Forty-Fourth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
Sept 27-29, 2006

Belief Propagation Is Asymptotically Equivalent to MAP Estimation for Sparse Linear Systems

Chih-Chun Wang and Dongning Guo

Abstract— **This paper addresses the relationship between belief propagation (BP) and the optimal maximum *a posteriori* (MAP) detection for large linear systems with Gaussian noise. Assuming an input vector of independent symbols with arbitrary distribution, it is proved that BP is asymptotically equivalent to MAP as long as the linear system is "sparse" and "semi-regular" in some sense where the probability of seeing short cycles in the factor graph which describes the system vanishes, and that the aspect ratio of the linear system is below a certain threshold. Moreover, the problem of estimating each input symbol through the sparse linear system is shown to be asymptotically equivalent to that of estimating the same symbol through a scalar Gaussian channel with some degradation in the signal-to-noise ratio (SNR), known as the efficiency of the system. The efficiency has been previously shown to satisfy a fixed-point equation by Guo and Verdú using non-rigorous statistical physics techniques in the context of code-division multiple access (CDMA) with no requirement of sparsity. This paper establishes a rigorous proof of the result in the special case of sparse systems, and henceforth fully characterizes the performance of such systems.**

I. Introduction

Consider the estimation of a vector input signal where the observation is a linear transformation of the input which is subsequently corrupted by Gaussian noise. This simple model is widely used in communications, control, and signal processing. The estimation problem has been well studied, especially in the context of multiuser detection in code-division multiple access (CDMA) [1]. Of particular interest in this work are the optimal maximum *a posteriori* (MAP) estimator and low-complexity suboptimal estimators based on belief propagation (BP), as well as their relationship.

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 relatively small. In case of a large randomly generated linear transformation, the MSE can be obtained using random matrix theory, because the empirical distribution of the singular values of the linear transformation converges to a deterministic law in the large-system limit (e.g., [1]–[4]).

For general (non-Gaussian) inputs, both the best known implementation and the performance evaluation of the optimal detection entail 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 MAP detection in the case of binary inputs was obtained using the replica method [5]. The result has been generalized by Guo and Verdú to arbitrary inputs with arbitrary powers and a family of suboptimal detectors in [6], where it is found that, in terms of estimating each symbol, 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 multiuser efficiency, is determined by a fixed-point equation [6]. Unfortunately, the replica method has not been fully justified mathematically. Hence the results in [5] and [6] are subject to doubt.

Estimators based on belief propagation have received a lot of attention due to their low complexity and good performance [7]–[9]. Citing wisdom in statistical physics and neural networks, several works suggest that the large-system performance of BP is described by the meta-stable solutions of the fixed-point equations of [5], [6].

A recent work [10] by Montanari and Tse is the first successful attempt to rigorously justify Tanaka's result in the special case of "sparse" spreading matrix with a small system load. The proof outlined in [10] suggests that for binary inputs, BP achieves the optimal performance in some *large-sparse-system limit*, and that the fixed-point equation satisfied by BP is identical to the one satisfied by the (optimal) MAP detector.

In a previous work [11], we generalized the results of [10] to arbitrary *a priori* input distributions and SNRs. Using the generalized density evolution [12], [13] for analyzing the BP algorithm, it is shown that the detection of each input symbol to the linear system is equivalent to detecting the same symbol as input to a scalar Gaussian channel with degradation in the SNR. The SNR degradation, known as the multiuser efficiency, satisfies the fixed-point equation of Guo and Verdú [6]. The mean-square optimality of BP in terms of estimating the linearly transformed input is then established by a sandwiching argument using a fundamental integral relationship between the mutual information and the MMSE [14]. The question of whether BP is optimal in mean-square sense for estimating each input symbol (rather than for the linear transformation of the input vector) is, however, not answered in [11].

Building upon the foundation of [11] and considering

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

Dongning Guo is with the Department of Electrical Engineering & Computer Science, Northwestern University, Evanston, IL 60208, USA.

the same large-sparse-system limit, this paper proves the optimality of BP in the strongest, posterior-distribution sense for estimating the original symbols and for estimating the linearly transformed symbols respectively, which supersedes all previous optimality results in [10], [11]. The proofs of the optimality of BP rely on the *relaxed BP* algorithm first introduced in [11], which performs the successive interference cancellation during each iteration. In particular, the following results are established in the large-sparse-system limit and all the equivalence statements are in the posterior-distribution sense: (i) The relaxed BP is equivalent to the classical BP for estimating each symbol as well as estimating the linear transformation of the input vector; (ii) BP is equivalent to MAP estimation if the load is below a certain threshold so that the solution to the fixed-point equation is unique; and (iii) the problem of detecting each symbol/chip in the linear system is equivalent to that of detecting the same symbol/chip as input to a scalar Gaussian channel, where the SNR suffers a degradation, which is determined by the fixed-point equation of Guo and Verdú [6]. The relaxed BP, based on the successive canceler and the scalar Gaussian channel perspective, admits efficient implementation. Results (i)–(iii) show that the relaxed BP performs as well as the optimal MAP detector when applied to sparse, large, linear systems, which is a major contribution of this work.

We note also that the special case of Gaussian inputs has been studied in [15], where the Tse-Hanly formula [2] is proved without using random matrix theory.

The remainder of this paper is organized as follows. Section II describes the sparse linear system and defines an MMSE transform. The main results are summarized in Section III. Section IV studies the BP algorithm for solving the detection problem. Section V completes the proof of the main results. Conclusions are drawn in Section VII.

II. Model and Formulation

A. Linear System

Consider the linear system described by

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

where $\boldsymbol{X} = [X_1, \dots, X_K]^\top$ denotes the input vector of dimension K, $\boldsymbol{S}$ is an $L \times K$ matrix which represents a linear transformation known to the receiver, and $\boldsymbol{N} \sim \mathcal{N}(0, \boldsymbol{I})$ consists of independent standard Gaussian random variables. Given $\boldsymbol{S}$ and the prior distribution of $\boldsymbol{X}$, the job of a receiver is to produce an estimate of $\boldsymbol{X}$ using the output $\boldsymbol{Y}$ of the linear system.

The linear system (1) describes in general a multi-input multi-output (MIMO) channel with additive Gaussian noise, such as what models a multi-antenna system. Another important application of the model is to describe a fully-synchronous K-user CDMA system with spreading factor L and $\boldsymbol{S} = [A_1\boldsymbol{s}_1, \cdots, A_k\boldsymbol{s}_k]$, where user k modulates the symbol X_k onto a spreading sequence $\boldsymbol{s}_k$ with positive amplitude A_k. In this paper, we will frequently use the language of CDMA, although the results encompass all linear systems of the same type. We also assume that the symbols X_k are independently identically distributed (i.i.d.) random variables with distribution (probability measure) P_X, which has zero mean and finite variance. Let the spreading sequence of user k be described by $\boldsymbol{s}_k = \frac{1}{\sqrt{\Lambda_k}}[s_{1k}, s_{2k}, \dots, s_{Lk}]^\top$ where $\sqrt{\Lambda_k}$ is a normalization factor. The output at one arbitrary chip $l \in \{1, 2, \dots, L\}$ is

$$Y_l = \sum_{k=1}^{K} \frac{s_{lk}}{\sqrt{\Lambda_k}} A_k X_k + N_l \tag{2}$$

$$= W_l + N_l \tag{3}$$

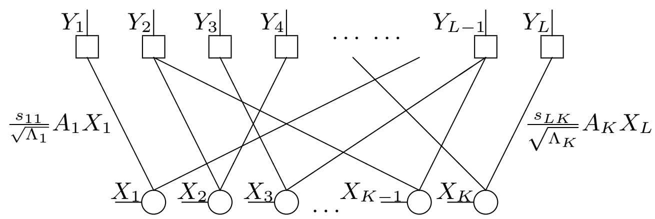

Fig. 1. Factor graph for the sparse CDMA system

where W_l is used to denote the noiseless signal in a chip. Reference [11] studies the mean-square optimal estimation of $\boldsymbol{W} = \boldsymbol{S}\boldsymbol{X}$ which is non-trivial because the prior distribution of $\boldsymbol{W}$ is quite complicated. In this paper we will also address the estimation of $\boldsymbol{X}$, which has a simple prior distribution but is observed through a correlated vector channel. It is often assumed that all X_k and $\boldsymbol{s}_k$ have unit energy, so that A_k^2 represents the SNR of user k. In this paper, however, we produce general results that do not require such unit-energy assumptions, where A_k^2 can be regarded as the gain in the SNR. Note also that the amplitudes are absorbed in the spreading matrix $\boldsymbol{S}$ in (1) for notational convenience.

