IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Survey propagation for the cascading Sourlas code

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2006 J. Phys. A: Math. Gen. 39 10659

(http://iopscience.iop.org/0305-4470/39/34/005)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 171.66.16.106
The article was downloaded on 03/06/2010 at 04:47

Please note that terms and conditions apply.

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **39** (2006) 10659–10672 doi:10.1088/0305-4470/39/34/005

Survey propagation for the cascading Sourlas code

J P L Hatchett[1] **and Y Kabashima**[2]

[1] 96 Highbury Hill, London N5 1AT, UK
[2] Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama 226 8502, Japan

E-mail: hatchettman@hotmail.com and kaba@dis.titech.ac.jp

Received 11 April 2006, in final form 12 July 2006
Published 9 August 2006
Online at stacks.iop.org/JPhysA/39/10659

Abstract

We investigate how insights from statistical physics, namely survey propagation, can improve decoding of a particular class of sparse error correcting codes. We show that a recently proposed algorithm, time averaged belief propagation, is in fact intimately linked to a specific survey propagation for which Parisi's replica symmetry breaking parameter is set to zero, and that the latter is always superior to belief propagation in the high-connectivity limit. We briefly look at further improvements available by going to the second level of replica symmetry breaking.

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

1. Introduction

Error correcting codes are an important device for communicating information through noisy channels. We are living in what some social commentators have called the 'information age', with apparently ever increasing appetite for broadband Internet, mobile phone data transmission, satellite television, computational power, etc. Hence, it is no surprise that such codes are of significant practical use.

A family of error-correcting codes based on an identification with the ever popular Ising model from statistical mechanics was proposed by Sourlas [1], and novel coding methods were examined [2, 3]. A further important advance, in terms of practical significance, was the move to finite rate, finite length Sourlas-type codes [4]—which used interactions between K bits and C interactions per bit. While these codes showed good theoretical performance, the number of successfully transmitted bits was significantly higher (for certain channel noises) when $K > 2$. However, in this case the basin of attraction (BOA) for a local decoding algorithm is very small meaning that decoding is computationally extremely challenging for reasonable code lengths. A more directly useful version of the above theory involves cascading Sourlas codes [5] and other irregular low density parity check codes [6, 7].

 © 2006 IOP Publishing Ltd Printed in the UK

The cascading approach [5] utilizes the large BOA for $K = 2$ and was first to introduce $K = 1$ as an important ingredient of the cascading method. These are combined with the high overlap possible for larger K. Explicitly, multiple values of K are used in the interaction graph. The first stage is to perform decoding restricted to the energy landscape given only by the lowest values of K. This landscape will ideally have a large BOA at the cost of a smaller final overlap. However, once decoding has finished, the decoded state is used as an initial state for decoding using the next larger value of K. Although with the larger value of K the BOA is expected to be smaller, the algorithm is not starting from a random initial condition but from a state which could have non-trivial overlap with the true solution. Ideally, it is within the BOA for the landscape comprising the interactions from both values of K. Simply using two different values of K can be sufficient to obtain a useful algorithm but more different values can be used and each higher value of K is introduced in turn: at each stage the algorithm is seeded with the output from the previous stage's decoding.

Belief propagation (BP) [8], which is closely related to the Bethe approximation and the cavity (replica) method in finite connectivity disordered systems [9, 10], has been the principle algorithm used for decoding the finite rate Sourlas codes including the case of the cascade version. For certain code/channel instances the quenched disorder in this problem, in the form of corrupted bits which cause frustration in the graph of bit interactions, brings about a certain disruption to the BP dynamics, which can lead to deterioration of error correction performance. For an infinite system size we expect that the failure of BP to converge (and hence failure to decode) is linked to the de Almeida–Thouless (AT) transition [18] in the equilibrium system, although we cannot prove this statement and a strict interpretation of the AT transition for finite code lengths (i.e. finite system sizes) is also lacking.

Recently, a finite temperature survey propagation algorithm (SP) [11–14] has been applied to a biased 3-body Sourlas code [15]. The latter found that this more advanced algorithm did lead to a shift in the critical channel noise for computationally feasible decoding when BP did not converge. They also noted that averaging the beliefs in time, or time averaged belief propagation (TABP) [16, 17] could improve decoding in such cases as well. However, the reason for this improvement has not been fully clarified yet.

The objective of this paper is to pinpoint the major contributing factors for such improvement in practical decoding. For this, we will investigate how the overlap between the original and estimated messages in the first state of the cascading codes, for which only $K = 2$ interactions are considered, is influenced by the employed decoding algorithms when the Almeida–Thouless (AT) stability is broken [18]. For simplicity of numerical experiments, we will mainly focus on the zero-temperature cases. Although the model that we will deal with does not directly accord with those in the preceding work [15], the two systems have a common feature that the AT stability is broken and, therefore, insights gained from our simpler system are probably useful for understanding what occurs in the biased 3-body interaction system at finite temperature. By applying SP to the first stage of a cascading Sourlas code, we will find that we can increase the overlap with the original message below the AT transition, which in turn improves the chance of being within the BOA for good decoding once we introduce higher body interactions. It has also recently been shown that the first step of replica symmetry breaking (1RSB) approach leads to higher values of the magnetization for disordered ferromagnetically biased $\pm J$ 2-spin Ising systems in finite connectivity random graphs [19]. Thus, we strongly speculate that the higher first stage overlap gained by introduction of the RSB ansatz in the cascading scheme is the main factor for the improvement of the error correcting performance. This speculation also suggests that algorithms based on increasing levels of replica symmetry breaking (RSB) would improve matters further, albeit at a high computational cost. We make an initial investigation of this by a brief analysis of a 2RSB

survey propagation (2SP: although this could rightly be called a survey of surveys propagation). We will also argue that TABP corresponds to survey propagation of a specific type for which the Parisi parameter, which specifies the size of subgroups of replicas in the 1RSB ansatz, is set to zero and is never bettered by the simple BP algorithm, as is observed in CDMA multiuser detection problems [20, 21].

