EUROPHYSICS LETTERS 1 December 1998

Europhys. Lett., **44** (5), pp. 668-674 (1998)

Belief propagation *vs.* TAP for decoding corrupted messages

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 5 February 1998; accepted in final form 7 October 1998)

PACS. 89.70+c – Information science.
PACS. 89.90+n – Other areas of general interest to physicists.
PACS. 02.50−r – Probability theory, stochastic processes, and statistics.

Abstract. – We employ two different methods, based on belief propagation and TAP, for decoding corrupted messages encoded by employing Sourlas's method, where the code word comprises products of K bits selected randomly from the original message. We show that the equations obtained by the two approaches are similar and provide the same solution as the one obtained by the replica approach in some cases ($K = 2$). However, we also show that for $K \geq 3$ and unbiased messages the iterative solution is sensitive to the initial conditions and is likely to provide erroneous solutions; and that it is generally beneficial to use Nishimori's temperature, especially in the case of biased messages.

Belief networks [1], also termed Bayesian networks, and influence diagrams are diagrammatic representations of joint probability distributions over a set of variables. The set of variables is usually represented by the vertices of a graph, while arcs between vertices represent probabilistic dependences between variables. Belief propagation provides a convenient mathematical tool for calculating iteratively joint probability distributions of variables, and have been used in a variety of cases to assess conditional probabilities and interdependences between variables in complex systems. One of the most recent uses of belief propagation is in the field of error-correcting codes, especially for decoding corrupted messages [2] (for a review of graphical models and their use in the context of error-correcting codes see [3]).

Error-correcting codes provide a mechanism for retrieving the original message after corruption due to noise during transmission. A new family of error-correcting codes, based on insights gained from statistical mechanics, has recently been suggested by Sourlas [4]. These codes can be mapped onto the many-body Ising spin problem and can thus be analysed using methods adopted from statistical physics [5-9].

In this letter we will examine the similarities and differences between the belief propagation (BP) and TAP approaches, used as decoders in the context of error-correcting codes. We will then employ these approaches to examine a few specific cases and compare the results to the solutions obtained using the replica method [8]. This will enable us to draw some conclusions on the efficacy of the TAP/BP approach in the context of error-correcting codes.

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

In a general scenario, a message represented by an N-dimensional binary vector $\boldsymbol{\xi}$ is encoded by a vector $\boldsymbol{J}^0$ which is then transmitted through a noisy channel with some flipping probability p per bit. The received message $\boldsymbol{J}$ is then decoded to retrieve the original message. The family of codes, suggested by Sourlas [4], is based on an encoded message of the form $J^0_{i_1,i_2\ldots i_K} = \xi_{i_1}\xi_{i_2}\ldots\xi_{i_K}$, taking the product of K message sites. The original message is then retrieved by exploring the ground state of the related Hamiltonian

$$\mathcal{H} = -\sum_{\langle i_1,i_2\ldots i_K\rangle} \mathcal{A}_{\langle i_1,i_2\ldots i_K\rangle}\; J_{\langle i_1,i_2\ldots i_K\rangle}\; S_{i_1}S_{i_2}\ldots S_{i_K} - F/\beta\sum_k S_k \;, \tag{1}$$

where $\boldsymbol{S}$ is the N-dimensional vector of binary dynamical variables and $\mathcal{A}$ is a sparse tensor with C unit elements per index (setting the rest of the elements to zero), used for constructing the code word $\boldsymbol{J}^0$ by selecting K message sites per code word bit. The last term on the right is required in the case of sparse or biased messages and will require assigning a certain value to the additive field F/β. Codes of $K=2$ and $K\to\infty$ have been analysed [4, 5] in the case of extensive connectivity with $C \sim \binom{N-1}{K-1}$ and code-rate $R = K/C \to 0$, corresponding to the SK [10] and Random Energy [11] models, respectively; the intensive case with finite and infinite K, which is of greater practical significance ($R \neq 0$) and which we will consider here, has only recently been analysed [8].