B. Factor Graph

A bipartite factor graph of the system (2) is illustrated in Fig. 1, where symbol node X_k and chip node Y_l are connected by an edge if $s_{lk} \neq 0$. The chip nodes (the squares in Fig. 1) are defined over the reals in contrast to the check nodes of low-density parity-check (LDPC) codes which are typically defined over a Galois field. Since now Y_l only gives information about the chip node, the closest analogy from the coding perspective is the low-density generating-matrix code [16] and the rateless digital fountain code [17].

C. The Limiting Ensemble of Linear Transformations

Rather than focusing on a particular transformation $\boldsymbol{S}$, this paper considers ensembles of linear transformations. Specifically, for any fixed K and L, let $\boldsymbol{S}$ be randomly constructed as follows. First, an $L \times K$ binary incidence matrix $\boldsymbol{H} = (H_{lk})$ is randomly picked from a certain ensemble to be described shortly. For all (l, k) with $H_{lk} = 0$, set $S_{lk} = 0$. Here we use S_{lk} in lieu of s_{lk} to represent a chip of the matrix $\boldsymbol{S}$ when the randomness is stressed. For all (l, k) with $H_{lk} = 1$, S_{lk} are drawn i.i.d. with distribution

P_S. The normalization factor for each spreading sequence $\boldsymbol{s}_k$ is $\sqrt{\Lambda_k}$, where $\Lambda_k = \sum_{l=1}^{L} H_{lk}$ is the symbol degree of X_k. Let $\Gamma_l := \sum_{k=1}^{K} H_{lk}$ denote the chip degree of Y_l and $\bar{\Gamma} := \frac{1}{L}\sum_{l=1}^{L}\Gamma_l$ the average chip degree. The factor graph depicted in Fig. 1 can be regarded as a random graph, depending on the (random) choice of $\boldsymbol{S}$.

For rigorous analysis, we consider only a family of ensembles satisfying the following conditions. The sequence of taking limit is also addressed as follows.

- For fixed (K, L), the ensembles of interest can be further indexed by $\bar{\Gamma}$, its averaged check node degree.
- For any $(K, L, \bar{\Gamma})$, the random configuration according to $\boldsymbol{H}$ is indistinguishable under any index permutation in the symbol, the chip, and the edge perspective. Namely, we consider the common LDPC graph ensembles for which the density evolution analysis [12] holds.
- We use the *large-sparse-system limit* to refer to that $K, L, \bar{\Gamma} \to \infty$ with $K/L \to \beta$, which is a given positive number, and that for any given t, the probability of having a loop through X_k of girth t or less vanishes. For most sparse ensembles, $\bar{\Gamma} = o(K^{1/2})$ guarantees that there is no short cycles through X_k in probability. The density of non-zero entries, $\bar{\Gamma}/K$, vanishes in the large-sparse-system limit. Throughout this paper, we use the following shorthand to denote the large-sparse-system limit under no-short-cycle condition: $\lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}\,\to\,\infty,\,\mathrm{N.S.C.}}}$.
- The family of ensembles should satisfy
$$\forall\, l, \forall \epsilon > 0, \lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}\,\to\,\infty,\,\mathrm{N.S.C.}}} \mathsf{P}\left\{|\Gamma_l - \bar{\Gamma}| > \epsilon\bar{\Gamma}\right\} = 0, \quad (4)$$
i.e., the degrees of all chip nodes concentrate to their average in probability. We call such an ensemble a *chip-semi-regular* ensemble.
- It is not required that the symbol-node degrees of each ensemble converge to a degree distribution of any sort in the large system limit. Nonetheless, we require that the ensemble satisfies an additional mild condition which we call the *balanced symbol degree* condition; that is, $\exists\, a < \infty$ such that
$$\forall k, \quad \lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}\,\to\,\infty,\,\mathrm{N.S.C.}}} \mathsf{P}\left\{\Lambda_k > a\bar{\Gamma}\right\} = 0. \quad (5)$$
In other words, for random $\boldsymbol{S}$ in the ensemble indexed by $\bar{\Gamma}$, the degrees of all symbol nodes are upper bounded by $a\bar{\Gamma}$ in probability.

The results in this paper apply to all LDPC-code-based, chip-semi-regular ensembles with balanced symbol degrees, including the following special cases:

C1) The classical graph-based ensembles from the LDPC coding literature, which assigns uniform probability to all bipartite graphs with regular chip node degrees and regular/irregular symbol node degree distributions [18].

C2) The symbol-irregular, chip-Poisson ensembles: For each symbol X_k, its degree Λ_k is i.i.d. chosen from a predefined distribution P_Λ. Then X_k is connected to Λ_k uniformly randomly selected chip nodes. In the large-system limit, the chip node degree becomes Poisson distributed which satisfies the semi-regularity condition in (4). Results in [10] are based on this ensemble.

C3) The doubly Poisson ensembles, where both the symbol and chip node degrees are Poisson distributed in the large-system limit for given $\bar{\Gamma}$. For example, H_{lk} are i.i.d. Bernoulli with $\mathsf{P}\{H_{lk} = 1\} = \bar{\Gamma}/L$. The ensemble in this example is also symbol-semi-regular.

We also 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. In fact, P_A can be understood as the received amplitude profile.

Finally, the input distribution P_X, the chip distribution P_S, the amplitude distribution P_A and the load β are fixed in the large-sparse-system limit. The load is assumed to not exceed a threshold which is dependent on P_X, P_S, and P_A. As we shall see, when β is within the constraint, BP is asymptotically optimal in mean-square sense. Beyond the threshold BP is generally not optimal.

D. The MMSE

The minimum mean-square error is a key measure in this work. In general, we use

$$\mathcal{E}\left(\boldsymbol{U}\,|\boldsymbol{V}\,\right) = \frac{1}{K}\mathsf{E}\left\{\|\boldsymbol{U} - \mathsf{E}\{\boldsymbol{U} \mid \boldsymbol{V}\}\|^2\right\} \quad (6)$$

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

The special scenario of estimating a scalar product AX from its scalar noisy observation with the side information A is of particular importance. In general, we define the *MMSE transform* as the following based on (6):

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

where $(X, A) \sim P_{XA}$ and N is independent standard Gaussian. This is the average MMSE of estimating the input of a Gaussian channel with random amplitude A known to the receiver. Throughout this paper, X and A are independent, and N is used to denote an independent standard Gaussian random variable without further notice, so is its standard vector counterpart $\boldsymbol{N}$. Hence the MMSE transform $\mathcal{E}_{X,A}\left(\gamma\right)$ depends on P_X and P_A, and is a decreasing function of γ.