This paper is organized as follows. In the next section, we introduce the model that we will study. In sections 3 and 4, two decoding schemes based on BP and SP are presented in the case of vanishing temperature. In sections 5 and 6, we will show numerically that the SP-based algorithm with zero Parisi parameter (SP0) is, in practice, equivalent to TABP and provides a better performance than BP when the AT stability is broken. This relationship of SP0 to other algorithms is provided analytically in the vanishing code rate limit in the appendix. The general proof of this relationship, unfortunately, has not been managed yet. We also show numerically that 2SP can lead to an improvement of the decoding performance in the situation where the AT stability broken. This implies that a further improvement can be expected by introducing a higher level RSB ansatz although the computational cost for decoding will be sacrificed. The final section is devoted to a summary.

2. Model definitions

We assume that we have some message source vector $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_N\} \in \{-1, 1\}^N$ (for simplicity of exposition we work within the Ising formalism). We encode this source vector via a transformation to an M-bit vector $\mathbf{J}^0$ where we have $M > N$ to introduce redundancy into the code to enable the possibility of error correction. The rate R of such a code is given by N/M. The structure of the code is set by the size $n_a = |a|$ for M sets a. A given code is formed by populating each set a with n_a distinct indices from $\{1, \ldots, N\}$ subject to the constraint that each index appears in C_K different K body interactions (sets of size K). Finally, the encoded source is given by $J_a^0 = \prod_{i \in a} \xi_i a = 1, \ldots, M$. We consider a binary symmetric channel so that when the encoded message vector $\mathbf{J}^0$ is passed through the channel each bit is flipped independently with probability p. The received message is denoted by $\mathbf{J}$. The object of any decoding algorithm is to estimate $\boldsymbol{\xi}$ with some $\boldsymbol{\sigma}$ only using information from $\mathbf{J}$. A natural performance measure is the overlap between our estimate and the true message $M = \frac{1}{N} \sum_i \xi_i \sigma_i$. One way to proceed with decoding is to look for the ground state of the following Hamiltonian:

$$\mathcal{H}(\boldsymbol{\sigma}) = -\sum_{a=1}^{M} J_a \prod_{i \in a} \sigma_i. \tag{1}$$

Since this Hamiltonian is invariant under the transformation $\sigma_i \to \sigma_i \xi_i$ and $J_a \to J_a \prod_{i \in a} \xi_i$ we can gauge away the original message and are left with a multispin ferromagnetically biased $\pm J$ spin glass model.

For the cascading approach here we consider a code with both 2-body interactions and K-body interactions. We define a new auxiliary Hamiltonian $\mathcal{H}'(\boldsymbol{\sigma}) = -\sum_{a=1}^{M} J'_a \prod_{i \in a} \sigma_i$ where $J'_a = J_a$ if $|a| = 2$ and $J'_a = 0$ otherwise. The initial decoding step is to find minima of $\mathcal{H}'(\boldsymbol{\sigma})$. This is relatively easier than finding minima of $\mathcal{H}$. We then use this minimum of $\mathcal{H}'$ as an initial condition for local search algorithms looking for minima of the full Hamiltonian $\mathcal{H}$ from equation (1).

3. Solution at the level of replica symmetry

An approximate method of finding the ground state is available via belief propagation, which is equivalent to the cavity method at the level of replica symmetry for locally treelike graphs. The present code is locally treelike with only a finite number of exceptions in the long code length limit. The same results can be found in the ensemble of graphs in the thermodynamic (long code length) limit by use of the replica method, however, since we are interested in decoding particular examples the former method is preferable. The method has been discussed to a great deal in a recent literature [8, 9, 10, 13] so we will be relatively brief. We define cavity fields $h_{i\to a}$ to be the field on site i (decoded estimator σ_i) in the absence of link a (received bit J_a) and the message $u_{a\to i}$ to be the message from received bit J_a to decoded bit σ_i. We are working with a discrete energy model and a self-consistent solution to these equations is given by restricting the values of $\{u, h\}$ to integers. Then, the belief propagation (BP) update equations describing these variables are:

$$h_{i\to a}^{t+1} = \sum_{b\in\mathcal{N}(i)\setminus a} u_{b\to i}^{t}, \tag{2}$$

$$u_{a\to i}^{t} = \operatorname{sign}\left(J_a \prod_{j\in a\setminus i} h_{j\to a}^{t} \right) \tag{3}$$

where $\mathcal{N}(i)$ denotes the set $\{a : i \in a\}$ and we take the convention $\operatorname{sign}(0) = 0$. Other than the trivial paramagnetic state ($h_{i\to a} = u_{a\to i} = 0\ \forall i, a$), fixed points of the message passing equations (3) correspond to local minima of the energy of the Hamiltonian (1). However, it is certainly not guaranteed that equations (3) will converge at all. If the messages converge, or after some specified number of updates if they do not, the estimator for the original message bits is given by

$$\sigma_i = \operatorname{sign}\left(\sum_{a\in\mathcal{N}(i)} u_{a\to i} \right). \tag{4}$$

Consider a specific example, that of a code with $C_2 = 3$ and $C_3 = 3$ so each original message bit ξ_i is contained within three 2-body interactions and three 3-body interactions. Due to the 3-body interactions we find that without using the cascading approach we never find the state with high overlap—even at zero channel noise—due to the higher body interactions, which is similar to the result found for purely ferromagnetic 3-body interactions in [22], where even for these purely ferromagnetic interactions the structure of the interactions led to glassiness. However, if we first allow the system to equilibrate using the energy landscape given by only the 2-body interactions (so perform BP on the auxiliary Hamiltonian $\mathcal{H}'$ found by setting the exchange energy of all 3-body interactions to zero in $\mathcal{H}$) we find that for a certain range of p we obtain non-zero values of M. Since for purely 2-body interactions there is a reflection symmetry in the Hamiltonian there is no bias at this stage between $\pm M$. However, when the 3-body interactions are added to the Hamiltonian it is possible to infer which of these two states gives a better overlap overall, namely the state with lower energy. BP is then performed on the Hamiltonian with the 3-body interactions added, but *not* from a random initial condition. Since the initial state for the combined system is the one with finite M, it is closer to the ground state with high overlap than a random initial condition so BP has a better chance of reaching the ground state. In practice, we add the 3-body interactions in three batches, at each batch we add one 3-body interaction to the Hamiltonian for each spin (so $N/3$ interactions).

