Average error exponent in Gallager low-density parity-check codes

N.S. Skantzos[†], J. van Mourik[†] Y. Kabashima[‡] and D. Saad[†]
[†]Neural Computing Research Group, Aston University, Birmingham B4 7ET, UK
[‡]Dept. of Computational Intelligence & Systems Science,
Tokyo Institute of Technology, Yokohama 2268502, Japan

November 5, 2002

Abstract

We present a theoretical method for a *direct* evaluation of the average error exponent in Gallager error-correcting codes using methods of statistical physics. Results for the binary symmetric channel (BSC) are presented for codes of both finite and infinite connectivity.

1 Introduction

Low-density parity-check codes (LDPC) have attracted significant interest in recent years due to their simplicity and exceptionally high performance [1]. Their simplicity and inherent randomness make them amenable to analysis using established methods in the area of statistical physics. These have been employed in a number of papers [2]-[8] to gain insight into the properties of LDPC codes and to evaluate their performance.

These studies include the evaluation of critical noise levels for given codes [2], an exact calculation of weight and magnetization enumerators [4], the performance of irregular codes [3], properties of codes in real-valued channels [5], and the derivation of bounds for the reliability exponent [6], to name but a few. These studies also represent the interdisciplinary nature of this research area and illustrate the successful interaction between researchers in the two disciplines.

The evaluation of error exponents has been a long-standing problem in information theory [10, 11]. Efforts to obtain exact expressions and/or bounds to the error exponent resulted in partial success; although tight bounds have been derived in the case of random codes and LDPC with infinite connectivity [10], only limited results have been obtained for sparsly connected codes. Main stream techniques to tackle the problem include sphere-packing and union-bound arguments [11, 10]. Below a certain code-rate value, the estimated bounds also become loose and require using the 'expurgated exponent' techniques [10] for obtaining a tighter bound.

In this paper, we employ methods of statistical physics to evaluate directly the typical (average) error exponent in Gallager LDPC codes. This can be carried out by averaging the error exponent over the ensemble of randomly generated LDPC codes of given rate and connectivity; this results in the emergence of macroscopic properties, representative of the ensemble properties, that can be obtained numerically and used to calculate the average error exponent. Solutions have been obtained for both finite and infinite connectivity vector ensembles.

As a reference point to test our theory, we use known results obtained in simple solvable limits (e.g. codes of infinite connectivity), and find that our method reproduces them exactly. Perhaps not surprisingly, we also find that at fixed noise level and code rate, the reliability exponent for codes of finite connectivity is always upper-bounded by that of the infinite-connectivity case.

Before we proceed, the distinction between the statistical physics based bounds [6] and the current calculation should be clarified. In the former, one employs methods of statistical physics to calculate the typical value of a *bound* based on inequalities introduced by Gallager; while in the current calculation, a direct estimation of the average error exponent, rather than a bound, is sought. An additional advantage of the current approach is that it can be extended to provide *reliability exponent* values for LDPC codes by restricted averages over codes of high performance.

The paper is organized as follows: In section 2, we introduce the general coding framework and the technique used. In sections 3 and 4 we present an outline of the derivation and the solutions obtained in both finite and infinite connectivity cases respectively. Discussion and conclusions are presented in section 5.

2 Definitions

A regular (k,j) Gallager error-correcting code is defined by the binary $(N-K)\times N$ (parity check) matrix $A=[C_1|C_2]$, which is known to both sender and receiver. The $(N-K)\times(N-K)$ matrix C_2 is taken to be invertible. The number of non-zero elements in each row of A is given by k, while the number of non-zero elements per column is given by $j\equiv k(N-K)/N$.

Gallager's encoding scheme consists of generating a codeword $\boldsymbol{t}\in\{0,1\}^N$ from an information (message) vector $\boldsymbol{s}\in\{0,1\}^K$ (with $N>K$) via the linear operation $\boldsymbol{t}=G^T\boldsymbol{s}$ (mod 2) where G is the generator matrix defined by $G=[I|C_2^{-1}C_1]$ (mod 2). The code rate is then given by $R\equiv K/N=1-j/k$, and measures the information redundancy of the transmitted vector.

