Randomly Spread CDMA: Asymptotics via Statistical Physics

Dongning Guo and Sergio Verdú

Submitted to IEEE Transactions on Information Theory, June 20, 2003.
Revised, October 9, 2004.

Abstract

This paper studies randomly spread code-division multiple access (CDMA) and multiuser detection in the large-system limit using the replica method developed in statistical physics. Arbitrary input distributions and flat fading are considered. A generic multiuser detector in the form of the posterior mean estimator is applied before single-user decoding. The generic detector can be particularized to the matched filter, decorrelator, linear MMSE detector, the jointly or the individually optimal detector, and others. It is found that the detection output for each user, although in general asymptotically non-Gaussian conditioned on the transmitted symbol, converges as the number of users go to infinity to a deterministic function of a "hidden" Gaussian statistic independent of the interferers. Thus the multiuser channel can be decoupled: Each user experiences an equivalent single-user Gaussian channel, whose signal-to-noise ratio suffers a degradation due to the multiple-access interference. The uncoded error performance (e.g., bit-error-rate) and the mutual information can then be fully characterized by this degradation, also known as the multiuser efficiency, which can be obtained by solving a pair of coupled fixed-point equations which we identify in this paper. Based on a general linear vector channel model, the results are also applicable to MIMO channels such as in multiantenna systems.

Index Terms: Code-division multiple access (CDMA), multiuser detection, channel capacity, multiuser efficiency, statistical physics, free energy, replica method, multiple-input multiple-output (MIMO) channel.

1 Introduction

Consider a multidimensional Euclidean space in which each user (or data stream) randomly picks a "signature vector" and modulates its own information-bearing symbols onto it for transmission. The received signal is then a superposition of all users' signals corrupted by Gaussian noise. Such a multiuser scheme, best described by a vector channel model, is very versatile and is widely used in applications that include code-division multiple access (CDMA), as well as certain multi-input multi-output (MIMO) systems. With knowledge of all signature vectors, the goal is to estimate the transmitted symbols and eventually recover the information intended for all or a subset of the users.

This paper focuses on a paradigm of multiuser channels, known as randomly spread CDMA. In such a CDMA system, a number of users share a common media to communicate to a single receiver

*Acknowledgement: Prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

simultaneously over the same bandwidth. Each user employs a pseudo random spreading sequence (signature waveform) with a large time-bandwidth product that results in many advantages particularly in wireless communications: frequency diversity, robustness to channel impairments, ease of resource allocation, etc. The price to pay is multiple-access interference (MAI) due to non-orthogonal spreading sequences from all users. Numerous multiuser detection techniques have been proposed to mitigate the MAI to various degrees. This work is concerned with the performance of such multiuser systems in two aspects: One is the *multiuser efficiency*, which in general measures the quality of multiuser detection outputs assuming uncoded transmission; the other is the *spectral efficiency*, which is the total information rate normalized by the dimension of the multiuser channel achievable by coded transmission.

1.1 Gaussian or Non-Gaussian?

The most efficient use of a multiuser channel is through jointly optimal decoding, which is an NP-complete problem [1]. A common suboptimal strategy is to apply a multiuser detector front end with polynomial complexity to generate a decision statistic for each user for subsequent independent single-user decoding. The quality of the detection output fed to decoders is of great interest.

In [2, 3, 4], Verdú first used the concept of multiuser efficiency to refer to the degradation of the output signal-to-noise ratio (SNR) relative to a single-user channel calibrated at the same bit-error-rate (BER) in binary (antipodal) uncoded transmission. The multiuser efficiencies of the matched filter, decorrelator, and linear minimum mean-square error (MMSE) detector were found as functions of the correlation matrix of the spreading sequences. Particular attention has been given to the asymptotic multiuser efficiency in the more tractable region of high SNR. Expressions for the optimum (high-SNR) asymptotic multiuser efficiency were found in [4, 5].

In the large-system limit, where the number of users and the spreading factor both tend to infinity with a fixed ratio, the dependence of system performance on the sequences vanishes, and random matrix theory proves to be an excellent tool for analyzing linear detectors. The limiting multiuser efficiency of the matched filter is trivial [6]. The large-system multiuser efficiency of the MMSE detector is obtained explicitly in [6] for the equal-power case (perfect power control), and in [7] for the case with flat fading as the solution to the Tse-Hanly fixed-point equation. The efficiency of the decorrelator is also known [8, 9]. The success with the wide class of linear detectors hinges on the fact that 1) the detection output is a sum of independent components: the desired signal, the MAI and Gaussian background noise, e.g., for user k,

$$\langle X_k \rangle = X_k + I_k + N_k; \tag{1}$$

and 2) the multiple-access interference (I_k) is asymptotically Gaussian (e.g., [10]). As far as linear multiuser detectors are concerned, regardless of the input distribution, the performance is fully characterized by the noise enhancement associated with the MAI variance. Indeed, by regarding the multiuser detector as part of the channel, an individual user experiences essentially a single-user Gaussian channel with an SNR degradation which is the multiuser efficiency.

The performance of nonlinear detectors such as the optimal ones is a hard problem. The difficulty here is inherent to nonlinear operations. That is, the detection output cannot be decomposed as a sum of independent components associated with the desired signal, the interferences and the noise respectively. Moreover, the detection output is in general asymptotically non-Gaussian conditioned on the input. An extreme case is the maximum-likelihood multiuser detector for binary transmission, the hard decision output of which takes only two values. The difficulty remains if we consider soft detection outputs. Hence, unlike for a Gaussian output, the conditional variance of a general detection output does not lead to simple characterization of the multiuser efficiency

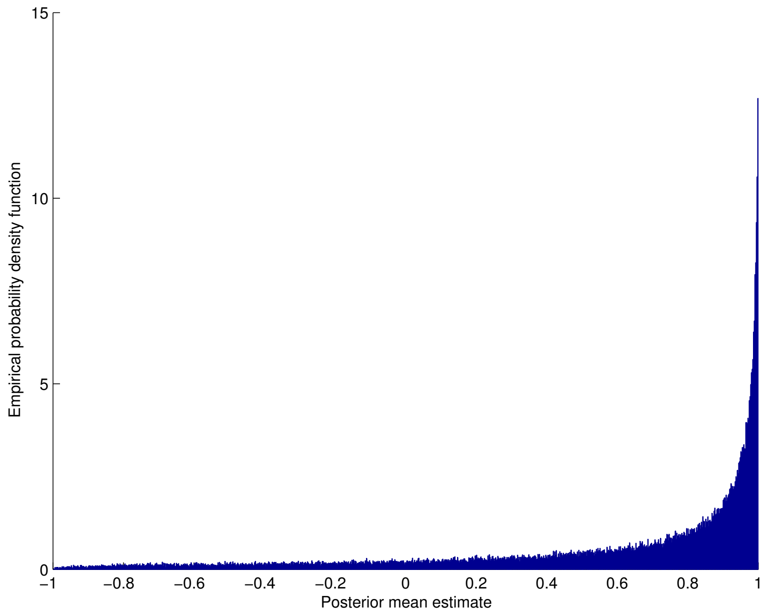


Figure 1: The probability density function obtained from the histogram of an individually optimal soft detection output conditioned on +1 being transmitted. The system has 8 users, the spreading factor is 12, and SNR=2 dB.

or error performance. For illustration, Figure 1 plots the probability density function of the soft output statistic of the individually optimal detector conditioned on +1 being transmitted. The simulated system has 8 users with binary inputs, a spreading factor of 12, and SNR=2 dB. Note that negative decision values correspond to decision error; hence the area under the curve on the negative half plane gives the BER. The distribution shown in Figure 1 is far from Gaussian. Thus the usual notion of output SNR fails to capture the essence of system performance. In fact, much literature is devoted to evaluating the error performance by Monte Carlo simulation.

This paper makes a contribution to the understanding of multiuser detection in the large-system regime. It is found under certain assumptions that the output decision statistic of a nonlinear detector, such as the one whose distribution is depicted by Figure 1, converges in fact to a very simple monotone function of a "hidden" conditionally Gaussian random variable, i.e.,

$$\langle X_k \rangle \to f(Z_k) \tag{2}$$

where $Z_k = X_k + N_k'$ and N_k' is Gaussian. One may contend that it is always possible to monotonically map a non-Gaussian random variable to a Gaussian one. What is surprisingly simple and useful here is that 1) the mapping f neither depends on the instantaneous spreading sequences, nor on the transmitted symbols which we wish to estimate in the first place; and 2) the statistic Z_k consists of the desired signal and an independent Gaussian noise. Indeed, a few parameters of the system easily determine the function f. By applying an inverse of this function to the detection output $\langle X_k \rangle$, an equivalent conditionally Gaussian statistic Z_k is recovered, so that we are back to the familiar ground where the output SNR (defined for the Gaussian statistic) completely characterizes the system performance. The multiuser efficiency is simply obtained as the ratio of the output and input SNRs. Since, after applying f^{-1}, each user enjoys asymptotically an equivalent single-user Gaussian channel with an enhanced Gaussian noise (independent of signals from other users) in lieu of the MAI, we will refer to this result as the "decoupling principle".

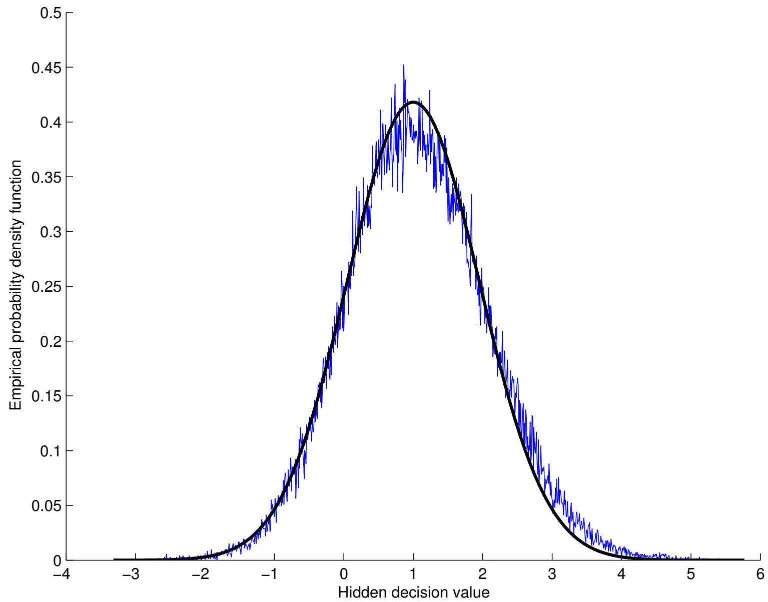


Figure 2: The probability density function obtained from the histogram of the hidden equivalent Gaussian statistic conditioned on +1 being transmitted. The system has 8 users, the spreading factor is 12, and SNR=2 dB. The asymptotic Gaussian distribution is also plotted for comparison.

