Asymptotic Mean-Square Optimality of Belief Propagation for Sparse Linear Systems

Dongning Guo
Dept. of Electrical Engineering & Computer Science
Northwestern University, Evanston, IL 60208, USA

Chih-Chun Wang
School of Electrical & Computer Engineering
Purdue University, West Lafayette, IN 47907, USA

***Abstract*— This paper studies the estimation of a high-dimensional vector signal where the observation is a known "sparse" linear transformation of the signal corrupted by additive Gaussian noise. A paradigm of such a linear system is code-division multiple access (CDMA) channel with sparse spreading matrix. Assuming a "semi-regular" ensemble of sparse matrix linear transformations, where the bi-partite graph describing the system is asymptotically cycle-free, it is shown that belief propagation (BP) achieves the minimum mean-square error (MMSE) in estimating the transformation of the input vector in the large-system limit. The result holds regardless of the the distribution and power of the input symbols. Furthermore, the mean squared error of estimating each symbol of the input vector using BP is proved to be equal to the MMSE of estimating the same symbol through a scalar Gaussian channel with some degradation in the signal-to-noise ratio (SNR). The degradation, called the efficiency, is determined from a fixed-point equation due to Guo and Verdú, which is a generalization of Tanaka's formula to arbitrary prior distributions.**

I. Introduction

Consider the estimation of a vector signal where the observation is a known linear transformation of the signal which is subsequently corrupted by Gaussian noise. The model is widely used in communications, control and signal processing, and has been well studied, especially in the context of code-division multiple access (CDMA) [1].

If the input is Gaussian distributed, the optimal estimator in mean-square sense is linear. In fact the linear minimum mean-square error (MMSE) estimator is often used in case of non-Gaussian inputs due to its simplicity, even though it is then suboptimal. The corresponding mean squared error (MSE) depends only on the second-order statistics of the input, and can be easily computed if the system size is small. In case of a large randomly generated linear transformation, the MSE can be obtained using random matrix theory, the central dictate of which is that the empirical distribution of the singular values of the random linear transformation converges to a deterministic law in the large-system limit (e.g., [1]–[5]).

For general (non-Gaussian) inputs, the performance evaluation of "optimal" detection entails infeasible exponential complexity in the system dimension. Random matrix theory is not applicable because the performance cannot be expressed in the singular values of the linear transformation. A breakthrough in large-system performance analysis was made by Tanaka using statistical physics techniques, where the minimum error probability achieved by optimal maximum *a posteriori* probability (MAP) detection in the case of equal-power binary inputs was obtained using the replica method [6]. The result has been generalized by Guo and Verdú to arbitrary inputs and a family of suboptimal detectors in [7], where it is found that the linear system with optimal detection is equivalent to a bank of scalar Gaussian channels with degradation in the signal-to-noise ratio (SNR). This degradation, known as the multiuser efficiency, is determined by a fixed-point equation [7], which is a generalization of Tanaka's formula in [6]. Unfortunately, the replica method has not been fully justified mathematically. Hence the results in [6] and [7] are subject to doubt, although they lead to good numerical results.

A recent paper [8] by Montanari and Tse is the first attempt to justify Tanaka's result in the special case of "sparse" spreading matrix with moderate load (less than 1.49). The proof outlined in [8] suggests that belief propagation (BP) achieves optimal performance in some large-system limit, and that the fixed-point equation describing the performance of BP is identical to Tanaka's formula, which is believed to be satisfied by the (optimal) MAP detector [6].

This work generalizes the results of [8] to *arbitrary* prior input distributions and powers. Extending [8], we propose a BP detector for non-binary (possibly continuous) inputs, which assumes Gaussian interference in each node. For an ensemble of large *sparse* linear systems, it is found that density evolution leads to Guo and Verdú's formula for the multiuser efficiency [7]. Unique to this paper is the characterization of the single-input multiple-output channel for each individual user using BP detection (i.e., the subtree with the user as the root) as an equivalent scalar Gaussian channel. It is shown that BP using the (MMSE-based) Gaussian approximation of interference suffers no loss in contrast to other Gaussian approximations based on the message-mean [9], error probability [10], and extrinsic information [11]. Interestingly, the Gaussian approximation leads to the update equation of the parallel interference canceler suggested in [12] as a specious interpretation of the statistical physics results; indeed the puzzle in [12] is solved.

