Multiuser Detection of Sparsely Spread CDMA

Dongning Guo, *Member, IEEE,* and Chih-Chun Wang, *Member, IEEE*

***Abstract*—Code-division multiple access (CDMA) is the basis of a family of advanced air interfaces in current and future generation networks. The benefits promised by CDMA have not been fully realized partly due to the prohibitive complexity of optimal detection and decoding of many users communicating simultaneously using the same frequency band. From both theoretical and practical perspectives, this paper advocates a new paradigm of CDMA with sparse spreading sequences, which enables near-optimal multiuser detection using belief propagation (BP) with low-complexity. The scheme is in part inspired by capacity-approaching low-density parity-check (LDPC) codes and the success of iterative decoding techniques. Specifically, it is shown that BP-based detection is optimal in the large-system limit under many practical circumstances, which is a unique advantage of sparsely spread CDMA systems. Moreover, it is shown that, from the viewpoint of an individual user, the CDMA channel is asymptotically equivalent to a scalar Gaussian channel with some degradation in the signal-to-noise ratio (SNR). The degradation factor, known as the multiuser efficiency, can be determined from a fixed-point equation. The results in this paper apply to a broad class of *sparse*, *semi-regular* CDMA systems with arbitrary input and power distribution. Numerical results support the theoretical findings for systems of moderate size, which further demonstrate the appeal of sparse spreading in practical applications.**

***Index Terms*—Code-division multiple access (CDMA), sparse spreading, multiuser detection, belief propagation.**

I. Introduction

In a code-division multiple access (CDMA) system, a number of users communicate with a base station simultaneously over the same frequency band. Typically, each user is assigned a randomly generated spreading sequence (signature) with a large time-bandwidth product. The sequence is used to directly modulate the transmitted waveform in time in direct sequence (DS) CDMA, to control the frequency shift in frequency hopping (FH) CDMA, or is translated to the phase and amplitude of subcarriers in multicarrier (MC) CDMA (see e.g. [1]). The spread of user signals in bandwidth and the ease of frequency planning provide many advantages particularly in wireless applications. However, the potential benefits of CDMA have not been fully realized in practice partly due to the computational complexity of effective *multiuser detection* in presence of multiple-access interference (MAI).

The MAI arises in many CDMA systems in which it is impossible to maintain orthogonality of the spreading sequences of all users. Optimal detection in such systems is equivalent to a test of an exponential number of hypotheses about the data symbols of all users, which is NP-complete and hence prohibitively complex for all but systems with very few users [2]. Numerous suboptimal multiuser detectors with lower complexity have been proposed to mitigate the MAI to various degrees. Most notable ones include the decorrelating detector, the linear minimum mean-square error (MMSE) detector, and the successive and parallel interference cancelers [2].

This work investigates a new paradigm of CDMA, called *sparsely spread CDMA* or simply *sparse CDMA*, which recently emerged in the literature [3], [4]. The key technique is to use sparse spreading sequences whose fraction of nonzero entries is small so that near-optimal performance is provably achievable by the linear-complexity belief propagation (BP) algorithm and its variants. Multiuser detection based on modified BP with heuristic Gaussian approximation was originally proposed for the usual dense CDMA [5]–[7]. The use of sparsity is in part inspired by the success of iterative decoding techniques for low-density parity-check (LDPC) codes. In [4], Yoshida and Tanaka proposed a family of sparse CDMA and analyzed the performance of such schemes in the large-system limit, which refers to the regime where the number of users and the spreading factor both tend to infinity with their ratio fixed. The heuristic statistical physics technique used in [4], known as the replica method, has previously been used to study dense CDMA in the large-system limit. In particular, Tanaka obtained the first analytical error performance for large CDMA systems [8], and Guo and Verdú showed that the CDMA channel followed by multiuser detection can essentially be decoupled into a bank of scalar Gaussian channels, one for each user [9], [10]. Concurrent to [4], Montanari and Tse proposed in [3] a different ensemble of sparse CDMA, and justified Tanaka's formula [8] for the first time in a special case without resorting to replicas. The work [3] has since been generalized to arbitrary input and noisy channels in the context of general sparse linear systems [11]–[13]. More recently, Raymond and Saad [14] studied sparse CDMA of the regular LDPC code type using the replica method and verified their findings numerically.

It is straightforward to apply sparse spreading sequences in CDMA systems. In sparse FH-CDMA and MC-CDMA, sparsity takes place in the frequency domain and is particularly robust to channel impairment. In sparse DS-CDMA, although extensive and dense delay profile generally destroys sparsity of the spreading sequences, sparsity may be preserved to some extent in many wireless dispersive channels subject to a few dominant paths (see, e.g., [15]). Moreover, sparsity is

Manuscript received May 30, 2007; revised October 26, 2007. This material is based upon work supported in part by the NSF CAREER Award CCF-0644344 and DARPA IT-MANET program under Grant W911NF-07-1-0028. This paper was presented in part at the 2007 IEEE Information Theory Workshop in Chengdu, China, and the 44th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA.

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

Chih-Chun Wang is with the Center for Wireless Systems and Applications (CWSA), School of Electrical & Computer Engineering, Purdue University, West Lafayette, IN 47907, USA.

Digital Object Identifier 10.1109/JSAC.2008.080402.

preserved regardless of whether the system is synchronous or asynchronous. One major consequence of sparse spreading in DS-CDMA and FH-CDMA is the increase in peak power because of zero transmission for some fraction of the time.

The CDMA systems considered in this work are described in Section II in full generality. A specific ensemble of sparse spreading matrices is also defined. Section III is devoted to the exposition of BP-based multiuser detection and also serves as a recipe for implementing the BP algorithm.

The main results in this paper are summarized in Section IV. The performance of sparse CDMA with several types of detectors is investigated. Consider first the performance of iterative detection using BP. Under the sparsity condition, detection of each symbol is essentially equivalent to statistical inference on a tree (or tree-like graph) of some depth determined by the number of iterations. The quality of the statistics obtained by BP can be evaluated using density evolution for any given ensemble. An alternative to density evolution is to use the replica method to obtain a set of coupled equations which describe the performance [14]. For the ensemble described in Section II, the node degrees increase without bound in the large-system limit, so that the aggregate likelihood ratio converges to a Gaussian random variable due to the central limit theorem. This allows the asymptotic performance to be determined rigorously using a fixed-point equation. It is interesting to note that density evolution with heuristic Gaussian approximation has been applied to obtain similar fixed-point equations which characterize the performance of dense CDMA systems [16], [17].

In Section V, it is shown that the quality of detection can be simply described by the variance of the likelihood ratio. Interestingly, in the viewpoint of each user, the CDMA channel combined with the BP detector is asymptotically equivalent to a scalar Gaussian channel, where the collective impact of interfering users is a degradation in the signal-to-noise ratio (SNR) of the desired user. The degradation, known as the multiuser efficiency, can be obtained from an iterative formula. In Section V, we also show that BP effectively produces the *a posteriori* probability of the input given the observed channel output in the large-system limit, if the iterative formula for the multiuser efficiency of BP has a unique fixed point. In other words, BP computes a sufficient statistic and is therefore optimal in such cases. The optimality of BP is not sustained if the iterative formula has more than one solution.

This paper adopts the following notational convention unless noted otherwise. Deterministic and random variables are denoted by lowercase and uppercase letters respectively, and scalars, vectors and matrices are distinguished using normal, bold and underlined bold fonts respectively. For any random variable X, let P_X denote its cumulative distribution function (cdf), and p_X denote the corresponding probability mass or density function.

II. System Model

A. CDMA System

Consider a fully-synchronous CDMA system with K users and spreading factor L. User k modulates the symbol X_k onto a spreading sequence $\boldsymbol{S}_k$ with positive amplitude A_k. 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

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

where $N_l \sim \mathcal{N}(0,1)$ are independent standard Gaussian random variables representing noise. Let $\boldsymbol{X} = [X_1, \dots, X_K]^\top$ denote the input vector and $\underline{\boldsymbol{S}} = [\boldsymbol{S}_1, \cdots, \boldsymbol{S}_k]$ denote the spreading matrix. The CDMA channel can be succinctly described by a linear system in Gaussian noise:

$$\boldsymbol{Y} = \underline{\boldsymbol{S}}\underline{\boldsymbol{A}}\boldsymbol{X} + \boldsymbol{N}$$

where $\boldsymbol{N} \sim \mathcal{N}(0, \boldsymbol{I})$, $\underline{\boldsymbol{A}} = \text{diag}(A_1, \dots, A_K)$, and $\{A_k\}$ are independent and identically distributed (i.i.d.) with distribution P_A of a finite fourth-order moment. Note that the amplitudes A_k may be different for different users and also vary over time. Thus fading is inherently considered in the model.