Under certain assumptions, we show the decoupling principle to hold for not only optimal detection, but also a broad family of generic multiuser detectors, called the posterior mean estimators (PME), which compute the mean value of the input conditioned on the observation assuming a certain postulated posterior probability distribution. In case the postulated posterior is identical to the one induced by the actual multiuser channel and input, the PME is a soft version of the individually optimal detector. The postulated posterior, however, can also be chosen such that the resulting PME becomes one of many other detectors, including but not limited to the matched filter, decorrelator, linear MMSE detector, as well as the jointly optimal detector. Moreover, the decoupling principle holds for not only binary inputs, but arbitrary input distributions with finite power.

For illustration of the new findings, Figure 2 plots the probability density function of the conditionally Gaussian statistic obtained by applying f^{-1} to the non-Gaussian detection output in Figure 1. The theoretically predicted Gaussian density function is also shown for comparison. The "fit" is good considering that a relatively small system of 8 users with a processing gain of 12 is considered. Note that in case the multiuser detector is linear, the mapping f is also linear, and (2) reduces to (1).

By virtue of the decoupling principle, the mutual information between the input and the output of the generic detector for each user converges to the input-output mutual information of the equivalent single-user Gaussian channel under the same input, which admits a simple analytical expression. Hence the large-system spectral efficiency of several well-known linear detectors, first found in [11] and [12] with and without fading respectively, can be recovered straightforwardly using the decoupling principle. New results on the spectral efficiency of nonlinear detection and arbitrary inputs under both joint and separate decoding are also obtained. Furthermore, the additive decomposition of optimal spectral efficiency as a sum of single-user efficiencies and a joint decoding gain [12] applies under more general conditions than originally thought.

It should be pointed out here that the large-system results are close representatives of practical system of moderate size. As we mentioned, a system with as few as 8 users can often be well approximated as a large system.

1.2 Random Matrix vs. Spin Glass

Much of the early success in the large-system analysis of linear detectors relies on the fact that the multiuser efficiency of a finite-size system can be written as an explicit function of the eigenvalues of the correlation matrix of the random signature waveforms, the empirical distributions of which converge to a known function in the large-system limit [13, 14]. As a result, the large-system multiuser efficiency can be obtained as an integral with respect to the limiting eigenvalue distribution. Indeed, this random matrix technique is applicable to any performance measure that can be expressed as a function of the eigenvalues. Based on an explicit expression for CDMA channel capacity in [15], Verdú and Shamai quantified the optimal spectral efficiency in the large-system limit [11, 12] (see also [16, 17]). The expression found in [11] also solved the capacity of single-user narrowband multiantenna channels as the number of antennas grows—a problem that was open since the pioneering work of Foschini [18] and Telatar [19]. Unfortunately, few explicit expressions of the efficiencies in terms of eigenvalues are available beyond the above cases. Much less success has been reported in the application of random matrix theory in problems of non-Gaussian input or nonlinear detection.

A major consequence of random matrix theory is that the dependence of performance measures on the spreading sequences vanishes as the system size increases without bound. In other words, the performance measures are "self-averaging." This property is nothing but a manifestation of a fundamental law that the fluctuation of macroscopic properties of certain many-body systems vanishes in the thermodynamic limit, i.e., when the number of interacting bodies becomes large. This falls under the general scope of statistical physics, whose principal goal is to study the macroscopic properties of physical systems starting from knowledge about microscopic interactions. Indeed, the asymptotic eigenvalue distribution of certain correlation matrices can be derived via statistical physics (e.g., [20]). Tanaka pioneered statistical physics concepts and methodologies in multiuser detection and obtained the large-system uncoded minimum BER (hence the optimal multiuser efficiency) and spectral efficiency with equal-power binary inputs [21, 22, 23, 24]. In [9] we further elucidated the relationship between CDMA and statistical physics and generalized Tanaka's results to the case of unequal powers. Inspired by [24], Müller and Gerstacker [25] studied the channel capacity under separate decoding and noticed that the additive decomposition of the optimum spectral efficiency in [12] holds also for binary inputs. Müller thus further conjectured the same formula to be valid regardless of the input distribution [26].

In this paper, we build upon Tanaka's ground-breaking contribution [24] and present a unified treatment of Gaussian CDMA channels and multiuser detection assuming an arbitrary input distribution and flat fading characteristic. A wide class of multiuser detectors, optimal as well as suboptimal, are treated under a uniform framework of posterior mean estimation. The central results are the decoupling principle and the characterization of multiuser efficiency for a generic multiuser detector via a pair of nonlinear equations.

The key technique in statistical physics, the replica method, has its origin in spin glass theory [27]. Analogies between statistical physics and neural networks, coding, image processing, and communications have long been noted (e.g., [28, 29]). There have also been many recent activities applying statistical physics wisdom to sparse-graph error-correcting codes [30, 31, 32]. The first application of the replica method to multiuser detection was made by Tanaka [24]. In this paper, we draw a parallel between the general statistical inference problem in multiuser communications and

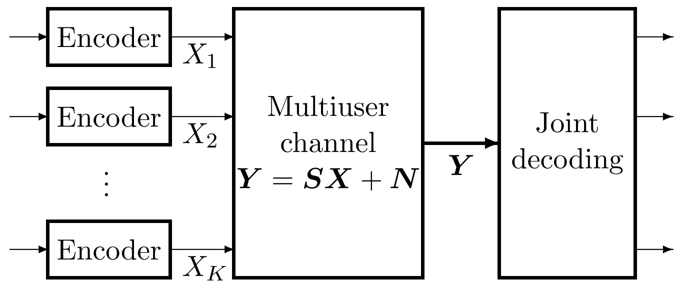


Figure 3: A multiuser system with joint decoding.

the problem of determining the configuration of random spins subject to quenched randomness. For the purpose of analytical tractability, we will invoke common assumptions in the statistical physics literature: 1) the self-averaging property applies, 2) the "replica trick" is valid, and 3) replica symmetry holds. These assumptions have been used successfully in many problems in statistical physics as well as in neural networks and coding theory, to name a few, while a complete justification of the replica method is a notoriously difficult challenge in mathematical physics, which has seen some important progress recently [33]. The results in this paper are based on these assumptions and therefore the rigorousness is pending on breakthroughs in those problems. In fact, what is not known is a set of easy-to-check sufficient and necessary conditions under which the assumptions hold so that the claims in this paper are rigorous. In the cases that the assumptions fail to hold, results obtained using the replica method may still capture many of the qualitative features of the system performance. Indeed, such results are often found as good approximations in some cases where not all of the assumptions are true [34, 35]. Furthermore, the decoupling principle carries great practicality and may find convenient uses as long as the analytical predictions are close to the reality even if not exact.

The remainder of this paper is organized as follows. Section 2 gives the model and summarizes major results. Relevant statistical physics concepts and methodologies are introduced in Section 3. Calculations based on a real-valued channel are presented in Section 4. Complex-valued channels are discussed in Section 5, followed by some numerical examples in Section 6.

2 Model and Summary of Results

2.1 System Model

Consider the synchronous K-user CDMA system with spreading factor L as depicted in Figure 3. Each encoder maps its message into a sequence of channel symbols. All users employ the same type of signaling so that at each interval the K symbols are independent identically distributed (i.i.d.) random variables with distribution (probability measure) P_X, which has zero mean and unit variance. Let $\boldsymbol{X} = [X_1, \dots, X_K]^\top$ denote the vector of input symbols from the K users in one symbol interval. For notational convenience in the analysis, it is assumed that either a probability density function or a probability mass function of P_X exists, and is denoted by p_X.[1] Let also $p_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{k=1}^{K} p_X(x_k)$ denote the joint (product) distribution.

Let user k's instantaneous SNR be denoted by snr_k and $\boldsymbol{\Gamma} = \mathrm{diag}\{\sqrt{\mathsf{snr}_1}, \dots, \sqrt{\mathsf{snr}_K}\}$. Denote the spreading sequence of user k by $\boldsymbol{s}_k = \frac{1}{\sqrt{L}}[S_{1k}, S_{2k}, \dots, S_{Lk}]^\top$, where S_{nk} are i.i.d. random

[1] The results in this paper hold in full generality and do not depend on the existence of a probability density or mass function.

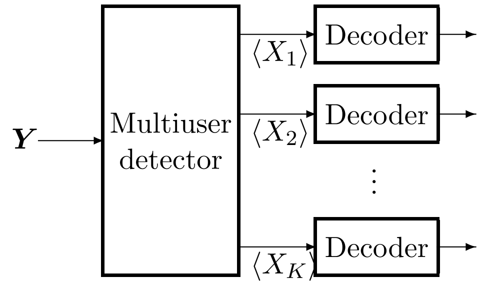


Figure 4: Multiuser detection followed by independent single-user decoding.

variables with zero mean and finite moments. Let the symbols and spreading sequences be randomly chosen for each user and not dependent on the SNRs. The $L \times K$ channel state matrix is denoted by $\boldsymbol{S} = [\sqrt{\mathsf{snr}_1}\, \boldsymbol{s}_1, \dots, \sqrt{\mathsf{snr}_K}\, \boldsymbol{s}_K]$. The synchronism CDMA channel with flat fading is described by:

$$\boldsymbol{Y} = \sum_{k=1}^{K} \sqrt{\mathsf{snr}_k}\, \boldsymbol{s}_k X_k + \boldsymbol{N} \quad (3)$$

$$= \boldsymbol{S}\boldsymbol{X} + \boldsymbol{N} \quad (4)$$

where $\boldsymbol{N}$ is a vector consisting of i.i.d. zero-mean Gaussian random variables. Depending on the domain that the inputs and spreading chips take values, the input-output relationship (4) describes either a real-valued or a complex-valued fading channel.

The linear system (4) is quite versatile. In particular, with $\mathsf{snr}_k = \mathsf{snr}$ for all k, it models the canonical MIMO channel in which all propagation coefficients are i.i.d. An example is single-user communication with K transmit antennas and L receive antennas, where the channel coefficients are not known to the transmitter.

2.2 Posterior Mean Estimation

The fundamental problem of communication can be described as follows. The information-bearing symbol (vector) $\boldsymbol{X}$ is drawn according to the prior distribution $p_{\boldsymbol{X}}$. The channel response to the input $\boldsymbol{X}$ is an output $\boldsymbol{Y}$ generated according to a conditional probability distribution $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ where $\boldsymbol{S}$ is the channel state. Upon receiving $\boldsymbol{Y}$, the estimator would like to infer the transmitted symbol $\boldsymbol{X}$ using knowledge about the state $\boldsymbol{S}$, and eventually reproduce the intended information reliably.

The most efficient use of the multiuser channel (4) in terms of information capacity is achieved by optimal joint decoding as depicted in Figure 3. The input-output mutual information of the multiuser channel given the channel state $\boldsymbol{S}$ is $I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S})$. Due to the complexity of joint decoding, one often breaks the process into multiuser detection followed by separate single-user error-control decoding as shown in Figure 4. A multiuser detector front end with no knowledge of the error-control codes used by the encoder outputs an estimate of the transmitted symbols given the received signal and the channel state. Each decoder only takes the decision statistic of a single user of interest to decoding without awareness of the existence of any other users (in particular, without knowledge of the spreading sequences). By adopting this separate decoding approach, the channel together with the multiuser detector front end is viewed as a single-user channel for each user. The detection output sequence for an individual user is in general not a sufficient statistic for decoding even this user's own information.