We will now present two approaches for decoding the corrupted received message based on the Bayesian framework and on a statistical-mechanics analysis; the two approaches stem from the same probabilistic framework and can be easily linked [4].

Decoding the received message $\boldsymbol{J}$ in the Bayesian framework can be carried out by calculating the marginal posterior probability $\mathcal{P}(S_l|\boldsymbol{J}) = \mathrm{Tr}_{\{S_{k\neq l}\}}\mathcal{P}(\boldsymbol{S}|\boldsymbol{J}) \sim \mathrm{Tr}_{\{S_{k\neq l}\}}\prod_\mu \mathcal{P}(J_\mu|\boldsymbol{S})\;\mathcal{P}_0(\boldsymbol{S})$ for each spin site l, where μ runs over the message components and $\mathcal{P}_0(\boldsymbol{S})$ represents the prior; note the similarities to the statistical-mechanics formulation as the logarithms of the likelihood and prior terms are directly related to the first and second components of the Hamiltonian (eq. (1)), respectively. Knowing the posterior, one can calculate the typical retrieved message elements and their alignment with ± 1, which correspond to the Bayes-optimal decoding; however, this turns out to be rather difficult in general and we therefore resort to the methods of belief propagation, aimed at providing a good approximation to the marginal posterior. This approach, which is quite similar to the practical approach employed in the case of Gallager codes [2], assumes a two-layer system corresponding to the elements of the corrupted message $\boldsymbol{J}$ and the dynamical variables $\boldsymbol{S}$, respectively, and focuses on the calculation of conditional probabilities between elements from the two layers when some elements of the system are set to specific values or removed. Through this process one defines sets of conditional probabilities relating elements in the two layers (following the general framework of [1] or the more specific treatments of refs. [2, 3]):

$$\begin{aligned} q^x_{\mu l} &= \mathcal{P}\left(S_l = x \mid \{J_{\nu\neq\mu}\}\right) , \\ r^x_{\mu l} &= \mathcal{P}\left(J_\mu \mid S_l = x, \{J_{\nu\neq\mu}\}\right) = \mathrm{Tr}_{\{S_{k\neq l}\}}\mathcal{P}\left(J_\mu \mid S_l = x, \{S_{k\neq l}\}\right)\; \mathcal{P}\left(\{S_{k\neq l}\}\,|\{J_{\nu\neq\mu}\}\right) \;, \end{aligned} \tag{2}$$

where the index μ represents *an element* of the multidimensional tensor $\boldsymbol{J}$ which is connected to the corresponding index of $\boldsymbol{S}$ (l in the first equation), *i.e.* for which the corresponding element $\mathcal{A}_{\langle i_1,\ldots,l,\ldots i_K\rangle}$ is non-zero; the notation $\{S_{k\neq l}\}$ refers to all elements of the vector $\boldsymbol{S}$, excluding the l-th element, which are connected to the corresponding index of $\boldsymbol{J}$ (μ in this case for the second equation); the index x can take values ± 1. The conditional probabilities $q^x_{\mu l}$ and $r^x_{\mu l}$ will enable us, through recursive calculations, to obtain an approximated expression to the posterior.

Employing Bayes rule and the assumptions that the dependence of S_l on an element J_ν is factorizable and vice versa (which are quite reasonable as variables from the same layer are not expected to be *directly* dependent):

$$\mathcal{P}\left(S_{l_1}, S_{l_2} \ldots S_{l_K} \middle| \{J_{\nu\neq\mu}\}\right) = \prod_{k=1}^{K} \mathcal{P}\left(S_{l_k} \middle| \{J_{\nu\neq\mu}\}\right) \quad \text{and}$$
$$\mathcal{P}\left(\{J_{\nu\neq\mu}\} \mid S_l = x\right) = \prod_{\nu\neq\mu} \mathcal{P}\left(J_\nu \mid S_l = x, \{J_{\sigma\neq\nu}\}\right) \ , \tag{3}$$

one can write a set of coupled equations for $q_{\mu l}^{\pm 1}$ and $r_{\mu l}^{\pm 1}$ of the form