E. The Classical BP

The classical BP is an efficient message-passing algorithm for computing the exact posterior distribution based on the observations on a local subtree with its root at the target symbol X_k (or chip W_l), assuming the underlying Markov inference network contains no short cycles. The latter assumption is satisfied in the large-sparse-system limit. Any sufficient statistic can be used as "messages" and the most common choice for binary symbols is the log-likelihood ratio (LLR) messages. Let $V_{k\to l}^{(t)}$ and $U_{l\to k}^{(t)}$ represent the LLR messages passed from X_k to W_l and from W_l to X_k in the t-th iteration respectively. Each message thus represents the *extrinsic* information of X_k in its LLR form. (See [19] and references therein.) The update equations at the t-th ($t > 0$) iteration are given as:

$$U_{l\to k}^{(t)} = \log \frac{\mathsf{P}\left\{X_k = +1 \,|\, Y_l, \boldsymbol{S}, \boldsymbol{A}, \{V_{i\to l}^{(t-1)}\}_{i\in\partial l\setminus k}\right\}}{\mathsf{P}\left\{X_k = -1 \,|\, Y_l, \boldsymbol{S}, \boldsymbol{A}, \{V_{i\to l}^{(t-1)}\}_{i\in\partial l\setminus k}\right\}} \quad (7)$$

$$V_{k\to l}^{(t)} = V_k^{(0)} + \sum_{j\in\partial k\setminus l} U_{j\to k}^{(t)}. \quad (8)$$

where $\boldsymbol{A} = \{A_i\}_{i=1,\cdots,K}$ denotes the amplitudes of all users, $\partial l = \{i \,|\, S_{li} \neq 0\}$ denotes the set of i such that X_i and Y_l are connected. With slight abuse of notation, $\partial k = \{j \,|\, S_{jk} \neq 0\}$. Here, the posterior probability $\mathsf{P}\left\{X_k = x \,|\, Y_l, \boldsymbol{S}, \boldsymbol{A}, \{V_{i\to l}^{(t-1)}\}_{i\in\partial l\setminus k}\right\}$ is defined as the probability of $X_k = x$ given Y_l, $\boldsymbol{S}$, $\boldsymbol{A}$, and the extrinsic information obtained from the collection $\{V_{i\to l}^{(t-1)}\}_{i\in\partial l\setminus k}$, which are the extrinsic LLRs of $\{X_i\}_{i\in\partial l\setminus k}$. The initial messages $V_{k\to l}^{(0)} \triangleq V_k^{(0)}$ correspond to the prior distribution P_X. After the final iteration, the decision at each variable node k is made according to the following LLR

$$V_k^{(t)} = V_k^{(0)} + \sum_{l\in\partial k} U_{l\to k}^{(t)}. \quad (9)$$

Formulas (7) to (9) are the BP update equations when the underlying messages are LLRs. In its most general form, the messages of BP can be any sufficient statistic of the extrinsic information, as long as the update equations are modified accordingly. For non-binary input symbols, using sufficient statistics other than the LLR is sometimes beneficial.

Remark: One of the most important analytical tools of BP is the density evolution, which focuses on the evolution of the distributions of the passing messages. Similarly, in its most general form, one only needs to focus on the evolving distributions of any sufficient statistics.

III. Main Results

This section presents the main theorems of this work, for which detailed guidelines of the proof are relegated to Sections IV and V.

In view of the linear system described by equations (1) and (2), we study two related problems of estimating: (i) the input vector $\boldsymbol{X}$, and (ii) its linear transformation $\boldsymbol{W} = \boldsymbol{SX}$ respectively. Both BP and MAP detection are considered.

Throughout this section, we assume that $X_k \sim P_X$ are i.i.d., $A_k \sim P_A$ are i.i.d, and the spreading matrix is randomly chosen from the chip-semi-regular ensembles with balanced symbol degrees and symmetrically distributed sequences P_S. Moreover, P_X and P_S are of zero mean and unit variance and P_A has finite moments of any order. We also assume that the load β is such that the following equation

$$\eta^{-1} = 1 + \beta\, \mathcal{E}_{X,A}(\eta) \quad (10)$$

has a unique fixed point, which is referred to as the *power efficiency* or simply *efficiency* of the linear system for reasons to be clear shortly.

A. MMSE and Mutual Information

Theorem 1 (The symbol-node MMSE): For every k,

$$\lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}\to\infty,\,\mathrm{N.S.C.}}} \mathcal{E}\left(X_k \,|\, \boldsymbol{SX}+\boldsymbol{N}, \boldsymbol{S}\right) = \mathcal{E}\left(X \,|\, \sqrt{\eta}AX + N, A\right)$$

where the efficiency η satisfies (10).

Theorem 2 (The chip-nodes MMSE): Following the same settings as in Theorem 1, we have

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

where the efficiency η satisfies (10).

In general, the MMSE of estimating X_k (or $\boldsymbol{X}$) given the observation of the linear transformation $\boldsymbol{SX}$ corrupted by additive Gaussian noise is dependent on $\boldsymbol{S}$. So is the MMSE of estimating $\boldsymbol{SX}$. Theorems 1 and 2 state that, as the system size becomes large, not only the dependence on $\boldsymbol{S}$ diminishes, but the large-sparse-system MMSE can be expressed as the MMSE for a scalar Gaussian channel, which is straightforward to compute. The efficiency is easy to determine from (10).

The fixed-point equation (10) for the efficiency was first obtained by Guo and Verdú in [6] as a generalization of Tanaka's formula [5]. In fact, the efficiency is the key to the error performance of optimal multiuser detection as well as the optimal spectral efficiency.

Theorem 3: The input–output mutual information per dimension of the linear system converges in the large-sparse-system limit:

$$\lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}\to\infty,\,\mathrm{N.S.C.}}} \frac{1}{K} I(\boldsymbol{X};\boldsymbol{Y}) = I(X; \sqrt{\eta}\,AX + N|A) + \frac{\eta - 1 - \log\eta}{2\beta} \quad (11)$$

where η satisfies (10).

The same formula (11) is claimed in [6] for general (dense) linear systems without rigorous justification.

B. Optimality of Belief Propagation

Consider the bipartite graph (Fig. 1) which describes the linear system (1). An iterative BP estimator can be devised based on the graph, which essentially updates the posterior distribution (or "belief") for each X_k conditioned on the observations within the reach of the local subtree $\boldsymbol{Y}^{(t)}$ with X_k as the root [19], [20]. Precisely, $\boldsymbol{Y}^{(t)}$ consists of all chips of $\boldsymbol{Y}$ that has its distance to X_k on the bipartite graph less than or equal to $2t-1$. Different X_k and $X_{k'}$ correspond to different $\boldsymbol{Y}^{(t)}$ while the roots of different $\boldsymbol{Y}^{(t)}$ should be clear from the context.

Let $p_{X_k}(\cdot|\boldsymbol{Y},\boldsymbol{S})$ denote the posterior distribution of X_k given the vector observation $\boldsymbol{Y}$ and the system description $\boldsymbol{S}$. Without loss of generality, we further assume that X_k takes values in a finite set so that $p_{X_k}(\cdot|\boldsymbol{Y},\boldsymbol{S})$ becomes a probability mass function (pmf). One can thus view $p_{X_k}(\cdot|\boldsymbol{Y},\boldsymbol{S})$ as a finite-dimensional vector with its value depending on the realization of $\boldsymbol{Y}$ and $\boldsymbol{S}$. For all practical systems, the set of constellation points of X_k is finite. The constraint that the domain of X_k is finite can be relaxed by the Kolmogorov extension theorem if a more general X_k is of interest.

Similarly, let $p^{(t)}_{X_k}(\cdot|\boldsymbol{Y}^{(t)},\boldsymbol{S})$ denote the posterior pmf of X_k when only local subtree observation $\boldsymbol{Y}^{(t)}$ is available. One can easily shown that $p^{(t)}_{X_k}(\cdot|\boldsymbol{Y}^{(t)},\boldsymbol{S})$ is the corresponding posterior pmf computed by BP after t iterations in the large system limit. We can also use $p_{W_l}(\cdot|\boldsymbol{Y},\boldsymbol{S})$ and $p^{(t)}_{W_l}(\cdot|\boldsymbol{Y}^{(t)},\boldsymbol{S})$ to denote the posterior pmf vector of W_l conditioned on $\boldsymbol{S}$, $\boldsymbol{Y}$, and $\boldsymbol{Y}^{(t)}$ respectively. The asymptotic optimality of BP, namely, the equivalence of BP and MAP, is then stated in the following theorems.