To capture the intended suboptimal structure, one has to restrict the capability of the multiuser detector; otherwise the detector could in principle encode the channel state and the received signal $(\boldsymbol{S}, \boldsymbol{Y})$ into a single real number as its output to each user, which is a sufficient statistic for all users. A plausible choice is the (canonical) *posterior mean estimator*, which computes the mean value of the posterior probability distribution $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$, hereafter denoted by angle brackets $\langle\cdot\rangle$:

$$\langle \boldsymbol{X} \rangle = \mathsf{E}\left\{ \boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{S} \right\}. \tag{5}$$

Also known as the conditional mean estimator, this estimator achieves the minimum mean-square error for each user. In fact it is also the soft-output version of the individually optimal multiuser detector (assuming uncoded transmission). The posterior probability distribution $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ is induced from the input distribution $p_{\boldsymbol{X}}$ and the conditional Gaussian density function $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ of the channel (4) by the Bayes formula:

$$p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = \frac{p_{\boldsymbol{X}}(\boldsymbol{x}) p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S})}{\int p_{\boldsymbol{X}}(\boldsymbol{x}) p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S}) \, \mathrm{d}\boldsymbol{x}}. \tag{6}$$

The PME can be understood as an "informed" optimal estimator which is supplied with the posterior $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$. A generalization of the canonical PME is conceivable: Instead of informing the estimator with the actual posterior $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$, we can supply at will any other well-defined conditional distribution $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$. Given $(\boldsymbol{Y}, \boldsymbol{S})$, the estimator can nonetheless perform "optimal" estimation based on this postulated measure q. We call this the *generalized posterior mean estimation*, which is conveniently denoted as

$$\langle \boldsymbol{X} \rangle_q = \mathsf{E}_q\left\{ \boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{S} \right\} \tag{7}$$

where $\mathsf{E}_q\{\cdot\}$ stands for the expectation with respect to the postulated measure q. For the sake of brevity, we will also refer to (7) by the name of the posterior mean estimator, or simply the PME. In view of (5), the subscript in (7) can be dropped if the postulated measure q coincides with the actual one p.

In general, postulating $q \neq p$ causes degradation in detection performance. Such a strategy may be either due to lack of knowledge of the true statistics or a particular choice that corresponds to a certain estimator of interest. In principle, any deterministic estimation can be regarded as a PME since we can always choose to put a unit mass at the desired estimation output given $(\boldsymbol{Y}, \boldsymbol{S})$. We will see in Section 2.3 that by choosing an appropriate measure q, it is easy to particularize the PME to many important multiuser detectors. As will also be shown in this paper, the generic representation (7) allows a uniform treatment of a large family of multiuser detectors which results in a simple performance characterization for all of them.

It is enlightening to introduce a new concept: the *retrochannel*, which is defined for a given channel and input as a companion channel characterized by a posterior distribution. For example, given the multiuser channel $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ with an input $p_{\boldsymbol{X}}$, we have a (canonical) retrochannel defined by $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ (6), which, upon an input $(\boldsymbol{Y}, \boldsymbol{S})$, generates a random output $\boldsymbol{X}$ according to $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$. A retrochannel in the single-user setting is similarly defined. In general, any valid posterior distribution $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ can be regarded as a retrochannel. Note that the retrochannel samples from the Bayesian posterior distribution (in general, the postulated one) in such a way that, conditioned on the observation, the input to the channel and the output of the retrochannel are independent. It is clear that the PME output $\langle \boldsymbol{X} \rangle_q$ is the expected value of the output of the retrochannel $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ given $(\boldsymbol{Y}, \boldsymbol{S})$.

In this paper, the posterior $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ supplied to the PME is assumed to be the one that corresponds to a postulated CDMA system, where the input distribution is an arbitrary q_X, and the input-output relationship of the postulated channel differs from the actual channel (4) by only the noise variance. Precisely, the postulated channel is characterized by

$$\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{X} + \sigma \boldsymbol{N}' \tag{8}$$

where the channel state matrix $\boldsymbol{S}$ is identical to that of the actual channel (4), and $\boldsymbol{N}'$ is statistically the same as the Gaussian noise $\boldsymbol{N}$ in (4). Easily, $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ is determined by q_X and $q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ according to the Bayes formula (cf. (6)). Here, σ serves as a control parameter. Also, the postulated input distribution q_X is assumed to have zero-mean and finite moments. In general, the assumed information about the channel state $\boldsymbol{S}$ could also be different from the actual instances, but this is out of the scope of this work. In short, this paper studies the family of multiuser detectors that can be represented as the PME parameterized by the postulated input and noise level (q_X, σ).

We note that PME under postulated posterior is known in the Bayes statistics literature. This technique was introduced to multiuser detection by Tanaka in the special case of equal-power users with binary or Gaussian inputs under the name of marginal-posterior-mode detectors [21, 22, 24]. In this paper we pursue further that direction to treat arbitrary input, arbitrary power distribution, and multiuser detection in a general scope.

2.3 Specific Detectors

The rest of this section assumes a real-valued system model. The inputs X_k, the spreading chips S_{nk}, and all entries of the noise $\boldsymbol{N}$ take real values and have unit variance. The characteristic of the actual channel is

$$p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S}) = (2\pi)^{-\frac{L}{2}} \exp\left[-\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right], \tag{9}$$

and that of the postulated channel is

$$q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S}) = \left(2\pi\sigma^2\right)^{-\frac{L}{2}} \exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right]. \tag{10}$$

We identify specific choices of the postulated input distribution q_X and noise level σ under which the PME is particularized to several well-known multiuser detectors.

2.3.1 Linear Detectors

Let the postulated input be standard Gaussian, $q_X \sim \mathcal{N}(0, 1)$. The posterior probability density function becomes

$$q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = [Z(\boldsymbol{y},\boldsymbol{S})]^{-1} \exp\left[-\frac{1}{2}\|\boldsymbol{x}\|^2 - \frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right] \tag{11}$$

where $Z(\boldsymbol{y}, \boldsymbol{S})$ is a normalization factor. Since (11) is a conditional Gaussian density, its mean is a linear filtering of the received signal $\boldsymbol{Y}$:

$$\langle \boldsymbol{X} \rangle_q = \left[\boldsymbol{S}^\top \boldsymbol{S} + \sigma^2 \boldsymbol{I}\right]^{-1} \boldsymbol{S}^\top \boldsymbol{Y}. \tag{12}$$

If $\sigma \to \infty$, the PME estimate (12) is consistent with the matched filter output:

$$\sigma^2 \, \langle X_k \rangle_q \longrightarrow \boldsymbol{s}_k^\top \boldsymbol{Y}, \quad \text{in } L^2 \text{ as } \sigma \to \infty. \tag{13}$$

If $\sigma = 1$, (12) is exactly the soft output of the linear MMSE detector. If $\sigma \to 0$, (12) converges to the soft output of the decorrelator.

In general, the PME represents a linear detector by postulating Gaussian inputs. The control parameter σ can then be tuned to choose from the matched filter, decorrelator, MMSE detector, etc.

2.3.2 Optimal Detectors

Let the postulated q_X be identical to the true one, p_X. The posterior is then

$$q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = [Z(\boldsymbol{y},\boldsymbol{S})]^{-1}\, p_{\boldsymbol{X}}(\boldsymbol{x}) \exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{x}\|^2\right] \tag{14}$$

where $Z(\boldsymbol{y},\boldsymbol{S})$ is a normalization factor.

Suppose that the postulated noise level $\sigma \to 0$, then the probability mass of the distribution $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ is concentrated on a vector that minimizes $\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{x}\|$, which also maximizes the likelihood function $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S})$. The PME $\lim_{\sigma\to 0}\langle\boldsymbol{X}\rangle_q$ is thus equivalent to that of jointly optimal (or maximum-likelihood) detection [6].

Alternatively, if $\sigma = 1$, then the postulated measure coincides with the actual measure, i.e., $q = p$. The PME output $\langle\boldsymbol{X}\rangle$ is the mean of the marginal of the conditional posterior probability distribution, and is referred to as the soft output of the individually optimal detector [6]. Indeed, for m-PSK (resp. binary) inputs, hard decision of the soft output gives the most likely value of the transmitted symbol (resp. bit).

Also worth mentioning here is that, if $\sigma \to \infty$, the PME reduces to the matched filter. Indeed, (13) can be shown to hold by noticing from (14) that

$$q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = p_{\boldsymbol{X}}(\boldsymbol{x})\left[1-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{x}\|^2 + \mathcal{O}\left(\sigma^{-4}\right)\right]. \tag{15}$$

2.4 Main Results

This subsection represents the main results of this paper assuming the real-valued system model. The replica analysis carried out to obtain these results is relegated to Sections 3 and 4. Results for a complex-valued model will be presented in Section 5.

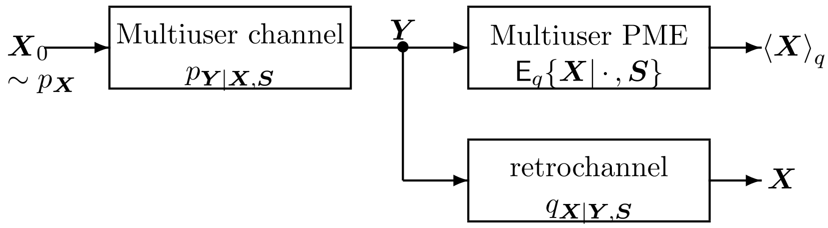
Consider the multiuser channel $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ given by (9) with input $\boldsymbol{X} \sim p_{\boldsymbol{X}}$, and the posterior mean estimator (7) parameterized by (q_X,σ). Section 2.3 illustrated the versatility of the PME encompassing many well-known detectors. The goal here is to quantify the optimal spectral efficiency $\frac{1}{L}I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S})$, the quality of the detection output $\langle X_k\rangle_q$ for each user k, as well as the input-output mutual information $I(X_k;\langle X_k\rangle_q|\boldsymbol{S})$.

Although these measures are all dependent on the realization of the channel state, such dependence vanishes in the large-system asymptote. A *large system* refers to the limit that both the number of users and the spreading factor tend to infinity but with their ratio, known as the *system load*, converging to a positive number, i.e., $K/L \to \beta$, which may or may not be smaller than 1. It is also assumed that the SNRs of all users, $\{\mathsf{snr}_k\}_{k=1}^K$, are i.i.d. with distribution P_{snr}, hereafter referred to as the *SNR distribution*. All moments of the SNR distribution are assumed to be finite. Clearly, the empirical distributions of the SNRs converge to the same distribution P_{snr} as $K \to \infty$. Note that this SNR distribution captures the (flat) fading characteristics of the channel.