A key result in this work is the relationship between the MMSE of estimating the linear transformation of a random vector observed in Gaussian noise and the MMSE of a scalar random variable in Gaussian noise. The fixed-point equation of Guo and Verdú is rigorously proved under the sparse spreading assumption. As is also seen in [8], the proof hinges on the fundamental MMSE-mutual information relationship in Gaussian channels [13]. We note also that the special case of Gaussian inputs has been studied in [14], where the Tse-Hanly formula [2] is proved without using random matrix theory.

1-4244-0067-8/06/$20.00 ©2006 IEEE.

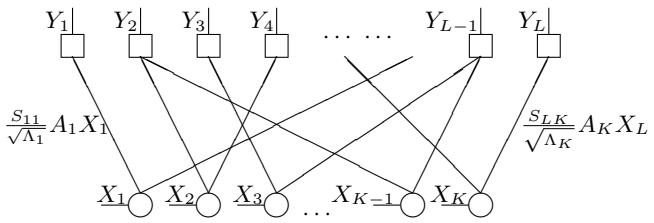

Fig. 1. Factor graph for the sparse CDMA system

The remainder of this paper is organized as follows. Section II introduces the sparse linear system, BP and the MMSE transform. The main results are summarized in Section III. Sections IV and V prove the main results. All the optimality results of BP discussed herein are in the MSE sense, which will be further strengthened in a follow-up paper [15].

II. System Model

Consider the linear system described by

$$\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{X} + \boldsymbol{N} \tag{1}$$

where $\boldsymbol{X} = [X_1, \ldots, X_K]^\top$ denotes the input K-vector, $\boldsymbol{S}$ is an $L \times K$ matrix that represents a known linear transformation, and $\boldsymbol{N} \sim \mathcal{N}(0, \boldsymbol{I})$ consists of independent standard Gaussian random variables. The system (1) describes in general a multi-input multi-output channel such as in multi-antenna or orthogonal frequency-division multiplexing (OFDM) systems.

An important application of the model is to describe a fully-synchronous K-user CDMA system with spreading factor L, where user k modulates symbol X_k onto a spreading sequence $\boldsymbol{s}_k$ with amplitude $A_k > 0$. This paper mainly uses the CDMA terminologies. The symbols X_k are assumed to be independently identically distributed (i.i.d.) with distribution (probability measure) P_X, which has zero mean and finite variance. Let the spreading sequence of user k be described by $\boldsymbol{s}_k = \frac{1}{\sqrt{\Lambda_k}}[S_{1k}, S_{2k}, \ldots, S_{Lk}]^\top$ where $1/\sqrt{\Lambda_k}$ is a normalization factor. The chip-wise representation of the model,

$$Y_l = \sum_{k=1}^{K} \frac{S_{lk}}{\sqrt{\Lambda_k}} A_k X_k + N_l, \quad l \in \{1, 2, \ldots, L\} \tag{2}$$

is equivalent to (1) with $\boldsymbol{S} = [A_1\boldsymbol{s}_1, \ldots, A_K\boldsymbol{s}_K]$. A bipartite factor graph of the system is illustrated in Fig. 1, where symbol X_k and chip Y_l are connected by an edge if $S_{lk} \neq 0$.

A. The Ensemble of Linear Transformations

The random transformation $\boldsymbol{S}$ is constructed as follows. First, a binary incidence matrix $\boldsymbol{H}_{L\times K} = (h_{lk})$ is "randomly" generated. For all (l, k) with $h_{lk} = 0$, set $S_{lk} = 0$. For all (l, k) with $h_{lk} = 1$, S_{lk} are i.i.d. with distribution P_S, which has zero mean and unit variance. The normalization factor for each sequence $\boldsymbol{s}_k$ is defined by $1/\sqrt{\Lambda_k}$, in which $\Lambda_k = \sum_{l=1}^{L} h_{lk}$ is the symbol degree of X_k as depicted in Fig. 1.