A multiuser detector assumes the symbols $\{X_k\}$ to be i.i.d. and to take values in the alphabet $\chi \subset \mathbb{R}$, which may be discrete or continuous. Let P_X denote the cdf of X_k, which is of zero mean and finite variance. Given $\underline{\boldsymbol{S}}$, $\underline{\boldsymbol{A}}$, and P_X, the job of the multiuser detector is to produce an estimate of $\boldsymbol{X}$ using the output $\boldsymbol{Y}$ of the CDMA system. Oftentimes, the multiuser detector is followed by a decoder for error-control codes, so that a soft estimate of the symbols is most desirable. The *optimal multiuser detector* in general produces the posterior distribution of X_k, i.e., $P_{X_k|\boldsymbol{Y},\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}}(\cdot|\boldsymbol{Y}, \underline{\boldsymbol{S}}\underline{\boldsymbol{A}})$, which is a sufficient statistic for X_k and based on which all classical decision rules can be derived. For example, MMSE estimator outputs the conditional mean $\mathsf{E}\{X_k|\boldsymbol{Y}, \underline{\boldsymbol{S}}\underline{\boldsymbol{A}}\}$ and the maximum *a posteriori* (MAP) detector finds $\hat{X}_k = \arg\max_{x\in\chi} \mathsf{P}\{X_k = x|\boldsymbol{Y}, \underline{\boldsymbol{S}}\underline{\boldsymbol{A}}\}$ in case of discrete alphabet χ. In practice, the alphabet is often very simple and the soft output of the multiuser detector admits many equivalent forms. For example, if the inputs are modulated in binary phase-shift keying (BPSK), where $\chi = \{+1, -1\}$, the optimal soft output is simply the *a posteriori* probabilities of $X_k = +1$ (or -1) given the output $\boldsymbol{Y}$ and the channel matrix $\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}$, and an equivalent and frequent form of the soft output is the log-likelihood ratio (LLR) between the two *a posteriori* probabilities.

B. Factor Graph Representation

Let $\underline{\boldsymbol{s}}$ be a specific realization of the spreading matrix $\underline{\boldsymbol{S}}$. We can construct the corresponding factor graph representation of the CDMA system as depicted in Figure 1. Each user symbol X_k is represented by a circle, called the symbol node, and each received signal entry Y_l corresponds to a square, called the chip node. For any (k,l), symbol node k and chip node l are connected by an edge if $s_{lk} \neq 0$, and each edge connecting k and l is associated with the corresponding gain factor $\frac{s_{lk}}{\sqrt{\Lambda_k}} A_k$. The purpose of the factor graph representation is twofold. First, the probability law of the CDMA system can be completely described using the factor graph, which renders multiuser detection equivalent to statistical inference on the graph. The BP algorithm [18], also known as the sum-product algorithm [19], [20], can thus be derived from

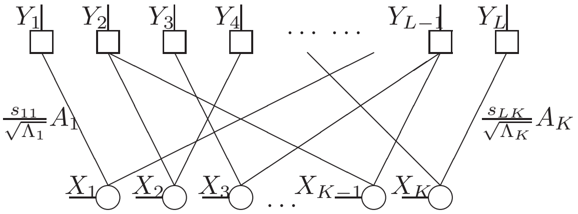


Fig. 1. The Forney-style factor graph for the CDMA system.

the factor graph representation, which will be elaborated in Section III. Secondly, the factor graph representation facilitates discussions of graph-based concepts, e.g., the neighborhood of a node and the graph-based system ensemble. The factor graph representation is very general. In particular, CDMA with usual dense spreading corresponds to a complete bipartite graph.

C. The Limiting Ensemble of Sparse Spreading Matrices

In a sparse CDMA system, all but a small fraction of the entries in the spreading sequence for each user are zero. Each symbol X_k is thus spread over a relatively small subset of the L chips, which is different from traditional CDMA where each symbol is spread over all L chips. A common approach for the analysis of CDMA is to consider the *large-system limit* where $K = \beta L \to \infty$ and allow random selection of the spreading matrix $\underline{\boldsymbol{S}}$, user power profile $\underline{\boldsymbol{A}}$, and user symbol vector $\boldsymbol{X}$. There are at least three types of *sparse system limits*: (i) each symbol is spread onto a *fixed* number of chips; (ii) each symbol is spread onto a random number of chips, the average of which is a fixed number regardless of K and L; and (iii) each symbol X_k is spread onto Λ_k chips where $\Lambda_k \to \infty$ but $\Lambda_k = o\big(L^{1/(4t)}\big)$ as $L \to \infty$ where $t \geq 1$ is a constant to be explained shortly. In any one of the three cases, the ratio of the number of nonzero chips for each user and L vanishes. Types (i) and (ii) are the subject of [14] and [4] respectively, whereas the analysis of this work focuses on type (iii). Numerical results in Section VI show that the asymptotic results are representative for systems with moderate size, where Λ_k are quite small.

Specifically, we consider a sequence of ensembles indexed by the user number K. For any fixed K and L as a function of K, let $\underline{\boldsymbol{S}}$ be randomly constructed as follows. First, an $L \times K$ binary incidence matrix $\underline{\boldsymbol{H}}_{L\times K} = (H_{lk})$ is randomly picked from a certain ensemble to be described shortly. For all (l, k) with $H_{lk} = 0$, set $S_{lk} = 0$. For all (l, k) with $H_{lk} = 1$, S_{lk} are drawn i.i.d. with distribution P_S, which is of zero mean, unit variance, and finite fourth-order moment. 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 (node) degree of X_k, which is the number of chips over which X_k is spread. We define $\bar{\Lambda} = \frac{1}{K}\sum_{k=1}^{K}\Lambda_k$ as the average symbol degree. Similarly, the chip (node) degree and the average chip degree are defined as $\Gamma_l = \sum_{k=1}^{K} H_{lk}$ and $\bar{\Gamma} = \frac{1}{L}\sum_{l=1}^{L}\Gamma_l$ respectively.

Two common classes of ensembles for $\underline{\boldsymbol{H}}_{L\times K}$ is the doubly Poisson ensembles and the regular bipartite graph ensembles. For a doubly Poisson ensemble, the entries $\{H_{lk}\}$ are i.i.d. Bernoulli distributed. A regular bipartite graph ensemble, however, consists of all bipartite graphs with K symbol nodes and L chip nodes such that all symbol nodes are of identical degree and so are all chip nodes, where $\underline{\boldsymbol{H}}$ is the incidence matrix of a uniformly randomly picked graph from the ensemble. Both classes of ensembles satisfy the properties:

1) The ensemble is automorphism under any permutation of the symbol or chip node indices.
2) If the expected average chip degree $\mathsf{E}\bar{\Gamma}$ grows as slow as $o\big(L^{1/(4t)}\big)$ with L, then for every k, the probability that X_k is involved in a cycle of length shorter than t approaches zero as $K = \beta L \to \infty$. This is often called the *asymptotic cycle-free* (A.C.F.) property [21].
3) As $\mathsf{E}\bar{\Gamma} \to \infty$, the chip degrees concentrate around their expected average, i.e., for every l and $\epsilon > 0$,

$$\lim_{K=\beta L\to\infty} \mathsf{P}\left\{|\Gamma_l - \mathsf{E}\bar{\Gamma}| > \epsilon\, \mathsf{E}\bar{\Gamma}\right\} = 0. \qquad (2)$$

We call an ensemble *chip-semi-regular* if it satisfies (2).

The analysis in this paper applies to any class of ensembles satisfying the above three properties, which hold for most popular ensembles including symbol-irregular bipartite graph ensembles. See [22] for more examples.

Once the ensemble of $\underline{\boldsymbol{H}}$ is determined, which in turn determines the ensemble of $\underline{\boldsymbol{S}}$, we consider the large-system limit with $K = \beta L, \mathsf{E}\bar{\Gamma} \to \infty$ such that $\mathsf{E}\bar{\Gamma}$ grows sufficiently slowly so that the asymptotic cycle-free property is ensured. Throughout this paper, we use $\lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}\to\infty,\,\text{A.C.F.}}}$ to denote this large-system limit for a sparse system ensemble satisfying the A.C.F. property, which is also simply referred to as the *large-sparse-system limit*.

Finally, we assume that the input distribution P_X, the user power profile P_A, the chip distribution P_S, and the system load β remain fixed regardless of the system size (K, L).

III. Belief Propagation for Multiuser Detection

BP is an efficient iterative message-passing algorithm for computing the marginal posterior distributions, which is devised based on the factor graph of the underlying Bayesian inference networks. Each node in the factor graph sends “messages” to its neighbors during each “iteration” and after several iterations, inference can be made based on the messages exchanged in the final round. If applied to a factor graph free of cycles, BP produces the exact marginal posterior distribution and is thus *optimal* in information-theoretic sense. BP is also frequently applied to graphs with cycles and is known to produce good approximation of the posterior marginals in practice, even though its (suboptimal) performance is notoriously difficult to quantify on loopy graphs [23], [24]. Nonetheless, for practical sparse CDMA (generally with a loopy factor graph), our main results characterizes rigorously its performance with simplicity, a unique feature of sparse CDMA when compared to other loopy inference problems. In the following we first provide the iterative formulas of BP for ready implementation in practical CDMA systems. The underlying theory will then be discussed to justify the algorithm and set the stage for the analytical results.

A. The BP Algorithm for Multiuser Detection

Consider the CDMA system described in Section II, where the spreading matrix $\underline{\boldsymbol{S}}$ is randomly picked from an ensemble of sparse matrices. Let the realization of $\underline{\boldsymbol{S}}$ and the amplitudes $\underline{\boldsymbol{A}}$ be denoted by $\underline{\boldsymbol{s}}$ and $\underline{\boldsymbol{a}}$ respectively, so that conditioning on the channel state is dropped in the remainder of this section.

Consider the bipartite graph depicted in Figure 1. In each iteration of BP, messages are first sent from symbol nodes to chip nodes; each chip node then computes messages to send back to the symbol nodes based on the previously received messages. These chip-to-symbol messages will then be used to generate the new symbol-to-chip messages in the next iteration. Let $\left\{V_{k\to l}^{(t)}(x)\right\}_{x\in\chi}$ or the shorthand $\left\{V_{k\to l}^{(t)}(\chi)\right\}$ represent the message from symbol node k to chip node l and $\left\{U_{l\to k}^{(t)}(\chi)\right\}$ represent the message in the reverse direction at the t-th iteration. The messages represent generally the *extrinsic* information of X_k in some form (see e.g., [19]).

