with $N = 5000$ and a maximum of 500 iterations, and is represented by $+$, while theoretical RS solutions are represented by $\diamond$. The full line indicates Shannon's limit. Symbols are larger than the actual error bars

can be further improved, even though the improvement shown in Figure 4 is already fairly significant.

The analytical solution obtained in Section 3 for $K \geq 3$ or $L \geq 3$, $K > 1$ and unbiased messages $p = 1/2$ implies that the system is bistable for arbitrary code constructions when these parameters are chosen. The spinodal point noise level is then $f_s = 0$ in this case and cannot be improved by adding irregularity to the construction. Up to the noise level f_c the ferromagnetic solution is the global minimum of the free energy, and therefore Shannon's limit is potentially saturated, however, the bistability makes these constructions unsuitable for practical decoding with a TAP/BP algorithm when unbiased messages are considered.

The situation improves when biased messages are used. Fixing the matrices $\boldsymbol{C_n}$ and $\boldsymbol{C_s}$ one can determine how the spinodal point noise level f_s depends on the bias p. In Figure 5 we compare simulation results with the theoretical predictions of f_s as a function of p. The spinodal point noise level f_s collapses to zero as p increases towards the unbiased case. It obviously suggests the use of biased messages for practical use of MN codes with parameters $K \geq 3$ or $L \geq 3$, $K > 1$ under TAP/BP decoding.

For biased messages with $K \geq 3$ or $L \geq 3$, $K > 1$ the qualitative picture of the energy landscape differs from the unbiased coding presented in [17, 18]. In Figure 6 this landscape is sketched as a function of the noise level f for a given bias. Up to the spinodal point f_s the landscape is totally dominated by the ferromagnetic state F. At the spinodal point another suboptimal state F' emerges, dominating the decoding

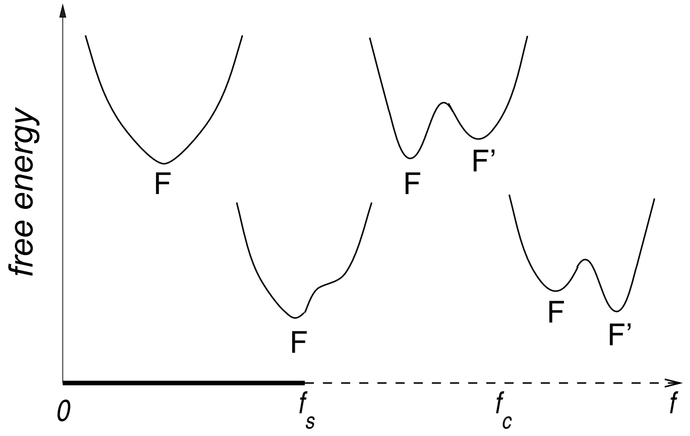

Figure 6. Pictorial representation of the free energy landscape as a function of the noise level f. Up to the spinodal point f_s there is only the ferromagnetic state F. At f_s another state F' appears dominating the decoding dynamics. The thermodynamical critical noise level f_c indicates the point where the state F' becomes the global minimum.

dynamics. At f_c the suboptimal state F' becomes the global minimum. The bold horizontal line represents the region where the ferromagnetic solution with $m = 1$ dominates the decoding dynamics. In the region represented by the dashed line decoding dynamics is dominated by suboptimal $m < 1$ solutions.

6.2. Unbiased coding

For the remaining parameter choices, namely general L, $K = 1$ and $K = L = 2$, it was shown in [17, 18] that unbiased coding is generally possible yielding close to Shannon's limit performance. The free energy landscape of the $K = 1$ was shown to behave in a similar way to the one depicted in Figure 6 while the landscape of the case $K = L = 2$ and unbiased messages shows a different behaviour where some regions include three stable states plus their mirror symmetries.

In the same way as in the $K \geq 3$ case the practical performance is defined by the spinodal point noise level f_s. The addition of irregularity also changes f_s in these cases.