4. Solution at the level of replica symmetry breaking

Past some critical channel noise we expect there to be an ergodicity breaking transition corresponding to replica symmetry breaking, at which point the update equations (3) will no longer be sufficient to correctly characterize the system. For example, if we consider the initial stage of the algorithm where messages are passed purely between sites and 2-body links then it is known that for $C_2 = 3$ (three 2-body interactions) the AT transition occurs at $p = \frac{1}{12}$ [19, 23]. Below this point a survey propagation algorithm [11, 14] should improve at least the initial overlap M (the AT-point for the full system will not be the same as that for the initial system). This means that at the second stage the initial condition is closer to the high overlap state. This improvement will occur at most until the transition from the mixed phase (glassy ferromagnetic pure states) to the spin glass phase. By considering the finite temperature replica symmetric transition lines away from the paramagnetic phase, a bifurcation analysis of the belief propagation equations yields critical temperatures for the $C_2 = 3$ regular Sourlas code:

$$P \to F: \qquad 1 = 2(1-2p)\tanh(\beta) \tag{5}$$

$$P \to SG: \qquad 1 = 2\tanh^2(\beta) \tag{6}$$

which implies the triple point is at $p = \frac{1}{2}\left(1 - \sqrt{\frac{1}{2}}\right) = 0.1464\ldots$. According to the Parisi–Toulouse hypothesis of no re-entrance this gives us the zero-temperature transition point between the mixed phase and the spin glass phase.

At the first level of replica symmetry breaking, survey propagation with Parisi's replica symmetry breaking parameter equal to x (SPx) rather than solving for a single message from sites to links (and links to sites) we have to solve for a distribution of messages. This distribution is over the messages in different ergodic sectors. While true ergodicity breaking will not occur in a finite length message, the breaking effects are sufficiently strong to mean that it does occur to all intents and purposes on any practical timescale for a message of reasonable length. The key idea behind the SPx is that when performing an iteration of our messages, the energies in different ergodic sectors will vary, and as they vary the probability of finding such a given state also varies. The SPx equations keep track of these changes assuming that the true nature of the states is 1RSB (which is not necessarily the case). If we again restrict ourselves to integer values for the messages (so in fact the integrals are really sums over the appropriate domain) then the iteration equations are:

$$P^{t+1}_{i\to a}(h) \propto \int \left[\prod_{b\in\mathcal{N}(i)\setminus a} \mathrm{d}u_b\, Q^t_{b\to i}(u_b)\right] \delta\left[h - \sum_b u_b\right] \mathrm{e}^{x|\sum_b u_b|} \tag{7}$$

$$Q^t_{a\to i}(u) \propto \int \left[\prod_{j\in a\setminus i} \mathrm{d}h_j\, P^t_{j\to a}(h_j)\right] \delta\left[u - \mathrm{sign}\left(J_a \prod_j h_j\right)\right] \mathrm{e}^{x\delta_{u,0}}. \tag{8}$$

The above pair of equations are iterated until convergence (or for some sufficient number of iterations). At that point we can construct our estimate for the decoded message via:

$$\sigma_i = \mathrm{sign}\left(\int \mathrm{d}H\, P_i(H)\mathrm{sign}(H)\right) \tag{9}$$

$$P_i(H) \propto \int \left[\prod_{b\in\mathcal{N}(i)} \mathrm{d}u_b\, Q_{b\to i}(u_b)\right] \delta\left[H - \sum_b u_b\right] \mathrm{e}^{x|\sum_b u_b|}. \tag{10}$$

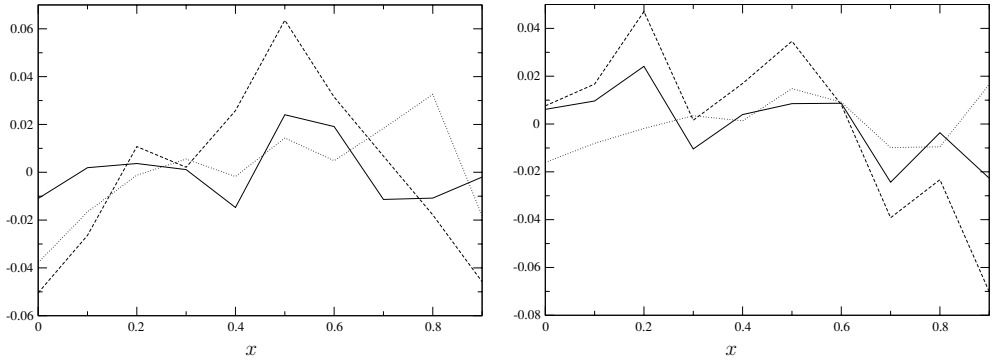

Figure 1. Variation (from the mean) in the overlap (solid line), energy (dashed line) and (zero-temperature) free energy (dotted line) versus the zero-temperature Parisi parameter x. The data are averages over 25 runs with code lengths $N = 1000$. The left figure is for $p = 0.14$ while the right figure is for $p = 0.12$. In both figures we see that variation in the overlap follows variation in the energy to a reasonable extent.

