EUROPHYSICS LETTERS 15 November 2001

Europhys. Lett., **56** (4), pp. 610–616 (2001)

Statistical mechanics of low-density parity check error-correcting codes over Galois fields

K. NAKAMURA[1](*), Y. KABASHIMA[1](**) and D. SAAD[2](***)

[1] *Department of Computational Intelligence and Systems Science*
Tokyo Institute of Technology - Yokohama 2268502, Japan
[2] *The Neural Computing Research Group, Aston University - Birmingham B4 7ET, UK*

(received 10 October 2000; accepted in final form 22 August 2001)

PACS. 89.90.+n – Other topics in areas of applied and interdisciplinary physics.
PACS. 89.70.+c – Information science.
PACS. 05.50.+q – Lattice theory and statistics (Ising, Potts, etc.).

Abstract. – A variation of low-density parity check (LDPC) error-correcting codes defined over Galois fields ($GF(q)$) is investigated using statistical physics. A code of this type is characterised by a sparse random parity check matrix composed of C non-zero elements per column. We examine the dependence of the code performance on the value of q, for finite and infinite C values, both in terms of the thermodynamical transition point and the practical decoding phase characterised by the existence of a unique (ferromagnetic) solution. We find different q-dependence in the cases of $C = 2$ and $C \geq 3$; the analytical solutions are in agreement with simulation results, providing a quantitative measure to the improvement in performance obtained using non-binary alphabets.

Error correction mechanisms are essential for ensuring reliable data transmission through noisy media. They play an important role in a wide range of applications from magnetic hard disks to deep space exploration, and are expected to become even more important due to the rapid development in mobile phones and satellite-based communication.

The error-correcting ability comes at the expense of information redundancy. Shannon showed in his seminal work [1] that error-free communication is theoretically possible if the code rate, representing the fraction of informative bits in the transmitted codeword, is below the channel capacity. Here, we focus on the case of unbiased messages transmitted through a Binary Symmetric Channel (BSC), characterized by a bit flip rate p. In this case, the maximal code rate $R = N/M$ which allows for an error-free communication satisfies $R < 1 - H_2(p)$, when both lengths of the original message N and codeword M become infinite and $H_2(p) = -p\log_2 p – (1-p)\log_2(1-p)$. The maximal rate is often termed *Shannon's limit*. Unfortunately, Shannon's derivation is non-constructive and the quest for practical codes which saturate this limit has been one of the central topics in information theory ever since.

Low-density parity check (LDPC) codes are based on the transmission of parity checks on top of the message itself, from which errors, which occur during transmission, could be identified and corrected. These codes were introduced by Gallager already in 1962 [2] but have been only recently rediscovered [3] and suggested as practical high-performance codes. They appear to offer the best performance to date and are likely to play an increasingly more

(*) E-mail: knakamur@fe.dis.titech.ac.jp
(**) E-mail: kaba@dis.titech.ac.jp
(***) E-mail: saadd@aston.ac.uk

TABLE I – *(a)* $GF(4 = 2^b)$ *numbers can be expressed as* $b = 2$ *bits Boolean sequences or* $b - 1$ *degree polynomials composed of Boolean coefficients. (b) Sum (left) and product (right) in* $GF(4)$.

(a)

$GF(4)$	0	1	2	3
Boolean	00	01	10	11
Polynomial	0	1	x	$x+1$

(b)

$\bigoplus$	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

$\bigotimes$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	3	1	2

important role in the future. Fortunately, these codes are amenable to statistical-mechanics–based analysis, by mapping the code onto a diluted Ising spin system, which provides insight into their typical properties and performance [4, 5].