The *large-system limit* is defined as $K, L \to \infty$ with the system load K/L converging to a positive number β. Let $\Gamma_l = \sum_{k=1}^{K} h_{lk}$ denote the chip degree of Y_l and $\bar{\Gamma} = \frac{1}{L}\sum_{l=1}^{L}\Gamma_l$ denote the average chip degree. We call an ensemble of random spreading matrices *chip-semi-regular* if it satisfies

$$\lim_{\bar{\Gamma}\to\infty} \lim_{K=\beta L\to\infty} \mathsf{P}\left\{|\Gamma_l - \bar{\Gamma}| > \epsilon\bar{\Gamma}\right\} = 0, \quad \forall\, \epsilon > 0, \forall l, \tag{3}$$

i.e., that the chip degrees of all nodes concentrate to their average in probability. This paper focuses on the *large-sparse-system limit*, where we take $K, L \to \infty$ first and $\bar{\Gamma} \to \infty$ afterwards. The results in this paper apply to all chip-semi-regular ensembles, including the following special cases:

1) The classic ensembles with regular chip degrees and regular or irregular symbol degrees [16].
2) The symbol-irregular chip-Poisson ensembles in which $\Lambda_1, \ldots, \Lambda_K$ are i.i.d. with distribution P_Λ. For every k, X_k is connected to Λ_k uniformly randomly selected chip nodes. The chip degree is asymptotically Poisson, which satisfies (3). This ensemble is assumed in [8].
3) The doubly Poisson ensembles (Ensemble G in [17]).

We assume the amplitudes A_k to be i.i.d. with distribution P_A, which has finite moments of any order. Clearly, as $K \to \infty$, the empirical distribution of $\{A_k\}$ converges to P_A, which can be understood as the received amplitude profile. Thus flat fading is incorporated in the model. If (1) describes an OFDM channel, the results in this paper can be generalized effortlessly to frequency-selective fading (see [18]).

B. The BP Algorithm

Consider the bipartite graph (Fig. 1) that describes the linear system. An iterative BP estimator can be devised based on the graph, which essentially updates the posterior distribution for each X_k conditioned on the observations within the reach of the local subtree with X_k as the root [19], [20]. The BP algorithm is best described by its corresponding "message" maps (or updates) at the symbol and the chip nodes.

For every (l, k) with $S_{lk} \neq 0$, let $V_{k\to l}^{(t)}$ denote the message from symbol node k to chip node l and $U_{l\to k}^{(t)}$ denotes the message in the reverse direction at iteration t. Any sufficient statistic for X_k can be used as the "message," while the most common choice for binary symbols is the log-likelihood ratios (LLR) of X_k given the corresponding observations. The update equations at the t-th ($t > 0$) iteration are:

$$U_{l\to k}^{(t)} = \log \frac{\mathsf{P}\{X_k = +1 | Y_l, \boldsymbol{S}, A_k, \{A_i, V_{i\to l}^{(t-1)}\}_{i\in\partial l\setminus k}\}}{\mathsf{P}\{X_k = -1 | Y_l, \boldsymbol{S}, A_k, \{A_i, V_{i\to l}^{(t-1)}\}_{i\in\partial l\setminus k}\}} \tag{4}$$

$$V_{k\to l}^{(t)} = V_{k\to l}^{(0)} + \sum_{j\in\partial k\setminus l} U_{j\to k}^{(t)} \tag{5}$$

where $\partial l = \{i \,|\, S_{li} \neq 0\}$, and with slight abuse of notation, $\partial k = \{j \,|\, S_{jk} \neq 0\}$. Here, the posterior probability $\mathsf{P}\{X_k = x \,|\, Y_l, \boldsymbol{S}, A_k, \{A_i, V_{i\to l}^{(t-1)}\}_{i\in\partial l\setminus k}\}$ is defined as the probability of $X_k = x$ given Y_l, $\boldsymbol{S}$, $\{A_i\}_{i\in\partial l}$ and the extrinsic LLRs

of $\{X_i\}_{i\in\partial l\backslash k}$, which is described in $\{V_{i\to l}^{(t-1)}\}_{i\in\partial l\backslash k}$. The initial message $V_{k\to l}^{(0)}$ is determined by the prior distribution P_X. After the final iteration ($t=\tau$), the decision at each variable node k is made according to the following LLR

$$V_k^{(\tau)} = V_{k\to l}^{(0)} + \sum_{l\in\partial k} U_{l\to k}^{(\tau)}.$$

The BP algorithm can be extended to non-binary symbols by using the LLR with respect to a reference symbol or the posterior distribution in lieu of the LLR (more on this in Section IV).