The approach thus far has still left the choice of x. In physical problems, the general approach is to extremize the generalized free energy w.r.t. x [12] as these states are the ground states of the system. However, despite using a physical approach, we are interested in the error correcting abilities of this system and it does not necessarily hold that extremizing the zero-temperature free energy corresponds to maximizing the overlap M. In fact, we have found that minimizing the energy $\mathcal{H}(\boldsymbol{\sigma})$ gives a good guide to choice of x which is reasonable since our initial aim was to decode via minimizing the energy of the Hamiltonian (1). If the computational effort required to fit the model is prohibitive, choosing a particular value of x for SPx (we found about 0.15 was reasonable) still gives a marked improvement over BP in decoding across a range of channel noises p. In figure 1 we plot variation in the overlap from its mean value to that of the energy and the free energy (we take ten times the variation of the free energy so that the scales are similar). The energy (which is observable) seems to give a reasonable indication of the overlap (which is not observable in practical situations).

It is to be expected that to correctly describe state space up to the transition from a mixed (glassy ferromagnetic) phase to a spin glass phase full RSB needs to be taken into account. However, this is not feasible computationally. This does make the next obvious step a 2RSB algorithm, 2SPxy, where the messages will be surveys of surveys. For a more detailed discussion on applications to k-SAT see [24]. We restrict our discussion for simplicity to the first stage of the decoding procedure (where all interactions are 2-body). Then we have update equations;

$$\mathcal{P}_{i\to a}^{t+1}[P] \propto \int \prod_{b\in\mathcal{N}(i)\backslash a} \mathrm{d}\mathcal{Q}_{b\to i}^{t}[Q_{b\to i}] z[\{Q\};y]^{x/y} \delta[P - P[\{Q\};y]] \tag{11}$$

$$\mathcal{Q}_{a\to i}^{t}[Q] = \int \mathrm{d}\mathcal{P}_{j\to a}^{t}[P]\delta[Q - Q[P;a]] \tag{12}$$

$$P[\{Q\};y] = \frac{1}{z[\{Q\};y]} \int \prod_{b\in\mathcal{N}(i)\backslash a} \mathrm{d}u_b Q_b(u_b) \delta\left[h - \sum_b u_b\right] \mathrm{e}^{y(|\sum_b u_b| - \sum_b |u_b|)} \tag{13}$$

$$Q[P;a](u) = \int \mathrm{d}h\, P(h)\delta[u - \mathrm{sign}(J_a h)] \tag{14}$$

where $z[\{Q\}; y]$ is the normalizing constant for $P[\{Q\}; y]$ and x and y are the zero-temperature Parisi parameters. The integrals over u, h are sums over the appropriate integers while the integrals over measures are over distributions on the integers. We will compare these different approaches numerically in the following.

5. Time averaged belief propagation and SP0

In general, there is no guarantee that BP will converge on a graph with loops. However, experimentally it appears that for large disordered random graphs the transition to non-convergence is closely linked to the AT transition [26]. One can also rationalize this in terms of correlation lengths. If the correlation length is less than $\log(N)$ then locally the graph appears to be a tree (apart from at a finite number of sites), so BP should converge provided the graph is sufficiently large (of course this is very large as the length scale of loops is $\log(N)$). Below the AT transition due to the proliferation of pure states there are necessarily correlations on all length scales throughout the graph, so one would not expect BP to converge as it is affected by the loops. One simple, robust approach to deal with estimation when BP has not converged is to use time averaged BP (TABP), which corresponds to taking averages in time of the messages, and performing inference with these average messages [15–17].

Another algorithm which improves on BP but is simpler than full SPx is SP0, as used for the CDMA problem in [20, 21]. SP0 does not resort to any reweighing of messages as the messages are passed but gains its improvement from a 'majority rules' approach. I.e. when decoding in a region where BP does not converge, inference is performed at each bit by using the majority decision of the different instances. For the fully connected CDMA problem, the local fields are Gaussian and it is possible to show analytically that SP0 improves inference compared to BP analytically [21]. In the present situation due to the finite connectivity nature of the problem the effective fields are not Gaussian distributed; thus we have not been able to show that SP0 improves on BP analytically (basically we do not have sufficiently general control of $\{Q(u), P(h)\}$). However, in the high C limit our present model reduces to the SK model [25] and in this case we again find that SP0 is strictly better in inference terms than BP below the AT line, which is shown in the appendix.

In fact, TABP and SP0 are intuitively quite similar. The first averages over BP states in time, while the latter averages over different BP states in parallel from different initial conditions and with mixing of different solutions at each update step. Provided the dynamics of BP are sufficiently chaotic that it explores all states one could expect both averages to give similar (or even identical) results. To test this hypothesis we compare the distribution of local fields given by SP0 to the distribution given by TABP. More specifically we examine the order parameter

$$D(x) = \frac{1}{N} \sum_i \sum_{a \in N(i)} D_H\left[P_{i\to a}^{SPx} \,\|\, P_{i\to a}^{TABP}\right] \tag{15}$$

with the Hellinger distance $D_H(P\|Q) \equiv 2\int \mathrm{d}x(\sqrt{P(x)} - \sqrt{Q(x)})^2$. In figure 2, we look at $D(x)$ versus x and see that we find a very good agreement for the hypothesis that SP0 is equivalent to TABP which gives an alternative explanation to its improved performance over BP and provides a new method for implementing SP0.