Theorem 4 (Symbol-wise optimality): For every k, x,

$$\lim_{t\to\infty}\limsup_{\substack{K=\beta L\to\infty\\ \bar{\Gamma}\,\to\,\infty,\,\mathrm{N.S.C.}}}\left|p^{(t)}_{X_k}(x|\boldsymbol{Y}^{(t)},\boldsymbol{S})-p_{X_k}(x|\boldsymbol{Y},\boldsymbol{S})\right|=0$$

in probability.

Theorem 5 (Chip-wise optimality): For every l, w,

$$\lim_{t\to\infty}\limsup_{\substack{K=\beta L\to\infty\\ \bar{\Gamma}\,\to\,\infty,\,\mathrm{N.S.C.}}}\left|p^{(t)}_{W_l}(w|\boldsymbol{Y}^{(t)},\boldsymbol{S})-p_{W_l}(w|\boldsymbol{Y},\boldsymbol{s})\right|=0$$

in probability.

The sketch of the proofs for Theorems 4 and 5 is provided in Section V.

One may argue that when t becomes sufficiently large, $\boldsymbol{Y}^{(t)}$ is very close to $\boldsymbol{Y}$ and Theorems 4 and 5 are trivial. Nonetheless, with the large-system limit taken first for any finite t, the size of $\boldsymbol{Y}$ is infinitely larger than $\boldsymbol{Y}^{(t)}$. The essence of Theorems 4 and 5 is that for sufficiently large t, a finite portion of the infinite vector $\boldsymbol{Y}$, represented by $\boldsymbol{Y}^{(t)}$, contains virtually all the information about X_k and W_l.

Theorems 4 and 5 establish the asymptotic equivalence of belief propagation and the MAP detector for sparse linear systems. As straightforward corollaries, when $t\to\infty$, the conditional means of the posterior distributions obtained from BP approach the MMSE estimators and achieve the MMSEs specified in Theorems 1 and 2.

C. Equivalent Scalar Gaussian Channels

Not only do MAP and BP achieve the MMSE of a scalar Gaussian channel with the same input as is seen in Theorems 1 and 2, but the problem of estimating each symbol (or chip) becomes asymptotically equivalent to that of estimating the same symbol (or chip) through a scalar Gaussian channel.

Consider the following scalar Gaussian channel:

$$Y=AX+\frac{1}{\sqrt{\eta}}N$$

where $X\sim P_X$, $N\sim\mathcal{N}(0,1)$, and η is the fixed point of (10). Let $p_{X|Y,A}(\cdot|y,a)$ denote the posterior pmf of X with the scalar observation $Y=y$ and the channel gain $A=a$ known to the receiver. We then have the following theorems:

Theorem 6 (Symbol-wise equivalent channel): For all k, x,

$$\lim_{t\to\infty}\limsup_{\substack{K=\beta L\to\infty\\ \bar{\Gamma}\,\to\,\infty,\,\mathrm{N.S.C.}}}\left|p_{X_k}(x|\boldsymbol{Y},\boldsymbol{S})-p_{X|Y,A}(x|f^{(t)}_k(\boldsymbol{Y}),A_k)\right|=0$$

in probability, where $f^{(t)}_k(\cdot)$ is a function obtained by successive cancellation and admits an iterative representation.

Consider another scalar Gaussian channel:

$$Y=W+N,$$

where W is Gaussian distributed with mean zero and variance $\beta\mathcal{E}_{X,A}(\eta)$ and $N\sim\mathcal{N}(0,1)$ is independently distributed. Let $p_{W|Y}$ denote the conditional (Gaussian) distribution of W given the output Y.

Theorem 7 (Chip-wise equivalent channel): For all l, w,

$$\lim_{t\to\infty}\limsup_{\bar{\Gamma}\to\infty}\limsup_{K=\beta L\to\infty}|p_{W_l}(w|\boldsymbol{Y},\boldsymbol{S})-p_{W|Y}(w|g^{(t)}_l(\boldsymbol{Y}))|=0$$

in probability, where $g^{(t)}_l(\cdot)$ is a function obtained by successive cancellation and admits an iterative representation.

Theorems 6 and 7 complete Theorems 1–5, and the results reveal the fact that the BP estimate of each symbol (or chip) is not only asymptotically equivalent to that of the MAP detector, but is also asymptotically identical to that of estimating the symbol (or chip) through a scalar Gaussian channel with properly degraded SNRs. The associated decision functions $f^{(t)}_k(\cdot)$ and $g^{(t)}_l(\cdot)$ can be computed efficiently and iteratively. In [6], η is called the multiuser efficiency since it corresponds to the ratio of the effective SNR and the actual SNR for each user (symbol).

We note that the fixed-point equation (10) has at least one solution for η because of the following. If η varies from 0 to 1, then the right hand side (RHS) of (10) decreases continuously to $1+\beta\mathcal{E}_{AX,A}(1)>1$, so that the two sides intersect as functions of η. Depending on P_X, P_A and β, there may exist more than one solutions to (10). For small enough β, however, the slope of the RHS as a function of η can be arbitrarily small, so that the solution to (10) is unique.

The sketch of the proofs includes the following steps. We first show that a (sub-optimal) relaxed BP achieves the equivalent channel perspectives stated in Theorems 6 and 7. We then prove that in the large-sparse-system limit, the relaxed BP is asymptotically identical to the classical BP, which in turn proves BP-based versions of Theorems 6 and 7. A sandwiching argument is then presented showing the equivalence of MAP and BP detectors in probability as stated in Theorems 4 and 5, which in turn proves Theorems 6 and 7 for MAP detectors. Theorems 1 and 2 can then be derived as direct corollaries of Theorems 4 to 7. An alternative proof of Theorem 2 using the fundamental relationship between the

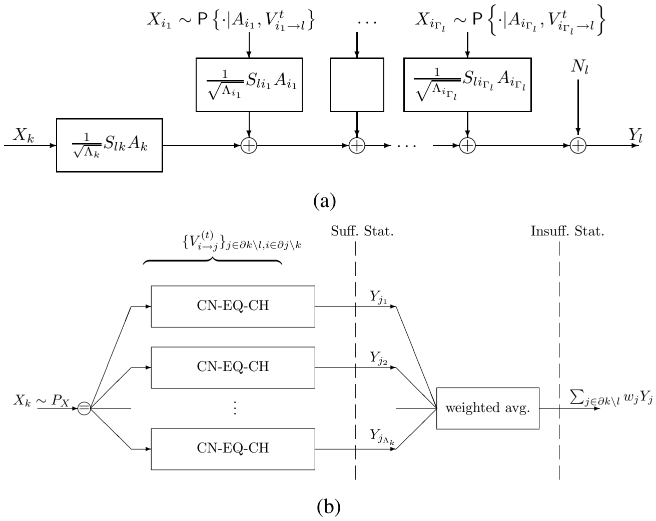

Fig. 2. Equivalent channels for the relaxed BP. (a) The chip node equivalent channel (CN-EQ-CH). (b) The perspective combining symbol and chip nodes.

mutual information and the MMSE is also provided.

It is worth noting that the normalization factor $\sqrt{\Lambda_k}$ is taken on a user-by-user basis, which is essential when considering symbol irregular ensembles. For instance, P_Λ in the special case (C2) may assign 1/2 probability to $\{\Lambda_k = 2\}$ while the other 1/2 probability is assigned $\{\Lambda_k = \lambda\}$ for some λ growing linearly with respect to $\bar{\Gamma}$. This ensemble is chip-semi-regular with balanced symbol degrees but exhibits non-negligible irregularity at the symbol node. If a common normalization factor is applied to all symbols, a more involved fixed point equation will be derived and all theorems have to be changed accordingly.