$$\begin{cases} q_{\mu l}^{x} = a_{\mu l}\ p_l^x \prod_{\nu\neq\mu} r_{\nu l}^{x} \,, \\ r_{\mu l}^{x} = \mathrm{Tr}_{\{S_{k\neq l}\}} \mathcal{P}\left(J_\mu \mid S_l = x, \{S_{k\neq l}\}\right) \ \prod_{k\neq l} q_{\mu k}^{S_k} \ , \end{cases} \tag{4}$$

where $a_{\mu l}$ is a normalising factor such that $q_{\mu l}^{1}+q_{\mu l}^{-1}=1$ and $p_l^x=\mathcal{P}\left(S_l=x\right)$ are our prior beliefs in the value of the source bits S_l.

This set of equations can be solved iteratively [2] by updating a closed coupled set of difference equations for $\delta q_{\mu l} = q_{\mu l}^{1} - q_{\mu l}^{-1}$ and $\delta r_{\mu l} = r_{\mu l}^{1} - r_{\mu l}^{-1}$, derived for this specific model, making use of the fact that the variables $r_{\mu l}^{x}$, and subsequentially the variables $q_{\mu l}^{x}$, can be calculated by exploiting the relation $r_{\mu l}^{\pm 1} = (1 \pm \delta r_{\mu l})/2$ and eqs. (4). At each iteration we can also calculate the pseudo-posterior probabilities $q_l^x = a_l\ p_l^x\ \prod_\nu r_{\nu l}^{x}$, where a_l are normalising factors, to determine the current typical value of S_l and consequently the decoded message.

Three points that are worthwhile noting: Firstly, the iterative solution makes use of the normalisation $r_{\mu l}^{1} + r_{\mu l}^{-1} = 1$, which is *not* derived from the basic probability rules and makes implicit assumptions about the probabilities of obtaining $S_l = \pm 1$ for all elements l. Secondly, the iterative solution would have provided the true posterior probabilities q_l^x if the graph connecting the message $\boldsymbol{J}$ and the encoded bits $\boldsymbol{S}$ would have been free of cycles, *i.e.* a tree with no recurrent dependences among the variables. The fact that the framework does provide adequate practical solutions has only recently been explained [12]. Thirdly, it is important to consider the complexity of this decoding scheme as it is of significant practical relevance. Such analysis has been carried out in ref. [2] resulting in an $\mathcal{O}(K/R)$ operations per decoded bit with a prefactor which depends on the number of iterations required and is typically around 100, which clearly renders this decoding scheme practical.

We will now turn to an alternative approach, showing that for this particular problem it is possible to obtain a similar set of equations from the corresponding statistical-mechanics framework based on Bethe approximation [13] or the TAP approach [14] to diluted systems ([1]). In this approach we assign a Boltzmann weight to each set comprising an encoded message bit J_μ and a dynamical vector $\boldsymbol{S}$,

$$w_B\left(J_\mu|\boldsymbol{S}\right) = e^{-\beta\ g\left(J_\mu|\boldsymbol{S}\right)} \ , \tag{5}$$

such that the first term of the system's Hamiltonian (eq. (1)) can be rewritten as $\sum_{\mu=1}^{L} g(J_\mu|\boldsymbol{S})$, where the index $\mu = 1$ runs over the L non-zero sites in the multidimensional tensor $\mathcal{A}$

([1]) Note that the terminology in the case of diluted systems is slightly vague as an expansion with respect to the large Onsager fields is meaningless; here we follow the conventional terminology for the Bethe approximation when applied to disordered systems subject to mean-field–type random interactions.

(which multiplies $\boldsymbol{J}$). We will now employ two straightforward assumptions to obtain a set of coupled equations for the mean field $q_{\mu l}^{S_l} = \mathcal{P}(S_l | \{J_{\nu \neq \mu}\})$, which may be identified as the same variable as in the belief network framework (eq. (2)), and the effective Boltzmann weight $w_{\text{eff}}(J_\mu | S_l, \{J_{\nu \neq \mu}\})$:

1) we assume a mean-field behaviour for the dependence of the dynamical variables $\boldsymbol{S}$ on a certain realization of the message sites $\boldsymbol{J}$, *i.e.* the dependence is factorizable and may be replaced by a product of mean fields.