One particularly powerful code is an irregular construction using a non-binary alphabet based on Galois fields, which provides one of the best error correction performance to date [6]. This construction, based on improving the parity check matrix and the alphabet used by trial and error, instigated the current work, aimed at clarifying the role played by the alphabet used in obtaining this outstanding performance. To separate the effect of code irregularity from that of the alphabet used we focus here on the dependence of *regular* constructions on the chosen alphabet. To some extent this complements our previous investigation on the impact of code irregularity on the system's performance [5] in the case of binary alphabets. As there is no directly equivalent physical model to the non-binary coding system under examination, one should design a model specifically for this task, which makes the analysis more interesting.

A Galois field $GF(q = 2^b)$ represents a closed set of q elements $\{0, 1, \ldots, q-1\}$ which can be added and multiplied in modulo. Employing the binary representation, each element is expressed as a sequence of b-bit Boolean number; this can be identified with a $b-1$ degree polynomial, the coefficients of which are defined by the Boolean elements of this number (table I(a)). Over this set, the sum and product operations are defined by a b degree irreducible polynomial which may be represented by Boolean coefficients. For instance, the irreducible polynomial for $GF(2^2 = 4)$ is $x^2 + x + 1$. Then, employing the polynomial expression, $3 \oplus 1 = (x+1) + 1 \pmod 2 = x = 2$ and $3 \otimes 2 = (x+1) \times x \pmod 2 = x^2 + x \pmod 2 = -1 \pmod 2 = 1$, setting $x^2 + x + 1 = 0 \pmod 2$, where modulo 2 operations are employed for the Boolean coefficients. Table I(b) summarises the sum and product operations in the Galois field $GF(4)$. More details about Galois fields can be found in [7].

In a general scenario, a non-binary alphabet based on the Galois field $GF(q)$ is used to define an encoder and decoder as follows: The sender first converts the Boolean message vector $\boldsymbol{\xi}^{\mathrm{B}}$ of dimensionality N, where $\xi_i^{\mathrm{B}} \in (0, 1)$, $\forall i$, to an N/b-dimensional vector of $GF(q = 2^b)$ elements, where each segment of b consecutive bits is mapped onto a $GF(q)$ number([1]). The $GF(q)$ vector is then encoded to an M/b dimensional $GF(q)$ codeword $\boldsymbol{z}_0$, in the manner described below, which is then reconverted to an M-dimensional Boolean codeword $\boldsymbol{z}_0^{\mathrm{B}}$, transmitted via a noisy channel. Corruption during transmission can be modelled by the noise vector $\boldsymbol{\zeta}^{\mathrm{B}}$, where corrupted bits are marked by the value 1 and all other bits are zero, so that the received corrupted codeword takes the form $\boldsymbol{z}^{\mathrm{B}} = \boldsymbol{z}_0^{\mathrm{B}} + \boldsymbol{\zeta}^{\mathrm{B}} \pmod 2$. The received corrupted Boolean message is then converted back to a $GF(q)$ vector $\boldsymbol{z}$, and decoded in the $GF(q)$ representation; finally the message estimate is interpreted as a Boolean vector.

An LDPC code in the $GF(q)$ representation is based on a randomly constructed sparse parity check matrix $\boldsymbol{A}$ of dimensionality $(M - N)/b \times M/b$. In regular codes, which we

([1])Binary vectors will be denoted by a superscript B; other vectors are in the $GF(q)$ representation.

focus on here, this matrix is characterised by fixed numbers C and K of non-zero $GF(q)$ elements per column/row. In irregular constructions the number of non-zero elements per column/row may vary. The choice of C, K is linked to the code rate R, obeying the relation $C/K = 1 - R$. Non-zero elements in each row are independently and randomly selected from a specific distribution that maximises the marginal entropy for each component of $\boldsymbol{A\zeta}$ (all vector operations in the $GF(q)$ representation will be carried out as defined for this field; for brevity we do not introduce different symbols to denote these operations) when $\boldsymbol{\zeta}$ is the $GF(q)$ representation of the binary random noise vector $\boldsymbol{\zeta}^{\mathrm{B}}$. Then, one constructs a dense $M/b \times N/b$ generator matrix $\boldsymbol{G}^T$ satisfying $\boldsymbol{AG}^T = 0$ [6].