Upon transmission of the codeword $\boldsymbol{t}$ via a noisy channel (taken here be a BSC) the vector $\boldsymbol{r}=\boldsymbol{t}+\boldsymbol{n}^0$ (mod 2) is received, where $\boldsymbol{n}^0\in\{0,1\}^N$ is the true channel noise. The statistics of the BSC is fully determined by the flip rate $p\in[0,1]$:

$$P(n_i^0)=(1-p)\,\delta_{n_i^0,0}+p\,\delta_{n_i^0,1} \tag{1}$$

Decoding is carried out by multiplying $\boldsymbol{r}$ by A to produce the syndrome vector $\boldsymbol{z}=A\boldsymbol{r}=A\boldsymbol{n}^0$, since $AG^T=0$ by construction. In order to reconstruct the original message $\boldsymbol{s}$, one has to obtain an estimate $\boldsymbol{n}$ for the true noise $\boldsymbol{n}^0$. First we select the parity check set of A and $\boldsymbol{n}^0$, i.e. all $\boldsymbol{n}$ that satisfy the parity check equations: $\mathcal{I}_{pc}(A,\boldsymbol{n}^0)\equiv\{\boldsymbol{n}\mid A\boldsymbol{n}=A\boldsymbol{n}^0\}$. Since all operations are performed in modulo 2 arithmetic, $\mathcal{I}_{pc}(A,\boldsymbol{n}^0)$ typically contains $\exp[NR\ln(2)]$ candidates for the true noise vector $\boldsymbol{n}^0$.

It was shown (see e.g. [2, 6, 8] for technical details) that this problem can be cast into a statistical mechanics formulation, by replacing the field $(\{0,1\},+\mathrm{mod}(2))$ by $(\{1,-1\},\times)$, and by adapting the parity checks correspondingly. ¿From the parity check matrix A we construct the binary tensor $\mathcal{A}=\{\mathcal{A}_{\langle i_1\cdots i_k\rangle},1\leq i_1<i_2\cdots<i_k\leq N\}$, where $\mathcal{A}_{\langle i_1\cdots i_k\rangle}=1$ if A has a row in which the elements $\{i_c,c=1\cdots k\}$ are all 1 (i.e. when the bits $\langle i_1\cdots i_k\rangle$ are involved in the same parity check), and 0 otherwise. The fact that each bit $i_1=1\cdots N$ is involved in exactly j parity checks is then expressed by $\sum_{i_2<\cdots<i_k}\mathcal{A}_{\langle i_1\cdots i_k\rangle}=j,\ \ \forall\ i_1=1,\ldots,N$ and the parity check equations become $\prod_{c=1}^k n^{i_c}=\prod_{c=1}^k n_{i_c}^0,\ \ \forall\mathcal{A}_{\langle i_1\cdots i_k\rangle}=1$.

Decoding now consists in selecting an $\boldsymbol{n}$ from $\mathcal{I}_{pc}(\mathcal{A},\boldsymbol{n}^0)$, on the basis of its noise statistics, which are fully described by its *magnetization* $m(\boldsymbol{n})=1/N\sum_i n_i$ (corresponding to the weight in the information theory literature). Note that the number $n_-(\boldsymbol{n})$ of flipped bits in a candidate noise vector $\boldsymbol{n}$ is given by $n_-(\boldsymbol{n})=N(1-m(\boldsymbol{n}))/2$. Therefore, we introduce a Hamiltonian or cost function for each noise candidate that is negatively proportional to its magnetisation:

$$H(\boldsymbol{n})=-F\sum_i n_i=-FNm(\boldsymbol{n}) \tag{2}$$

where we take $F=\frac{1}{2}\log\frac{1-p}{p}$, such that up to normalisation $\exp(-H(\boldsymbol{n}))$ yields the correct prior for candidate noise vectors generated by the BSC [12]. Then, a vector $\boldsymbol{n}$ from $\mathcal{I}_{pc}(\mathcal{A},\boldsymbol{n}^0)$ with the highest magnetization (lowest weight) is selected as a solution; this corresponds to MPM decoding.