For convenience, the alphabet χ is assumed to be discrete and finite so that each message can be represented as a vector with dimension equal to the cardinality of χ. Let $p_X(x)$ denote the probability mass function of the symbol distribution P_X and the statement "$V(x) \propto u(x)$" means that $\forall x \in \chi$, $V(x) = c \cdot u(x)$ for some constant c such that $\sum_{x\in\chi} V(x) = 1$. Let ∂l (resp. ∂k) denote the subset of symbols (resp. chips) connected directly to chip l (resp. symbol k), called its neighborhood. Also, let $\partial l\backslash k$ denote the neighborhood of chip node l excluding symbol node k.

The iterative BP algorithm for computing the (approximate) *a posteriori* distribution of all symbols is described as follows.

```
Input: $\underline{\boldsymbol{s}}$, $\underline{\boldsymbol{a}}$, $\boldsymbol{y}$.
Initialization:
for all $k = 1,\ldots,K$ and $l = 1,\ldots,L$ do
    $U_{l\to k}^{(0)}(x) \leftarrow 1$ for all $x \in \chi$.
Main Iterations:
for $t = 1$ to $T$ (typically ranging from 20 to 40) do
    for all $k$, $l$ with $s_{lk} \neq 0$ and $x \in \chi$ do
        $V_{k\to l}^{(t)}(x) \propto p_X(x) \prod_{j\in\partial k\backslash l} U_{j\to k}^{(t-1)}(x)$    (3a)
    for all $k$, $l$ with $s_{lk} \neq 0$ and $x \in \chi$ do
        (Let $\sum_{(x_i)_{\partial l\backslash k}}$ denote sum over all possible values of $x_i \in \chi$ for all $i \in \partial l\backslash k$.)
        $U_{l\to k}^{(t)}(x) \propto \sum_{(x_i)_{\partial l\backslash k}} \exp\left[-\frac{1}{2}\left(y_l - \frac{s_{lk}}{\sqrt{\Lambda_k}}a_k x - \sum_{i\in\partial l\backslash k}\frac{s_{li}}{\sqrt{\Lambda_i}}a_i x_i\right)^2\right] \prod_{i\in\partial l\backslash k} V_{i\to l}^{(t)}(x_i)$    (3b)
return $V_k(x) \propto p_X(x) \prod_{j\in\partial k} U_{j\to k}^{(T)}(x)$, $x\in\chi$, $k=1,\ldots K$.
```

At termination, $\{V_k(x)\}_{x\in\chi}$ is the (approximate) posterior probability density or probability mass function of symbol X_k produced by the BP algorithm for each k.

The complexity of the algorithm depends on the number of nonzero entries in the spreading matrix $\underline{\boldsymbol{s}}$, and is in fact exponential in the chip degree Γ_l because of summation in (3b). Nonetheless, with relatively small values of Γ_l and $|\chi|$, the complexity per symbol is essentially linear with respect to the spreading factor L (hence also K) and the number of iterations. This makes the algorithm particularly suitable for sparse CDMA with binary or quaternary modulation and small node degrees (e.g., $\Gamma_l < 10$). At the cost of degraded performance, the above algorithm can be further simplified for detection of dense CDMA with a complexity also linear in $\bar{\Gamma}$ [5]–[7]. This is achieved by noting that the summation (3b) can be reduced if each $\left\{V_{i\to l}^{(t)}(x)\right\}$ is replaced by its *Gaussian approximation* and the sum over $(x_i)_{i\in\partial l\backslash k}$ is replaced by an expectation over a Gaussian density.

B. The BP Algorithm: Justification and Implementation

In the following, we demonstrate how to derive the message-passing algorithm from the viewpoint of statistical inference on graph assuming that the spreading matrix is sparse. For pedagogical reasons, the messages are represented in the form of the log-likelihood ratio, which will eventually be reduced to the unnormalized posteriors seen in (3). Let us fix a reference point $x_0 \in \chi$ throughout the paper. In general, the LLR function of some observation $Z = z$ about X is defined as[1]

$$\mathcal{L}_{Z|X}(z|x) = \log\frac{p_{Z|X}(z|x)}{p_{Z|X}(z|x_0)} = \log\frac{p_{X|Z}(x|z)}{p_{X|Z}(x_0|z)} - \log\frac{p_X(x)}{p_X(x_0)}$$

for all $x \in \chi$, where $p_{Z|X}$ is the conditional probability density function of Z given X. Clearly, the LLR $\mathcal{L}_{Z|X}(z|\cdot)$ and the *a posteriori* distribution $P_{X|Z}(\cdot|z)$ determine each other for every z. The LLR is a sufficient statistic of Z for X and contains the extrinsic information about X.

Consider first the problem of inferring about an arbitrary symbol X_k using only one chip Y_l which is immediately connected to X_k. A sufficient statistic is the *a posteriori* probability of X_k, which is expressed as $P_{X_k|Y_l}(x|Y_l)$, $x \in \chi$, or equivalently, the LLR function $\mathcal{L}_{Y_l|X_k}$. This statistic is readily computed from the model (1), where all the other neighboring X_i of Y_l, $i \in \partial l\backslash k$ are i.i.d. with the prior distribution P_X:

$$\begin{aligned}\mathcal{L}_{Y_l|X_k}(y|x) &= \log\frac{p_{Y_l|X_k}(y|x)}{p_{Y_l|X_k}(y|x_0)}\\ &= \log\frac{\mathsf{E}\left\{p_{Y_l|\boldsymbol{X}}(y|\boldsymbol{X}) \mid X_k = x\right\}}{\mathsf{E}\left\{p_{Y_l|\boldsymbol{X}}(y|\boldsymbol{X}) \mid X_k = x_0\right\}}.\end{aligned} \tag{4}$$

Thus the expectation in (4) can be evaluated as

$$\begin{aligned}&\mathsf{E}\left\{p_{Y_l|\boldsymbol{X}}(y|\boldsymbol{X}) \mid X_k = x\right\}\\ &= \sum_{\substack{(x_i)_{\partial l\backslash k}\\ x_k = x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(y - \sum_{i\in\partial l}\frac{s_{li}}{\sqrt{\Lambda_i}}a_i x_i\right)^2} \prod_{i\in\partial l\backslash k} p_X(x_i).\end{aligned} \tag{5}$$

Consider the problem of inferring about X_k using all chips which are immediately connected to X_k, i.e., all chip nodes with distance one to X_k on the factor graph. Let the chips be denoted by $Y_{l_1},\ldots,Y_{l_{\Lambda_k}}$ as depicted in Figure 2. An LLR is obtained for each Y_l according to (4). We need a crucial assumption to facilitate the development of the BP algorithm, which is that the subsets of symbols connected to the Λ_k chips do not overlap except for X_k, so that they are conditionally independent given X_k. The posterior distribution of X_k can

[1] We assume natural logarithm throughout the paper.

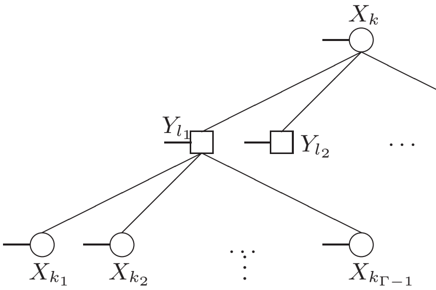


Fig. 2. Statistical inference over the corresponding tree structure.

be obtained from the LLRs (4) and is equivalent to the LLR

$$\mathcal{L}_{Y_{l_1},\dots,Y_{l_{\Lambda_k}}|X_k}(y_{l_1},\dots,y_{l_{\Lambda_k}}|x)=\sum_{l=l_1,\dots,l_{\Lambda_k}}\mathcal{L}_{Y_l|X_k}(y_l|x). \quad (6)$$

In view of Figure 2, the LLR (6) is obtained by inferring about X_k using all chips on a subtree of the factor graph of depth two with X_k as its root, where all leaf (symbol) nodes on the subtree are assumed to be distinct and i.i.d. with the prior p_X.

The above computation can be carried out to obtain the posterior distribution of every symbol given its neighboring chips. Now that we have a posterior about each symbol, this computation can be repeated for each symbol using the posteriors in lieu of the prior distribution p_X. In fact the LLR in (4) is computed in the same manner, except that the expectation (5) should be carried out with $p_X(x_i)$ in the equation replaced by the posterior distribution of X_i induced by the newly obtained LLRs. As a result, inference about each symbol X_k is based on all chips on a subtree of depth four with X_k as the root. Again, the inference is exact if all leaf (symbol) nodes on the depth-four subtree are distinct and i.i.d.