Using the matrix $\boldsymbol{G}$, encoding is carried out in the $GF(q)$ representation by taking the product $\boldsymbol{z}_0 = \boldsymbol{G}^T\boldsymbol{\xi}$; decoding is performed by taking the product of the parity check matrix $\boldsymbol{A}$ and the received corrupted message $\boldsymbol{z} = \boldsymbol{z}_0 + \boldsymbol{\zeta}$, which yields the *syndrome* vector $\boldsymbol{J} = \boldsymbol{Az} = \boldsymbol{A\zeta}$. The most probable estimate of the noise vector $\boldsymbol{n}$ is defined using the equation

$$\boldsymbol{An} = \boldsymbol{J} \ . \tag{1}$$

Belief propagation (BP) [3] is widely used to find the most probable vector $\boldsymbol{n}$. This has been linked, in the case of Boolean codes, to the TAP (Thouless, Anderson, Palmer)-based solution of a similar physical system [8], a relation which holds also in the case of $GF(q)$ codes.

The noise vector estimate is then employed to remove the noise from the received codeword and retrieve the original message $\boldsymbol{\xi}$ by solving the equation $\boldsymbol{G}^T\boldsymbol{\xi} = \boldsymbol{z} - \boldsymbol{n}$.

The similarity between error-correcting codes and physical systems was first pointed out by Sourlas [9], by considering a simple Boolean code, and by mapping the code onto well-studied Ising spin systems. We recently extended his work, which focused on extensively connected systems, to the case of finite connectivity [8]. Here, we generalise these connections to spin systems in which the interaction is determined using the $GF(q)$ algebra.

In order to facilitate the current investigation, we first map the problem to that of a "$GF(q)$ spin system" of finite connectivity. The syndrome vector $\boldsymbol{J}$ is generated by taking sums of the relevant noise vector elements $J_\mu = A_{\mu i_1}\zeta_{i_1} + \ldots + A_{\mu i_K}\zeta_{i_K}$, where $\boldsymbol{\zeta} = (\zeta_{i=1,\ldots,M/b})$ represents the true channel noise; the indices $i_1, \ldots, i_K$ correspond to the non-zero elements in the μ-th row of the parity check matrix $\boldsymbol{A} = (A_{\nu k})$. It should be noted that the noise components ζ_i are derived from a certain distribution $P_{pr}(\zeta_i)$, representing the nature of the communication channel; this will serve as our prior belief to the nature of the corruption process. This implies that the most probable solution of eq. (1) corresponds to the ground state of the Hamiltonian

$$\mathcal{H}(\boldsymbol{n}) = \sum_{\langle i_1, i_2, \ldots, i_K \rangle} \mathcal{D}_{\langle i_1, i_2, \ldots, i_K \rangle} \left(1 - \delta\left[J_{\langle i_1, i_2, \ldots, i_K \rangle}; A_{\mu i_1} n_{i_1} + \ldots + A_{\mu i_K} n_{i_K}\right]\right) - \frac{1}{\beta} \sum_{i=1}^{M/b} \ln P_{pr}(n_i), \tag{2}$$

in the zero-temperature limit $\beta = 1/T \to \infty$. Elements of the sparse tensor $\mathcal{D}_{\langle i_1, i_2, \ldots, i_K \rangle}$ take the value 1 if all the corresponding indices of parity matrix $\boldsymbol{A}$ are non-zero in some row, μ, and 0 otherwise. The last expression on the right relates to the prior probability of the noise vector elements. Note that operations between vectors/elements in the $GF(q)$ representation (*e.g.*, within the δ-function) are carried out as defined in this field.