C. The MMSE Transform

The MMSE is pivotal to this work. In general, we use

$$\mathcal{E}\left(\boldsymbol{Z}\,|\boldsymbol{W}\right) = \frac{1}{K}\mathsf{E}\left\{\|\boldsymbol{Z} - \mathsf{E}\left\{\boldsymbol{Z}\mid\boldsymbol{W}\right\}\|^2\right\} \tag{6}$$

to denote the average MMSE per dimension of estimating an arbitrary K-dimensional random vector $\boldsymbol{Z}$ from any observation(s) $\boldsymbol{W}$, where the expectation is taken over the joint distribution $P_{\boldsymbol{ZW}}$. It is straightforward to generalize the definition to the case in which $\boldsymbol{W}$ is a collection of random variables taking values in any (abstract) space.

One special scenario of estimating a scalar product AX from its scalar observation $\sqrt{\gamma}\,AX+N$ with the side information A is of particular importance. We define the following MMSE transform based on (6):

$$\mathcal{E}_{X|A}\left(\gamma\right) = \mathcal{E}\left(AX\,|\sqrt{\gamma}AX+N, A\right)$$

where the symbol $X\sim P_X$ is unknown, the independent amplitude $A\sim P_A$ is known to the estimator, and the observation is corrupted by an independent standard Gaussian noise N. The MMSE transform $\mathcal{E}_{X|A}\left(\gamma\right)$ depends on P_X and P_A, and is a decreasing function of γ for given $P_X\times P_A$.

III. Main Results

Consider the following fixed-point equation:

$$\eta = \frac{1}{1+\beta\,\mathcal{E}_{X|A}\left(\gamma\eta\right)} \tag{7}$$

where the fixed point η is a function of γ. Note that (7) has at least one solution for every $\gamma\geq 0$ and the solution η is unique if β is sufficiently small, which can be easily proved using elementary calculus. Throughout this paper, it is assumed that β is such that (7) has a unique solution for every $\gamma\geq 0$. This solution is referred to as the *power efficiency* or simply *efficiency* of the linear system for reasons to be clear shortly.

Theorem 1: Consider the chip-semi-regular ensemble of linear systems with the symbol and amplitude distributions P_X and P_A. Let β be such that the solution to (7) is unique for every $\gamma\geq 0$. Then

$$\lim_{\bar{\Gamma}\to\infty}\lim_{K=\beta L\to\infty}\mathcal{E}\left(\boldsymbol{SX}\,|\boldsymbol{SX}+\boldsymbol{N},\boldsymbol{S}\right) = \beta\,\eta\,\mathcal{E}_{X|A}\left(\eta\right) = 1-\eta$$

where the efficiency η satisfies (7) with $\gamma=1$, i.e.,

$$\eta = \frac{1}{1+\beta\,\mathcal{E}_{X|A}\left(\eta\right)}. \tag{8}$$

In general, the MMSE of estimating $\boldsymbol{X}$ in (1) is dependent on the linear transformation $\boldsymbol{S}$. Theorem 1 states that, as the system size becomes large, not only the dependence on $\boldsymbol{S}$ diminishes, but the large-sparse-system MMSE can be expressed using the MMSE for a scalar Gaussian channel which is straightforward to compute. The efficiency is easy to determine from the fixed-point equation (8). The equation was first obtained in its general form in [7] as a generalization of Tanaka's formula [6].

Let $\hat{X}_k^{(t)}$ denote the conditional mean estimate of X_k obtained using the BP algorithm after t iterations.