The above computation can be repeated any number of times, each time using the LLR obtained in a previous iteration to improve on the estimate of the new LLRs, and thereby enlarging the subset of chips used for inference about each symbol. In order for proper Bayesian inference, we require that while computing the message from one node to another, the previous message from the destination node is not included in computing the new message intended for the same node. This algorithm can be summarized in terms of the LLRs as the following. Let $L^{(t)}_{k\to l}$ and $L^{(t)}_{l\to k}$ represent the LLR messages passed between symbol k and chip l at the t-th iteration. Then

$$L^{(t)}_{k\to l}(x)=\sum_{j\in\partial k\setminus l}L^{(t-1)}_{j\to k}(x). \quad (7a)$$

and

$$L^{(t)}_{l\to k}(x) = \log\frac{\mathsf{E}\left\{p_{Y_l|\boldsymbol{X}}(y_l|\boldsymbol{X})\mid X_k=x,\{L^{(t)}_{i\to l}(\chi)\}_{i\in\partial l\setminus k}\right\}}{\mathsf{E}\left\{p_{Y_l|\boldsymbol{X}}(y_l|\boldsymbol{X})\mid X_k=x_0,\{L^{(t)}_{i\to l}(\chi)\}_{i\in\partial l\setminus k}\right\}} \quad (7b)$$

In particular, the expectations conditioned on the set of $L^{(t)}_{i\to l}$ in (7b) can be regarded as expectation with respect to $X_i\sim P^{(t)}_{X_i}$ where the probability mass function $p^{(t)}_{X_i}(x)\propto p_X(x)\exp\left[L^{(t)}_{i\to l}(x)\right]$, $\forall x\in\chi$.

Formulas (7) are not particularly convenient for computation, e.g., due to the normalization in (7b) in order to obtain the LLRs. In fact normalization is not necessary in intermediate stages. Consider passing messages proportional to $\exp[L]$, which denote unnormalized posteriors, instead of using L, (7) can be rewritten in the form of (3) in Algorithm 1 which is more efficient in terms of implementation.

The iterative formulas (7) perform exact marginalization of each symbol X_k given the entire observation $\boldsymbol{Y}$ in at most K iterations if the factor graph contains no cycle. In the CDMA model where the average node degree is always greater than 2, cycles are inevitable, and it is impossible to maintain that all leaf nodes of each subtree of depth $2t$ be distinct for large t. Thus, the BP algorithm performs approximate inference by *assuming* that the leaf nodes are always distinct and i.i.d. Note that sparse spreading implies that the graph is locally tree-like, which generally leads to little performance degradation due to myopia of the algorithm. Such a phenomenon has previously been observed in the BP decoding of LDPC codes (cf. the concentration theorem in [21]).

Finally, we remark that the finite-alphabet constraint on χ can in principle be dropped (e.g., P_X can be Gaussian), while the BP algorithm still computes approximate posteriors, except that the messages defined on χ will have infinite dimension, in which case the BP algorithm in the form of (3) is not practical.

IV. Main Results

This section presents several large-system results on multiuser detection in sparse CDMA systems, the proof of which are relegated to Section V. Of particular interest is the quality of the estimate obtained by optimal detection as well as BP-based suboptimal detection. Throughout this section, we assume that the symbols $X_k\sim P_X$ are i.i.d., the amplitudes $A_k\sim P_A$ are i.i.d., and the spreading matrix is randomly chosen from the ensemble described in Section II-C.

A. The Asymptotic Performance of BP

Consider the problem of inferring about an individual symbol X_k using the BP algorithm. After t iterations, the output of BP is a posterior distribution for X_k computed based on all observations at chip nodes within distance $2t-1$ to X_k on the factor graph, denoted by $\boldsymbol{Y}^{(t)}_k$. With slight abuse of notation, let $P^{\text{bp}}_{X_k}(\cdot|\boldsymbol{Y}^{(t)}_k,\underline{\boldsymbol{SA}})$ denote the output cdf of BP, which is the approximate posterior of X_k given $\boldsymbol{Y}^{(t)}_k$ and the channel matrix. For simplicity, we omit the adjective "approximate" as long as it is clear from the context that the output of the BP-based detection is referred to. A key result in this paper states that the posterior computed for each symbol using BP after t iterations essentially converges to the posterior of a scalar Gaussian channel as the size of the CDMA system increases.

Let us introduce the canonical scalar Gaussian channel:

$$Z=\sqrt{g}X+N \quad (8)$$

where $X\sim P_X$ and $N\sim\mathcal{N}(0,1)$ are independent, and g denotes the gain of the channel in SNR. Throughout this paper, we use $P_{X|Z;g}(\cdot|z;g)$ to denote the cdf of the posterior distribution of the input X given $Z=z$, according to the Gaussian model (8), parameterized by g.

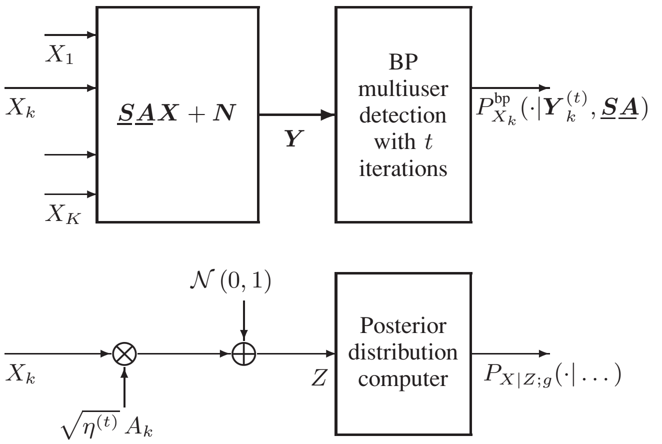

X_1
X_k
X_K
$\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}\boldsymbol{X}+\boldsymbol{N}$
$\boldsymbol{Y}$
BP multiuser detection with t iterations
$P^{\text{bp}}_{X_k}(\cdot|\boldsymbol{Y}^{(t)}_k,\underline{\boldsymbol{S}}\underline{\boldsymbol{A}})$
$\mathcal{N}(0,1)$
X_k
$\sqrt{\eta^{(t)}}\,A_k$
Z
Posterior distribution computer
$P_{X|Z;g}(\cdot|\dots)$

Fig. 3. Upper diagram: Multiuser channel and BP detection. Lower diagram: The asymptotically equivalent scalar Gaussian channel.

Theorem 1: Fix the number of iterations t. For every k and x where $P_X(x)$ is continuous,

$$P^{\text{bp}}_{X_k}\big(x\big|\boldsymbol{Y}^{(t)}_k,\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}\big)\to P_{X|Z;g}\big(x\big|h\big(\boldsymbol{Y}^{(t)}_k,\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}\big);\eta^{(t)}A_k^2\big)$$

in probability in the large-sparse-system limit for some $\eta^{(t)}\in[0,1]$ and some function $h(\cdot)$ such that $h\big(\boldsymbol{Y}^{(t)}_k,\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}\big)\sim\mathcal{N}\big(\sqrt{\eta^{(t)}}\,ax,1\big)$ conditioned on $X_k=x$ and $A_k=a$. Moreover, $\eta^{(0)}=0$ and $\eta^{(t)}$, $t=1,2,\dots$, are determined by the following recursion:

$$\frac{1}{\eta^{(t)}}=1+\beta\,\overline{\mathsf{var}}\left\{AX\,\middle|\,\sqrt{\eta^{(t-1)}}\,AX+N,A\right\}\qquad(9)$$

where $N\sim\mathcal{N}(0,1)$ and

$$\overline{\mathsf{var}}\{U\,|V\}=\mathsf{E}\left\{(U-\mathsf{E}\{U\mid V\})^2\right\}$$

stands for the average conditional variance (or equivalently the MMSE) of estimating random variable U given observation V.

Theorem 1 states that the problem of estimating each individual symbol X_k using t iterations of belief propagation is asymptotically equivalent to that of estimating the same symbol through a scalar Gaussian channel with SNR equal to $\eta^{(t)}A_k^2$, e.g., with a degradation $\eta^{(t)}$ in the input SNR. The parameter $\eta^{(t)}$ is termed the *multiuser efficiency* of the BP detector after t iterations. This type of information equivalence is referred to as the *decoupling principle* (see, e.g., [9], [10]), because the collective effect of the noise and the interference from other users to the desired user is equivalent to an additive Gaussian noise of zero-mean and known variance. Note that formulas similar to (9) have been obtained in the context of iterative decoding of coded CDMA using an empirically inspired Gaussian approximation [16], [17], whereas the derivation of (9) is rigorous.

Interestingly, formula (9) involves the MMSE of estimating AX as the input to the scalar Gaussian channel shown in Figure 3. The function $h(\cdot)$ finds essentially a Gaussian sufficient statistic for the inference problem. We relegate discussion of the function h to Section V.

We refer to Figure 3 for an interpretation of the result. As a consequence of Theorem 1, if an observer has access only to the posterior of X_k computed by BP after t iterations (the upper information flow) and the posterior distribution of X_k computed through a scalar Gaussian channel (the lower information flow), the observer is not able to distinguish these two models from each other for a large system.

B. Optimal Detection and Its Relation to BP

Let $P_{X_k|\boldsymbol{Y},\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}}(\cdot|\boldsymbol{Y},\underline{\boldsymbol{S}}\underline{\boldsymbol{A}})$ denote the actual posterior cdf of X_k given the received signal $\boldsymbol{Y}$ and the channel state $\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}$.

Theorem 2: Suppose the following equation

$$\frac{1}{\eta}=1+\beta\,\overline{\mathsf{var}}\{AX\,|\sqrt{\eta}AX+N,A\}\qquad(10)$$

with $N\sim\mathcal{N}(0,1)$ has a unique fixed point η. Then for every k and x where $P_X(x)$ is continuous,

$$P_{X_k|\boldsymbol{Y},\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}}\big(x\big|\boldsymbol{Y},\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}\big)\to P_{X|Z;g}\big(x\big|h(\boldsymbol{Y},\underline{\boldsymbol{S}}\underline{\boldsymbol{A}});A_k^2\eta\big)$$

in probability in the large-sparse-system limit, where $h(\cdot)$ is such that $h(\boldsymbol{Y})\sim\mathcal{N}(\sqrt{\eta}\,ax,1)$ conditioned on $X_k=x$ and $A_k=a$.