The delta-function provides 1 if the contribution for the selected site $A_{\mu i_1} n_{i_1} + \ldots + A_{\mu i_K} n_{i_K}$ is in agreement with the corresponding syndrome value $J_{\langle i_1, i_2, \ldots, i_K \rangle}$, and 0 otherwise. Notice that this term is not frustrated as there are M/b degrees of freedom while only $(M - N)/b$ constraints arise from eq. (1), and its contribution can therefore vanish at sufficiently low temperatures. The choice of $\beta \to \infty$ imposes the restriction (1), limiting the solutions to

those for which the first term of (2) vanishes, while the second term, representing the prior information about the noise, survives.

The optimal estimator, minimising the expectation of discrepancy per noise bit, is of the form $\widehat{n}_i = \mathrm{argmax}_{a\in GF(q)} \langle \delta(n_i, a)\rangle_{\beta\to\infty}$. This is known as the *marginal posterior maximiser* (MPM) [10] and corresponds to the finite-temperature decoding at Nishimori's temperature studied in other codes [9,11,12]. Notice that here, due to the hard constraints imposed on the dynamical variables, decoding at zero temperature is optimal, as the true posterior distribution (given $\boldsymbol{J}$) relates to the ground state of Hamiltonian (2), similar to other LDPC codes [4]. The macroscopic quantity $m = (b/M)\left\langle \sum_{i=1}^{M/b} \delta(\widehat{n}_i, \zeta_i)\right\rangle_{\{\mathcal{D},A,\boldsymbol{\zeta}\}}$ serves as the performance measure.

To eliminate the dependence of the syndrome $J_{\langle i_1,\ldots,i_K\rangle}$ on the noise vector $\boldsymbol{\zeta}$ we employ the gauge transformation $n_i \to n_i + \zeta_i$, $J_{\langle i_1,\ldots,i_K\rangle} \to 0$. Rewriting eq. (2) in this gauge moves the dependence on $\boldsymbol{\zeta}$ to the second term where it appears in a decoupled form $(1/\beta)\ln P_{pr}(n_i+\zeta_i)$. The remaining difficulty comes from the complicated site dependence caused by non-trivial $GF(q)$ algebra in the first term. However, one can rewrite this dependence in the simpler form

$$\delta\left[0; A_{\mu i_1} n_{i_1} + \ldots + A_{\mu i_K} n_{i_K}\right] = \sum_{A_1,\ldots,A_K,a_1,\ldots,a_K=0}^{q-1} \delta\left[0; A_1 a_1 + \ldots + A_K a_K\right] \times \\ \times\, \delta(A_1, A_{\mu i_1}) \ldots \delta(A_K, A_{\mu i_K}) \times \delta(a_1, n_{i_1}) \ldots \delta(a_K, n_{i_K}), \quad (3)$$

by introducing Kroncker's δ and the dummy variables $A_1, \ldots, A_K$ and $a_1, \ldots, a_K$.

Since codes of this type are usually used for long messages with $N = 10^3$–10^5, it is natural to analyse their properties using the methods of statistical mechanics. The random selection of a sparse tensor $\mathcal{D}$ identifies the non-zero elements of $\boldsymbol{A}$, and the noise vector $\boldsymbol{\zeta}$ introduces quenched disorder to the system. We then calculate the partition function $\mathcal{Z}(\mathcal{D}, A, \boldsymbol{\zeta}) = \mathrm{Tr}_{\boldsymbol{n}} \exp\left[-\beta\mathcal{H}\right]$ averaged over the disorder using the replica method [4,8]. Taking $\beta \to \infty$ gives rise to a set of order parameters

$$\mathcal{Q}_{a_1,a_2,\ldots,a_n} = \frac{b}{M} \sum_{i=1}^{M/b} \left\langle Z_i \prod_{\alpha=1}^{n} \langle \delta\left(a_\alpha, n_{i\alpha}\right)\rangle_{\beta\to\infty} \right\rangle_{\mathcal{D},A,\boldsymbol{\zeta}}, \quad (4)$$

where $\alpha = 1, \ldots, n$ represents the replica index and a_α runs from 0 to $q-1$, and the variables Z_i come from enforcing the restriction of C connections per index i

$$\delta\left(\sum_{\langle i_2,\ldots,i_K\rangle} \mathcal{D}_{\langle i,i_2,\ldots,i_K\rangle} - C\right) = \oint \frac{\mathrm{d}Z}{2\pi} Z^{\sum_{\langle i_2,\ldots,i_K\rangle} \mathcal{D}_{\langle i,i_2,\ldots,i_K\rangle} - (C+1)}. \quad (5)$$