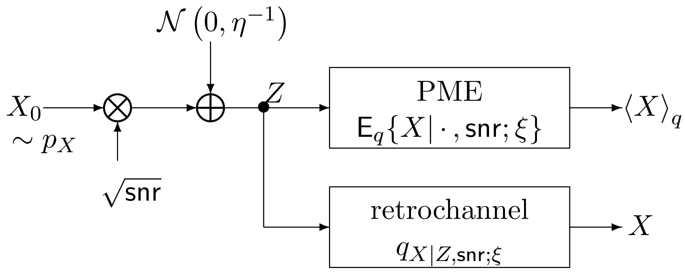
Given the parameters $(\beta, P_{\mathsf{snr}}, p_X, q_X, \sigma)$, we express in these parameters the large-system limit of the multiuser efficiency and spectral efficiency under both separate and joint decoding.

2.4.1 The Decoupling Principle

The multiuser channel $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ and the multiuser posterior mean estimator parameterized by (q_X,σ) are depicted in Figure 5(a), together with the companion (multiuser) retrochannel $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$. Here the input to the multiuser channel is denoted by $\boldsymbol{X}_0$ to distinguish from the output $\boldsymbol{X}$ of the retrochannel. For an arbitrary user k, the SNR is snr_k, and X_{0k}, X_k and $\langle X_k\rangle_q$ denote the input symbol, the retrochannel output and the PME output, all for user k.



(a)



(b)

Figure 5: (a) The multiuser channel, the (multiuser) PME, and the companion (multiuser) retrochannel. (b) The equivalent single-user Gaussian channel, PME and retrochannel.

In order to show the decoupling result, let us also consider the composition of a Gaussian channel, a PME and a companion retrochannel in the single-user setting as depicted in Figure 5(b). The input and output of the Gaussian channel are related by:

$$Z = \sqrt{\mathsf{snr}}\, X_0 + \frac{1}{\sqrt{\eta}} N \tag{16}$$

where the input $X_0 \sim p_X$, snr is the *input SNR*, $N \sim \mathcal{N}(0,1)$ the noise independent of X_0, and $\eta > 0$ the *inverse noise variance*. The conditional distribution associated with the channel is

$$p_{Z|X,\mathsf{snr};\eta}(z|x,\mathsf{snr};\eta) = \sqrt{\frac{\eta}{2\pi}} \exp\left[-\frac{\eta}{2}\left(z - \sqrt{\mathsf{snr}}\,x\right)^2\right]. \tag{17}$$

Let $q_{Z|X,\mathsf{snr};\xi}$ represent a Gaussian channel akin to (16), the only difference being that the inverse noise variance is ξ instead of η:

$$q_{Z|X,\mathsf{snr};\xi}(z|x,\mathsf{snr};\xi) = \sqrt{\frac{\xi}{2\pi}} \exp\left[-\frac{\xi}{2}\left(z - \sqrt{\mathsf{snr}}\,x\right)^2\right]. \tag{18}$$

Similar to that in the multiuser setting, by postulating the input distribution to be q_X, a posterior probability distribution $q_{X|Z,\mathsf{snr};\xi}$ is induced by the Bayes rule. Thus we have a single-user retrochannel defined by $q_{X|Z,\mathsf{snr};\xi}$, which outputs a random variable X given the channel output Z. A (generalized) single-user PME is defined naturally as:

$$\langle X\rangle_q = \mathsf{E}_q\left\{X \mid Z, \mathsf{snr};\xi\right\}. \tag{19}$$

The probability law of the composite system is determined by snr and two parameters η and ξ. We define the mean-square error of the PME as

$$\mathcal{E}(\mathsf{snr};\eta,\xi) = \mathsf{E}\left\{\left(X_0 - \langle X\rangle_q\right)^2 \,\middle|\, \mathsf{snr};\eta,\xi\right\}, \tag{20}$$

and also define the variance of the retrochannel as

$$\mathcal{V}(\mathsf{snr};\eta,\xi) = \mathsf{E}\left\{ \left(X - \langle X \rangle_q \right)^2 \,\middle|\, \mathsf{snr};\eta,\xi \right\}. \tag{21}$$

The following is claimed.[2]

Claim 1 *Consider the multiuser channel (4) with input distribution p_X and SNR distribution P_{snr}. Let its output be fed into the posterior mean estimator (7) and a retrochannel $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$, both parameterized by the postulated input q_X and noise level σ. Fix $(\beta, P_{\mathsf{snr}}, p_X, q_X, \sigma)$. Let X_{0k}, X_k, and $\langle X_k \rangle_q$ be the input, the retrochannel output and the posterior mean estimate for user k with input signal-to-noise ratio snr_k. Then,*

(a) The joint distribution of $(X_{0k}, X_k, \langle X_k \rangle_q)$ conditioned on the channel state $\boldsymbol{S}$ converges in probability as $K \to \infty$ and $K/L \to \beta$ to the joint distribution of $(X_0, X, \langle X \rangle_q)$, where $X_0 \sim p_X$ is the input to the single-user Gaussian channel (17) with inverse noise variance η, X is the output of the single-user retrochannel parameterized by (q_X, ξ), and $\langle X \rangle_q$ is the corresponding posterior mean estimate (19), with $\mathsf{snr} = \mathsf{snr}_k$.

(b) The parameter η, known as the multiuser efficiency, satisfies together with ξ the coupled equations:

$$\eta^{-1} = 1 + \beta\, \mathsf{E}\left\{ \mathsf{snr} \cdot \mathcal{E}(\mathsf{snr};\eta,\xi) \right\}, \tag{22a}$$

$$\xi^{-1} = \sigma^2 + \beta\, \mathsf{E}\left\{ \mathsf{snr} \cdot \mathcal{V}(\mathsf{snr};\eta,\xi) \right\}, \tag{22b}$$

where the expectations are taken over P_{snr}. In case of multiple solutions to (22), (η, ξ) is chosen to minimize the free energy expressed as

$$\begin{aligned} \mathcal{F} = &- \mathsf{E}\left\{ \int p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta) \log q_{Z|\mathsf{snr};\xi}(z|\mathsf{snr};\xi)\, \mathrm{d}\, z \right\} + \frac{1}{2\beta}[(\xi-1)\log e - \log \xi] \\ &- \frac{1}{2}\log\frac{2\pi}{\xi} - \frac{\xi}{2\eta}\log e + \frac{\sigma^2\xi(\eta-\xi)}{2\beta\eta}\log e + \frac{1}{2\beta}\log(2\pi) + \frac{\xi}{2\beta\eta}\log e. \end{aligned} \tag{23}$$

Claim 1 reveals that, from an individual user's viewpoint, the input-output relationship of the multiuser channel, PME and companion retrochannel is increasingly similar to that under a simple single-user setting in the large-system limit. In other words, given the three (scalar) input and output statistics, it is not possible to distinguish whether the underlying system is in the (large) multiuser or the single-user setting as depicted in Figure 5. It is also interesting here that the (asymptotically) equivalent single-user system takes a similar structure as the multiuser one, but with different noise levels.

Obtained using the replica method, the coupled equations (22) may have multiple solutions. This is known as phase coexistence in statistical physics. Among those solutions, the thermodynamically dominant solution is the one that gives the smallest value of the free energy (23). This is the solution that carries relevant operational meaning in the communication problem. In general, as the system parameters (such as the load) change, the dominant solution may switch from one of the coexisting solutions to another. This is known as *phase transition* (refer to Section 6 for numerical examples).

[2]Since as explained in Section 1, some of the key statistical physics tools (essentially the replica method) are not rigorously justified, some of the results in this paper are referred to as claims. Proofs are provided in Section 4 based on those statistical physics tools.

The single-user PME (19) is merely a decision function applied to the Gaussian channel output, which can be expressed explicitly as

$$\mathsf{E}_q\left\{X \mid Z, \mathsf{snr}; \xi\right\} = \frac{q_1(Z, \mathsf{snr}; \xi)}{q_0(Z, \mathsf{snr}; \xi)} \tag{24}$$

where we define the following important functions:

$$q_i(z, \mathsf{snr}; \xi) = \mathsf{E}_q\left\{ X^i \, q_{Z|X,\mathsf{snr};\xi}(z|X, \mathsf{snr}; \xi) \,\middle|\, \mathsf{snr} \right\}, \quad i = 0, 1, \ldots, \tag{25}$$

where the expectation is taken over q_X. Note that $q_0(z, \mathsf{snr}; \xi) = q_{Z|\mathsf{snr};\xi}(z|\mathsf{snr}; \xi)$. The decision function (24) is in general nonlinear. Due to Claim 1, although the multiuser PME output $\langle X_k \rangle_q$ is in general non-Gaussian, it is in fact asymptotically a function (the decision function (24)) of a conditionally Gaussian random variable Z centered at the actual input X_k scaled by $\sqrt{\mathsf{snr}_k}$ with a variance of η^{-1}.

Corollary 1 *In the large-system limit, the channel between the input X_{0k} and the multiuser posterior mean estimate $\langle X_k \rangle_q$ for user k is equivalent to the Gaussian channel $p_{Z|X,\mathsf{snr};\eta}$ concatenated with the one-to-one decision function (24) with $\mathsf{snr} = \mathsf{snr}_k$, where η is the multiuser efficiency determined by Claim 1.*

As shown in Section 4.2, for fixed snr and ξ, the decision function (24) is strictly monotone increasing in Z. Therefore, in the large-system limit, given the detection output $\langle X_k \rangle_q$, one can apply the inverse of the decision function to recover an equivalent conditionally Gaussian statistic Z. Note that $\eta \in [0, 1]$ from (22a). It is clear that, in the large-system limit, the multiple-access interference is consolidated into an enhancement of the thermal noise by η^{-1}, i.e., the effective SNR is reduced by a factor of η, hence the term *multiuser efficiency*. Equal for all users, the multiuser efficiency solves the coupled fixed-point equations (22). In case of multiple solutions, only the parameters (η, ξ) that minimize the free energy (23) conform with the operational meaning of the communication system. Indeed, the multiuser channel can be decoupled under the PME front end into a bank of single-user Gaussian channels with degradation in the SNR. This is referred to as the decoupling principle.

Since the decision function is one-to-one, it is inconsequential in both detection- and information-theoretic viewpoints. Hence the following result:

Corollary 2 *In the large-system limit, the mutual information between input symbol and the output of the multiuser posterior mean estimator for a particular user is equal to the input-output mutual information of the equivalent single-user Gaussian channel with the same input distribution and SNR, and an inverse noise variance η as the multiuser efficiency given by Claim 1.*

According to Corollary 2, the mutual information $I\left(X_k; \langle X_k \rangle \,|\boldsymbol{S}\right)$ for a user with signal-to-noise ratio $\mathsf{snr}_k = \mathsf{snr}$ converges to a special multiuser information function defined as

$$I(\eta \, \mathsf{snr}) = \mathsf{D}\left(p_{Z|X,\mathsf{snr};\eta} \| p_{Z|\mathsf{snr};\eta} | p_X\right), \tag{26}$$

where $\mathsf{D}\left(\cdot \| \cdot | \cdot\right)$ stands for conditional (Kullback-Leibler) divergence, and $p_{Z|\mathsf{snr};\eta}$ is the marginal distribution of the output of the channel (16). The overall spectral efficiency under separate decoding is the sum of the single-user mutual informations divided by the dimension of the multiuser channel (spreading factor L), which is simply

$$\mathsf{C}_{\text{sep}}(\beta) = \beta \, \mathsf{E}\left\{I(\eta \, \mathsf{snr})\right\}, \tag{27}$$

where the expectation is over P_{snr}.