Theorem 2: Assuming the same settings as in Theorem 1, BP achieves the MMSE as the number of iterations increases:

$$\lim_{t\to\infty}\lim_{\bar{\Gamma}\to\infty}\lim_{K=\beta L\to\infty}\mathsf{E}\left\{\left\|\boldsymbol{S}\left(\boldsymbol{X}-\hat{\boldsymbol{X}}^{(t)}\right)\right\|^2\right\} = 1-\eta.$$

Together with Theorem 1, Theorem 2 establishes the asymptotic mean-square optimality of BP for large sparse linear systems in terms of estimating $\boldsymbol{SX}$, generalizing the corresponding statements for binary symbols [8] to the case of non-binary symbols with arbitrary user power profiles. The most straightforward approach to proving Theorem 2 relies on the density evolution method, assuming the *density* of the messages of BP to be Gaussian. See equations (4) and (5) and [9], [16] for more references.

Suppose the minimum symbol node degree $\min_k\Lambda_k$ approaches infinity as $\bar{\Gamma}\to\infty$. Since the LLR messages entering the symbol node are independent and of diminishing mean and variance, the central limit theorem guarantees the asymptotic normality of the outgoing symbol-to-chip messages in (5), as first pointed out in [8]. However, the scaling law of the mean and variance of the incoming chip-to-symbol non-Gaussian messages cannot be easily determined due to the competition of convergence speeds. Instead of making some fallible Gaussian assumption, we modify the message *maps* of BP and introduce a (sub-optimal) relaxed BP. The benefits of the relaxed BP are two-fold: First, for non-binary symbols, the relaxation leads to much simpler message maps and more insights; secondly, many classic results on degraded channels can be applied naturally to the relaxed BP. This paper shows that the relaxed BP achieves the MMSE of estimating the linear transformation of the input in the large-sparse-system limit, thereby also validates the MSE optimality of BP by sandwiching arguments.

As another major contribution of this work, we prove rigorously for the first time the equivalence between the detection over a degraded scalar Gaussian channel and the detection over a vector Gaussian noise in the large-sparse-system limit. Let $\Pi_{X_k|\boldsymbol{Y}}^{(t)}$ denote the posterior distribution of X_k computed by the relaxed BP after t iterations, which itself is a random distribution dependent on the observed random vector $\boldsymbol{Y}$. This paper shows that $\Pi_{X_k|\boldsymbol{Y}}^{(t)}$ converges weakly to the posterior distribution of the input of a scalar Gaussian channel. Precisely, consider the following Gaussian channel:

$$Y' = aX' + N'/\sqrt{\eta}$$

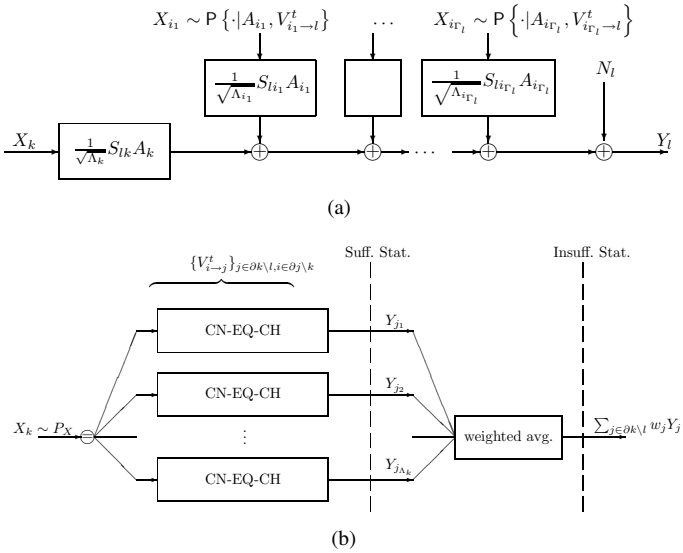

Fig. 2. Equivalent channels for the relaxed BP. (a) The chip node equivalent channel (CN-EQ-CH). (b) The symbol and chip node perspective.

where $X' \sim P_X$, $N' \sim \mathcal{N}(0,1)$, and $\eta \in (0,1]$. Let $\Pi_{X'|Y'}$ denote the (random) posterior distribution of X'.