To proceed further, one has to make an assumption about the symmetry of order parameters. The assumption made here is that of replica symmetry reflected in the representation of the order parameters and of the related conjugate variables:

$$\mathcal{Q}_{a_1,a_2,\ldots,a_n} = a_{\mathcal{Q}} \int \mathrm{d}\boldsymbol{P}\; \pi(\boldsymbol{P}) \prod_{\alpha=1}^{n} P_{a_\alpha}, \quad \widehat{\mathcal{Q}}_{a_1,a_2,\ldots,a_n} = a_{\widehat{\mathcal{Q}}} \int \mathrm{d}\widehat{\boldsymbol{P}}\; \widehat{\pi}(\boldsymbol{P}) \prod_{\alpha=1}^{n} \widehat{P}_{a_\alpha}, \quad (6)$$

where $a_{\mathcal{Q}}$ and $a_{\widehat{\mathcal{Q}}}$ are normalisation coefficients; $\pi(\boldsymbol{P})$ and $\widehat{\pi}(\widehat{\boldsymbol{P}})$ represent probability distributions for q-dimensional vectors $\boldsymbol{P} = (P_0, \ldots, P_{q-1})$ and $\widehat{\boldsymbol{P}} = (\widehat{P}_0, \ldots, \widehat{P}_{q-1})$, respectively. Unspecified integrals are performed over the region $P_0 + \ldots + P_{q-1} = 1$, $P_{a=0,\ldots,q-1} \geq 0$ or

$\widehat{P}_0 + \ldots + \widehat{P}_{q-1} = 1$, $\widehat{P}_{a=0,\ldots,q-1} \geq 0$. Extremising the averaged expression with respect to the probability distributions, one obtains the following free energy per spin:

$$-\frac{b}{M}\langle \ln \mathcal{Z} \rangle_{\mathcal{D},A,\boldsymbol{\zeta}} = -\underset{\{\pi,\widehat{\pi}\}}{\text{Ext}} \left\{ \int \prod_{l=1}^{C} \mathrm{d}\widehat{\boldsymbol{P}}^{\,l}\ \widehat{\pi}\left(\widehat{\boldsymbol{P}}^{\,l}\right) \left\langle \ln\left(\sum_{a=0}^{q-1} \prod_{l=1}^{C} \widehat{P}_a^{\,l}\ P_{pr}\left(a+\zeta\right)\right)\right\rangle_{\zeta} + \right.$$
$$+\ \frac{C}{K} \int \prod_{l=1}^{K} \mathrm{d}\boldsymbol{P}^{l} \pi\left(\boldsymbol{P}^{l}\right) \left\langle \ln\left(\sum_{a_1,\ldots,a_K=0}^{q-1} \delta\left[0; A_1 a_1 + \ldots + A_K a_K\right] \prod_{l=1}^{K} P_{a_l}^{l}\right)\right\rangle_{A} -$$
$$\left. -\ C \int \mathrm{d}\boldsymbol{P}\ \mathrm{d}\widehat{\boldsymbol{P}}\ \pi\left(\boldsymbol{P}\right)\ \widehat{\pi}(\widehat{\boldsymbol{P}}) \ln\left(\sum_{a=0}^{q-1} P_a \widehat{P}_a\right)\right\}\ , \tag{7}$$