In general, it is straightforward to determine the multiuser efficiency η (and the inverse noise variance ξ) by solving the joint equations (22). Define the following functions akin to (25):

$$p_i(z, \mathsf{snr}; \eta) = \mathsf{E}\left\{ X^i \, p_{Z|X,\mathsf{snr};\eta}(z \,|\, X, \mathsf{snr}; \eta) \,\middle|\, \mathsf{snr} \right\}, \quad i = 0, 1, \dots \tag{28}$$

Some algebra leads to

$$\mathcal{E}(\mathsf{snr}; \eta, \xi) = 1 + \int p_0(z, \mathsf{snr}; \eta) \frac{q_1^2(z, \mathsf{snr}; \xi)}{q_0^2(z, \mathsf{snr}; \xi)} - 2 p_1(z, \mathsf{snr}; \eta) \frac{q_1(z, \mathsf{snr}; \xi)}{q_0(z, \mathsf{snr}; \xi)} \, \mathrm{d}z \tag{29}$$

and

$$\mathcal{V}(\mathsf{snr}; \eta, \xi) = \int p_0(z, \mathsf{snr}; \eta) \frac{q_0(z, \mathsf{snr}; \xi) q_2(z, \mathsf{snr}; \xi) - q_1^2(z, \mathsf{snr}; \xi)}{q_0^2(z, \mathsf{snr}; \xi)} \, \mathrm{d}z. \tag{30}$$

Numerical integrations can be applied to evaluate (29) and (30) in general. It is then viable to find solutions to the joint equations (22) numerically. In case of multiple sets of solutions, the ambiguity is resolved by choosing the one that minimizes the free energy (23). Note that the mean-square error and variance often admit simpler expressions than (29) and (30) under certain practical inputs, which may ease the computation significantly (see examples in Section 2.5).

2.4.2 Optimal Detection and Spectral Efficiency

Among all multiuser detection schemes, the individually optimal detector has particular importance since it achieves the best error performance. As we shall see, the optimal spectral efficiency achievable by joint decoding is also tightly related to the multiuser efficiency of optimal detection.

As shown in Section 2.3.2, the individually optimal detector can be regarded as a PME with a postulated measure that is exactly the same as the actual measure, i.e., $q = p$. Consider the channel, PME and retrochannel in the multiuser setting as depicted in Figure 5(a). It is clear that in case of optimal detection, the input $\boldsymbol{X}_0$ to the multiuser channel and the retrochannel output $\boldsymbol{X}$ are independent and identically distributed given $(\boldsymbol{Y}, \boldsymbol{S})$. In fact the retrochannel becomes unnecessary. The decoupling principle stated in Claim 1 can be particularized in case of $q = p$. Easily, the multiuser efficiency and the postulated inverse noise variance satisfy joint equations:

$$\eta^{-1} = 1 + \beta \, \mathsf{E}\left\{ \mathsf{snr} \cdot \mathcal{E}(\mathsf{snr}; \eta, \xi) \right\}, \tag{31a}$$

$$\xi^{-1} = 1 + \beta \, \mathsf{E}\left\{ \mathsf{snr} \cdot \mathcal{V}(\mathsf{snr}; \eta, \xi) \right\}. \tag{31b}$$

Due to the replica symmetry assumption, and noting that $\mathcal{E}(\mathsf{snr}; x, x) = \mathcal{V}(\mathsf{snr}; x, x)$ for all x, we take the solution $\eta = \xi$. It should be cautioned that (31) may have other solutions with $\eta \neq \xi$ in the unlikely case that replica symmetry does not hold for the optimal detection.

In the equivalent single-user setting (Figure 5(b)), the above arguments imply that the postulated channel is also identical to the actual channel, and X and X_0 are i.i.d. given Z. The posterior mean estimate of X given the output Z is

$$\langle X \rangle = \mathsf{E}\left\{ X | Z, \mathsf{snr}; \eta \right\}. \tag{32}$$

Clearly, $\langle X \rangle$ is also the (nonlinear) MMSE estimate, since it achieves the minimum mean-square error:

$$\mathsf{mmse}(\eta \, \mathsf{snr}) = \mathsf{E}\left\{ (X - \langle X \rangle)^2 \,\middle|\, \mathsf{snr}; \eta \right\}. \tag{33}$$

Indeed,

$$\mathcal{E}(\mathsf{snr}; x, x) = \mathcal{V}(\mathsf{snr}; x, x) = \mathsf{mmse}(x \, \mathsf{snr}), \quad \forall x. \tag{34}$$

The following is a special case of Corollary 1 for the individually optimal detector.

Claim 2 *In the large-system limit, the distribution of the output $\langle X_k \rangle$ of the individually optimal detector for the multiuser channel (4) conditioned on $X_k = x$ being transmitted with signal-to-noise ratio snr_k is identical to the distribution of the posterior mean estimate $\langle X \rangle$ of the single-user Gaussian channel (16) conditioned on $X_0 = x$ being transmitted with $\mathsf{snr} = \mathsf{snr}_k$, where the optimal multiuser efficiency η satisfies a fixed-point equation:*

$$\eta^{-1} = 1 + \beta \, \mathsf{E} \{ \mathsf{snr} \cdot \mathsf{mmse}(\eta \, \mathsf{snr}) \} . \tag{35}$$

The single-user PME (32) is a (nonlinear) decision function that admits an expression as (24) with q replaced by p. The MMSE can be computed as

$$\mathsf{mmse}(\eta \, \mathsf{snr}) = 1 - \int \frac{p_1^2(z, \mathsf{snr}; \eta)}{p_0(z, \mathsf{snr}; \eta)} \, \mathrm{d}z. \tag{36}$$

Solutions to the fixed-point equation (35) can be found without much difficulty. There are cases that (35) has more than one solution. The ambiguity is resolved by taking the one that gives the minimum of the free energy (23) with $\xi = \eta$, or equivalently, as we shall see next, the optimal spectral efficiency.

The single-user mutual information is given by (27) due to Corollary 2, where the multiuser efficiency is now given by Claim 2. The optimal spectral efficiency under joint decoding is greater than (27), where the increase is given by the following:

Claim 3 *The spectral efficiency gain of optimal joint decoding over individually optimal detection followed by separate decoding of the multiuser channel (4) is determined, in the large-system limit, by the optimal multiuser efficiency η as*

$$\mathsf{C}_{\text{joint}}(\beta) - \mathsf{C}_{\text{sep}}(\beta) = \frac{1}{2} [(\eta - 1) \log e - \log \eta] = \mathsf{D} \left(\mathcal{N}(0, \eta) \| \mathcal{N}(0, 1) \right) . \tag{37}$$

In other words, the spectral efficiency under joint decoding is

$$\mathsf{C}_{\text{joint}}(\beta) = \beta \, \mathsf{E} \{ I(\eta \, \mathsf{snr}) \} + \frac{1}{2} [(\eta - 1) \log e - \log \eta]. \tag{38}$$

In case of multiple solutions to (35), the optimal multiuser efficiency η is the one that gives the smallest $\mathsf{C}_{\text{joint}}$.

Indeed, Müller's conjecture on the mutual information loss [26] is true for arbitrary inputs and SNRs. Incidentally, the loss is identified as a divergence between two Gaussian distributions in (37).

Equal-power Gaussian input is the first known case that admits a closed-form solution for the multiuser efficiency [6] and thus also the spectral efficiencies. The spectral efficiencies under joint and separate decoding were found for Gaussian inputs with fading in [12], and then found implicitly in [24] and later explicitly [25] for equal-power users with binary inputs. Formulas (27) and (38) are the first general results for arbitrary input distributions and received powers.

Interestingly, the spectral efficiencies under joint and separate decoding are also related by an integral equation.

Theorem 1 *For every load $\beta > 0$, every input and power distribution,*

$$\mathsf{C}_{\text{joint}}(\beta) = \int_0^\beta \frac{1}{\beta'} \mathsf{C}_{\text{sep}}(\beta') \, \mathrm{d}\beta'. \tag{39}$$

Proof: Since $\mathsf{C}_{\text{joint}}(0) = 0$ trivially, it suffices to show

$$\beta \frac{\mathrm{d}}{\mathrm{d}\beta} \mathsf{C}_{\text{joint}}(\beta) = \mathsf{C}_{\text{sep}}(\beta). \tag{40}$$

By (37) and (38), it is enough to show

$$\beta \frac{\mathrm{d}}{\mathrm{d}\beta} \mathsf{E}\left\{I(\eta\,\mathsf{snr})\right\} + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\beta}[(\eta - 1)\log e - \log \eta] = 0. \tag{41}$$

Noticing that the multiuser efficiency η is a function of the system load β, (41) is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}\eta} \mathsf{E}\left\{I(\eta\,\mathsf{snr})\right\} + \frac{1}{2\beta}\left(1 - \eta^{-1}\right)\log e = 0. \tag{42}$$

By a recent formula that links the mutual information and MMSE in Gaussian channels [36, 37],

$$\frac{1}{\log e}\frac{\mathrm{d}}{\mathrm{d}\eta} I(\eta \mathsf{snr}) = \frac{\mathsf{snr}}{2}\,\mathsf{mmse}(\eta\,\mathsf{snr}). \tag{43}$$

Thus (42) holds as η satisfies the fixed-point equation (35). ■

Theorem 1 is an outcome of the chain rule of mutual information, which holds for all inputs and arbitrary number of users:

$$I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}) = \sum_{k=1}^{K} I(X_k;\boldsymbol{Y}|\boldsymbol{S}, X_{k+1}, \dots, X_K). \tag{44}$$

The left hand side of (44) is the total mutual information of the multiuser channel. Each mutual information in the right hand side of (44) is a single-user mutual information over the multiuser channel conditioned on the symbols of previously decoded users. As argued in the following, the limit of (44) as $K \to \infty$ becomes the integral equation (39).