Theorem 3: Given $X_k = x$ and $A_k = a$, the conditional distribution $\Pi^{(t)}_{X_k|\boldsymbol{Y}}$, obtained by the BP algorithm after t iterations, converges weakly in the large-sparse-system limit to the conditional distribution of $\Pi_{X'|Y'}$ given $X' = x$, where the efficiency η is the solution to (8).

Theorem 3 points out that the statistics of the relaxed BP estimate for each input symbol X_k is asymptotically identical to that of estimating the same input through a scalar Gaussian channel, the SNR of which is degraded by a factor of η.

By the channel degradation argument, Theorem 3, describing the behavior of the relaxed BP, can serve also as a lower bound on the achievable MSE performance of estimating X_k by the classic BP algorithm and by the MMSE detector. Proving the tightness of this lower bound involves establishing a stronger asymptotic equivalence between the relaxed BP and the classic BP and is deferred to a follow-up of this work [15].

IV. The Relaxed BP

Using clever heuristics, efficient BP algorithms have been proposed in [21]–[23], which invariably take some interference cancellation structure. The performance can usually be obtained through solving a set of coupled equations (e.g., [22]). In contrast to these existing results, we introduce a relaxed BP for non-binary symbols, which is motivated by the equivalent channel perspective in [7], [24].

Let us use the posterior distributions instead of the LLRs as messages. The iterative BP estimator updates the posterior distribution for each symbol conditioned on the observations within the reach of the local subtree with the symbol as its root. The chip node message map, analogous to (4), can be obtained by solving the posterior distribution of the detection problem in Fig. 2(a). Similarly, the symbol node message map corresponding to (5) can be derived by considering a repetition channel. Combining both the symbol and the chip node maps, the $V^{(t-1)}_{k\to l} \mapsto V^{(t)}_{k\to l}$ map can be obtained by solving the posterior distribution of the detection problem in Fig. 2(b) if the sufficient statistics $\{Y_j\}_{j\in\partial k\setminus l}$ are employed. The relaxed BP is best explained by solving the same detection problem in Fig. 2(b) with the following insufficient statistic, a weighted sum:

$$\sum_{j\in\partial k\setminus l} \frac{1}{\sqrt{\Lambda_k}} S_{jk} Y_j. \tag{9}$$

Consider the asymptotic performance of the relaxed BP in the large-sparse-system limit: $\lim_{\bar{\Gamma}\to\infty}\lim_{K=\beta L\to\infty}$. Since the perfect projection condition holds naturally for all linear systems described by (1), the generalized density evolution analysis [25] can be applied to the relaxed BP even though one cannot assume the all-one vector $\boldsymbol{X}$ being transmitted when non-binary symbols are considered.

With the near-optimal weight selection in (9), after t iterations of the relaxed BP, the generalized density evolution shows that each user is facing a scalar Gaussian channel

$$X_k \mapsto A_k X_k + N/\sqrt{\eta^{(t)}} \tag{10}$$

where $N \sim \mathcal{N}(0,1)$ and $\eta^{(t)} \in (0,1]$ is the corresponding channel degradation coefficient. With the help of the chip-semi-regularity condition, $\eta^{(t)}$ can be computed by the following simple iterative update equation:

$$\eta^{(t+1)} = \frac{1}{1+\beta\mathcal{E}_{X|A}\left(\eta^{(t)}\right)}. \tag{11}$$

Removing the chip-semi-regularity condition would result in a more involved denominator in (11).

Noticeably, the noise for all users is of *identical* power $1/\eta^{(t)}$, and the performance of user k depends only on A_k but not on the effective spreading length Λ_k and the spreading sequence $\boldsymbol{s}_k$. Denote the limit of $\eta^{(t)}$ by η, the fixed-point equation of Guo and Verdú (8) is thus obtained. In view of the equivalent single-user channel (10), Theorem 3 is proved.[1]

Not surprisingly, (11) is also the update equation for the parallel interference canceler with the conditional mean as the soft decision function. In fact (11) was noted as a specious interpretation of the fixed-point equation (7) in [12], which is now justified in the large-sparse-system limit.