2) Boltzmann weights for a specific site S_l are factorizable with respect to the message sites J_μ.

One may argue that these assumptions will provide a reasonable approximation due to the lack of direct dependence between elements of $\boldsymbol{S}$ and similarly between elements of $\boldsymbol{J}$ ([2]). The resulting set of equations are of the form

$$\begin{cases} w_{\text{eff}}(J_\mu \mid S_l, \{J_{\nu \neq \mu}\}) \, \text{Tr}_{\{S_{k \neq l}\}} \; w_B(J_\mu \mid \boldsymbol{S}) \; \prod_{k \neq l} q_{\mu l}^{S_k} \, , \\ q_{\mu l}^{S_l} = \tilde{a}_{\mu l} \; p_l^{S_l} \; \prod_{\nu \neq \mu} w_{\text{eff}}(J_\nu \mid S_l, \{J_{\sigma \neq \nu}\}) \, , \end{cases} \tag{6}$$

where $\tilde{a}_{\mu l}$ is a normalisation factor and $p_l^{S_l}$ represents our prior knowledge of the source's bias. Replacing the effective Boltzmann weight by a normalised field, which may be identified as the variable $r_{\mu l}^{S_l}$ in the belief network framework (eq. (2)), we obtain

$$r_{\mu l}^{S_l} = \mathcal{P}(S_l \mid J_\mu, \{J_{\nu \neq \mu}\}) = a_{\mu l} \; w_{\text{eff}}(J_\mu \mid S_l, \{J_{\nu \neq \mu}\}) \, , \tag{7}$$

i.e. a set of equations equivalent to eqs. (4). The explicit expressions of the normalisation coefficients, $a_{\mu l}$ and $\tilde{a}_{\mu l}$, are

$$a_{\mu l}^{-1} = \text{Tr}_{\{S\}} \; w_B(J_\mu | \boldsymbol{S}) \prod_{k \neq l} q_{\mu l}^{S_k} \quad \text{and} \quad \tilde{a}_{\mu l}^{-1} = \text{Tr}_{\{S_l\}} \; p_l^{S_l} \prod_{\nu \neq \mu} r_{\nu l}^{S_l} \, . \tag{8}$$

The somewhat arbitrary use of the differences $\delta q_{\mu l} = \langle S_l^\mu \rangle_q$ and $\delta r_{\mu l} = \langle S_l^\mu \rangle_r$ in the BP approach becomes clear form the TAP formulation, where they represent the expectation values of the dynamical variables with respect to the fields. The statistical-mechanics formulation also provides a partial answer to the successful use of the BP methods to loopy systems, as we consider a finite number of steps on an infinite lattice [15]. However, it does not provide an explanation in the case of small loopy systems which should be examined using other methods.

The formulation so far has been rather general and enabled us to show the similarity between the set of iterative equations obtained by the BP and TAP approaches. We will now make use of this set of equations to study the efficacy and usefulness of these methods to the problem at hand, *i.e.* decoding corrupted messages encoded using Sourlas's code. In this case we can make use of the explicit expression for the function g (from eq. (1)) to derive the relation between $q_{\mu l}^{S_l}$, $r_{\mu l}^{S_l}$, $\delta q_{\mu l}$ and $\delta r_{\mu l}$,

$$q_{\mu l}^{S_l} = \frac{1}{2}(1 + \delta q_{\mu l} S_l) \quad \text{and} \quad r_{\mu l}^{S_l} = \frac{1}{2}(1 + \delta r_{\mu l} S_l) \, , \tag{9}$$

([2]) Obviously, the TAP approach is an approximation in this case and these assumptions will be validated later on by comparing the solutions to those obtained by a different method.