IV. BP and the Iterative Formula

This section shows the connection between BP and (10). BP-based detection in linear systems has been studied in [7], [8]. Using clever heuristics, several efficient detection schemes have been proposed, which invariably take some interference cancellation structure [7], [9], [21]. The performance can usually be obtained through solving a set of coupled equations (e.g., [7]).

For chip-semi-regular ensembles, a relaxed BP will be introduced under which the performance of large sparse linear systems can be described by a scalar iterative formula. The balanced symbol degree condition and the symmetry of P_S are then incorporated to prove the asymptotic equivalence between the relaxed BP and the classical BP.

A. The Relaxed BP—The Iterative Formula

We directly derive the relaxed BP based on the setting of non-binary symbols. For non-binary symbols, one common choice is to use the posterior distributions instead of the LLRs as messages, which generally involves infinite-dimensional functional analysis. In addition to using the posterior distributions, we sometimes use the "entire observed value(s)" as the sufficient statistic for the passing messages.

The message map at the chip node, analogous to (7), can be obtained by solving the posterior distribution of the detection problem in Fig. 2(a). Similarly, the symbol node message map corresponding to (8) can be derived by considering repetition channels. As suggested in [22], we can combine both halves of the iteration, one at the symbol node and one at the chip node, so that the overall $V_{k\to l}^{(t-1)} \mapsto V_{k\to l}^{(t)}$ message map can be obtained by solving the message $V_{k\to l}^{(t)}$, a sufficient statistic (generally the posterior distribution) of X_k, from the detection problem in Fig. 2(b) assuming the entire observation vector $\{Y_j\}_{j\in\partial k\backslash l}$ is employed. It is worth emphasizing that the message $V_{k\to l}^{(t)}$ can be any type of sufficient statistic.

The relaxed BP is best explained by solving the same detection problem in Fig. 2(b) based on the following degraded observation

$$\sum_{j\in\partial k\backslash l} \frac{1}{\sqrt{\Lambda_k}} S_{jk} Y_j, \tag{12}$$

a weighted average of the original observation vector $\{Y_j\}_{j\in\partial k\backslash l}$. The detection problem based on this weighted average can be analyzed as follows. Let $\mathsf{SC}(k,l)$ denote the successive canceler of estimating X_k from Y_l. Namely,

$$\mathsf{SC}(k,l) = \sum_{i\in\partial l\backslash k} \frac{1}{\sqrt{\Lambda_i}} S_{li} A_i \mathsf{E}\left\{ X_i \mid A_i, V_{i\to l}^{(t-1)} \right\},$$

which can be computed without randomness given the messages $\{V_{i\to l}^{(t-1)}\}$ of the $(t-1)$-th iteration. Let $\mathsf{MAI}(k,l)$ denote the remaining random component of the corresponding multiple access interference such that

$$\mathsf{MAI}(k,l) = \sum_{i\in\partial l\backslash k} \frac{1}{\sqrt{\Lambda_i}} S_{li} A_i \left(X_i - \mathsf{E}\left\{ X_i \mid A_i, V_{i\to l}^{(t-1)} \right\}\right).$$

Since subtracting a known number does not change the sufficiency of (12) as an observation, we can subtract each Y_j with its $\mathsf{SC}(k,j)$:

$$\begin{aligned} &Y_j - \mathsf{SC}(k,j) \\ &= \left(\frac{1}{\sqrt{\Lambda_k}} S_{jk} A_k\right) X_k + \mathsf{MAI}(k,j) + N_l. \end{aligned} \tag{13}$$

By performing this type of substraction on all CN-EQ-CHs, the relaxed BP corresponds to estimating X_k from the weighted average (12) with shifts as described in (13). Furthermore, for the relaxed BP, we choose this shifted weighted average as the message $V_{k\to l}^{(t)}$:

$$V_{k\to l}^{(t)} \triangleq \sum_{j\in\partial k\backslash l} \frac{1}{\sqrt{\Lambda_k}} S_{jk} Y_j - \sum_{j\in\partial k\backslash l} \frac{1}{\sqrt{\Lambda_k}} S_{jk} \mathsf{SC}(k,j) \tag{14}$$

$$\begin{aligned} &= \sum_{j\in\partial k\backslash l} \frac{1}{\Lambda_k} S_{jk}^2 A_k X_k \\ &\qquad + \sum_{j\in\partial k\backslash l} \frac{1}{\sqrt{\Lambda_k}} S_{jk} \left(\mathsf{MAI}(k,j) + N_j\right). \end{aligned} \tag{15}$$

Equation (14) defines the $V_{k\to l}^{(t-1)} \mapsto V_{k\to l}^{(t)}$ message map of our relaxed BP. Similarly, for the final decoding stage $t = \tau$, the decision is obtained by detecting X_k from the decoding message

$$V_k^{(\tau)} \triangleq \sum_{j\in\partial k} \frac{1}{\sqrt{\Lambda_k}} S_{jk} Y_j - \sum_{j\in\partial k} \frac{1}{\sqrt{\Lambda_k}} S_{jk} \mathsf{SC}(k,j) \quad (16)$$

$$= \sum_{j\in\partial k} \frac{S_{jk}^2}{\Lambda_k} A_k X_k + \sum_{j\in\partial k} \frac{S_{jk}}{\sqrt{\Lambda_k}} \left(\mathsf{MAI}(k,j) + N_j\right) \quad (17)$$

in which (16) is analogous to (9).

Remark: Equations (14) and (16) are used for explicitly computing the messages as in a real decoding process, while (15) and (17) are auxiliary equations explaining the relationship between $V_{k\to l}^{(t)}$ (or V_k^{τ}) and X_k.

B. The Relaxed BP—The Density Evolution Analysis

Consider the asymptotic performance of the relaxed BP in the large-sparse-system limit where the incidence matrix of the bipartite factor graph is drawn form the chip-semi-regular ensembles. The predefined symbol distribution P_X, the amplitude distribution P_A, and the spreading sequence distribution P_S all have finite moments of any order. We also require P_S and P_X to have zero means and unit variances. Unlike the case of binary symbols and the uniform power distribution, the all-one $\{X_k\}_{k=1,\cdots,K}$ cannot be assumed and the traditional density evolution does not hold. Nonetheless, since the perfect projection condition [13] holds naturally for all linear systems described by (1), the modified density evolution analysis [13] can be applied, provided the symbol and amplitude configurations $\{X_k, A_k\}_{k=1,\cdots,K}$ being properly averaged, which is the direction we should proceed in this work.

Due to the aforementioned order of limits, we can assume there is no cycle of any finite size without loss of generality. That is, after any finite t iterations, there is no repeated Y_l on the local subtree with its root at the target symbol X_k or the chip W_l. For the following, we first assume that all symbols X_k are with $\Lambda_k \to \infty$ as $\bar{\Gamma} \to \infty$. (This assumption will be relaxed later.) By the weak law of large numbers, the $\left(\sum_j \frac{1}{\Lambda_k} S_{jk}^2\right)$ in (15) and in (17) converge to 1 in probability. By the central limit theorem, when $\bar{\Gamma} \to \infty$, the

$$\sum_j \frac{1}{\sqrt{\Lambda_k}} S_{jk} \left(\mathsf{MAI}(k,j) + N_j\right)$$

in (15) and in (17) converge weakly to a scalar Gaussian random variable with zero mean and variance

$$1 + \beta \mathsf{E}\left\{A_i^2 \mathsf{var}\left\{X_i \,|\, A_i, V_{i\to j}^{(t-1)}\right\}\right\}. \quad (18)$$