where $\langle\cdot\rangle_A$ and $\langle\cdot\rangle_\zeta$ denote averages over the distribution of non-zero units per row in constructing the matrix $\boldsymbol{A}$ and over $P_{pr}(\zeta)$, respectively. One calculates the free energy via the saddle point method, as is done for the binary codes [4]. Solving the equations obtained by varying eq. (7) is generally difficult. However, it can be shown analytically that a successful solution

$$\pi(\boldsymbol{P}) = \delta(P_0 - 1) \prod_{a=1}^{q-1} \delta(P_a), \quad \widehat{\pi}(\widehat{\boldsymbol{P}}) = \delta(\widehat{P}_0 - 1) \prod_{a=1}^{q-1} \delta(\widehat{P}_a), \tag{8}$$

which implies perfect decoding $m = 1$, extremises the free energy for $C \geq 2$. For $C \to \infty$, an unsuccessful solution, which provides $m < 1$, is also obtained analytically:

$$\pi(\boldsymbol{P}) = \left\langle \prod_{a=0}^{q-1} \delta\left(P_a - P_{pr}(a+\zeta)\right)\right\rangle_{\zeta}, \quad \widehat{\pi}(\widehat{\boldsymbol{P}}) = \prod_{a=0}^{q-1} \delta\left(\widehat{P}_a - \frac{1}{q}\right). \tag{9}$$

Inserting these solutions into (7) it is found that the solution (8) becomes thermodynamically dominant with respect to (9) for $R < 1 - H_2(p)$ independently of q; which implies that the code saturates Shannon's limit for $C \to \infty$ as reported in the information theory literature [6].

Finding additional solutions analytically is difficult, we therefore resorted to numerical methods. Approximating the distributions $\pi(\boldsymbol{P})$ and $\widehat{\pi}(\widehat{\boldsymbol{P}})$ by 5×10^3–3×10^4 sample vectors of $\boldsymbol{P}$ and $\widehat{\boldsymbol{P}}$ we obtained solutions by updating the saddle point equations (100–500 iterations) for codes of connectivity $C = 2, \ldots, 6$ and $GF(q)$ representation $q = 2, 4, 8$ and for both BSC and Gaussian channels. Less than 50 iterations were typically sufficient for the solutions to converge. Due to lack of space we present here results only for the case of the BSC; results for the case of Gaussian channels are qualitatively similar and will be presented elsewhere.

Since the suggested properties are different for $C \geq 3$ and $C = 2$, we describe the results separately for the two cases. For $C \geq 3$, it turns out that eq. (8) is always locally stable. However, an unsuccessful solution, approaching (9) as $C \to \infty$, becomes thermodynamically dominant for a sufficiently large flip rate p. As the noise level is reduced, the solution (8) becomes thermodynamically dominant at a certain flip rate $p = p_{\mathrm{t}}$, and remains dominant until $p \to 0$. This implies that perfect decoding $m = 1$ is feasible for $p < p_{\mathrm{t}}$. However, the unsuccessful solution remains as well above a certain noise level (the spinodal point) $p_{\mathrm{s}} (\leq p_{\mathrm{t}})$. Note that the entropy of this solution vanishes below p_{t}, exhibiting replica symmetry breaking. Nevertheless, the spinodal point p_{s} obtained by this solution provides an accurate estimate to the practical decoding limit when BP is employed for $C \geq 3$, as is observed in [4].