Consider an interference canceler with PME front ends against yet undecoded users that decodes the users successively in which reliably decoded symbols are used to reconstruct the interference for cancellation. Since the error probability of intermediate decisions vanishes with code block-length, the interference from decoded users are asymptotically completely removed. Assume without loss of generality that the users are decoded in reverse order, then the PME for user k sees only $k-1$ interfering users. Hence the performance for user k under such successive decoding is identical to that under multiuser detection with separate decoding in a multiuser system with k instead of K users. Nonetheless, the equivalent single-user channel for each user is Gaussian by Corollary 1. The multiuser efficiency experienced by user k, $\eta(k/L)$, is a function of the load k/L seen by the PME for user k. By Corollary 2, the single-user mutual information for user k is therefore

$$I\left(\eta(k/L)\,\mathsf{snr}_k\right). \tag{45}$$

Since snr_k are i.i.d., the overall spectral efficiency under successive decoding converges almost surely:

$$\frac{1}{L}\sum_{k=1}^{K} I\left(\eta(k/L)\,\mathsf{snr}_k\right) \to \mathsf{E}\left\{\int_0^{\beta} I(\beta'\,\mathsf{snr})\,\mathrm{d}\beta'\right\}. \tag{46}$$

Note that the above result on successive decoding is true for arbitrary input distribution and PME detectors. In the special case of the individually optimal detection, for which the postulated system is identical to the actual one, the right hand side of (46) is equal to $\mathsf{C}_{\text{joint}}(\beta)$ by Theorem 1. We can summarize this principle as:

Claim 4 *In the large-system limit, successive decoding with an individually optimal detection front end against yet undecoded users achieves the optimal CDMA channel capacity under arbitrary input distributions.*

Claim 4 is a generalization of the previous result that a successive canceler with a linear MMSE front end against undecoded users achieves the capacity of the CDMA channel under Gaussian inputs [38, 39, 11, 40, 41, 42]. In the special case of Gaussian inputs, however, the optimality is known to hold for any finite number of users [38, 11].

2.5 Recovering Known Results

As shown in 2.3, several well-known multiuser detectors can be regarded as appropriately parameterized PMEs. Thus many previously known results can be recovered as special case of the new findings in Section 2.4.

2.5.1 Linear Detectors

Let the postulated prior q_X be standard Gaussian so that the PME represents a linear multiuser detector. Since the input Z and output X of the retrochannel are jointly Gaussian (refer to Figure 5(b)), the single-user PME is simply a linear attenuator:

$$\langle X \rangle_q = \frac{\xi\sqrt{\mathsf{snr}}}{1+\xi\mathsf{snr}} Z. \tag{47}$$

¿From (20), the mean-square error is

$$\begin{aligned} \mathcal{E}(\mathsf{snr};\eta,\xi) &= \mathsf{E}\left\{ \left[X_0 - \frac{\xi\sqrt{\mathsf{snr}}}{1+\xi\mathsf{snr}} \left(\sqrt{\mathsf{snr}} X_0 + \frac{1}{\sqrt{\eta}} N \right) \right]^2 \right\} & (48) \\ &= \frac{\eta + \xi^2\mathsf{snr}}{\eta(1+\xi\mathsf{snr})^2}. & (49) \end{aligned}$$

Meanwhile, the variance of X conditioned on Z is independent of Z. Hence the variance (21) of the retrochannel output is independent of η:

$$\mathcal{V}(\mathsf{snr};\eta,\xi) = \frac{1}{1+\xi\mathsf{snr}}. \tag{50}$$

¿From Claim 1, one finds that ξ is the solution to

$$\xi^{-1} = \sigma^2 + \beta\,\mathsf{E}\left\{ \frac{\mathsf{snr}}{1+\xi\mathsf{snr}} \right\}, \tag{51}$$

and the multiuser efficiency is determined as

$$\eta = \xi + \xi\left(\sigma^2 - 1\right) \left[1 + \beta\,\mathsf{E}\left\{ \frac{\mathsf{snr}}{(1+\xi\mathsf{snr})^2} \right\} \right]^{-1}. \tag{52}$$

Clearly, the large-system multiuser efficiency of such a linear detector is independent of the input distribution.

Suppose also that the postulated noise level $\sigma \to \infty$. The PME becomes the matched filter. One finds $\xi\sigma^2 \to 1$ by (51) and consequently, the multiuser efficiency of the matched filter is [6]

$$\eta^{(\mathrm{mf})} = \frac{1}{1+\beta\,\mathsf{E}\{\mathsf{snr}\}}. \tag{53}$$

In case $\sigma = 1$, one has the linear MMSE detector. By (52), $\eta = \xi$ and by (51), the multiuser efficiency $\eta^{(\mathrm{mmse})}$ satisfies

$$\eta^{-1} = 1 + \beta\,\mathsf{E}\left\{\frac{\mathsf{snr}}{1+\eta\mathsf{snr}}\right\}, \tag{54}$$

which is exactly the Tse-Hanly equation [7, 11]. The fixed-point equation (54) has a unique positive solution.

By letting $\sigma \to 0$ one obtains the decorrelator. If $\beta < 1$, then (51) gives $\xi \to \infty$ and $\xi\sigma^2 \to 1-\beta$, and the multiuser efficiency is found as $\eta = 1 - \beta$ by (52) regardless of the SNR distribution (as shown in [6]). If $\beta > 1$, and assuming the generalized form of the decorrelator as the Moore-Penrose inverse of the correlation matrix [6], then ξ is the unique solution to

$$\xi^{-1} = \beta\,\mathsf{E}\left\{\frac{\mathsf{snr}}{1+\xi\mathsf{snr}}\right\} \tag{55}$$

and the multiuser efficiency is found by (52) with $\sigma = 0$. In the special case of identical SNRs, an explicit expression can be found [8, 9]

$$\eta^{(\mathrm{dec})} = \frac{\beta - 1}{\beta + \mathsf{snr}(\beta-1)^2}, \qquad \beta > 1. \tag{56}$$

By Corollary 1, the mutual information with input distribution p_X for a user with snr under linear multiuser detection is the input-output mutual information of the single-user Gaussian channel (16) with the same input:

$$I(X; \langle X\rangle_q | \mathsf{snr}) = I(\eta\,\mathsf{snr}), \tag{57}$$

where η depends on which type of linear detector is in use. Gaussian priors are known to achieve the capacity:

$$\mathsf{C}(\mathsf{snr}) = \frac{1}{2}\log(1 + \eta\,\mathsf{snr}). \tag{58}$$

By Corollary 3, the total spectral efficiency under Gaussian inputs is expressed in terms of the linear MMSE multiuser efficiency:

$$\mathsf{C}_{\mathrm{joint}}^{(\mathrm{Gaussian})} = \frac{\beta}{2}\,\mathsf{E}\left\{\log\left(1+\eta^{(\mathrm{mmse})}\mathsf{snr}\right)\right\} + \frac{1}{2}\left[\left(\eta^{(\mathrm{mmse})} - 1\right)\log e - \log\eta^{(\mathrm{mmse})}\right]. \tag{59}$$

This is exactly Shamai and Verdú's result for fading channels [12].

2.5.2 Optimal Detectors

Using the actual input distribution p_X as the postulated prior of the PME results in optimum multiuser detectors. In case of the jointly optimal detector, the postulated noise level $\sigma = 0$, and (22) becomes

$$\eta^{-1} = 1 + \beta\,\mathsf{E}\left\{\mathsf{snr}\cdot\mathcal{E}(\mathsf{snr};\eta,\xi)\right\}, \tag{60a}$$

$$\xi^{-1} = \beta\,\mathsf{E}\left\{\mathsf{snr}\cdot\mathcal{V}(\mathsf{snr};\eta,\xi)\right\}, \tag{60b}$$

where $\mathcal{E}(\cdot)$ and $\mathcal{V}(\cdot)$ are given by (29) and (30) respectively with $q_i(z,\mathsf{snr};x) = p_i(z,\mathsf{snr};x)$, $\forall x$. The parameters can then be solved numerically.

In case of the individually optimal detector, one sets $\sigma = 1$ so that $q = p$. The optimal multiuser efficiency η is the solution to the fixed-point equation (35) given in Claim 2.

It is of practical interest to find the spectral efficiency under the constraint that the input symbols are antipodally modulated as in the popular BPSK. In this case, the probability mass function $p_X(x) = 1/2$, $x = \pm 1$, maximizes the mutual information. It is not difficult to show that

$$\mathsf{mmse}(\eta\,\mathsf{snr}) = 1 - \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tanh\left(\eta\mathsf{snr} - z\sqrt{\eta\mathsf{snr}}\right) \mathrm{d}z. \tag{61}$$

By Claim 2, The multiuser efficiency, $\eta^{(\mathrm{b})}$, where the superscript (b) stands for binary inputs, is a solution to the fixed-point equation [9]:

$$\eta^{-1} = 1 + \beta\,\mathsf{E}\left\{\mathsf{snr}\left[1 - \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tanh\left(\eta\mathsf{snr} - z\sqrt{\eta\mathsf{snr}}\right) \mathrm{d}z\right]\right\}, \tag{62}$$

which is a generalization of an earlier result assuming equal-power users due to Tanaka [24]. The single-user channel capacity for a user with signal-to-noise ratio snr is the same as that obtained by Müller and Gerstacker [25] and is given by

$$\mathsf{C}^{(\mathrm{b})}(\mathsf{snr}) = \eta^{(\mathrm{b})}\,\mathsf{snr}\log e - \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \log\cosh\left(\eta^{(\mathrm{b})}\mathsf{snr} - z\sqrt{\eta^{(\mathrm{b})}\mathsf{snr}}\right) \mathrm{d}z. \tag{63}$$

The total spectral efficiency of the CDMA channel subject to binary inputs is thus

$$\begin{aligned}\mathsf{C}^{(\mathrm{b})}_{\mathrm{joint}} =& \beta\,\mathsf{E}\left\{\eta^{(\mathrm{b})}\,\mathsf{snr}\log e - \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \log\cosh\left(\eta^{(\mathrm{b})}\mathsf{snr} - z\sqrt{\eta^{(\mathrm{b})}\mathsf{snr}}\right) \mathrm{d}z\right\} \\ &+ \frac{1}{2}\left[\left(\eta^{(\mathrm{b})} - 1\right)\log e - \log\eta^{(\mathrm{b})}\right],\end{aligned} \tag{64}$$

which is also a generalization of Tanaka's implicit result in [24].

3 Multiuser Communications and Statistical Physics

This section prepares the reader with concepts and methodologies that will be needed to prove the results summarized in Section 2.4. Although one can work with the mathematical framework only and avoid foreign concepts, we believe it is more enlightening to draw an equivalence in between multiuser communications and many-body problems in statistical physics. Such an analogy is first seen in a primitive form in [24] and will be developed to a full generality here.

3.1 A Note on Statistical Physics

Let the microscopic state of a system be described by the configuration of some K variables as a vector $\boldsymbol{x}$. The *Hamiltonian* is a function of the configuration, denoted by $H(\boldsymbol{x})$. The state of the system evolves over time according to some physical laws, and after long enough time it reaches thermal equilibrium. The time average of an observable quantity can be obtained by averaging over the ensemble of the states. In particular, the *energy* of the system is

$$\mathcal{E} = \sum_{\boldsymbol{x}} p(\boldsymbol{x}) H(\boldsymbol{x}) \tag{65}$$

where $p(\boldsymbol{x})$ is the probability of the system being found in configuration $\boldsymbol{x}$. In other words, as far as the macroscopic properties are concerned, it suffices to describe the system statistically instead of solving the exact dynamic trajectories. The disorder of the thermodynamic system is measured by the notion of *entropy*, which is defined as

$$\mathcal{S} = -\sum_{\boldsymbol{x}} p(\boldsymbol{x}) \log p(\boldsymbol{x}). \tag{66}$$

One particularly useful macroscopic quantity of the thermodynamic system is the *free energy*:

$$\mathcal{F} = \mathcal{E} - T\,\mathcal{S} \tag{67}$$

where $T \geq 0$ is the *temperature*.