Therefore, given A_k, the problem of estimating X_k from $V_{k\to l}^{(t)}$, namely, estimating X_k by a t-iteration relaxed BP detector, converges weakly to a detection problem of estimating the same X_k through a scalar Gaussian channel with channel gain A_k and noise variance as in (18). Let $\eta^{(t)}$ denote the SNR degradation factor $1/(1 + \beta \mathsf{E}\left\{A_i^2 \mathsf{var}\left\{X_i \,|\, A_i, V_{i\to j}^{(t-1)}\right\}\right\})$ after t iterations. We can derive the iterative equation

$$\begin{aligned} \frac{1}{\eta^{(t)}} &= 1 + \beta \mathsf{E}\left\{A_i^2 \mathsf{var}\left\{X_i \,|\, A_i, V_{i\to j}^{(t-1)}\right\}\right\} \\ &= 1 + \beta \mathcal{E}\left(AX \,\middle|\, AX + \frac{1}{\sqrt{\eta^{(t-1)}}} N, A\right) \\ &= 1 + \beta \mathcal{E}_{X,A}\left(\eta^{(t-1)}\right), \end{aligned}$$

with the initial value $\eta^{(0)} = 0$ representing the case that there is no extrinsic information during the first iteration. Noticeably, the noise for all users is of *identical* power $1/\eta^{(t)}$, and the performance of user k depends on A_k but not on the effective spreading length Λ_k and the sequence $\boldsymbol{s}_k$. Denote the limit of $\eta^{(t)}$ by η (which may or may not be unique), the fixed-point equation of Guo and Verdú (10) is thus obtained. Theorem 6 is thus proved if we substitute the MAP decoder used therein with the relaxed BP.

A similar version of Theorem 7 for the relaxed BP can be proved straightforwardly by focusing on the chip node W_l instead, which would involve the same iterative equation (10). The end results show that the MMSE estimate of the chip node W_l from the relaxed BP becomes

$$\frac{\beta \mathcal{E}_{X,A}(\eta)}{1 + \beta \mathcal{E}_{X,A}(\eta)} Y_l$$

and the MSE of which is

$$\beta \eta \mathcal{E}_{X,A}(\eta) = 1 - \eta. \quad (19)$$

We close this subsection by removing the assumption that all $\Lambda_k \to \infty$ as $\bar{\Gamma} \to \infty$ and by discussing the individual normalization factor $\sqrt{\Lambda_k}$.

Regardless whether Λ_k is bounded from infinity as $\bar{\Gamma} \to \infty$ or not, by the central limit theorem, the term $\mathsf{MAI}(k,j) + N_j$ in (15) (and in (17) respectively) becomes a scalar Gaussian random variable with variance equal (18). There is no problem when Λ_k tends to infinity when $\bar{\Gamma} \to \infty$ as discussed previously. However, with Λ_k bounded away from infinity, different values of $\frac{1}{\Lambda_k}\sum_{j\in\partial k\backslash l} S_{jk}^2$ will result in different SNRs for the corresponding scalar channels in (15) and in (17). However, the *edge-based* iterative nature of message-exchanging algorithms ensures that only the effects on a "significant portion of the edges" will be passed to the next iteration. Since the edge-wise contribution of those X_k with bounded Λ_k is asymptotically zero, it can be thus proved that during iterations, i.e., when (15) is considered, one can assume Λ_k approaches infinity without loss of generality. For the detection stage (17), the SNR of the corresponding scalar Gaussian channel is determined by the random realization of $\frac{1}{\Lambda_k}\sum_{j\in\partial k} S_{jk}^2$. Theorem 6 still holds if we consider the actual power $A_{k,\mathrm{actual}}^2 \triangleq A_k^2 \sum_l S_{lk}^2/\Lambda_k$ and the corresponding perfectly normalized $\boldsymbol{s}_{k,\mathrm{actual}} \propto \boldsymbol{s}_k$ satisfying $\frac{1}{\Lambda_k}\sum_l S_{lk,\mathrm{actual}}^2 = 1$. From the above reasonings, the assumption that $\Lambda_k \to \infty$ can be relaxed.

Individual normalization $\sqrt{\Lambda_k}$ is important for *symbol-irregular* ensembles when compared to the schemes in which a global normalization factor $\sqrt{\bar{\Lambda}}$ is used for all X_k. One example of symbol-irregular ensembles is when a half of symbol nodes of degree $\Lambda_k = d$ and the other half of symbol nodes of degree $\Lambda_k = 2d$ for some varying d. Although all symbol degrees are unbounded when $\bar{\Gamma} \to \infty$, they are not concentrated around its mean $\bar{\Lambda} = 3d/2$. The individual-normalization scheme adopted in this paper ensures that all our results hold for this symbol-irregular ensemble. Similar analysis can be performed on the scheme using global normalization and the results show that for each user X_k in the t-th iteration, the global-normalization scheme is equivalent to estimating X_k from a scalar Gaussian channel with gain $\sqrt{\frac{\Lambda_k}{\bar{\Lambda}}}A_k$ and noise variance $1/\eta_{\bar{\Lambda}}^{(t-1)}$. The new efficiency $\eta_{\bar{\Lambda}}^{(t)}$ satisfies a different iterative equation:

$$\frac{1}{\eta_{\bar{\Lambda}}^{(t)}} = 1 + \beta\mathcal{E}\left(AX \,\middle|\, \sqrt{\frac{\Lambda}{\bar{\Lambda}}}AX + \frac{1}{\sqrt{\eta_{\bar{\Lambda}}^{(t-1)}}}N, A, \Lambda\right)$$

in which we have to use the edge-based symbol-degree distribution of Λ. For the global-normalization scheme, the performance for different users depends not only on its A_k^2 but also on the effective length Λ_k of its spreading sequence.

C. The Equivalence of the Classical and the Relaxed BPs

Based on the degraded observation $\sum \frac{1}{\sqrt{\Lambda_k}} S_{jk} Y_j$, the relaxed BP has the degradation properties commonly used in the information theory literature. Furthermore, the simple expressions and proofs of the relaxed BP also shed many insights on the corresponding detection problem in the asymptotic limit. In this subsection, we will generalize the same results for the classical BP by switching back to messages of the LLR form, which will be defined rigorously for non-binary input symbols later, and by invoking the central limit theorem during iterations (7) and (8). The relaxed BP, although sub-optimal for finite $\bar{\Gamma}$, is thus asymptotically equivalent to the classical BP when $\bar{\Gamma} \to \infty$. The two additional assumptions necessary for this subsection are that P_S is symmetric with respect to $S = 0$ and the chip-semi-regular ensembles are with balanced symbol degrees. Both conditions can be easily satisfied for most practical applications.

The first challenge to deal with is the symbols with Λ_k bounded away from infinity, for which the central limit theorem at the symbol node message map (8) fails. Second, for those X_k with Λ_k approaching infinity, the normality of the *extrinsic* LLR message $\left(\sum_j U_{j\to k}^{(t)}\right)$ in (8) is granted by the central limit theorem. Nonetheless, to explicitly quantify the SNR of the corresponding equivalent scalar Gaussian channel, one has to obtain the coefficients of the scaling laws for the mean and the variance of the incoming LLR messages $\{U_{j\to k}^{(t)}\}_j$ with respect to Λ_k. Precisely, the incoming messages $U_{j\to k}^{(t)}$ can be obtained by solving the detection problem in (13), or equivalently, by solving the problem of detecting X_k from the chip node equivalent channel (CN-EQ-CH):

$$\begin{aligned} &(S_{jk}A_k)\, X_k + \sqrt{\Lambda_k}\,(\mathsf{MAI}(k,j) + N_j) \\ =\ &\sqrt{\Lambda_k}\,(Y_j - \mathsf{SC}(k,j))\,. \end{aligned} \quad (20)$$

Although the $\mathsf{MAI}(k,j) + N_j$ term in of (20) converges weakly to a Gaussian random variable, the simultaneously growing term $\sqrt{\Lambda_k}$ may very well spoil the normality of the CN-EQ-CH that X_k is experiencing and we have a convergence speed competition problem.