We are now interested in the probability that other candidate noise vectors are selected from the parity check set $\mathcal{I}_{pc}(\mathcal{A},\boldsymbol{n}^0)$, other than the correct (i.e. true) noise vector $\boldsymbol{n}^0$, for any given combination $\{\boldsymbol{n}^0,\mathcal{A}\}$; this is termed the *block error probability.* In order to calculate this probability, we introduce an indicator function:

$$\Delta(\boldsymbol{n}^0,\mathcal{A})=\lim_{\beta_{1,2}\to\infty}\ \lim_{\lambda_{1,2}\to\pm\lambda}\left[Z_1^{\lambda_1}(\boldsymbol{n}^0,\mathcal{A};\beta_1)\ Z_2^{\lambda_2}(\boldsymbol{n}^0,\mathcal{A};\beta_2)\right]\Big|_{\beta_1=\beta_2=\beta} \tag{3}$$

where

$$Z_1(\boldsymbol{n}^0,\mathcal{A};\beta_1)=\sum_{\boldsymbol{n}\in\mathcal{I}_{pc}(\boldsymbol{n}^0,\mathcal{A})\backslash\boldsymbol{n}^0}e^{-\beta_1 H(\boldsymbol{n})}\qquad Z_2(\boldsymbol{n}^0,\mathcal{A};\beta_2)=\sum_{\boldsymbol{n}\in\mathcal{I}_{pc}(\boldsymbol{n}^0,\mathcal{A})}e^{-\beta_2 H(\boldsymbol{n})} \tag{4}$$

The two partition functions $Z_1(\boldsymbol{n}^0,\boldsymbol{A};\beta_1)$ and $Z_2(\boldsymbol{n}^0,\boldsymbol{A};\beta_2)$ differ only in the exclusion of $\boldsymbol{n}^0$ from Z_1. If the true noise $\boldsymbol{n}^0$ has the highest magnetization of all candidates in the parity check set (decoding success), the Boltzmann factor $\exp[-\beta H(\boldsymbol{n}^0)]$ will dominate the sum over states in Z_2 in the limit $\beta \to \infty$, and $\Delta(\boldsymbol{n}^0, \mathcal{A}) = 0$. Alternatively, if some other vector $\boldsymbol{n} \neq \boldsymbol{n}^0$ has the highest magnetization of all candidates in the parity check set (decoding failure), its Boltzmann factor will dominate both Z_1 and Z_2 and $\Delta(\boldsymbol{n}^0, \mathcal{A}) = 1$. Note that the separate temperatures β_1 and β_2, which are put to be equal to β in the end, and the powers $\lambda_{1,2}$ which are taken to be $\pm\lambda$ in the end, have been introduced in order to allow us to determine whether obtained solutions are physical or not.

To derive the *average error exponent*, we take the logarithm of the indicator function average with respect to all possible realisations of true noise vectors $\boldsymbol{n}^0$, and the ensemble of regular (k, j) codes $\mathcal{A}$:

$$Q = \lim_{N\to\infty} \frac{1}{N} \log \left\langle \langle \Delta(\boldsymbol{n}^0, \mathcal{A}) \rangle_{\boldsymbol{n}^0} \right\rangle_{\mathcal{A}} \tag{5}$$

where

$$\langle f(\boldsymbol{n}^0) \rangle_{\boldsymbol{n}^0} = \frac{1}{(2\cosh F)^N} \sum_{\boldsymbol{n}^0} \exp(F \sum_i n_i^0) \; f(\boldsymbol{n}^0) \tag{6}$$

and

$$\langle f(\mathcal{A}) \rangle_{\mathcal{A}} = \frac{\sum_{\mathcal{A}} \prod_{i_1=1}^{N} \delta[\sum_{i_2<\cdots<i_k} \mathcal{A}_{\langle i_1 \cdots i_k \rangle} - j] \; f(\mathcal{A})}{\sum_{\mathcal{A}} \prod_{i_1=1}^{N} \delta[\sum_{i_2<\cdots<i_k} \mathcal{A}_{\langle i_1 \cdots i_k \rangle} - j]}. \tag{7}$$