It is assumed that the system is not isolated and may interact with the surroundings. As a result, at thermal equilibrium, the temperature and energy of the system remain constant, the entropy is the maximum possible, and the free energy is at its minimum. It can be shown using the Lagrange multiplier method that the equilibrium probability distribution that maximizes the entropy is the Boltzmann distribution:

$$p(\boldsymbol{x}) = Z^{-1} \exp\left[-\frac{1}{T} H(\boldsymbol{x})\right] \tag{68}$$

where

$$Z = \sum_{\boldsymbol{x}} \exp\left[-\frac{1}{T} H(\boldsymbol{x})\right] \tag{69}$$

is the *partition function*, and the temperature T is determined by the energy constraint (65). Hence the system is found in each configuration with a probability that is negative exponential in the Hamiltonian associated with the configuration, and the most probable configuration is the ground state which has the minimum Hamiltonian. Using (65)–(69), one finds that the free energy at equilibrium can also be expressed as

$$\mathcal{F} = -T\, \log Z. \tag{70}$$

The free energy is often the starting point for calculating macroscopic properties of a thermodynamic system.

3.2 Multiuser Communications and Spin Glasses

The communication problem faced by the detector is to infer statistically the information-bearing symbols given the received signal and knowledge about the channel state. Naturally, the posterior probability distribution plays a central role. In the multiple-access channel (4), the channel state consists of the spreading sequences and the SNRs, collectively represented by the matrix $\boldsymbol{S}$. The channel is described by the Gaussian density $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ given by (9). By postulating an input q_X and a channel (10) which differs from the actual one only in the noise level, the postulated posterior distribution can be obtained by using the Bayes formula (cf. (6)) as

$$q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{S}) = \left[q_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\right]^{-1} q_{\boldsymbol{X}}(\boldsymbol{x})\ \left(2\pi\sigma^2\right)^{-\frac{L}{2}} \exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right] \tag{71}$$

where

$$q_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S}) = \left(2\pi\sigma^2\right)^{-\frac{L}{2}}\ \mathsf{E}_q\left\{\exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{X}\|^2\right] \,\middle|\, \boldsymbol{S}\right\} \tag{72}$$

and the expectation in (72) is taken conditioned on $\boldsymbol{S}$ over $\boldsymbol{X}$ with distribution $q_{\boldsymbol{X}}$.

Interestingly, one can associate the posterior probability distribution (71) with the characteristics of a thermodynamic system called spin glass. In certain special cases, this connection is found in Tanaka's important paper [24], but we believe this work is the first to draw this analogy in the general setting. A *spin glass* is a system consisting of many directional spins, in which the interaction of the spins is determined by the so-called *quenched random variables* whose values are determined by the realization of the spin glass.[3] Let the microscopic state of a spin glass be denoted by a K-dimensional vector $\boldsymbol{x}$, and the quenched random variables by $(\boldsymbol{y}, \boldsymbol{S})$. The system can be understood as K random spins sitting in quenched randomness $(\boldsymbol{y}, \boldsymbol{S})$, and its statistical physics described as in Section 3.1 with a parameterized Hamiltonian $H_{\boldsymbol{y},\boldsymbol{S}}(\boldsymbol{x})$.

Indeed, if the temperature $T = 1$ and that the Hamiltonian is defined as

$$H_{\boldsymbol{y},\boldsymbol{S}}(\boldsymbol{x}) = \frac{L}{2} \log\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2} \|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2 - \log q_{\boldsymbol{X}}(\boldsymbol{x}), \tag{73}$$

then $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{S}}$ is a Boltzmann distribution and $q_{\boldsymbol{Y}|\boldsymbol{S}}$ the corresponding partition function (cf. (68) and (69)). In other words, by defining an appropriate Hamiltonian, the posterior probability distribution associated with a multiuser communication system is identical to the configuration distribution of the spin glass at equilibrium. Precisely, the probability that the transmitted symbol is $\boldsymbol{X} = \boldsymbol{x}$ under the postulated model, given the observation $\boldsymbol{Y}$ and the channel state $\boldsymbol{S}$, is equal to the probability that the spin glass is at configuration $\boldsymbol{x}$, given the quenched random variables $(\boldsymbol{Y}, \boldsymbol{S})$. It is interesting to note that Gaussian distribution is a natural Boltzmann distribution with squared Euclidean norm as the Hamiltonian.

The quenched randomness $(\boldsymbol{Y}, \boldsymbol{S})$ takes a specific distribution in our problem, i.e., $(\boldsymbol{Y}, \boldsymbol{S})$ is a realization of the received signal and channel state matrix according to the prior and conditional distributions that underlie the "original" spins. Indeed, the communication system depicted in Figure 5(a) can be also understood as a spin glass $\boldsymbol{X}$ subject to physical law q sitting in the quenched randomness caused by another spin glass $\boldsymbol{X}_0$ subject to physical law p. The channel corresponds to the random mapping from a given spin glass configuration to an induced quenched randomness. Conversely, the retrochannel corresponds to the random mechanism that maps some quenched randomness into an induced spin glass configuration distribution.

The free energy of the thermodynamic (or communication) system normalized by the number of users is

$$-\frac{T}{K} \log Z(\boldsymbol{Y}, \boldsymbol{S}) = -\frac{1}{K} \log q_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{S}). \tag{74}$$

Due to the self-averaging assumption, the randomness of (74) vanishes as $K \to \infty$. As a result, the free energy per user converges in probability to its expected value over the distribution of the quenched random variables $(\boldsymbol{Y}, \boldsymbol{S})$ in the large-system limit, which is denoted by $\mathcal{F}$,

$$\mathcal{F} = -\lim_{K\to\infty} \mathsf{E}\left\{\frac{1}{K} \log q_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{S})\right\}. \tag{75}$$

Hereafter, by the free energy we refer to the large-system limit (75), which will be calculated in Section 4.

The reader should be cautioned that for disordered systems, thermodynamic quantities may or may not be self-averaging [43]. The self-averaging property remains to be proved or disproved in the CDMA context. This is a challenging problem on its own. Buttressed by numerical examples and

[3] An example is a system consisting molecules with magnetic spins that evolve over time, while the positions of the molecules that determine the amount of interactions are random (disordered) but remain fixed for each concrete instance as in a piece of glass (hence the name of spin glass).

associated results using random matrix theory, in this work the self-averaging property is assumed to hold.

The self-averaging property resembles the asymptotic equipartition property (AEP) in information theory [44]. An important consequence is that a macroscopic quantity of a thermodynamic system, which is a function of a large number of random variables, may become increasingly predictable from merely a few parameters independent of the realization of the random variables as the system size grows without bound. Indeed, such a macroscopic quantity converges in probability to its ensemble average in the thermodynamic limit.

In the CDMA context, the self-averaging property leads to the strong consequence that for almost all realizations of the received signal and the spreading sequences, macroscopic quantities such as the BER, the output SNR and the spectral efficiency, averaged over data, converge to deterministic quantities in the large-system limit. Previous work (e.g. [11, 7, 10]) has shown convergence of performance measures for almost all spreading sequences. The self-averaging property results in convergence of certain empirical performance measures, which holds for almost all realizations of the data and noise as well.

3.3 Spectral Efficiency and Detection Performance

Consider the multiuser channel, the multiuser PME and the companion retrochannel as depicted in Figure 5(a). Equipped with the statistical physics concepts introduced in the above, this subsection associates the spectral efficiency and detection performance of such a system with more tangible quantities for calculation.

3.3.1 Spectral Efficiency and Free Energy

For a fixed input distribution p_X, the total input-output mutual information of the multiuser channel is

$$I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}) = \mathsf{E}\left\{ \log \frac{p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S})}{p_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{S})} \,\middle|\, \boldsymbol{S} \right\} \tag{76}$$

$$= \mathsf{E}\left\{ \log p_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{S}) \,\middle|\, \boldsymbol{S} \right\} - \frac{L}{2}\log(2\pi e). \tag{77}$$

where the simplification to (77) is because $p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}$ given by (9) is an L-dimensional Gaussian density. To calculate (77) is formidable for an arbitrary realization of $\boldsymbol{S}$. However, due to the self-averaging property, it suffices to evaluate its expectation over the spreading sequences. In view of (75), the large-system spectral efficiency is affine in the free energy with a postulated measure q identical to the actual measure p:

$$\mathsf{C} = \frac{1}{L} I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}) \tag{78}$$

$$= -\beta\,\mathsf{E}\left\{ \frac{1}{K}\log p_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{S}) \,\middle|\, \boldsymbol{S} \right\} - \frac{1}{2}\log(2\pi e) \tag{79}$$

$$\to \beta\,\mathcal{F}|_{q=p} - \frac{1}{2}\log(2\pi e). \tag{80}$$

The relationship (80) is a full generalization of a previous observation [24] in some special cases. In fact, the analogy between free energy and information-theoretic quantities has also been noticed in belief propagation [45], coding [46] and optimization problems [47].

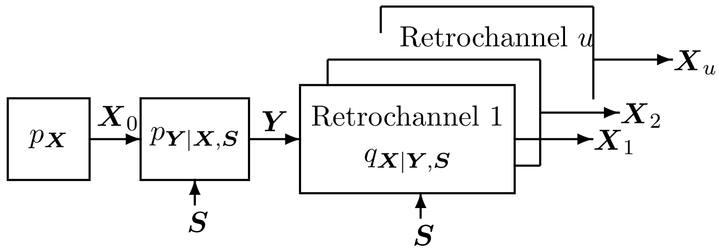


Figure 6: The replicas of the retrochannel.

3.3.2 Detection Performance and Moments

In case of a multiuser detector front end, one is interested in the quality of the detection output for each user, which is completely described by the distribution of the detection output conditioned on the input. Let us focus on an arbitrary user k, and let X_{0k}, $\langle X_k \rangle_q$ and X_k be the input, the PME output, and the retrochannel output in the multiuser setting (cf. Figure 5(a)). Instead of the conditional distribution $P_{\langle X_k \rangle_q | X_{0k}}$, we solve a more ambitious problem: What is the joint distribution of $(X_{0k}, \langle X_k \rangle_q, X_k)$ conditioned on the channel state $\boldsymbol{S}$ in the large-system limit?

Our approach is to calculate the joint moments

$$\mathsf{E}\left\{ X_{0k}^i \, X_k^j \, \langle X_k \rangle_q^l \,\middle|\, \boldsymbol{S} \right\}, \quad i, j, l = 0, 1, \dots \tag{81}$$

By the self-averaging property, the moments converge to the same values for almost all realizations of the channel state. Thus it suffices to calculate

$$\mathsf{E}\left\{ X_{0k}^i \, X_k^j \, \langle X_k \rangle_q^l \right\} \tag{82}$$

as $K \to \infty$, which is viable by studying the free energy associated with a modified version of the partition function (72). More on this later.