Fortunately, for the case in which Λ_k being bounded away from infinity, no such convergence speed competition exists. Each CN-EQ-CH, $X_k \mapsto Y_j$ becomes a Gaussian channel when $\bar{\Gamma} \to \infty$. Since each CN-EQ-CH is Gaussian, the degraded observation $\sum_j \frac{1}{\Lambda_k} S_{jk} Y_j$ becomes a *sufficient statistic* for the repetition symbol node, and the relaxed BP equals the classical BP.

For the case that $\Lambda_k \to \infty$, we first define the vector messages $\{V_{k\to l}^{(t)}(x)\}_x$ and $\{U_{l\to k}^{(t)}(x)\}_x$ as infinite dimensional LLR vectors indexed by x such that for all possible x,

$$V_{k\to l}^{(t)}(x) = \log \frac{\mathsf{P}\left\{X_k = x \mid \boldsymbol{S}, \boldsymbol{A}, \{\Lambda_i\}, \{U_{j\to k}^{(t)}(x)\}\right\}}{\mathsf{P}\left\{X_k = x_a \mid \boldsymbol{S}, \boldsymbol{A}, \{\Lambda_i\}, \{U_{j\to k}^{(t)}(x)\}\right\}}$$

$$U_{l\to k}^{(t)}(x) = \log \frac{\mathsf{P}\left\{Y_l \mid X_k = x, \boldsymbol{S}, \boldsymbol{A}, \{\Lambda_i\}, \{V_{i\to l}^{(t-1)}(x)\}\right\}}{\mathsf{P}\left\{Y_l \mid X_k = x_a, \boldsymbol{S}, \boldsymbol{A}, \{\Lambda_i\}, \{V_{i\to l}^{(t-1)}(x)\}\right\}}$$

for some predefined fixed x_a where $\boldsymbol{A} \triangleq \{A_i\}_{i=1,\cdots,K}$ and $\{\Lambda_i\} \triangleq \{\Lambda_i\}_{i=1,\cdots,K}$ denote the individual amplitudes and symbol degrees, and the extrinsic information are defined as

$$\begin{aligned} \{U_{j\to k}^{(t)}(x)\} &\triangleq \{U_{j\to k}^{(t)}(x)\}_{x, j\in\partial k\backslash l} \\ \{V_{i\to l}^{(t-1)}(x)\} &\triangleq \{V_{i\to l}^{(t-1)}(x)\}_{x, i\in\partial l\backslash k}. \end{aligned}$$

With these definitions of $V_{k\to l}^{(t)}(\cdot)$ and $U_{k\to l}^{(t)}(\cdot)$, the same message maps (7) and (8) hold verbatim. It can be easily shown that if the extrinsic message $\{\sum_j U_{j\to k}^{(t)}(x)\}_x$ corresponds to a scalar Gaussian channel with noise variance $1/\gamma$, then for all x_b's, $\{\sum_j U_{j\to k}^{(t)}(x)\}_x$ is a Gaussian stochastic process with mean and covariance functions

$$\mathsf{mean}(x) = \gamma x_b (x - x_a) \quad (21)$$

$$\mathsf{cov}(x', x'') = \gamma (x' - x_a)(x'' - x_a) \quad (22)$$

conditioning on $X_k = x_b$ being transmitted. Conversely, if for x_b, x', and x'', $\left(\sum_j U_{j\to k}^{(t)}(x')\right)$ and $\left(\sum_j U_{j\to k}^{(t)}(x'')\right)$ are jointly Gaussian distributed with the mean and covariance as in (21) and (22) assuming $X_k = x_b$ being transmitted, then the extrinsic information $\{\sum_j U_{j\to k}^{(t)}(x)\}_x$ corresponds to a scalar Gaussian with variance $1/\gamma$.

As discussed before, to quantify the SNR of the symbol node repetition channel while its normality is granted by the central limit theorem, the coefficients in the scaling laws must be identified, for which we need the following lemma.

Lemma 1 ([23, p. 533]): For $X_1, \cdots, X_n, \cdots$ being i.i.d. distributed with zero mean, unit variance, the third moment μ_3, and a common characteristic function φ. Suppose $|\varphi|^\nu$ is integrable for some $\nu \geq 1$. Then the normalized partial sum $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ has density f_n for

$n \geq \nu$, and for $n \to \infty$

$$f_n(x) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{\mu_3}{6\sqrt{n}}(x^3 - 3x)\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + o\left(\frac{1}{\sqrt{n}}\right),$$

uniformly in x.

Consider the CN-EQ-CH illustrated in Fig. 2(a) and described in (13). By decomposing N_j as a summation of $(\bar{\Gamma} - 1)$ independent small Gaussian random variables: $N_j = \sum_{i \in \partial k \setminus j} \left(\frac{1}{\sqrt{\bar{\Gamma}-1}} N_{ji}\right)$, we can rewrite (13) as

$$\begin{aligned} Y_j' &\triangleq Y_j - \mathsf{SC}(k,j) \\ &= \left(\frac{1}{\sqrt{\Lambda_k}} S_{jk} A_k\right) X_k + \sum_{i \in \partial j \setminus k} \left(\frac{1}{\sqrt{\bar{\Gamma}-1}} N_{ji} \right. \\ &\qquad \left. + \frac{1}{\sqrt{\Lambda_i}} S_{ji} A_i \left(X_i - \mathsf{E}\left\{X_i \mid A_i, V_{i \to j}^{(t-1)}\right\}\right)\right). \end{aligned} \tag{23}$$

With P_S being symmetric and the assumptions of finite moments of any order, one can verify that with the addition of small $\left(\frac{1}{\sqrt{\bar{\Gamma}-1}} N_{li}\right)$, each term in the summation of (23) is i.i.d. with $\mu_3 = 0$ and $|\varphi|^\nu$ integrable for some $\nu \geq 1$. Invoking Lemma 1, the conditional density of Y_l' given $X_k = x$ satisfies

$$\begin{aligned} &\mathsf{P}\left\{Y_j' \in dy | X_k = x, A_k, \Lambda_k, S_{jk}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_{\mathrm{MAI+N}}^2}} e^{-\frac{\left(y - \frac{1}{\sqrt{\Lambda_k}} S_{jk} A_k x\right)^2}{2\sigma_{\mathrm{MAI+N}}^2}} + o(1/\sqrt{\bar{\Gamma}})\right) dy, \end{aligned} \tag{24}$$

where $\sigma_{\mathrm{MAI+N}}^2$ is

$$\sigma_{\mathrm{MAI+N}}^2 = 1 + \beta \mathsf{E}\left\{A_i^2 \,\mathsf{var}\left\{X_i \mid A_i, V_{i \to j}^{(t-1)}\right\}\right\}.$$

Based on (24), with observation $Y_j' = y$, the LLR incoming vector messages $\{U_{j \to k}^{(t)}(x)\}_x$ become

$$U_{j \to k}^{(t)}(x) = \frac{1}{\sigma_{\mathrm{MAI+N}}^2} y \left(\frac{1}{\sqrt{\Lambda_k}} S_{jk} A_k\right)(x - x_a) \pm o\left(U_{j \to k}^{(t)}(x)\right). \tag{25}$$

Using (25), one can verify that the mean and the variance functions converge to the forms of (21) and (22) as $\bar{\Gamma} \to \infty$. Nonetheless, the above convergence does not imply normality for each individual CN-EQ-CH since the convergence happens only when both the mean and covariance functions are zeros for all x, x' and x''. On the other hand, (25) proves that the CN-EQ-CH has identical mean and variance *scaling laws* with respect to Λ_k as those of ideal Gaussian channels, the coefficients of which can be easily obtained. With the exact coefficients of the scaling laws and by invoking the vector-version of the central limit theorem at the symbol node iteration (8), one can quantify the SNR of the equivalent scalar Gaussian channel for the classical BP, which is identical to the efficiency η computed using the relaxed BP arguments. From the above reasonings for the two cases whether Λ_k is bounded away from infinity or not, Theorems 6 and 7 are proved for the classical BP as well.