Since there are only discrete degrees of freedom, physically meaningful solutions must have a non-negative entropy, requiring the disorder-averaged entropies of the two partition functions (4) to be non-negative. For general values of $\beta_{1,2}$ and $\lambda_{1,2}$, it can be shown that these disorder-averaged entropies are given by

$$\langle S_x \rangle = \frac{\partial Q}{\partial \lambda_x} - \frac{\beta_x}{\lambda_x} \frac{\partial Q}{\partial \beta_x} \geq 0, \qquad x = 1, 2 \tag{8}$$

which have to be positive.

3 General solution

Using standard statistical physics methods such as in [12], we perform the gauge transformation $n_i \to n_i n_i^0$, and the averages over true noise (6) and code constructions (7); we then assume the simplest replica symmetric scheme [9] to arrive at the following expression for the average error exponent:

$$Q(\beta_1, \beta_2, \lambda_1, \lambda_2) = \mathrm{Extr}_{\pi,\hat{\pi}} \left[\frac{j}{k} \log \; I_1[\pi] - j \log \; I_2[\pi, \hat{\pi}] + \log \; I_3[\hat{\pi}] \right] \tag{9}$$

where

$$I_1 = \int \left\{ \prod_{c=1}^{k} d\pi(x_c, y_c) \right\} \left(\frac{1 + \prod_{c=1}^{k} x_c}{2} \right)^{\lambda_+} \left(\frac{1 + \prod_{c=1}^{k} y_c}{2} \right)^{\lambda_-} \tag{10}$$

$$I_2 = \int \{ d\pi(x, y) \; d\hat{\pi}(\hat{x}, \hat{y}) \} \left(\frac{1 + x\hat{x}}{2} \right)^{\lambda_+} \left(\frac{1 + y\hat{y}}{2} \right)^{\lambda_-} \tag{11}$$

$$I_3 = \int \left\{ \prod_{c=1}^{j} \; d\hat{\pi}(\hat{x}_c, \hat{y}_c) \right\} \left\langle \left[\sum_{u=\pm 1} e^{\beta_1 F n^0 u} \prod_{c=1}^{j} \left(\frac{1 + u\hat{x}_c}{2} \right) \right]^{\lambda_+} \left[\sum_{v=\pm 1} e^{\beta_2 F n^0 v} \prod_{c=1}^{j} \left(\frac{1 + v\hat{y}_c}{2} \right) \right]^{\lambda_-} \right\rangle_{n^0} \tag{12}$$

where we have used the short-hand notation $df(x,y)$=$dxdy \;\; f(x,y)$. Functional extremisation of (9) with respect to the densities $\pi(x, y)$ and $\hat{\pi}(\hat{x}, \hat{y})$ results in a closed set of equations (reminiscent of 'density evolution' equations [1]):

$$\hat{\pi}(\hat{x}, \hat{y}) = \int \left[\prod_{c=1}^{k-1} d\pi(x_c, y_c) \right] \delta \left[\hat{x} - \prod_{c=1}^{k-1} x_c \right] \; \delta \left[\hat{y} - \prod_{c=1}^{k-1} y_c \right] \tag{13}$$