6. Comparison of the different decoding algorithms

We have used four different algorithms for the initial decoding, BP, TABP (or SP0), SPx and 2SPxy, corresponding to different levels of replica symmetry breaking and model fitting. For

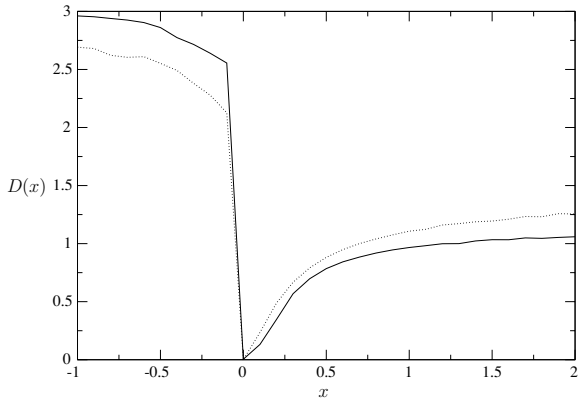

Figure 2. We see how the average distance $D(x)$ between the message distributions from SPx and the message distributions from TABP varies with x. The solid line is for channel noise $p = 0.12$ while the dotted line is for $p = 0.14$. We find good support for the hypothesis that TABP is equivalent to SP0.

each algorithm there is essentially only one pertinent question which is whether the overlap reached after initial decoding is sufficient to lead to good full decoding. While it is tempting to think of some critical M—in reality this is not sufficient to describe the state in full detail and other factors such as p and the decoding algorithms themselves will all contribute to changes in the M required (in addition to finite size fluctuations) to reach the high overlap state. However, due to issues of computational complexity we have tested all four algorithms on the initial decoding only. The 2SPxy algorithm has only been investigated very briefly there, for a single value of x and y rather than optimizing over them for each codeword received. We compare the results of these initial overlaps in figure 3. We see that each improvement in the algorithm does lead to an increased overlap after the initial stage, between the beginning of the mixed phase and the AT point. The 2SPxy does appear (as far as we can tell from the data) to give an improvement on the SPx algorithm for a limited range of channel noise.

Next we look in figure 4 at the final overlap reached after full decoding with BP, SP0 and SPx. We see that improving the quality of the algorithm leads to a marked improvement in the critical channel noise enlarging the space for which the cascading code approach will work. As we mentioned above, it is hard to draw firm conclusions from the improvement of M from the first stage as to the overall improvement of the algorithm, but in figure 4 we see that improvement at the first stage corresponds to a movement of the critical noise level p for overall decoding.

6.1. Comparison of computational complexity

It is clear from figure 4 that the suggested improvements in decoding algorithm shift the critical noise level past which the message cannot be decoded. However, the algorithms also increase in complexity so there is a computational cost to be paid for this improvement. A single iteration of BP involves the update of $NC_K K$ cavity messages for each K and NC cavity fields (where C is the total connectivity of each message bit). When BP converges, typically, it does so in $\mathcal{O}(10)$ iterations, so the overall complexity of BP is $\mathcal{O}\big(10N\big(C + \sum_K KC_K\big)\big)$.

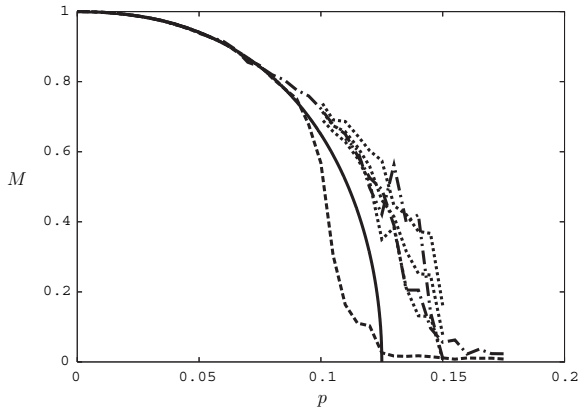

Figure 3. The value of the initial overlap M versus the channel noise p for several different algorithms. The solid line is the theoretical replica symmetric result in the infinite system sized limit. The dashed line is the average result (over 10 runs) of BP (note agreement is good after $p \approx 0.085$ the theoretical AT point). The dot–dashed line is TABP (averages over 10 runs) which improves significantly on BP for certain p. The three dashed lines are SPx (averaged over 10 runs and optimized over x) $\pm$ one standard deviation. The improvement of SPx on TABP is relatively modest. Finally the dot–dot–dashed line is a single run of 2SPxy with $x = 0.15$ and $y = 0.1$ (so not optimized)—it appears to improve on SPx for $p = 0.13, 0.135, 0.14$. Note that the predicted transition from a mixed phase to a spin-glass phase is at $p \approx 0.146$.

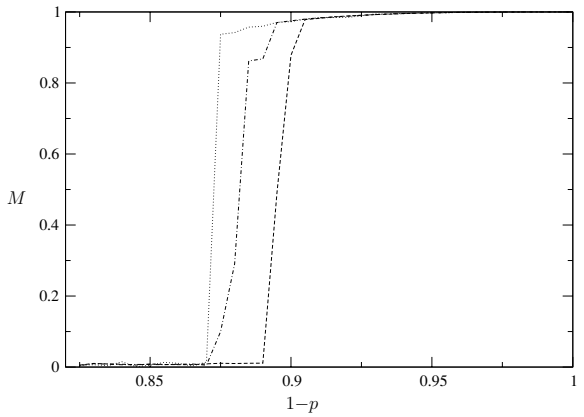

Figure 4. The final overlap M versus p averaged over 10 runs of code length $N = 10\,000$ for BP (dashed line), TABP (dot–dashed line) and SPx (dotted line). Each advance in algorithm leads to an improvement in the critical channel noise.

For TABP, BP does not converge and to get reasonable statistics for the messages/fields at a single site we took $\mathcal{O}(100)$ iterations before averaging, so in terms of the dominant contribution to the complexity TABP is approximately 10 times more demanding than BP.

SPx is more demanding again. When updating each of the NC cavity fields in (7) we have to sum over the three possible values of each message for $C - 1$ messages. In (8) we update a K-body message, which requires summing over the $2C - 1$ possible values of the field, $1 - C, \ldots, C - 1$, for $K - 1$ different fields. SPx also converges after $\mathcal{O}(10)$ iterations,

giving an overall complexity of $\mathcal{O}\big(10N(C3^{C-1}+\sum_K C_K K(2C-1)^{K-1})\big)$ which for realistic examples can easily be at least $\mathcal{O}(10^2)$ more demanding than BP if not more.