Theorem 2 states that the problem of estimating each X_k given the entire observation $\boldsymbol{Y}$ is also asymptotically equivalent to estimating the same symbol through a scalar Gaussian channel. The multiuser efficiency is determined by fixed-point equation (10), which corresponds to iterative formula (9), and was originally obtained in [9] for dense CDMA using the replica method.

It can be shown that $\lim_{t\to\infty}\eta^{(t)}=\eta$ as long as (10) has a unique solution. In this case, Theorems 1 and 2 together imply that the quality of the posterior obtained by BP is asymptotically as good as that obtained by optimal detection. Precisely, the following result is established in Section V.

Corollary 1: If (10) has a unique fixed point, then

$$\begin{aligned}\lim_{t\to\infty}\ \lim_{\substack{K=\beta L\to\infty\\ \bar{\Gamma}\,\to\,\infty,\,\text{A.C.F.}}}\ &\Big|P^{\text{bp}}_{X_k}\big(x\big|\boldsymbol{Y}^{(t)},\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}\big)\\ &-P_{X_k|\boldsymbol{Y},\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}}\big(x\big|\boldsymbol{Y},\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}\big)\Big|\to 0\end{aligned}\qquad(11)$$

in probability for every k and x where $P_X(x)$ is continuous.

Corollary 1 implies that essentially the same posterior about each symbol is obtained either using $\boldsymbol{Y}^{(t)}$ or using $\boldsymbol{Y}$ in large sparse systems as the number of iterations t becomes large, although $\boldsymbol{Y}^{(t)}$ is a much smaller vector compared to $\boldsymbol{Y}$ as the ratio of their dimensions vanishes for large systems.

It is important to note the order of limits taken in (11): The system size goes to infinity before the number of iterations. The two limits do not commute in general. In fact, the posterior obtained by BP may not converge if the number of iterations is sent to infinity first for a finite-size system.

The fixed-point equation (10) has at least one solution. As η varies from 0 to 1 ($1/\eta$ from ∞ to 1), the right-hand side (RHS) of (10) decreases continuously to a number greater than 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 sufficiently small β, however, the slope of the RHS as a function of η is negligible, so that the solution to (10) is unique.

Theorem 2 characterizes the optimal performance one can hope to achieve when the solution of (10) is unique. What if the solution to (10) is not unique, which generally corresponds to the scenario when the system load β is large? Let η_0 and η_1 denote the smallest and the largest fixed point of (10) respectively. The performance of BP is characterized by η_0 because iterative formula (9) leads to η_0 with $\eta^{(0)} = 0$. It is not known whether the decoupling principle still holds true in this case. In Section V, we show that the quality of optimal detection is inferior than the posterior of the scalar Gaussian channel $X \mapsto \sqrt{\eta_1} AX + N$.

C. Error Performance and Mutual Information

Practical measures of the performance of the CDMA system, such as the probability of error, the output signal-to-interference ratio (SIR), and the mutual information, can all be derived from the posterior distribution generated by BP or optimal detection. According to the decoupling principle stated in Theorems 1 and 2, these performance measures by optimal and BP-based detection in large sparse systems can be analytically expressed or numerically computed at ease using the equivalent scalar Gaussian channel as a proxy of the high-dimensional input-output system.

For example, the output SIR for user k achieved by t iterations of BP detection is equal to $a_k^2\,\eta^{(t)}$. If (10) has a unique fixed point η, then the output SIR of optimal detection is simply equal to $a_k^2\,\eta$. In case of binary inputs with $\mathsf{P}\{X_k = -1\} = \mathsf{P}\{X_k = 1\} = 1/2$ for all k, the bit-error probability achieved by t iterations of BP is

$$\mathsf{P}\left\{X_k \neq \hat{X}_{k,BP}\right\} = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{a_k^2\,\eta^{(t)}}}^{\infty} e^{-\frac{z^2}{2}}\, \mathrm{d}z. \tag{12}$$

Similar can be said about the bit-error probability of a maximum *a posteriori* (MAP) detector if (10) has a unique fixed point η, where $\eta^{(t)}$ in (12) is replaced by η.

Another performance measure of interest is the MMSE (i.e., the average conditional variance). By Theorem 1,

$$\overline{\mathsf{var}}\left\{X_k \,\middle|\, \boldsymbol{Y}^{(t)}, \underline{\boldsymbol{SA}}\right\} \to \overline{\mathsf{var}}\left\{X \,\middle|\, \sqrt{\eta^{(t)}} AX + N, A\right\} \tag{13}$$

in the large-sparse-system limit. By Theorem 2, (13) also holds literally with the superscript (t) removed from both sides.

Furthermore, Theorems 1 and 2 imply that the mutual information between each individual input symbol and the output of BP detection conditioned on the input power is

$$I\left(X_k; \boldsymbol{Y}^{(t)} | A_k\right) \to I\left(X; \sqrt{\eta^{(t)}} AX + N | A\right)$$

in the large-sparse-system limit, where $A_k, A \sim P_A$, $X \sim P_X$ and $N \sim \mathcal{N}(0,1)$. Similarly, for the optimal detector,

$$I\left(X_k; \boldsymbol{Y} | A_k\right) \to I\left(X; \sqrt{\eta} AX + N | A\right) \tag{14}$$

if (10) has a unique fixed point. Note that (14) specifies the maximum achievable rate for user k in a CDMA system with *joint detection* of all symbols but *separate decoding* of the information stream of each user. Higher throughput can be achieved by joint detection and decoding, which is given by the following result.

Theorem 3: Suppose (10) has a unique fixed point η. The input–output mutual information per dimension of the sparse CDMA system converges in the large-sparse-system limit:

$$\frac{1}{K} I(\boldsymbol{X}; \boldsymbol{Y}) \to I(X; \sqrt{\eta}\, AX + N | A) + \frac{\eta - 1 - \log \eta}{2\beta} \tag{15}$$

where the unit of information is nats.

V. Proof

A. The Asymptotic Performance of BP

In this subsection, we prove Theorem 1 by considering messages of the LLR form and by applying the central limit theorem to (7). We consider again those P_X with finite support χ and our results can be generalized to continuous cases using the Kolmogorov extension theorem. In deriving the asymptotic performance formula for the relaxed BP, we first make the following observations.

Observation 1: Consider the scalar Gaussian channel

$$Y = \sqrt{\gamma} X + N \tag{16}$$

where $N \sim \mathcal{N}(0,1)$ and $\gamma > 0$ is the SNR. Fix a reference symbol $x_0 \in \chi$. Conditioned on $X = x \in \chi$, the LLRs $\left\{\mathcal{L}_{Y|X}(Y|x_1) = \log \frac{p_{Y|X}(Y|x_1)}{p_{Y|X}(Y|x_0)}\right\}_{x_1 \in \chi}$ can be regarded as a random vector of dimension equal to the cardinality of χ. Then the random vector is Gaussian distributed. In particular, its mean and convariance functions are

$$\mathsf{E}\left\{\mathcal{L}_{Y|X}(Y|x_1)\right\} = \gamma x(x_1 - x_0) - \gamma(x_1^2 - x_0^2)/2$$

$$\mathsf{cov}\big(\mathcal{L}_{Y|X}(Y|x_1), \mathcal{L}_{Y|X}(Y|x_2)\big) = \gamma(x_1 - x_0)(x_2 - x_0).$$

Moreover, the converse also holds true. That is, for any channel $X \mapsto Z$, if the LLR vector is Gaussian with the mean and covariance identical to those given above, the channel must be statistically equivalent to the scalar Gaussian channel (16).

Proof: The direct part of the observation is due to

$$\mathcal{L}_{Y|X}(Y|x_1) = \sqrt{\gamma}\,(x_1 - x_0)Y - \gamma(x_1^2 - x_0^2)/2$$

obtained from $p_{Y|X}(y|x) = \exp\left[-\frac{1}{2}(y - \sqrt{\gamma}x)^2\right] / \sqrt{2\pi}$. The converse can be shown by reconstructing the channel characteristic based on the LLRs, which is omitted here. ■

Observation 2: For any number of iterations t, there is no cycle in the subgraph induced by all nodes within distance $2t-1$ from X_k with probability 1 in the large-system limit. Therefore, in view of (7), all $L_{j\to k}^{(t-1)}(x)$ are i.i.d. conditioned on $X_k = x_k$. By the central limit theorem, $\left\{L_{k\to l}^{(t)}(x)\right\}_{x\in\chi}$ is a Gaussian distributed vector.