$$\pi(x, y) = \frac{\int \left\{ \prod_{c=1}^{j-1} d\hat{\pi}(\hat{x}_c, \hat{y}_c) \right\} \left\langle D_+^{\lambda_+}(\hat{\boldsymbol{x}}; \beta_1) \; D_+^{\lambda_-}(\hat{\boldsymbol{y}}; \beta_2) \; \delta \left[x - \frac{D_-(\hat{\boldsymbol{x}};\beta_1)}{D_+(\hat{\boldsymbol{x}};\beta_1)} \right] \delta \left[y - \frac{D_-(\hat{\boldsymbol{y}};\beta_2)}{D_+(\hat{\boldsymbol{y}};\beta_2)} \right] \right\rangle_{n^0}}{\int \left\{ \prod_{c=1}^{j-1} d\hat{\pi}(\hat{x}_c, \hat{y}_c) \right\} \left\langle D_+^{\lambda_+}(\hat{\boldsymbol{x}}; \beta_1) D_+^{\lambda_-}(\hat{\boldsymbol{y}}; \beta_2) \right\rangle_{n^0}} \tag{14}$$

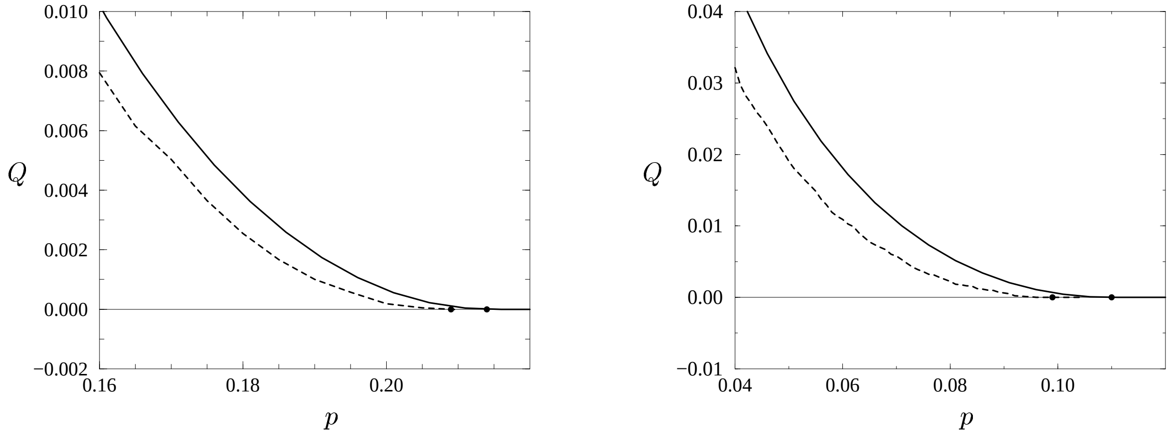

Figure 1: Average error exponent Q as function of the flip rate p for codes of $(k, j) = (4, 3)$ (left picture) and $(k, j) = (6, 3)$ (right picture). Dashed lines correspond to the finite (k, j) cases. Dots indicate critical flip rates where Q becomes zero. For comparison we also present (solid lines) the value of the average error exponent in the case of $k, j \to \infty$ with $R = 1/4$ (left) and $R = 1/2$ (right). Note that the transition from type I to type II solution occurs at small p values outside the range of this figure.

where

$$D_{\pm}(\boldsymbol{z};\beta) = [e^{\beta F n^0} \prod_{c=1}^{j-1} (1 + z_c)] \pm [e^{-\beta F n^0} \prod_{c=1}^{j-1} (1 - z_c)] \tag{15}$$

For given $(\beta_1, \beta_2, \lambda_1, \lambda_2)$ in general, solutions to (13) and (14) can only be obtained numerically. Inserting these solutions into (9) we then obtain $Q(\beta_1, \beta_2, \lambda_1, \lambda_2)$, which becomes the *average error exponent* for $\lambda_1 = -\lambda_2 = \lambda > 0$, and for $\beta_1 = \beta_2 = \beta \to \infty$.

We must recall, however, that physically meaningful solutions must satisfy the conditions (8) stating that the entropies related to the full and the restricted partition sums are non-negative.