The simple expression in (11) also requires the individual normalization factor $1/\sqrt{\Lambda_k}$ for each user X_k in the linear system model (2). If a global normalization $1/\sqrt{\bar{\Lambda}}$ factor is used instead, where $\bar{\Lambda} = \frac{1}{K}\sum_k \Lambda_k$, similar analysis can be performed and a different, more involved iterative update equation will be obtained.

V. Optimality of Belief Propagation

In the following we justify Theorems 1–3 assuming that the inputs to the linear system model (1) are discrete with

[1]In this paper, the symbol degree Λ_k is allowed to be bounded away from infinity when $\bar{\Gamma} \to \infty$. In this case, different realization of S_{lk} may changes the effective power of transmitting X_k. Theorem 3 still holds if we consider the actual power $A^2_{k,\text{actual}} = A^2_k \sum_l S^2_{lk}/\Lambda_k$ and the corresponding perfectly normalized spreading sequence $\boldsymbol{s}_{k,\text{actual}} \propto \boldsymbol{s}_k$ satisfying $\frac{1}{\Lambda_k}\sum_l S^2_{lk,\text{actual}} = 1$.

finite entropy. We omit the proof for the case with general inputs.

The following result is the key to the optimality of BP.

Lemma 1 ([13]): For any K-vector $\boldsymbol{Z}$ with $\mathsf{E}\|\boldsymbol{Z}\|^2 < \infty$ and independent $\boldsymbol{N} \sim \mathcal{N}(0, \boldsymbol{I})$ of identical dimension,

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} I(\boldsymbol{Z}; \sqrt{\gamma}\,\boldsymbol{Z} + \boldsymbol{N}) = \frac{K}{2} \mathcal{E}\left(\boldsymbol{Z} \,|\sqrt{\gamma}\,\boldsymbol{Z} + \boldsymbol{N}\right), \quad \gamma \geq 0.$$

The following entropy–MMSE relationship is straightforward,

$$\frac{1}{2}\int_0^\infty \mathcal{E}_{\boldsymbol{S}\boldsymbol{X}|\boldsymbol{S}}(\gamma)\,\mathrm{d}\gamma = \frac{1}{L} H(\boldsymbol{S}\boldsymbol{X}|\boldsymbol{S}). \tag{12}$$

In the large-sparse-system limit, the right hand side of (12) converges to $H(\boldsymbol{X})/L = \beta\, H(X)$ because $\boldsymbol{X}$ can be recovered from $\boldsymbol{S}\boldsymbol{X}$ with probability 1.

It can be proved that the MSE of estimating $\sum_k \frac{S_{lk}}{\sqrt{\Lambda_k}} A_k X_k$ by the relaxed BP converges to $\beta\,\eta\,\mathcal{E}_{X|A}(\eta)$ for all l. In the following, we show that the MSE $\beta\,\eta\,\mathcal{E}_{X|A}(\eta)$ integrates to the same entropy as in (12).

For every $\gamma \geq 0$, let η be the solution to the fixed-point equation (7). Define

$$C(\gamma) = \mathsf{E}\{I(X; \sqrt{\gamma\eta}\,AX + N|A)\} + \frac{\eta - 1 - \log(\eta)}{2\beta}.$$

Theorem 4: For every $\gamma \geq 0$,

$$C(\gamma) = \frac{1}{2}\int_0^\gamma \eta\,\mathcal{E}_{X|A}(\gamma\eta)\;\mathrm{d}\gamma.$$

Proof: Since $C(0) = 0$, $\eta = 1$ when $\gamma = 0$, and η is a differentiable function with respect to γ, showing

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} C(\gamma) = \frac{\eta}{2}\,\mathcal{E}_{X|A}(\eta\gamma) \tag{13}$$

is sufficient. Using Lemma 1, we obtain (13) because

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\gamma} C(\gamma) \\
&= \mathsf{E}\left\{\frac{\mathrm{d}}{\mathrm{d}\gamma} I(X; \sqrt{\gamma\eta}\,AX + N|A)\right\} + \frac{1}{2\beta}\frac{\mathrm{d}}{\mathrm{d}\gamma}(\eta - 1 - \log\eta) \\
&= \frac{1}{2}\,\mathcal{E}_{X|A}(\gamma\eta)\,\frac{\mathrm{d}}{\mathrm{d}\gamma}(\gamma\eta) + \frac{1}{2\beta}(1 - \eta^{-1})\eta' \\
&= \frac{\eta}{2}\,\mathcal{E}_{X|A}(\eta\gamma) + \frac{1}{2\beta}\left[\beta\gamma\,\mathcal{E}_{X|A}(\eta\gamma) + 1 - \eta^{-1}\right]\eta'
\end{aligned} \tag{14}$$