In the general L, $K = 1$ family we illustrate the effect of irregularity by the choice of $L = 2$, $C_o = 4$ and $C_e = 10$. In Figure 7 we show the transitions observed by performing 20 decoding experiments with messages of length $N = 5000$ and a maximal number of iterations set to 500 (+ for regular and $*$ for irregular). We compare the experimental results with theoretical predictions based on the RS saddle-point equations (12) ($\diamond$ for regular and $\circ$ for irregular). Shannon's limit is represented by a full line. The

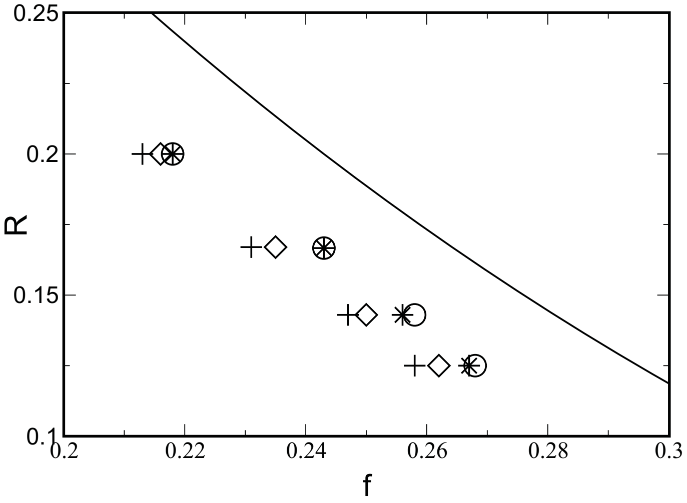

Figure 7. Spinodal point noise level f_s for regular and irregular codes. The constructions are of $K = 1$ and $L = 2$, irregular codes are parametrised by $C_o = 4$ and $C_e = 10$. TAP/BP decoding is carried out with $N = 5000$ and a maximum of 500 iterations ; they are denoted by $+$ (regular) and $*$ (irregular). Numerical solutions for RS equations are denoted by $\diamond$ (regular) and $\circ$ (irregular). The coding limit is represented by a line. Symbols are larger than the actual error bars.

improvement is modest, what is expected since regular codes already present close to optimal performance. Discrepancies between the theoretical and numerical results are due to finite size effects.

We also performed a set of experiments using $K = L = 2$ with $C_o = 3$ and $C_e = 8$, the same system size $N = 5000$ and maximal number of decoding iterations 500. The transitions obtained experimentally and predicted by theory are shown in Figure 8.

7. Conclusions

We showed that in the thermodynamic limit MN codes are equivalent to a multi-spin ferromagnet submitted to a random field. A replica calculation shows that a phase transition from an *errorless* (ferromagnetic) phase to a *failure* (either paramagnetic or suboptimal ferromagnetic) phase occurs as the noise level increases. The phase transition line can be analytically obtained in the case where constructions with $K, L \geq 3$, a minimum of two checks per bit and unbiased messages ($p = 1/2$) are used. It coincides with Shannon's coding limit and is independent of the code construction.

For other parameter choices the transition only can be obtained numerically and coincides with a simple upper bound, being necessarily below Shannon's limit.

The practical decoding using belief propagation is shown to attain inferior

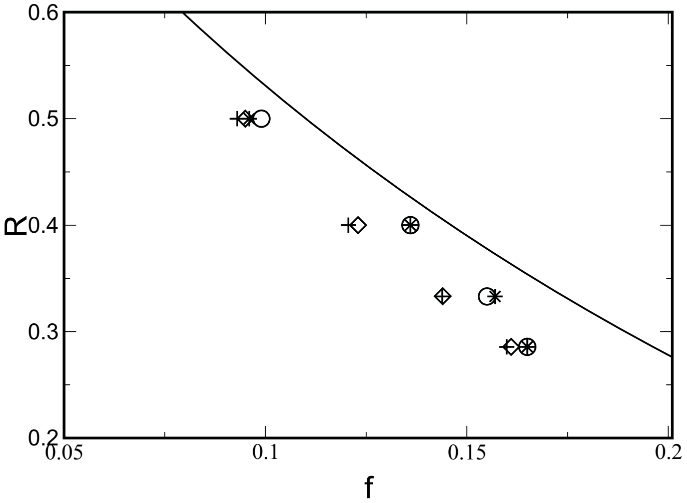

Figure 8. Spinodal point noise level values f_s for regular and irregular codes. Constructions are of $K = 2$ and $L = 2$, irregular codes are parametrised by $C_o = 3$ and $C_e = 8$. TAP/BP decoding is carried out with $N = 5000$ and a maximum of 500 iterations; they are denoted by $+$ (regular) and $*$ (irregular). Theoretical predictions are denoted by $\diamond$ (regular) and $\circ$ (irregular). The coding limit is represented by a line. Symbols are larger than the actual error bars.

performance to Shannon's limit due to the collapse of the ferromagnetic basin of attraction when new states emerge at the spinodal point noise level f_s. Irregularity increases f_s thus improving the code's performance. We show that the maximal noise level corrected by an MN code agrees with the replica theory prediction for the spinodal point noise level f_s.