The final algorithm we have used, 2SPxy, encodes a distribution of survey propagations through a population dynamics method. In practice we used 100 surveys at each message/field. The additional reweighing and sampling of surveys at each update is an additional cost, but a relatively small one. Thus we see that compared to plain BP, 2SPxy is at least $\mathcal{O}(10^4)$ more demanding numerically, depending on the structure of the code. This is not taking into account the fact that x and y are real variables which should be optimized. These computation costs should be borne in mind when deciding on a decoding regime.

7. Conclusion

We have examined how taking replica symmetry breaking into account can improve decoding of cascading Sourlas codes. The basis behind our idea is that below the AT transition, the true overlap at the minima of the Hamiltonian (1) is above its replica symmetric estimate so increasing steps of replica symmetry breaking will serve to improve the overlap between the estimate of the message and the true message. We have also shown that TABP is closely related (if not identical) to SP0—this throws up two obvious areas for further investigation, one to find under what conditions these algorithms are identical and secondly to show that both of them are strict improvements on BP below the AT transition generally. Another area to improve is in terms of the 2SP, where further optimization of the Parisi parameters should lead to further improvements, albeit at some considerable computational cost. The graph structure used here is not expected to be optimal—after the initial stage using higher body interactions (i.e. $K > 3$) should give improvements in the final overlap, optimizing over possible architectures should also improve matters, but the principles developed here will help in any case where the channel noise is sufficiently high that decoding is taking place below the AT transition.

Acknowledgments

JPLH would like to thank Tommaso Castellani and Bastian Wemmenhove for helpful discussion of their work. YK acknowledges support by Grants-in-Aid Nos. 14084205 and 17340116 from MEXT/JSPS, Japan.

Appendix. Superiority of SP0 to BP in dense connectivity limit

Here, we show analytically that SP0 is never bettered by BP in terms of the error correction performance when the number of connections tends to infinity. Such a property is likely to hold in the case of sparse connectivity as well, analytical proof of which has not yet been managed.

A.1. SPx for densely connected systems

For simplicity and generality, we focus on regular $K = 2$ systems at general *finite* temperature $T = \beta^{-1}$. Extension to other K values and the case of vanishing temperature is straightforward. SPx of this system can be expressed as

$$P_{i\to a}^{t+1}(h) \propto \int \left[\prod_{b\in\mathcal{N}(i)\setminus a} \frac{\mathrm{d}u_b\, Q_{b\to i}^{t}(u_b)}{(2\cosh(\beta u_b))^{\frac{x}{\beta}}}\right] \delta\left[h-\sum_b u_b\right]\left(2\cosh\left(\beta\sum_b u_b\right)\right)^{\frac{x}{\beta}}, \tag{A.1}$$

$$Q_{a\to i}^{t}(u) \propto \int \left[\prod_{j\in a\setminus i} \mathrm{d}h_j\, P_{j\to a}^{t}(h_j)\right] \delta[u-u(h_j, J_a;\beta)], \tag{A.2}$$

where $u(h_j, J_a;\beta) \equiv \beta^{-1}\tanh^{-1}(\tanh(\beta J_a)\tanh(\beta h_j))$. In order to obtain a simpler expression for $C_2 = C \to \infty$, we introduce appropriate scalings $|J_a| \to \frac{J}{\sqrt{C}}$ and $p = \frac{1}{2} - \frac{J_0}{2J\sqrt{C}}$, where $J_0 > 0$ and $J > 0$. This corresponds to situations of low code rate communication that copes with strong noise despite restricted signal power. Let us denote $\int \mathrm{d}h_i\, P_{i\to a}^{t}(h_i)\tanh(\beta h_i) = m_{i\to a}^{t}$, $\int \mathrm{d}h_i\, P_{i\to a}^{t}(h_i)\tanh^2(\beta h_i) = M_{i\to a}^{t}$ and characterize the system utilizing two macroscopic variables $C^{-1}\sum_{b\in\mathcal{N}(i)\setminus a} M_{i\to b}^{t} \sim C^{-1}\sum_{a\in\mathcal{N}(i)} M_{i\to a}^{t} \sim (NC)^{-1}\sum_{i,a\in\mathcal{N}(i)} M_{i\to a}^{t} \equiv Q_1^t$ and $C^{-1}\sum_{b\in\mathcal{N}(i)\setminus a}\left(m_{i\to b}^{t}\right)^2 \sim C^{-1}\sum_{a\in\mathcal{N}(i)}\left(m_{i\to a}^{t}\right)^2 \sim (NC)^{-1}\sum_{i,a\in\mathcal{N}(i)}\left(m_{i\to a}^{t}\right)^2 \equiv Q_0^t$ based on the law of large numbers. These, in conjunction with the central limit theorem, indicate that equation (A.1) is reduced to

$$P_{i\to a}^{t+1}(h) = \frac{(2\cosh(\beta h))^{\frac{x}{\beta}}\, \mathrm{e}^{-\frac{(h-\phi_{i\to a}^{t+1})^2}{2\Delta^{t+1}}}}{\int \mathrm{d}h (2\cosh(\beta h))^{\frac{x}{\beta}}\, \mathrm{e}^{-\frac{(h-\phi_{i\to a}^{t+1})^2}{2\Delta^{t+1}}}}, \tag{A.3}$$

where

$$\phi_{i\to a}^{t+1} = \sum_{b\in\mathcal{N}(t)\setminus a} J_b m_{j\to b}^{t}, \tag{A.4}$$

$$\Delta^{t+1} = \sum_{b\in\mathcal{N}(t)\setminus a} J_b^2\left(M_{j\to b}^{t} - \left(m_{j\to b}^{t}\right)^2\right) = J^2\left(Q_1^t - Q_0^t\right). \tag{A.5}$$