We restrict ourselves to regions below the thermodynamic transition where the average case is dominated by the ferromagnetic solution, such that we can safely fix the denominator to the ferromagnetic solution. This dominance is guaranteed if the following constraint is satisfied

$$\left.\frac{\partial Q}{\partial \beta}\right|_{\lambda_1 = -\lambda_2 = \lambda} \leq 0 \ . \tag{16}$$

It turns out that for given $\lambda > 0$, the largest value of β for which (16) is satisfied is given by $\beta = 1/(1+\lambda)$. Hence, in order to maximize β, we must look for the smallest value λ_* that satisfies the conditions on the non-negativity of the entropies (8). Unfortunately, in general this value λ_* can only be obtained numerically. The value obtained for the average error exponent by this analysis is then given by $Q(1/(1+\lambda_*), 1/(1+\lambda_*), \lambda_*, -\lambda_*)$ from (9).

In figure 1 we present the obtained average error exponent as a function of the flip rate for $(k, j) = (4, 3)$ and $(k, j) = (6, 3)$ codes. We observe that the error exponent indeed converges to zero, as it should, when the flip rate approaches its critical value.

Notice the similarity between the equations obtained here and in [6] in spite of the different starting points. It has been shown in [6] that the analysis should be refined in low rate regions by applying a more complex symmetry assumption in the derivation termed one step replica symmetry breaking (for more details see [12]). The refined analysis resulted in tight bounds of the error exponent even in the region of low code-rates, similar to those obtained using expurgated exponent methods. One can exploit the similarity between the equations obtained in [6] and in the current manuscript to derive similar results in the low-rate region.

4 An exactly solvable limit: $k, j \to \infty$

Whereas for finite density codes we were depending on a numerical analysis, in the limit of $k, j \to \infty$ (while keeping the rate $R = 1 - j/k$ finite) we obtain two types of analytic solutions to equations (13) and (14), which can be verified by substitution:
Type I:

$$\begin{aligned} \pi(x,y) &= \frac{1}{2}\left[\delta(x-1)+\delta(x+1)\right]\delta(y-1) \\ \hat{\pi}(\hat{x},\hat{y}) &= \frac{1}{2}\left[\delta(\hat{x}-1)+\delta(\hat{x}+1)\right]\delta(\hat{y}-1) \end{aligned} \tag{17}$$

Type II:

$$\begin{aligned} \pi(x,y) &= \Big[\, G_{+}(F(1{+}\beta_2\lambda_{+}))\ \delta(x{-}\tanh(\beta_1 F)) + G_{-}(F(1{+}\beta_2\lambda_{-}))\ \delta(x{+}\tanh(\beta_1 F))\Big]\ \delta(\hat{y}-1) \\ \hat{\pi}(\hat{x},\hat{y}) &= \delta(\hat{x})\ \delta(\hat{y}-1) \end{aligned} \tag{18}$$

with $G_{\pm}(x) = \frac{1}{2}[1 \pm \tanh(x)]$.

The average error exponent as obtained from the type I solution is given by

$$Q_I(\beta,\beta,\lambda,-\lambda) = -\frac{j}{k}\log 2 - \log\cosh F + \log\cosh(\beta F\lambda) + \log 2\cosh(F - \beta F\lambda)\ . \tag{19}$$

We find that the entropies (8) are always identically zero, and that the constraint (16) requires that $\beta = 1/2$, such that $\lambda = 1$ and

$$Q_I = -\frac{j}{k}\log 2 - \log\cosh F + \log[\cosh F + 1] \tag{20}$$

which is exactly the Bhattacharyya limit [11].

The average error exponent as obtained from the type II solution is given by

$$Q_{II}(\beta,\beta,\lambda,-\lambda) = \lambda\left[-\frac{j}{k}\log 2 + \log 2\cosh[\beta F]\right] + \log[2\cosh(F-\beta F\lambda)] - \log 2\cosh F \tag{21}$$

The condition on the entropy $\langle S_2\rangle \geq 0$ is satisfied for all $\beta > 0$ whereas the condition $\langle S_1\rangle \geq 0$ is violated below the critical (freezing) temperature $1/\beta^*$ obtained from

$$-\frac{j}{k}\log 2 - \beta^* F\tanh[\beta^* F] + \log 2\cosh[\beta^* F] = 0 \tag{22}$$