As $C \to \infty$, the transition point p_{t} converges to Shannon's limit $p_{\mathrm{c}} = H_2^{-1}(1 - R)$ from below, irrespective of the value of q. For finite C, p_{t} monotonically increases with q but does

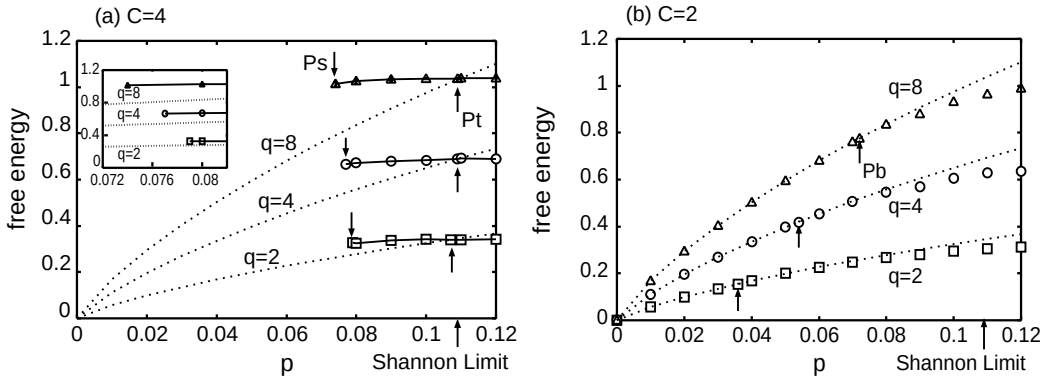

Fig. 1 – Extremised free energies (7) obtained for $q = 2, 4, 8$ as functions of the flip rate p (a) for connectivity $C = 4$ and (b) for $C = 2$ codes with the same code rate $R = 1 - C/K = 1/2$. In both codes, broken lines represent free energies of solution (8) corresponding to decoding success, while markers stand for failure solutions ($m < 1$). Monte Carlo methods with 5×10^3–3×10^4 samplings at each step are employed for obtaining the latter with statistical fluctuations smaller than the symbol size. For each q value, the solution having the lower free energy becomes thermodynamically dominant. In (a), crossing points provide the critical flip rate for $p_{\rm t}$ being $0.106, 0.108$ and 0.109 (within the numerical precision) for $q = 2, 4$ and 8, respectively, monotonically approaching Shannon's limit $p_{\rm c} = H^{-1}(1/2) = 0.109$. The inset focuses on the spinodal points $p_{\rm s}$, which determine the limit of successful practical decoding. This shows $p_{\rm s}$ to decrease with increasing q. (b) shows that $C = 2$ codes exhibit continuous transitions between the solutions of decoding success and failure. The critical flip rate $p_{\rm b}$, pointed by arrows, increases with q, while it is still far from Shannon's limit.

not saturate $p_{\rm c}$. This implies that the error-correcting ability of the codes when optimally decoded is monotonically improving as q increases.

The behaviour of the spinodal point $p_{\rm s}$ is quite different, as shown in fig. 1a, presenting the dependence of $p_{\rm t}$ and $p_{\rm s}$ on q for connectivity $C = 4$. It appears that $p_{\rm s}$ is generally decreasing with respect to q (except for pathological cases), which indicates a lower practical corruption limit for which BP/TAP decoding will still be effective. Above this limit BP/TAP dynamics is likely to converge to the unsuccessful solution due to its dominant basin of attraction [4]. In contrast, $C = 2$ codes exhibit a different behaviour; the solution (8) becomes the unique minimum of free energy (7) for sufficiently small noise levels, which implies that practical decoding dynamics always converges to the perfect solution. However, as the noise level increases, the solution loses its stability and bifurcates to a stable suboptimal solution. Unlike the case of $C \geq 3$, this bifurcation point $p_{\rm b}$, which monotonically increases with q, determines the limit of practical BP/TAP decoding. The practical limit obtained is considerably lower than both Shannon's limit and the thermodynamic transition point $p_{\rm t}$ for other $C \geq 3$ codes with the same q value (fig. 1b). Therefore, the optimal decoding performance of $C = 2$ codes is the worst within this family of codes.

However, $p_{\rm b}$ can become closer to, and even higher than, the spinodal point $p_{\rm s}$ of other $C \geq 3$ codes for large q values (table II), implying that the practical decoding performance of $C = 2$ codes is not necessarily inferior to that of $C \geq 3$ codes. This is presumably due to the decreasing solution numbers to eq. (1) for $C = 2$ as q increases, compared to the moderate logarithmic increase in the information content, tipping the balance in favour of the perfect solution. This may shed light on the role played by $C = 2$ elements in irregular constructions.