In turn, equation (A.3) provides expressions of microscopic variables

$$m_{i\to a}^{t} = \frac{\int \mathrm{d}h (2\cosh(\beta h))^{\frac{x}{\beta}}\, \mathrm{e}^{-\frac{(h-\phi_{i\to a}^{t})^2}{2\Delta^{t}}}\tanh(\beta h)}{\int \mathrm{d}h (2\cosh(\beta h))^{\frac{x}{\beta}}\, \mathrm{e}^{-\frac{(h-\phi_{i\to a}^{t})^2}{2\Delta^{t}}}}, \tag{A.6}$$

$$M_{i\to a}^{t} = \frac{\int \mathrm{d}h (2\cosh(\beta h))^{\frac{x}{\beta}}\, \mathrm{e}^{-\frac{(h-\phi_{i\to a}^{t})^2}{2\Delta^{t}}}\tanh^2(\beta h)}{\int \mathrm{d}h (2\cosh(\beta h))^{\frac{x}{\beta}}\, \mathrm{e}^{-\frac{(h-\phi_{i\to a}^{t})^2}{2\Delta^{t}}}}. \tag{A.7}$$

Equations (A.4)–(A.7) constitute the simpler expression of SPx in the dense connectivity limit. Utilizing $m_{i\to a}^{t}$ and Δ^t, the estimate for the decoded message at the tth update is provided as

$$\sigma_i = \mathrm{sign}\left(\int \mathrm{d}H\, P_i(H)\tanh(\beta H)\right), \tag{A.8}$$

$$P_i^t(H) = \frac{(2\cosh(\beta H))^{\frac{x}{\beta}}\, \mathrm{e}^{-\frac{(H-\phi_i^t)^2}{2\Delta^t}}}{\int \mathrm{d}H (2\cosh(\beta H))^{\frac{x}{\beta}}\, \mathrm{e}^{-\frac{(H-\phi_i^t)^2}{2\Delta^t}}}, \tag{A.9}$$

where $\phi_i^t = \sum_{a\in\mathcal{N}(i)} J_a m_{j\to a}^{t-1}$.

A.2. Macroscopic dynamics

The performance of the SPx decoding algorithm (A.4)–(A.7) can be examined by following time evolution of several relevant macroscopic variables [26–28]. The maximum entropy principle, in conjunction of the law of large numbers, makes it possible to regard the microscopic variable $\phi^t_{i\to a}$ as Gaussian random numbers, the average and variance of which are provided independently of site and check indices i, a as E^t and F^t, respectively. We assume that the gauge transformation $\sigma_i \to \sigma_i \xi_i$ and $J_a \to J_a \prod_{i\in a} \xi_i$ is already carried out, which implies that the original message is specified by $(1, 1, \ldots, 1)$. These indicate that Q^t_1, Q^t_0 and the macroscopic overlap to the original message $m^t = (NC)^{-1} \sum_{i,a\in\mathcal{N}(i)} m^t_{i\to a}$ are represented by

$$Q^t_1 = \int Dv \frac{\int Du (2\cosh(\beta\Phi^t(u,v)))^{\frac{x}{\beta}} \tanh^2(\beta\Phi^t(u,v))}{\int Du (2\cosh\beta\Phi^t(u,v))^{\frac{x}{\beta}}}, \tag{A.10}$$

$$Q^t_0 = \int Dv \left(\frac{\int Du (2\cosh(\beta\Phi^t(u,v)))^{\frac{x}{\beta}} \tanh(\beta\Phi^t(u,v))}{\int Du (2\cosh(\beta\Phi^t(u,v)))^{\frac{x}{\beta}}} \right)^2, \tag{A.11}$$

$$m^t = \int Dv \frac{\int Du (2\cosh(\beta\Phi^t(u,v)))^{\frac{x}{\beta}} \tanh(\beta\Phi^t(u,v))}{\int Du (2\cosh(\beta\Phi^t(u,v)))^{\frac{x}{\beta}}}, \tag{A.12}$$

respectively, where $Dz = \frac{\mathrm{d}z\, \mathrm{e}^{-z^2/2}}{\sqrt{2\pi}}$ and $\Phi^t(u,v) = \sqrt{\Delta^t}u + \sqrt{F^t}v + E^t$. On the other hand, the self-averaging property [26, 29] and equations (A.10), (A.11) and (A.12) imply that F^{t+1} and E^{t+1} are provided by

$$\begin{aligned} F^{t+1} &= N^{-1} \sum_{i,a\in\mathcal{N}(i)} \left(\overline{\left(J_a m^t_{j\to a}\right)^2} - \overline{\left(J_a m^t_{j\to a}\right)}^2 \right) \\ &\simeq N^{-1} \sum_{i,a\in\mathcal{N}(i)} \overline{\left(J_a m^t_{j\to a}\right)^2} \simeq J^2 Q^t_0, \end{aligned} \tag{A.13}$$

$$E^{t+1} = N^{-1} \sum_{i,a\in\mathcal{N}(i)} \overline{J_a m^t_{j\to a}} \simeq J_0 m^t, \tag{A.14}$$

respectively, where $\overline{\cdots}$ denotes the configuration average with respect to the channel noise and lattice configuration. The update rule for Δ^{t+1} is already provided by equation (A.5). Using these macroscopic variables, the error probability per bit for SPx, $P_b(x)$, is calculated from

$$\begin{aligned} P_b(x) &= \int Dv \Theta \left(-\frac{\int Du (2\cosh(\beta\Phi^t(u,v)))^{\frac{x}{\beta}} \tanh(\beta\Phi^t(u,v))}{\int Du (2\cosh(\beta\Phi^t(u,v)))^{\frac{x}{\beta}}} \right) \\ &= \int_{-\infty}^{-\frac{E^t}{\sqrt{F^t}}} Dv = \mathrm{erfc}\left(\frac{E^t}{\sqrt{F^t}} \right), \end{aligned} \tag{A.15}$$

where $\Theta(z) = 1$ and 0 for $z > 0$ and $z < 0$, respectively. It may be remarked that equations (A.5) and (A.10)–(A.14) accord with the macroscopic description of the SPx dynamics for the SK model [30].