as well as an explicit expression for $w_B\left(J_\mu|\boldsymbol{S},\beta\right)$,

$$w_B\left(J_\mu|\boldsymbol{S},\beta\right)=\frac{1}{2}\cosh\beta J_\mu\left(1+\tanh\beta J_\mu\prod_{l\in\mathcal{L}(\mu)}S_l\right)\ , \tag{10}$$

where $\mathcal{L}(\mu)$ is the set of all sites of $\boldsymbol{S}$ connected to J_μ, *i.e.* for which the corresponding element of the tensor $\mathcal{A}$ is non-zero. The explicit form of the equations for $\delta q_{\mu l}$ and $\delta r_{\mu l}$ becomes

$$\begin{cases}\delta r_{\mu l}=\tanh\beta J_\mu\prod_{l\in\mathcal{L}(\mu)/l}\delta q_{\mu l}\ ,\\ \delta q_{\mu l}=\tanh\left(\sum_{\nu\in\mathcal{M}(l)/\mu}\tanh^{-1}\delta r_{\nu l}+F\right)\ ,\end{cases} \tag{11}$$

where $\mathcal{M}(l)/\mu$ is the set of all indices of the tensor $\boldsymbol{J}$, excluding μ, which are connected to the vector site l; the external field F which previously appeared in the last term of eq. (1) is directly related to our prior belief of the message bias

$$p_l^{S_l}=\frac{1}{2}\left(1+\tanh FS_l\right)\ . \tag{12}$$

We will now employ eqs. (11) and the explicit expressions obtained above, by making use of differences $\delta q_{\mu l}$ and $\delta r_{\mu l}$, to obtain values of $q_{\mu l}^{\pm 1}$ and $r_{\mu l}^{\pm 1}$. After these differences are determined, the (approximated) marginal posterior $q_l^{S_l}=(1+\delta q_l S_l)/2$ can be calculated,

$$\delta q_l=\tanh\left(\sum_{\mu\in\mathcal{M}(l)}\tanh^{-1}\delta r_{\mu l}+F\right), \tag{13}$$

providing the Bayes-optimal decoding $\xi_l^B=\mathrm{sign}\,\langle S_i\rangle_T=\mathrm{sign}\left(\delta q_l\right)$. The magnetisation $M=1/N\ \sum_{i=1}^N\xi_i\xi_i^B$ serves as our performance measure.

We obtained numerical solutions for the cases $K=2,5$, corruption rate $0\leq p\leq 0.5$, two bias values (0.1, 0.5) and several temperatures, as shown in fig. 1, which will be compared to previously obtained solutions [8] using the replica method. The latter have been obtained by replica symmetric and one step replica symmetry-breaking calculations of the system's free energy for the ferromagnetic and paramagnetic phases and the spin-glass phase, respectively (expecting strong replica symmetry breaking only in the latter), following the work of Sherrington and Wong [15]; saddle-point equations have been solved both analytically and numerically by employing Monte Carlo techniques.

In the experiments, connectivity is set as $C=4,10$ for $K=2,5$, respectively, which provides the same code rate $R=1/2$ for both cases. For each run, 20000 bit code words $\boldsymbol{J}^{\mathrm{o}}$ are generated from 10000 bit message $\boldsymbol{\xi}$ using a fixed random sparse tensor $\mathcal{A}$. The noise-corrupted code word $\boldsymbol{J}$ was decoded according to eqs. (11) and (13) to retrieve the original message $\boldsymbol{\xi}$. Numerical solutions of 10 individual runs [16], for each value of the flip rate p starting from different initial conditions, obtained for the case $K=2$, different biases ($f=p_l^1=0.1,0.5$ —the probability of $+1$ bit in the original message $\boldsymbol{\xi}$) and temperatures ($T=0.26,T_{\mathrm{n}}$) are shown in fig. 1(a). The choice of $T=0.26$, rather than $T=0$, for representing solutions at low temperatures is in order to avoid computational difficulties. We obtain good agreement between the TAP/BP solutions and the theoretical values obtained using the methods of [8] (diamond symbols and dashed line, respectively). The results for biased patterns at $T=0.26$ presented in the form of mean values and standard deviation, show a suboptimal improvement in performance, as expected. Obtaining solutions under similar conditions but at Nishimori's