In summary, we have investigated the properties of LDPC codes defined over $GF(q)$ within

TABLE II – *The critical noise level, below which BP/TAP-based decoding works successfully, for different connectivity values C in the case of $q = 8$ and $R = 1 - C/K = 0.5$. This is determined as the spinodal point p_s and the bifurcation point p_b for $C \geq 3$ and $C = 2$, respectively. The critical noise for $C = 2$ becomes higher than that of $C \geq 5$.*

C	2	3	4	5	6
Critical noise	0.072	0.088	0.073	0.062	0.050

the framework of statistical mechanics. Employing the replica method, one can evaluate the typical performance of codes in the limit of infinite message length. It has been shown analytically that codes of this type saturate Shannon's limit as $C \to \infty$ irrespective of the value of q, in agreement with results reported in the information theory literature [6]. For finite C, numerical calculations suggest that these codes exhibit two different behaviours for $C \geq 3$ and $C = 2$. For $C \geq 3$, we show that the error-correcting ability of these codes, when optimally decoded, is monotonically improving as q increases, while the practical decoding limit, determined by the emergence of a suboptimal solution, deteriorates. On the other hand, $C = 2$ codes exhibit a continuous transition from optimal to suboptimal solutions at a certain noise level, below which practical decoding dynamics based on BP/TAP methods converges to the (unique) optimal solution. This critical noise level monotonically increases with q and becomes even higher than that of some codes of connectivity $C \geq 3$, while the optimal decoding performance is inferior to that of $C \geq 3$ codes with the same q value. This may elucidate the role played by $C = 2$ components in irregular constructions.

Future directions include extending the analysis to irregular Gallager codes as well as to regular and irregular MN code [3, 4] in the Galois representation.

* * *

We acknowledge support from the Grants-in-Aid (Nos. 13680400), the Japan-Anglo Collaboration Programme of the JSPS (YK), EPSRC (GR/N00562) and The Royal Society (DS).

REFERENCES

[1] SHANNON C. E., *Bell Sys. Tech. J.*, **27** (1948) 379; 623.
[2] GALLAGER R. G., *IRE Trans. Info. Theory*, **IT-8** (1962) 21.
[3] MACKAY D. J. C., *IEEE Trans. Info. Theory*, **45** (1999) 399; MACKAY D. J. C. and NEAL R. M., *Electronic Lett.*, **33** (1997) 457.
[4] KABASHIMA Y., MURAYAMA T. and SAAD D., *Phys. Rev. Lett.*, **84** (2000) 1355; MURAYAMA T., KABASHIMA Y., SAAD D. and VICENTE R., *Phys. Rev. E*, **62** (2000) 1577.
[5] VICENTE R., SAAD D. and KABASHIMA Y., *J. Phys. A*, **33** (2000) 1527; *Europhys. Lett.*, **51** (2000) 698.
[6] DAVEY M. C. and MACKAY D. J. C., *IEEE Comm. Lett.*, **2** (1998) 165.
[7] LIDL R. and NIEDERREITER H., *Introduction to Finite Fields and Their Applications* (Cambridge University Press, Cambridge) 1994.
[8] KABASHIMA Y. and SAAD D., *Europhys. Lett.*, **44** (1998) 668; **45** (1999) 97.
[9] SOURLAS N., *Nature*, **339** (1989) 693; *Europhys. Lett.*, **25** (1994) 159.
[10] IBA Y., *J. Phys. A*, **32** (1999) 3875.
[11] RUJÁN P., *Phys. Rev. Lett.*, **70** (1993) 2968.
[12] NISHIMORI H., *J. Phys. Soc. Jpn.*, **62** (1993) 2973.