From the above two observations, proving Theorem 1 is equivalent to showing that the mean and covariance of $L_{k\to l}^{(t)}(x)$ described by (7a) admits a form in Observation 1 and with $\gamma = A_k^2 \eta^{(t)}$ where $\eta^{(t)}$ satisfies (9). To this end, we focus on quantifying the *scaling law* of the mean and covariance of $L_{j\to k}^{(t-1)}(x)$ or $L_{j\to k}^{(t)}(x)$ with respect to $\frac{1}{\Lambda_k}$. We first recall the definition of $p_{X_i}^{(t)}(x) \propto p_X(x) \exp\left[L_{i\to j}^{(t)}(x)\right]$, $\forall x \in \chi$. Define

$$W_j = \sum_{i \in \partial j \setminus k} \frac{S_{ji}}{\sqrt{\Lambda_i}} A_i x_i \tag{18}$$

and let three functions $f_m(y)$ for $m = 0, 1, 2$,

$$f_m(y) = \sum_{(x_i)_{\partial j\backslash k}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-W_j)^2} \times \left[(y-W_j)^m - \delta_{m,2}\right] \prod_{i\in\partial j\backslash k} p_X^{(t)}(x_i) \tag{19}$$

where $\delta_{m,2}$ is equal to 1 if $m = 2$ and 0 otherwise. Then by the Taylor series expansion to the second order,

$$\begin{aligned}
&\mathsf{E}\left\{ p_{Y_j|\boldsymbol{X}}(y|\boldsymbol{X}) \;\middle|\; X_k = x, \underline{\boldsymbol{S}}, \underline{\boldsymbol{A}}, \{L^{(t)}_{i\to j}(\chi)\}_{i\in\partial j\backslash k} \right\} \\
&= \sum_{(x_i)_{\partial j\backslash k}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(y - W_j - \frac{S_{jk}}{\sqrt{\Lambda_k}}A_k x\right)^2} \prod_{i\in\partial j\backslash k} p_X^{(t)}(x_i) \\
&= f_0(y) + f_1(y)\frac{S_{jk}}{\sqrt{\Lambda_k}}A_k x + f_2(y)\frac{S_{jk}^2}{\Lambda_k}A_k^2 x^2 + o\left(\frac{1}{\Lambda_k}\right)
\end{aligned} \tag{20}$$

where $f_0(y)$ to $f_2(y)$ correspond to the zeroth to the second order terms of the expansion. Note that the third order derivative of $\exp\left[-(\cdot)^2/2\right]$ is bounded (i.e., independent of y_j). The remainder term can thus be moved out of the summation. Following (20) and Taylor series expansion of $\log(1+x)$, we have by (7b)

$$\begin{aligned}
L^{(t)}_{j\to k}(x) = &\frac{f_1(y_j)}{f_0(y_j)}\frac{S_{jk}}{\sqrt{\Lambda_k}}A_k(x - x_0) \\
&+ \frac{f_2(y_j)}{f_0(y_j)}\frac{S_{jk}^2}{\Lambda_k}A_k^2(x^2 - x_0^2) \\
&- \frac{1}{2}\frac{f_1^2(y_j)}{f_0^2(y_j)}\frac{S_{jk}^2}{\Lambda_k}A_k^2(x^2 - x_0^2) + o\left(\frac{1}{\Lambda_k}\right).
\end{aligned} \tag{21}$$

We are interested in the scaling law of the mean and the covariance of $L^{(t)}_{j\to k}(x)$ over the ensemble.

We discuss each term in (21) separately in the following. By (19) and Taylor series expansion to the first order,

$$\begin{aligned}
&\mathsf{E}\left\{ \frac{f_1(Y_j)}{f_0(Y_j)} \;\middle|\; X_k = x_k, \underline{\boldsymbol{S}}, \underline{\boldsymbol{A}}, \{L^{(t)}_{i\to j}(\chi)\}_{i\in\partial j\backslash k} \right\} \\
&= \int \sum_{(x_i)_{\partial j\backslash k}} \frac{f_1(y)}{f_0(y)} e^{-\frac{1}{2}\left(y - W_j - \frac{S_{jk}}{\Lambda_k}A_k x_k\right)^2} \prod_{i\in\partial j\backslash k} p^{(t)}_{X_i}(x)dy \\
&= \int f_1(y)dy + \int \frac{f_1^2(y)}{f_0(y)}dy\frac{S_{jk}}{\sqrt{\Lambda_k}}A_k x_k + o\left(\frac{1}{\sqrt{\Lambda_k}}\right)
\end{aligned} \tag{22}$$

where the integrability of the remainder term follows from the finiteness of χ and the properties of $\exp\left[-(\cdot)^2/2\right]$. Throughout, we assume all integrals with respect to y are from $-\infty$ to ∞. It is ease to verify that $\int f_1(y)dy = 0$. Therefore, based on (22), the first term in (21) contributes $\int \frac{f_1^2(y)}{f_0(y)}dy\frac{S_{jk}^2}{\Lambda_k}A_k^2 x_k(x - x_0)$ to the mean. By applying similar techniques and noting that $\int f_2(y)dy = 0$, it can be shown that the second term of (21) does not contribute to the mean while the third term contributes $\int \frac{f_1^2(y)}{f_0(y)}dy\frac{S_{jk}^2}{2\Lambda_k}A_k^2(x_0^2 - x^2)$. Similarly, to the covariance between $L^{(t)}_{j\to k}(x_1)$ and $L^{(t)}_{j\to k}(x_2)$, the first term contributes $\int \frac{f_1^2(y)}{f_0(y)}dy\frac{S_{jk}^2}{\Lambda_k}A_k^2(x_1 - x_0)(x_2 - x_0)$ while the other two terms contribute zero. In all, the scaling law of the mean and covariance of $L^{(t)}_{j\to k}(x)$ is decided and the mean and the covariance of the Gaussian vector $L^{(t)}_{k\to l}(x)$ (conditioned on $X_k = x_k$, $\underline{\boldsymbol{S}}$, $\underline{\boldsymbol{A}}$, and $\{L^{(t-1)}_{i\to j}(\cdot)\}$) is

$$\mathsf{E}\left\{L^{(t)}_{k\to l}(x)\right\} = \Theta\left(x_k(x - x_0) - (x^2 - x_0^2)/2\right)$$

$$\mathsf{cov}\left(L^{(t)}_{k\to l}(x_1), L^{(t)}_{k\to l}(x_2)\right) = \Theta(x_1 - x_0)(x_2 - x_0)$$

where

$$\Theta = \int \frac{f_1^2(y)}{f_0(y)}dy\frac{\sum_{j\in\partial k\backslash l} S_{jk}^2}{\Lambda_k}A_k^2.$$

In view of Observation 1, the mean and covariance are indeed those of a scalar Gaussian channel. By law of large numbers, $\sum_{j\in\partial k\backslash l} S_{jk}^2/\Lambda_k \to 1$ in probability when Λ_k is sufficiently large. The last piece of the proof of Theorem 1 is to quantify the corresponding SNR by showing that

$$\lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}\,\to\,\infty,\,\text{A.C.F.}}} \int \frac{f_1^2(y)}{f_0(y)}dy = \eta^{(t)}$$

satisfies recursion (9). By again invoking the central limit theorem that W_j is asymptotically Gaussian and integrating the limit forms of $f_0(y)$ and $f_1(y)$ respectively, we have

$$\lim_{\Gamma_j\to\infty} f_m(y) = \mathsf{E}\left\{\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-W)^2}{2}}(y - W)^m\right\}$$

where $m = 0, 1$ and $W \sim \mathcal{N}(\mu_W, \sigma_W^2)$. Evidently,

$$\lim_{\Gamma_j\to\infty}\frac{f_1^2(y)}{f_0(y)} = \frac{1}{\sqrt{2\pi}}\frac{(y-\mu_W)^2}{(1+\sigma_W^2)^{\frac{5}{2}}}\exp\left[-\frac{(y-\mu_W)^2}{2(1+\sigma_W^2)}\right]$$

where (μ_W, σ_W^2) are the mean and variance of W_j. The finiteness of χ enables us to exchange the order of the limit and integration, we thus have

$$\lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}\,\to\,\infty,\,\text{A.C.F.}}} \int \frac{f_1^2(y)}{f_0(y)}dy = \int \lim_{\substack{K=\beta L\to\infty \\ \bar{\Gamma}\,\to\,\infty,\,\text{A.C.F.}}} \frac{f_1^2(y)}{f_0(y)}dy = \frac{1}{1+\sigma_W^2}.$$

What remains is to associate σ_W^2 in iteration t with the quality of BP estimation in iteration $t-1$. In view of (18), using the law of large numbers for the case when Γ_j is large and using induction on t, one can show that $\sigma_W^2 = \beta\,\overline{\mathsf{var}}\left\{AX \,\middle|\, \sqrt{\eta^{(t-1)}}AX + N, A\right\}$, where the initial $\eta^{(0)}$ is set to zero, because no inference information is available before the first iteration. The proof is thus complete.

B. The Asymptotic Equivalence of BP and MAP

Recall that BP is an optimal detection rule based on the limited observations $\boldsymbol{Y}_k^{(t)}$ on the subtree of depth $2t-1$ rooted at the each desired symbol X_k. Define $\boldsymbol{X}_k^{(t)}$ as the symbols on the same subtree. Let us define a genie-aided BP (gBP) as the optimal detection based on $\boldsymbol{Y}^{(t)}$ where all entries of $\boldsymbol{X}$ not in $\boldsymbol{X}_k^{(t)}$ are revealed to the BP estimator by a genie. This effectively guarantees independence of the leaves on the subtree of depth $2t-1$. One can show that the a posteriori detector for symbol X_k based on $\boldsymbol{Y}$ is a *physically degraded* detection rule with respect to gBP while the classical BP is a physically degraded rule when compared to the MAP detector.