where the last term in (14) vanishes by (7). ∎

By Theorem 4,

$$\frac{1}{2}\int_0^\infty \beta\eta\mathcal{E}_{X|A}(\gamma\eta)\;\mathrm{d}\gamma = \beta\, C(\infty) = \beta\, H(X). \tag{15}$$

From (12) and (15), it is clear that the MMSE of the linear system and the MSE achieved by the relaxed BP integrate to the same entropy. Since the MSE is lower bounded by the MMSE for every SNR, they must be equal for all SNR. In other words, in the large-sparse-system limit, the relaxed BP detector approaches the MAP detector and achieves the MMSE. By the channel degradation and the sandwiching arguments, Theorems 1 and 2 are thus established.

REFERENCES

[1] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.

[2] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inform. Theory*, vol. 45, pp. 641–657, Mar. 1999.

[3] D. Guo, S. Verdú, and L. K. Rasmussen, "Asymptotic normality of linear multiuser receiver outputs," *IEEE Trans. Inform. Theory*, vol. 48, pp. 3080–3095, Dec. 2002.

[4] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[5] R. Müller, "Random matrix methods for design of multiuser communication systems," *Acta Physica Polonica B*, vol. 36, pp. 2733–2746, Sept. 2005.

[6] T. Tanaka, "A statistical mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2888–2910, Nov. 2002.

[7] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1982–2010, June 2005.

[8] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *Proc. IEEE Information Theory Workshop*, Punta del Este, Uruguay, 2006.

[9] S. Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 657–670, Feb. 2001.

[10] F. Lehmann and G. M. Maggio, "Analysis of the iterative decoding of LDPC and product codes using the Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2993–3000, Nov. 2003.

[11] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, pp. 1727–1737, Oct. 2001.

[12] D. Guo, *Gaussian Channels: Information, Estimation and Multiuser Detection*. PhD thesis, Department of Electrical Engineering, Princeton University, 2004.

[13] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1261–1282, Apr. 2005.

[14] A. Montanari, B. Prabhakar, and D. Tse, "Belief propagation based multi-user detection," in *Proc. 43rd Allerton Conf. Commun., Control and Computing*, Monticello, IL, USA, 2005.

[15] C.-C. Wang and D. Guo, "Belief propagation is asymptoticly equivalent to MAP detection for sparse linear systems," in *Proc. 44th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, 2006.

[16] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.

[17] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: Asymptotic distance distributions," *IEEE Trans. Inform. Theory*, vol. 48, pp. 887–908, Apr. 2002.

[18] D. Guo, "Performance of multicarrier CDMA in frequency-selective fading via statistical physics," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1765–1774, Apr. 2006.

[19] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[20] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[21] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *Journal of Physics A: Mathematical and General*, vol. 36, pp. 11111–11121, 2003.

[22] T. Tanaka and M. Okada, "Approximate belief propagation, density evolution, and statistical neurodynamics for CDMA multiuser detection," *IEEE Trans. Inform. Theory*, vol. 51, pp. 700–706, Feb. 2005.

[23] J. P. Neirotti and D. Saad, "Improved message passing for inference in densely connected systems," *Europhys. Lett.*, vol. 71, no. 5, pp. 866–872, 2005.

[24] C. C. Wang, S. R. Kulkarni, and H. V. Poor, "Finite-dimensional bounds on $\mathbf{Z}_m$ and binary LDPC codes with belief propagation decoders," *IEEE Trans. Inform. Theory*, 2006. to be published.

[25] C. C. Wang, S. R. Kulkarni, and H. V. Poor, "Density evolution for asymmetric memoryless channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 4216–4236, Dec. 2005.