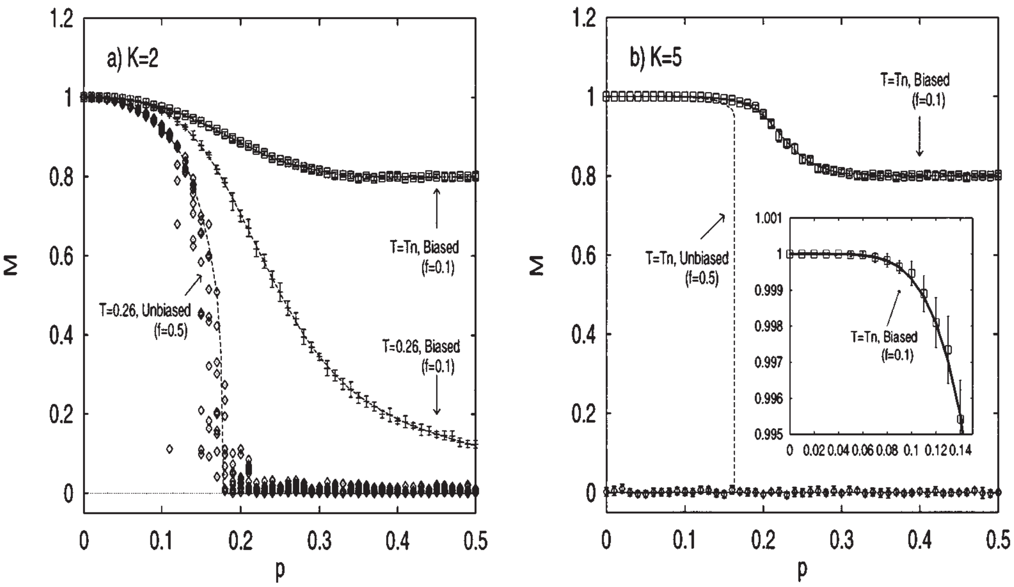

Fig. 1. – Numerical solutions for M and different flip rate p. (a) For the case $K=2$, different biases ($f=p_l^1=0.1, 0.5$) and temperatures ($T=0.26, T_{\mathrm{n}}$), we see good agreement between the TAP/BP solutions and the theoretical values. Results for the unbiased patterns are shown as raw data, *i.e.* results of 10 runs for each flip rate value p (diamond), while the theoretical solution is marked by the dashed line. Results for biased patterns are shown by their mean and standard deviation, showing a suboptimal improvement in performance as expected for $T=0.26$ and an optimal one at Nishimori's temperature $-T_{\mathrm{n}}$. Note that in the case of $T=T_{\mathrm{n}}$ the standard deviation is significantly smaller than the symbol size. (b) Shows results for the case $K=5$ and $T=T_{\mathrm{n}}$ in similar conditions to (a). Also here iterative solutions may generally drift away from the theoretical values where temperatures other than T_{n} are employed (not shown); using Nishimori's temperature alleviates the problem only in the case of biased messages and the results are in close agreement with the theoretical solutions (focusing on low p values in the inset).

temperature $-1/T_{\mathrm{n}}=1/2\ln[(1-p)/p]$ [17], we see that pattern sparsity is exploited optimally resulting in a magnetization $M\approx 0.8$ for high corruption rates, as T_{n} simulates accurately the loss of information due to channel noise [6, 7]; results for unbiased patterns (not shown) are not affected significantly by the use of Nishimori's temperature. The replica-based theoretical solutions [8] indicate a profoundly different behaviour for the $K=2$ case in comparison to other K values. We therefore obtained solutions for $K=5$ under similar conditions (which are representative for results obtained in other cases of $K\neq 2$); the results presented in fig. 1(b), in terms of means and standard deviation of 10 individual runs per flip rate value p, are less encouraging as the iterative solutions are highly sensitive to the choice of initial conditions and tend to converge to suboptimal values unless high sparsity and using the appropriate choice of temperature (T_{n}) forces them to the correct values, showing then good agreement with the theoretical results (solid line, see inset). This phenomenon is indicative of the fact that the ground state of the non-biased system is macroscopically degenerate with multiple equally good ground states.