This negative entropy is an artifact of a simplistic assumption about the symmetry between replicas, and can be remedied by considering a 'frozen RSB' ansatz [4]. In this ansatz, for all $\beta \geq \beta^*$, the (frozen) average error exponent is given by

$$Q_{II}^{fr}(\beta,\beta,\lambda,-\lambda) = \frac{\beta}{\beta^*}\lambda\left[-\frac{j}{k}\log 2 + \log 2\cosh[\beta^* F]\right] + \log 2\cosh[F - \beta F\lambda] - \log 2\cosh F \tag{23}$$

However, condition (16) is violated for $\beta > 1 - \beta^*$, such that the average error exponent is given by

$$Q_{II}^{fr} = F\tanh[\beta^* F] + \frac{j}{k}\log 2 - \log 2\cosh F \tag{24}$$

What remains is to determine whether the type I or the type II solution is physically dominant, by using Q as a generating function. Results for the case of $k, j \to \infty$ are presented in figure 2 for $p = 0.01$ and $p = 0.05$.

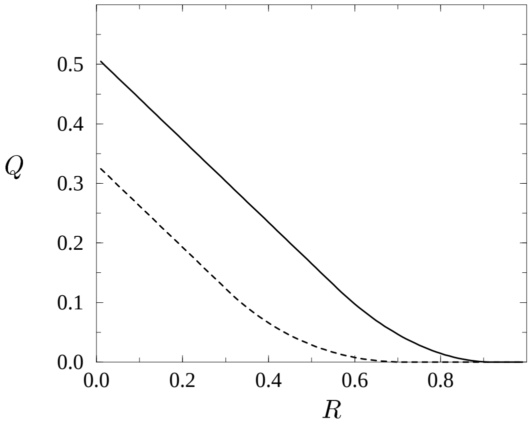

Figure 2: Reliability exponent Q as function of the code rate R for regular $k, j \to \infty$ Gallager codes for which analytical expressions can be derived; see (20) and (24) (solid: $p = 0.01$ and dashed: $p = 0.05$).

5 Discussion

In this paper we suggest a method for direct evaluation of the average error exponent over the ensemble of LDPC error-correcting codes of given rate and connectivity. An analytical solution has been obtained using methods of statistical physics, which is in perfect agreement with known results in the limit $k, j \to \infty$ (with R finite). Solutions obtained for codes of finite (k, j) values seem to be upper bounded by the $k, j \to \infty$ results.

An interesting feature of the present study is the similarity of our equations to those obtained in [6] in spite of the different approaches used. An important advantage offered by the current approach is a potential extension to restrict the averages to codes of high performance to obtain *reliability exponent* values for LDPC codes of both finite and infinite connectivity; this study is currently underway.

References

[1] T Richardson, A Shokrollahi and R Urbanke (2001), *IEEE Trans. on Info. Theory* **47** 619-637

[2] R Vicente, D Saad and Y Kabashima (1999) *Phys Rev E* **60** 5352-5366

[3] R Vicente, D Saad and Y Kabashima (2000) *J Phys A* **33** 6527-6542

[4] J van Mourik, D Saad and Y Kabashima (2002) *Phys Rev E* **66** 026705

[5] T Tanaka and D Saad (2002) submitted in *Phys Rev E*

[6] Y Kabashima, N Sazuka, K Nakamura and D Saad (2001) *Phys Rev E* (2001) **64** 046113

[7] A Montanari and N Sourlas *Eur Phys J B* **18** 107-119

[8] A Montanari *Eur Phys J B* **23** 121-136

[9] K Y M Wong and D Sherrington (1987) *J Phys A* **20** L793-L799

[10] R G Gallager (2001) 'Information theory and reliable communication' Wiley & Sons, NY

[11] A J Viterbi and J K Omura (1979) 'Principles of Digital Communication and Coding', McGraw-Hill Int Ed (Singapore)

[12] H Nishimori (2001) 'Statistical Physics of Spin Glasses and Information Processing', Oxford University Press, UK