This framework is currently being employed for optimising code constructions (recently studied in [12]), as well as for finding alternatives to the TAP/BP decoding scheme and for analysing the effect of using inaccurate priors.

Acknowledgments

We would like to thank the anonymous referees for their helpful comments. This work was partially supported by EPSRC grant GR/N00562, a Royal Society travel grant (RV and DS) and by the program "Research For the Future" (RFTF) of the Japanese Society for the Promotion of Science (YK).

References

[1] Shannon C 1948 *Bell Syst. Tech. J.* **27** 379-423

[2] Viterbi A J and Omura J K 1979 *Principles of Digital Communication and Coding* (Singapore: McGraw-Hill Book Co.)

[3] Cover T and Thomas J A 1991 *Elements of Information Theory* (New York, Wiley & Sons, Inc.)

[4] Berrou G, Glavieux A and Thitimajshima 1993 *Proc. IEEE Int. Conf. on Comm. (Geneva)* p 1064-70

[5] Davey M C 1998 *Record-breaking error correction using Low-Density Parity-Check Codes* (University of Cambridge, 1998 Hamilton Prize essay)

[6] Gallager R G 1962 *IRE Trans. Info. Theory* **8** 21-8

[7] Gallager R G 1963 *Low Density Parity Check Codes* (Cambidge, Mass., Research monograph series No.21 MIT Press)

[8] MacKay D J C and Neal R M 1996 *Electr. Lett.* **32** 1645-6

[9] MacKay D J C 1999 *IEEE Trans. Info. Theory* **45** 399-431

[10] MacKay D J C, Wilson S and Davey M C 1999 *IEEE Trans. on Comm.* **47** 1449-54

[11] Luby M *et. al.* 1998 *Digital SRC Technical Note* **8**

[12] Richardson T, Shokrollahi A and Urbanke R 1999 preprint

[13] Kanter I and Saad D 1999 *Phys. Rev. Lett.* **83** 2660-3; 2000 *J. Phys. A* **33** 1675-81

[14] Sourlas N 1989 *Nature* **339** 693-5

[15] Kabashima Y and Saad D 1999 *Europhys. Lett.* **45** 97-103

[16] Vicente R, Saad D and Kabashima Y 1999 *Phys. Rev. E* **60** 5352-66

[17] Kabashima Y, Murayama T and Saad D 2000 *Phys. Rev. Lett.* **84** 1355-8

[18] Murayama T, Kabashima Y, Saad D, Vicente R, cond-mat/0003121

[19] Montanari A. and Sourlas N. cond-mat/9909018 and Montanari A. cond-mat/0003218

[20] Thouless D J, Anderson P W and Palmer R G 1977 *Phil. Mag.* **35** 593-601

[21] Plefka T 1982 *J. Phys. A* **15** 1971-8

[22] Sourlas N 1994 *Europhys. Lett* **25** 159-64

[23] Ruján P 1993 *Phys. Rev. Lett.* **70** 2968-71

[24] Nishimori H 1993 *J. Phys. Soc. Jpn.* **62**, 2793-5

[25] Iba Y 1999 *J. Phys. A* **32** 3875-88

[26] Pearl J 1988 *Probabilistic Reasoning in intelligent Systems* (San Francisco CA: Morgan Kaufmann Publishers, Inc.)

[27] Cheng J F 1997 *Iterative Decoding* (PhD Thesis, California Institute of Technology, Pasadena CA)

[28] Kabashima Y and Saad D 1998 *Europhys. Lett.* **44** 668-74

[29] Kschischang F R and Frey B J 1998 *IEEE J. Selec. Areas in Comm.* **16** 1-11

[30] Davey M C and MacKay D J C 1998 *IEEE Comm. Lett.* **2** 165

[31] Richardson T and Urbanke R 1998 preprint

[32] MacKay D J C 1995 *Electronics Letters* **31** 446-7