A.3. Comparison between SP0 and SPβ(BP)

Now, we focus on two specific Parisi parameters $x = 0$ and $x = \beta$. A distinctive property of these two parameter choices is that only four out of the six variables are relevant in equations (A.5) and (A.10)–(A.14), which can be reduced to an identical set of four equations

$$G^{t+1} = J^2 Q^t, \tag{A.16}$$

$$E^{t+1} = J_0 m^t, \tag{A.17}$$

$$Q^t = \int Dz \tanh^2(\beta(\sqrt{G^t} z + E^t)), \tag{A.18}$$

$$m^t = \int Dz \tanh(\beta(\sqrt{G^t} z + E^t)), \tag{A.19}$$

where $G^t = \Delta^t + F^t$, $Q^t = Q_1^t$ for $x = 0$ while $G^t = F^t$, $Q^t = Q_0^t$ for $x = \beta$. As Δ^t is a non-negative variable, this implies that F^t, which indicates the variance of the local fields, for $x = \beta$ cannot be smaller than that for $x = 0$ while E^t, which represents the signal strength of the original message in the local fields, is identical in both cases. Applying this result to equation (A.15) leads to a conclusion that SP0 does not provide worse performance than SPβ.

In terms of the decoding problem, SPβ and BP provide identical performance. Therefore, the above argument indicates that SP0 is never overcome by BP with respect to the error correction performance. In order to show this, we characterize the auxiliary distributions $P_{i\to a}^t(h)$ and $Q_{a\to i}^t(u)$ utilizing the first moments of $\tanh(\beta h)$ and $\tanh(\beta u)$, respectively, as

$$H_{i\to a}^t = \frac{1}{\beta} \tanh^{-1}\left(\int \mathrm{d}h \, P_{i\to a}^t(h) \tanh(\beta h) \right), \tag{A.20}$$

$$U_{a\to i}^t = \frac{1}{\beta} \tanh^{-1}\left(\int \mathrm{d}h \, Q_{a\to i}^t(u) \tanh(\beta u) \right). \tag{A.21}$$

Although higher moments are necessary for completely describing general SPx dynamics, equations (A.1) and (A.2) indicate that for the specific choice of $x = \beta$, $H_{i\to a}^t$ and $U_{a\to i}^t$ constitute a closed update rule as

$$H_{i\to a}^{t+1} = \sum_{b\in\mathcal{N}(i)\backslash a} U_{b\to i}^t, \tag{A.22}$$

$$U_{a\to i}^t = u\big(H_{j\to a}^t, J_a; \beta\big), \tag{A.23}$$

which is nothing but the BP dynamics of finite temperature $T = \beta^{-1}$ and are reduced to equations (2) and (3) in the zero temperature limit $\beta \to \infty$. This implies that as long as only the first moment is concerned, which is the case for the current decoding problem, SPβ and BP are equivalent.

References

[1] Sourlas N 1989 *Nature* **339** 693
[2] Sourlas N 1994 *Europhys. Lett.* **25** 159
[3] Ruján P 1993 *Phys. Rev. Lett.* **70** 2968
[4] Kabashima Y and Saad D 1999 *Europhys. Lett.* **45** 97–103
[5] Kanter I and Saad D 2000 *Phys. Rev.* E **61** 2137-40

[6] Kanter I and Saad D 1999 *Phys. Rev. Lett.* **83** 2660
[7] Kanter I and Saad D 2000 *J. Phys. A: Math. Gen.* **33** 1675
[8] Pearl J 1988 *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (San Francisco: Morgan Kaufmann)
[9] Yedida J S, Freeman W T and Weiss Y 2001 *Advances in Neural Information Processing Systems* vol 13, ed T K Leen, T G Dietterich and V Treps (Cambridge, MA: MIT Press) p 689
[10] Kabashima Y and Saad D 1998 *Eur. Phys. Lett.* **44** 668
[11] Mézard M, Parisi G and Zecchina R 2002 *Science* **297** 812
[12] Mézard M and Parisi G 2001 *Eur. Phys. J.* B **20** 217
[13] Mézard M and Parisi G 2003 *J. Stat. Phys.* **111** 1
[14] Mézard M and Zecchina R 2002 *Phys. Rev.* E **66** 056126
[15] Wemmenhove B and Kappen H J 2005 *J. Phys. A: Math. Gen.* **39** 1265–83
[16] van Mourik J 2006 in preparation
[17] Bounkong S, van Mourik J and Saad D 2005 *Preprint* cond-mat 0507579
[18] de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983
[19] Castellani T, Krzakala F and Ricci-Tersenghi F 2005 *Eur. Phys. J.* B **47** 99–108
[20] Kabashima Y 2005 *Preprint* cs.IT/0506062
[21] Kabashima Y 2006 in preparation
[22] Franz S, Mézard M, Ricci-Tersenghi F, Weigt M and Zecchina R 2001 *Europhys. Lett.* **55** 465
[23] Kwon C and Thouless D J 1988 *Phys. Rev.* B **37** 7649
[24] Montanari A, Parisi G and Ricci-Tersenghi F 2004 *J. Phys. A: Math. Gen.* **37** 2073–91
[25] Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792–6
[26] Kabashima Y 2003 *J. Phys. Soc. Jpn.* **72** 1645–9
[27] Kabashima Y 2003 *J. Phys. A: Math. Gen.* **36** 11111–21
[28] Richardson T and Urbanke R 2001 *IEEE Trans. IT* **47** 599–618
[29] Mezard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[30] Kabashima Y 2005 *J. Phys. Soc. Jpn.* **74** 2133–6