Following the same scalar Gaussian channel analysis in the previous subsection, one can show that the asymptotic performance gBP can be described by the same iterative formula in Theorem 1 while the only difference for the gBP is that the initial $\eta^{(0)}$ is set to 1, which corresponds to the case of "perfect" initial extrinsic information.[2] Due to the uniqueness of the fixed point of (10) by assumption, BP and gBP will have the same asymptotic performance. Namely,

$$P^{\mathsf{bp}}_{X_k}(x|\boldsymbol{Y}^{(t)},\underline{\boldsymbol{SA}}) \xrightarrow{t\to\infty} P_{X|Z;g}(x|h_{\mathrm{BP}}(Y^{(t)},\underline{\boldsymbol{SA}}),\eta A_k^2)$$
$$P^{\mathsf{bp}}_{X_k}(x|\boldsymbol{Y}^{(t)},\underline{\boldsymbol{SA}},\boldsymbol{X}\backslash\boldsymbol{X}^{(t)}) \xrightarrow{t\to\infty} P_{X|Z;g}(x|h_{\mathrm{gBP}}(Y^{(t)},\underline{\boldsymbol{SA}},\boldsymbol{X}\backslash\boldsymbol{X}^{(t)}),\eta A_k^2)$$

in probability where h_{BP} and h_{gBP} are two functions computing the equivalent scalar Gaussian channel outputs for BP and gBP. Theorem 1 alone does not guarantee that h_{BP} and h_{gBP} generate the same equivalent scalar channel outputs when t tends to infinity, and the asymptotic equivalency between BP and gBP can be further strengthened as follows.

We claim that, in the large-sparse-system limit,

$$\left|h_{\mathrm{BP}}(Y^{(t)},\underline{\boldsymbol{SA}}) - h_{\mathrm{gBP}}(Y^{(t)},\underline{\boldsymbol{SA}},\boldsymbol{X}\backslash\boldsymbol{X}^{(t)})\right| \to 0$$

in probability as $t\to\infty$. Namely, in the large-sparse-system limit, BP and gBP not only produce outputs of identical quality, but also compute identical posterior distributions for almost all instances as well. This can be established by noting that BP is inferior to gBP for all instances of $\boldsymbol{X}\backslash\boldsymbol{X}^{(t)}$. Suppose for some event with strictly positive probability that $h_{\mathrm{BP}}(Y^{(t)},\underline{\boldsymbol{SA}}) \nrightarrow h_{\mathrm{gBP}}(Y^{(t)},\underline{\boldsymbol{SA}},\boldsymbol{X}\backslash\boldsymbol{X}^{(t)})$. Then since BP is a strict degraded detection rule of gBP for any event, BP will have strictly worse performance than gBP conditioned on that event. As a result, BP will again have a strictly worse performance than gBP in average, which contradicts the implications of Theorem 1 that their performance are identical.

Since the performances of BP and gBP sandwich the performance of optimal posterior probability detector, the optimal detector will also face a scalar Gaussian channel in the large-system limit as described in Theorem 2 as long as (10) has a unique fixed point. Moreover, by the similar argument as in the last paragraph, with probability one, the posterior distribution $P_{X_k|\boldsymbol{Y},\underline{\boldsymbol{SA}}}(\cdot|\boldsymbol{Y},\underline{\boldsymbol{SA}})$ are identical to the posterior distributions $\lim_{t\to\infty}P^{\mathsf{bp}}_{X_k}(\cdot|\boldsymbol{Y}^{(t)},\underline{\boldsymbol{SA}})$ and $\lim_{t\to\infty}P^{\mathsf{bp}}_{X_k}(\cdot|\boldsymbol{Y}^{(t)},\underline{\boldsymbol{SA}},\boldsymbol{X}\backslash\boldsymbol{X}^{(t)})$ computed by BP and gBP respectively. Corollary 1 is thus established.

C. Joint Decoding vs. Separate Decoding

Theorem 3 is an outcome of the chain rule of mutual information applied to the input–output mutual information of the CDMA channel

$$I(\boldsymbol{X};\boldsymbol{Y}|\underline{\boldsymbol{SA}}) = \sum_{k=1}^{K} I(X_k;\boldsymbol{Y}|\underline{\boldsymbol{SA}},X_{k+1},\dots,X_K). \qquad (23)$$

Each summand in the RHS is a single-user mutual information over the multiuser channel conditioned on the symbols of previously decoded users, where we assume without loss of generality that the users are decoded in reverse order.

Since the error probability of decoded symbols vanishes with code block-length, the interference from decoded users are asymptotically completely removed. Consequently user k sees only $k-1$ interfering users. Hence the posterior for user k under such successive decoding is identical to that under multiuser detection with separate decoding in a system with k instead of K users. The equivalent scalar channel for each user is Gaussian by Theorem 2. The multiuser efficiency experienced by user k is a function of the effective load $\beta_k = k/L$, and the mutual information converges to $I\big(X;\sqrt{\eta(\beta_k)}\,AX+N|A\big)$. In view of (23),

$$\frac{1}{K}I(\boldsymbol{X};\boldsymbol{Y}|\underline{\boldsymbol{SA}}) = \frac{1}{K}\sum_{k=1}^{K} I\big(X;\sqrt{\eta(\beta_k)}\,AX+N|A\big) \to \frac{1}{\beta}\int_0^\beta I\big(X;\sqrt{\eta(\xi)}\,AX+N|A\big)\,\mathrm{d}\xi$$

as $K\to\infty$ where the last equation is by definition of the Riemann integral.

As far as Theorem 3 is concerned, it suffices to prove that

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\left[\beta\, I\big(X;\sqrt{\eta(\beta)}\,AX+N|A\big) + \frac{\eta(\beta)-1-\log\eta(\beta)}{2}\right] = I\left(X;\sqrt{\eta(\beta)}\,AX+N|A\right)$$

or, equivalently,

$$\beta\frac{\mathrm{d}}{\mathrm{d}\beta}I\big(X;\sqrt{\eta(\beta)}\,AX+N\,|\,A\big) = \frac{\mathrm{d}}{\mathrm{d}\beta}\frac{\log\eta(\beta)-\eta(\beta)}{2}. \qquad (24)$$

Noticing that the multiuser efficiency η is a function of the system load β, (24) is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}\eta}I(X;\sqrt{\eta}\,AX+N\,|\,A) = \frac{1}{2\beta}\left(\eta^{-1}-1\right). \qquad (25)$$

The mutual information and the MMSE in Gaussian channels are related by the following formula [25, Theorem 1],

$$\frac{\mathrm{d}}{\mathrm{d}g}I(X;\sqrt{g}\,X+N) = \frac{1}{2}\,\mathsf{var}\left\{X\,|\sqrt{g}\,X+N\right\}, \quad \forall g.$$

Thus (25) holds as η satisfies the fixed-point equation (10).

VI. Numerical Results

This section investigates the quality of BP and its comparison with optimal detection in the asymptotic regime. Like many implications based on the central limit theorem and BP, the asymptotic performance matches that of practical systems even when the system size (K,L), the node degrees and the number of iterations are not very large. The performance of sparse CDMA for practical system parameters is confirmed by several numerical examples in this section.

Figure 4 shows the multiuser efficiency as a function of the SNR, which is obtained by the fixed point of (10) under various input and power distributions. The results apply to *any* chip distribution P_S as it does not affect the iterative equation (9). The iterative equation (9) provides an efficient method of probing the sparse CDMA performance for different SNRs without resorting to brute force simulation.

Using the theoretically computed multiuser efficiency (as illustrated in Figure 4 for various SNRs), one can efficiently compute the predicted asymptotic symbol-error-rate (SER) for the sparse CDMA system. For example, Figure 5 shows the

[2] An even more straightforward initial value is $\eta^{(0)}=\infty$, which corresponds to a Gaussian channel with infinite SNR and thus represents perfect extrinsic information. The first iteration leads to $\eta^{(1)}=1$.

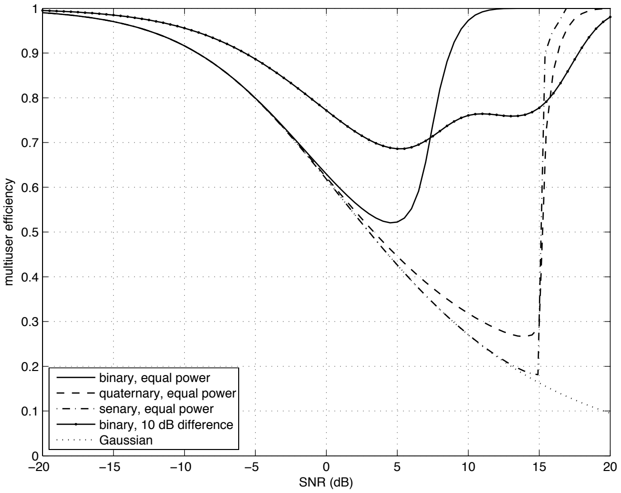


Fig. 4. Multiuser efficiency as a function of the SNR. For the first three curves, the symbol distribution P_X is equally likely among 2 or 4 or 6 evenly spaced scalar points while all users have the same power. We also consider a different power profile in the fourth curve for which half of the users are 10 dB higher than the rest. The multiuser efficiency for standard Gaussian input with equal power is plotted for comparison.

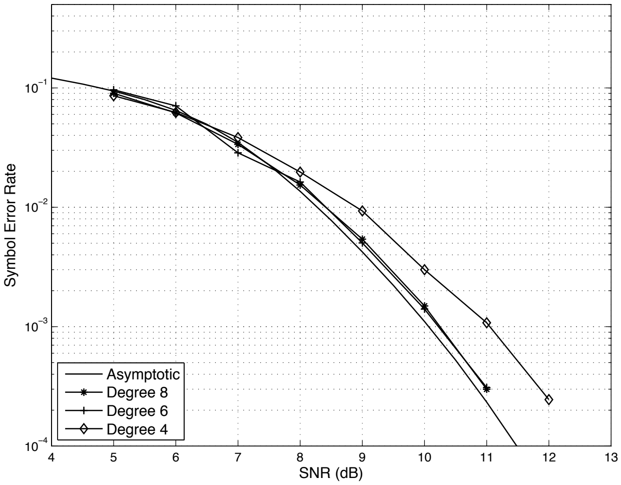


Fig. 5. The SER vs. SNR of BP multiuser detection with $K = L = 100$ users ($\beta = 1$). The spreading matrix is chosen from a regular bipartite graph ensemble with degrees 4, 6, and 8. The chip distribution P_S is evenly spread over four points with zero mean and unit variance. All input symbols are BPSK modulated with equal power.