The conclusion from these experiments is that the TAP/BP approach may be highly useful in the case of biased patterns but may lead to errors for unbiased patterns and $K\geq 3$, and that the use of the appropriate temperature, *i.e.* Nishimori's temperature, enables one to obtain improved results, in agreement with results presented elsewhere [5-7].

In this letter we compared the use of belief propagation to that of TAP for decoding corrupted messages encoded by using Sourlas's method. We have discovered that in this particular case the two methods provide an identical set of equations. We then employed these equations iteratively to derive solutions for particular scenarios and compared them to those obtained by the replica method. The solutions indicate that the method is particularly useful in the case of biased messages and that using Nishimori's temperature is highly beneficial; solutions obtained using other temperature values may be suboptimal. For unbiased messages and $K \geq 3$ we may obtain erroneous solutions using these methods.

It would be interesting to explore whether the similarity in the equations derived using TAP and BP is restricted to this particular case or whether there is a deeper general link between the two methods. Another important question that remains open is the generality of our conclusions on the efficacy of these methods for decoding corrupted messages, as they are currently being applied in a variety of state-of-the-art coding schemes (*e.g.*, [2, 3]). Understanding the limitations of these methods and the proper way to use them in general, especially in the context of error-correcting codes, may be highly beneficial to practitioners. These questions and others, on the relations between statistical mechanics and error-correcting codes, will be discussed in future publications.

***

This work was partially supported by the program "Research For The Future" (RFTF) of the Japanese Society for the Promotion of Science (YK) and by EPSRC grant GR/L19232 (DS).

REFERENCES

[1] Pearl J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann) 1988.
[2] MacKay D. J. C. and Neal R. M., *Elect. Lett.*, **33** (1997) 457; and MacKay D. J. C., to be published in *IEEE IT*.
[3] Frey B. J., *Graphical Models for Machine Learning and Digital Communication* (MIT Press) 1998.
[4] Sourlas N., *Nature*, **339** (1989) 693.
[5] Sourlas N., *Europhys. Lett.*, **25** (1994) 159.
[6] Ruján P., *Phys. Rev. Lett.*, **70** (1993) 2968.
[7] Nishimori H., *J. Phys. Soc. Jpn.* **62** (1993) 1169.
[8] Kabashima Y. and Saad D., preprint (1997).
[9] Saakian D. B., *JTEP Lett.*, **67** (1998) 440.
[10] Sherrington D. and Kirkpatrick S., *Phys. Rev. Lett.*, **35** (1975) 1792.
[11] Derrida B., *Phys. Rev. B*, **24** (1981) 2613.
[12] Weiss Y., MIT preprint CBCL155 (1997).
[13] Bethe H., *Proc. R. Soc. London, Ser. A*, **151** (1935) 540.
[14] Thouless D., Anderson P. W. and Palmer R. G., *Philos. Mag.*, **35** (1977) 593.
[15] Sherrington D. and Wong K. Y. M., *J. Phys. A*, **20** (1987) L785.
[16] Each run requires a number of iterations which is of the order of 10, where initial conditions are set as $\delta r_{\mu l} = 0$ and $\delta q_{\mu l} = \tanh F$ reflecting the prior belief; whenever the TAP/BP approach was successful in predicting the theoretical values we observed convergence in most runs corresponding to the ferromagnetic phase while almost all runs at low temperatures did not converged to a stable solution above the critical flip-rate (although the magnetization M did converge as one may expect). Solutions obtained for all cases did not show a significant dependence on the initial conditions, except for the case $K = 5$, unbiased patterns at low temperatures, where one expects high dependence due to the nature of the solution.
[17] Nishimori H., *J. Phys. C*, **13** (1980) 4071.