V. The Equivalence of BP and MAP

We first recall that the classical BP is an optimal detection rule based on the limited observations $\boldsymbol{Y}^{(t)}$ on the subtree of depth $2t - 1$ rooted at the target symbol X_k. Similarly, we can define $\boldsymbol{X}^{(t)}$ as the symbols on the same subtree. A genie-aided BP can be defined as the optimal detection based on $\boldsymbol{Y}^{(t)}$ as well as all entries of $\boldsymbol{X}$ not in $\boldsymbol{X}^{(t)}$, the latter of which are provided by a genie. Using simple degradation arguments, one can show that the MAP detector for symbol X_k based on $\boldsymbol{Y}$ is a degraded detection rule with respect to the genie-aided BP while the classical BP is a degraded rule when compared to the MAP detector.

Following the same scalar Gaussian channel analysis, one can show that the genie-aided BP admits the same iterative formula as in (10) while the only difference for the genie-aided BP is that the initial $\eta^{(0)}$ is set to 1. Due to the uniqueness of the fixed point of (10) by assumption, the classical and the genie-aided BP will perform identically, namely, the distributions of $p_{X_k}^{(t)}(\cdot, \boldsymbol{Y}^{(t)}, \boldsymbol{S})$ and $p_{X_k}^{(t)}(\cdot, \boldsymbol{Y}^{(t)}, \boldsymbol{S}, \boldsymbol{X} \setminus \boldsymbol{X}^{(t)})$, computed by the classical and the genie-aided BPs respectively, converge to that of a scalar Gaussian channel of identical noise variance $1/\eta$ as stated in Theorem 6. Since the classical BP is a degraded rule of the genie-aided BP that has the same performance, the posterior probabilities computed by the classical BP and by the genie-aided BP must coincide. By a sandwiching argument, the posterior probabilities computed by the classical BP and by the MAP detector must coincide in probability as well. Theorem 4 is proved.

By similar sandwiching arguments while focusing on the l-th chip W_l, Theorem 5 can also be proved. Theorems 6 and 7 now apply for MAP detectors as well, and Theorems 1 and 2 then become direct corollaries of Theorems 4 to 7.

VI. Relationship to the I-MMSE Formula

Let us introduce an additional parameter γ and consider the following fixed-point equation akin to (10):

$$\eta(\gamma) = \frac{1}{1 + \beta\, \mathcal{E}_{X,A}(\gamma\eta(\gamma))}. \tag{26}$$

The notation $\eta(\gamma)$ stresses the fact that the unique fixed point solution of (26) is a function of γ. Suppose further that (26) has a unique fixed point for all $\gamma \geq 0$ (rather than for $\gamma = 1$ only). An alternative proof of Theorem 2 is as follows.

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

Lemma 2 ([14]): For any $\boldsymbol{Z}$ with $\mathsf{E}\|\boldsymbol{Z}\|^2 < \infty$ and independent $\boldsymbol{N} \sim \mathcal{N}(0, \boldsymbol{I})$ of identical dimension,

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} I(\boldsymbol{Z}; \sqrt{\gamma}\,\boldsymbol{Z} + \boldsymbol{N}) = \frac{1}{2}\,\mathcal{E}\left(\boldsymbol{Z} \,|\, \sqrt{\gamma}\,\boldsymbol{Z} + \boldsymbol{N}\right), \quad \gamma \geq 0.$$

The following entropy–MMSE relationship is straightforward,

$$\frac{1}{2}\int_0^\infty \mathcal{E}\left(\boldsymbol{S}\boldsymbol{X} \,|\, \sqrt{\gamma}\,\boldsymbol{S}\boldsymbol{X} + \boldsymbol{N}\right) \mathrm{d}\gamma = \frac{1}{L} H(\boldsymbol{S}\boldsymbol{X} | \boldsymbol{S}). \tag{27}$$

In the large-sparse-system limit, the RHS of (27) converges to $H(\boldsymbol{X})/L = \beta\, H(X)$ because $\boldsymbol{X}$ can be recovered from $\boldsymbol{SX}$ with probability 1.

We have shown in (19) that the MSE of estimating the l-th chip W_l by BP converges to $\beta\,\eta\,\mathcal{E}_{X,A}(\eta)$. In the following, we show that this MSE integrates to the same entropy as in the right-hand side of (27).

For every $\gamma \geq 0$, let $\eta(\gamma)$ be the solution to the fixed-point equation (26). Define

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

Proposition 1: For every $\gamma \geq 0$,

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

Proof: Since $\eta(0) = 1$ and $C(0) = 0$, it suffices to show

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}C(\gamma) = \frac{\eta}{2}\,\mathcal{E}_{X,A}(\eta\gamma) \tag{28}$$

where we use η as a short hand for $\eta(\gamma)$, which is a differentiable function of γ. Using Lemma 2, we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\gamma}C(\gamma)\\
&= \frac{\mathrm{d}}{\mathrm{d}\gamma}I(X; \sqrt{\gamma\eta}\, AX + N|A) + \frac{1}{2\beta}\frac{\mathrm{d}}{\mathrm{d}\gamma}(\eta - 1 - \log\eta)\\
&= \frac{1}{2}\,\mathcal{E}_{X,A}(\gamma\eta)\,\frac{\mathrm{d}}{\mathrm{d}\gamma}(\gamma\eta) + \frac{1}{2\beta}(1-\eta^{-1})\eta'\\
&= \frac{\eta}{2}\,\mathcal{E}_{X,A}(\eta\gamma) + \frac{1}{2\beta}\left[\beta\gamma\,\mathcal{E}_{X,A}(\eta\gamma) + 1 - \eta^{-1}\right]\eta'. \qquad (29)
\end{aligned}$$

(28) follows because the last term in (29) vanishes by (26). ■

By Proposition 1,

$$\frac{1}{2}\int_0^\infty \beta\eta(\gamma)\mathcal{E}_{X,A}\left(\gamma\eta(\gamma)\right)\,\mathrm{d}\gamma = \beta\, C(\infty) = \beta\, H(X). \tag{30}$$

From (27) and (30), it is clear that when considering the problem of estimating individual chip values W_l, the MMSE of the linear system and the MSE achieved by 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 BP detector approaches the MAP detector and achieves the MMSE of estimating W_l. Theorem 2 is thus established.

Moreover, using the integral representation of mutual information implied by Lemma 2, Theorem 3 is proved by Proposition 1.

VII. Conclusion

The asymptotic equivalence of BP and MAP for sparse linear systems is established in the strongest sense of the posterior probability for arbitrary input distributions. The results suggest that BP is near-optimal for sparse linear systems of relatively large size. This phenomenon is contrary to the wisdom for LDPC codes with infinitely long codeword length that the MAP decoder is generally strictly better than BP.

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] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *IEEE Information Theory Workshop*, Chengdu, China, 2006.

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

[13] 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.

[14] 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.

[15] 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.

[16] J. Garcia-Frias and W. Zong, "Approaching near shannon performance by iterative decoding of linear codes with low-density generator matrix," *IEEE Commun. Letters*, vol. 7, pp. 266–268, June 2003.

[17] M. G. Luby, "LT codes," in *Proc. 43rd Annu. IEEE Symp. Foundations of Compuer Science (FOCS'02)*, pp. 271–280, Vancouver, BC, Canada, Nov. 2002.

[18] 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.

[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] T. Fabricius and O. Winther, "Correcting the bias of subtractive interference cancellation in CDMA: Advanced mean field theory," tech. rep., Informatics and Mathematical Modelling, Technical University of Denmark, 2003.

[22] 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.

[23] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II. John Wiley & Sons, Inc., 2nd ed., 1971.