SER achieved by BP detection for systems with $K = L$ equal-power users (thus $\beta = 1$) with BPSK modulation. In addition to the asymptotic SER, we also plot the performance of BP detection for three simulated systems with size $K = L = 100$ and fixed node degrees $\Lambda_k = \Gamma_l = 4$, 6, and 8 respectively. It can be seen that with degrees as small as 6, the finite-sized sparse CDMA system is closely approaching the asymptotic prediction based on (9) and (10). Figure 6 assumes a similar setting but considers that half of the users have double the power of the rest of the users (3 dB higher). Again, the finite-sized system performance fits the prediction pretty well.

Figure 7 demonstrates the empirical probability density function of the decision statistics of a BP detection on a finite system $K = L = 100$ with SNR=11 dB, equal-power users with BPSK modulation, the same quaternary P_S as in

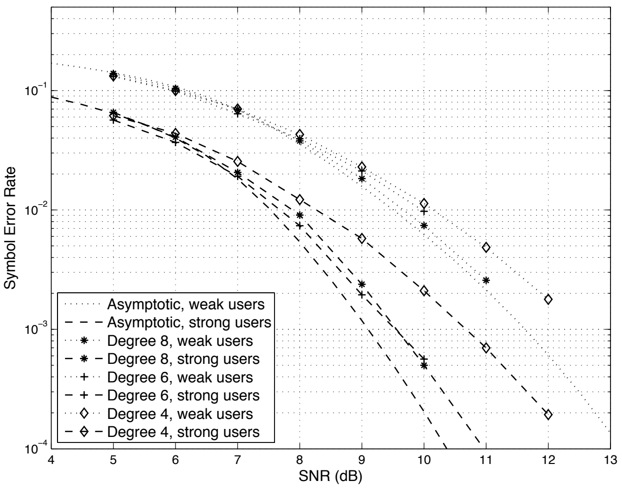


Fig. 6. The SER vs. average user SNR of BP multiuser detection with $K = L = 100$ users ($\beta = 1$). The spreading matrix is chosen from a regular bipartite graph ensemble with degrees 4, 6, and 8. The chip distribution P_S is evenly spread over four points with zero mean and unit variance. All input symbols are BPSK modulated. Half of the user has double the power (3 dB) over the rest. Users of high power are seen to have lower SER than those of lower power.

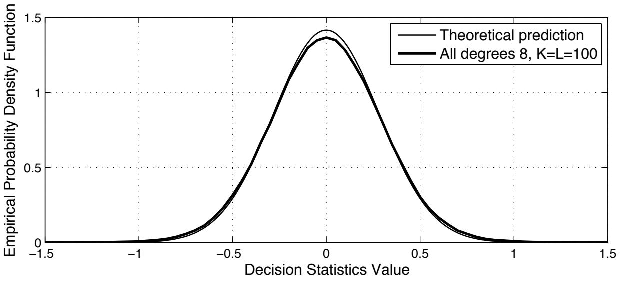


Fig. 7. The empirical distribution of the decision statistics of a finite-size system obtained by BP. Also shown is the theoretical large-system asymptotic distribution.

Figures 5 and 6. Since (10) focuses on the variances, we further shift the statistics to mean zero. Even for such a short system with very small degrees, the normality is apparent and the variance fits the predicted η very well.

Figure 8 shows the evolution of the multiuser efficiency with the number of iterations according to (9) in case of equal-power users with BPSK modulation. Clearly, the optimal η can be reached within 10 iterations. In all the numerical experiments we have conducted, the output posterior distribution of BP converges within 15–20 iterations, which demonstrates the efficiency of BP on sparse CDMA systems.

VII. Concluding Remarks

We have studied sparsely spread CDMA systems and low-complexity multiuser detection based on belief propagation. Assuming a chip-semi-regular ensemble of sparse spreading matrices, the posterior distribution for each symbol computed by BP is shown to be asymptotically equivalent to the posterior of a scalar Gaussian channel with the same input symbol. BP-based detection is shown to be asymptotically optimal as long as the load of the system is not too large, where the asymptotic equivalence of BP and a posterior detection is established in

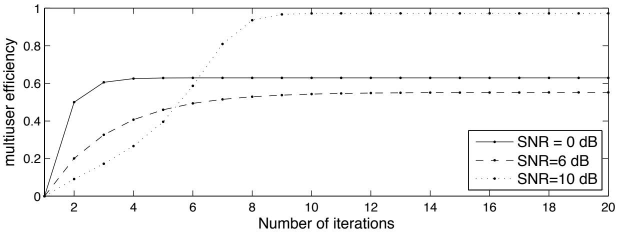


Fig. 8. The evolution of multiuser efficiency with the number of iterations.

the (strongest) sense of the posterior probability for arbitrary input distributions. This phenomenon is in contrast to the wisdom for LDPC codes with infinitely long codeword length that the MAP decoder is generally strictly better than BP.

An important question is how the performance of sparse CDMA compares with that of dense CDMA. The difference is believed to be small, as is buttressed by numerical results (see, e.g., [9], [14]), which adds to the practical appeal of sparse CDMA. In fact, the iterative formula obtained in Section V leads to exactly the fixed-point equation for the multiuser efficiency under dense CDMA obtained using the heuristic replica method [8]–[10]. A precise quantification of the difference between sparse and dense CDMA is not available, however, due to lack of an accurate and manageable expression of the optimal performance of dense CDMA.

REFERENCES

[1] J. G. Proakis, *Digital Communications*. McGraw-Hill, 4th ed., 2001.

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

[3] 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 Inform. Theory Workshop*, pp. 122–126, Punta del Este, Uruguay, Mar. 2006.

[4] M. Yoshida and T. Tanaka, "Analysis of sparsely-spread CDMA via statistical mechanics," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 2378–2382, Seattle, WA, USA, 2006.

[5] 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.

[6] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A: Math. Gen.*, vol. 36, pp. 11111–11121, 2003.

[7] 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.

[8] 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.

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

[10] D. Guo and T. Tanaka, "Generic multiuser detection and statistical physics," in *Advances in Multiuser Detection* (M. Honig, ed.), Wiley. to be published.

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

[12] 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*, pp. 926–935, Monticello, IL, USA, Oct. 2006.

[13] D. Guo and C.-C. Wang, "Random sparse linear systems observed via arbitrary channels: A decoupling principle," in *Proc. IEEE Int. Symp. Inform. Theory*, Nice, France, June 2007.

[14] J. Raymond and D. Saad, "Sparsely spread CDMA–a statistical mechanics-based analysis," *J. Phys. A: Math. Theor.*, vol. 40, pp. 12315–12333, 2007.

[15] T. S. Rappaport, S. Y. Seidel, and R. Singh, "900-mhz multipath propagation measurements for us digital cellular radio telephone," *IEEE Trans. Veh. Technol.*, vol. 39, pp. 132–139, May 1990.

[16] J. J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1772–1793, July 2002.

[17] G. Caire, R. R. Müller, and T. Tanaka, "Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1950–1973, Sept. 2004.

[18] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988. Revised 2nd printing.

[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] H. A. Loeliger, "Some remarks on factor graphs," in *Proc. 3rd Int'l. Symp. Turbo Codes & Related Topics*, pp. 111–115, Brest, France, 2003.

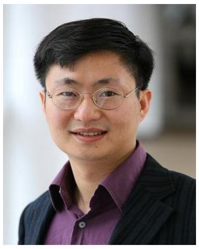
[21] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.

[22] 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.

[23] C. Measson, A. Montanari, T. Richardson, and R. Urbanke, "The generalized area theorem and some of its consequences," *IEEE Trans. Inform. Theory*. submission print: cs.IT/0511039.

[24] C. Measson, A. Montanari, and R. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding," *IEEE Trans. Inform. Theory*. submission print: cs.IT/0506083.

[25] 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.



Dongning Guo has been an Assistant Professor in the Department of Electrical Engineering & Computer Science at Northwestern University since 2004. He received the Ph.D. and M.Sc. degrees from Princeton University, the M.Eng. degree from the National University of Singapore and the B.Eng. degree from University of Science & Technology of China. He was an R&D Engineer in the Centre for Wireless Communications (now the Institute for Infocom Research), Singapore from 1998 to 1999. He was a Visiting Professor at Norwegian University of Science and Technology in summer 2006. He received the Huber and Suhner Best Student Paper Award in the International Zurich Seminar on Broadband Communications in 2000 and the National Science Foundation Faculty Early Career Development (CAREER) Award in 2007. His research interests are in information theory, communications and networking.

Chih-Chun Wang joined the School of Electrical and Computer Engineering in January 2006 as an Assistant Professor. He received the B.E. degree in E.E. from National Taiwan University, Taipei, Taiwan in 1999, the M.S. degree in E.E., the Ph.D. degree in E.E. from Princeton University in 2002 and 2005, respectively. He worked in Comtrend Corporation, Taipei, Taiwan, as a design engineer during in 2000 and spent the summer of 2004 with Flarion Technologies, New Jersey. In 2005, he held a post-doc position in the Electrical Engineering Department of Princeton University. His research interests are in optimal control, information theory, detection theory, coding theory, iterative decoding algorithms, and network coding.