The joint distribution becomes clear once all the moments (82) are determined, so does the relationship between the detection output $\langle X_k \rangle_q$ and the input X_{0k}. It turns out the large-system joint distribution of $(X_{0k}, \langle X_k \rangle_q, X_k)$ is exactly the same as that of the input, PME output and retrochannel output associated with a single-user Gaussian channel with the same input distribution but a degradation in the SNR. In short, a CDMA channel with a multiuser detector front end can be decoupled into a bank of equivalent single-user channels in the large-system limit. The mutual information between the input and the detection output for user k is given by

$$I(X_{0k}; \langle X_k \rangle_q \mid \boldsymbol{S}), \tag{83}$$

which can be derived once the input-output relationship is known. It will be shown that conditioning on the channel state $\boldsymbol{S}$ becomes superfluous as $K \to \infty$.

We have distilled our problems under both joint and separate decoding to finding some ensemble averages, namely, the free energy (75) and the joint moments (82). In order to calculate these quantities, we resort to a powerful technique developed in the theory of spin glass, the heart of which is sketched in the following subsection.

3.4 Replica Method

The replica method, originally developed in spin glass theory [27], was introduced to the field of multiuser detection by Tanaka [24] to analyze the optimal detectors under equal-power Gaussian or binary input (see also [48]). We now outline the method in a general setting.

The expected value of the logarithm in (75) can be reformulated as

$$\mathcal{F} = -\lim_{K\to\infty} \frac{1}{K} \lim_{u\to 0} \frac{\partial}{\partial u} \log \mathsf{E}\left\{Z^u(\boldsymbol{Y}, \boldsymbol{S})\right\} \tag{84}$$

where $Z(\boldsymbol{Y}, \boldsymbol{S}) = q_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{S})$. The equivalence of (75) and (84) can be easily verified by noticing that

$$\lim_{u\to 0} \frac{\partial}{\partial u} \log \mathsf{E}\left\{\Theta^u\right\} = \lim_{u\to 0} \frac{\mathsf{E}\left\{\Theta^u \log \Theta\right\}}{\mathsf{E}\left\{\Theta^u\right\}} = \mathsf{E}\left\{\log \Theta\right\}, \quad \forall \Theta > 0. \tag{85}$$

For an arbitrary integer replica number u, we introduce u independent replicas of the retrochannel (or the spin glass) with the same received signal $\boldsymbol{Y}$ and channel state $\boldsymbol{S}$ as depicted in Figure 6. The partition function of the replicated system is

$$Z^u(\boldsymbol{y}, \boldsymbol{S}) = \mathsf{E}_q\left\{ \prod_{a=1}^{u} q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{X}_a, \boldsymbol{S}) \,\middle|\, \boldsymbol{S} \right\} \tag{86}$$

where the expectation is taken over the replicas $\{X_{ak} | a = 1, \ldots, u,\ k = 1, \ldots, K\}$. Here, X_{ak} are i.i.d. (with distribution q_X) since $(\boldsymbol{Y}, \boldsymbol{S})$ are given. Hence we can evaluate

$$-\lim_{K\to\infty} \frac{1}{K} \log \mathsf{E}\left\{Z^u(\boldsymbol{Y}, \boldsymbol{S})\right\} \tag{87}$$

as a function of the integer u. The *replica trick* assumes that the resulting expression is also valid for an arbitrary real number u at the vicinity of 0 and finds the derivative at $u = 0$ as the free energy. Besides validity of continuing to non-integer values of the replica number u, it is also necessary to assume that the two limits in (84) can be interchanged.

It remains to calculate (87). Note that $(\boldsymbol{Y}, \boldsymbol{S})$ is induced by the transmitted symbols $\boldsymbol{X}_0$. By taking expectation over $\boldsymbol{Y}$ first and then averaging over the spreading sequences, but conditioned on $\boldsymbol{X}_0$ and $\boldsymbol{X}_a$, one finds that

$$\frac{1}{K} \log \mathsf{E}\left\{Z^u(\boldsymbol{Y}, \boldsymbol{S})\right\} = \frac{1}{K} \log \mathsf{E}\left\{ \exp\left[\beta^{-1} K\, G_K^{(u)}(\boldsymbol{\Gamma}, \underline{\boldsymbol{X}})\right]\right\} \tag{88}$$

where $G_K^{(u)}$ is some function of the SNRs and the transmitted symbols and their replicas, collectively denoted by a $K \times (u+1)$ matrix $\underline{\boldsymbol{X}} = [\boldsymbol{X}_0, \ldots, \boldsymbol{X}_u]$. By first conditioning on the correlation matrix $\boldsymbol{Q}$ of $(\boldsymbol{\Gamma}\underline{\boldsymbol{X}})$, the central limit theorem helps to reduce (88) to

$$\frac{1}{K} \log \int \exp\left[\beta^{-1} K\, G^{(u)}(\boldsymbol{Q})\right] \mu_K^{(u)}(\,\mathrm{d}\boldsymbol{Q}) + O\left(\frac{1}{K}\right) \tag{89}$$

where $G^{(u)}$ is some function (independent of K) of the $(u+1) \times (u+1)$ correlation matrix $\boldsymbol{Q}$, and $\mu_K^{(u)}$ is the probability measure of the random matrix $\boldsymbol{Q}$. Large deviations can be invoked to show that there exists a rate function $I^{(u)}$ such that the measure $\mu_K^{(u)}$ satisfies

$$-\lim_{K\to\infty} \frac{1}{K} \log \mu_K^{(u)}(\mathcal{A}) = \inf_{\boldsymbol{Q}\in\mathcal{A}} I^{(u)}(\boldsymbol{Q}) \log e \tag{90}$$

for all measurable set $\mathcal{A}$ of $(u+1) \times (u+1)$ matrices. Using Varadhan's theorem [49], (89) is found to converge as $K \to \infty$ to

$$\sup_{\boldsymbol{Q}} \left[\beta^{-1} G^{(u)}(\boldsymbol{Q}) - I^{(u)}(\boldsymbol{Q})\right] \log e. \tag{91}$$

Seeking the extremum over a $(u+1)^2$-dimensional space is a hard problem. The technique to circumvent this is to assume *replica symmetry*, namely, that the supremum in $\boldsymbol{Q}$ is symmetric over all replicated dimensions. The resulting supremum is then over merely a few parameters, and the free energy can be obtained.

The replica method is also used to calculate the moments (82). Clearly, $\boldsymbol{X}_0$—$(\boldsymbol{Y},\boldsymbol{S})$—$[\boldsymbol{X}_1,\dots,\boldsymbol{X}_u]$ is a Markov chain. The moments (82) are equivalent to some moments under the replicated system:

$$\lim_{K\to\infty} \mathsf{E}\left\{ X_{0k}^i\, X_{mk}^j \prod_{a=1}^{l} X_{ak} \right\} \tag{92}$$

where we choose $m > l$, which can be readily evaluated by working with a modified partition function akin to (86).

Following the replica recipe outlined in the above, a more detailed analysis of the real-valued channel is carried out in Section 4. The complex-valued counterpart is discussed in Section 5. As previously mentioned, while the replica trick and replica symmetry are assumed to be valid as well as the self-averaging property, their rigorous justification is still an open problem in mathematical physics.

4 Proofs Using the Replica Method

This section proves Claims 1–3 using the replica method. The free energy (75) is calculated first so that the spectral efficiency under joint decoding is derived. The joint moments (82) are then found and it is demonstrated that the multiuser channel can be effectively decoupled into single-user Gaussian channels.

For notational convenience, natural logarithms are assumed throughout this section. All results on information measures can be converted to indefinite unit by multiplying the logarithm of e for statement in other sections.

4.1 Free Energy

We will find the free energy by (84) and then the spectral efficiency is trivial by (80). From (9), (10) and (86),

$$\begin{aligned}
\mathsf{E}\left\{Z^u(\boldsymbol{Y},\boldsymbol{S})\right\} &= \mathsf{E}\left\{\int p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{X}_0,\boldsymbol{S}) \prod_{a=1}^{u} q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{X}_a,\boldsymbol{S})\,\mathrm{d}\boldsymbol{y}\right\} && (93)\\
&= \mathsf{E}\left\{\int (2\pi)^{-\frac{L}{2}} (2\pi\sigma^2)^{-\frac{uL}{2}} \exp\left[-\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{X}_0\|^2\right]\right.\\
&\qquad \left.\times \prod_{a=1}^{u} \exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{X}_a\|^2\right]\,\mathrm{d}\boldsymbol{y}\right\}. && (94)
\end{aligned}$$

where the expectations are taken over the channel state matrix $\boldsymbol{S}$, the original symbol vector $\boldsymbol{X}_0$ (i.i.d. entries with distribution p_X), and the replicated symbols $\boldsymbol{X}_a$, $a=1,\dots,u$ (i.i.d. entries with distribution q_X). Note that $\boldsymbol{S}$, $\boldsymbol{X}_0$ and $\boldsymbol{X}_a$ are independent. Let $\underline{\boldsymbol{X}} = [\boldsymbol{X}_0,\dots,\boldsymbol{X}_u]$. We glean from the fact that the L dimensions of the CDMA channel are independent and statistically identical, and write (94) as

$$\mathsf{E}\left\{Z^u(\boldsymbol{Y},\boldsymbol{S})\right\} = \mathsf{E}\Bigg\{\left[\left(2\pi\sigma^2\right)^{-\frac{u}{2}}\int \mathsf{E}\left\{\exp\left[-\frac{1}{2}\left(y-\tilde{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{X}_0\right)^2\right]\right.\right.$$
$$\left.\left.\times\prod_{a=1}^{u}\exp\left[-\frac{1}{2\sigma^2}\left(y-\tilde{\boldsymbol{S}}\boldsymbol{\Gamma}\boldsymbol{X}_a\right)^2\right]\middle|\,\boldsymbol{\Gamma},\underline{\boldsymbol{X}}\right\}\frac{\mathrm{d}y}{\sqrt{2\pi}}\right]^L\Bigg\} \tag{95}$$

where the inner expectation in (95) is taken over $\tilde{\boldsymbol{S}} = [S_1,\dots,S_K]$, a vector of i.i.d. random variables each taking the same distribution as the random spreading chips S_{nk}. Define the following variables:

$$V_a = \frac{1}{\sqrt{K}}\sum_{k=1}^{K}\sqrt{\mathsf{snr}_k}\,S_k X_{ak}, \quad a = 0,1,\dots,u. \tag{96}$$

Clearly, (95) can be rewritten as

$$\mathsf{E}\left\{Z^u(\boldsymbol{Y},\boldsymbol{S})\right\} = \mathsf{E}\left\{\exp\left[L\,G_K^{(u)}\left(\boldsymbol{\Gamma},\underline{\boldsymbol{X}}\right)\right]\right\} \tag{97}$$

where

$$G_K^{(u)}\left(\boldsymbol{\Gamma},\underline{\boldsymbol{X}}\right) = -\frac{u}{2}\log\left(2\pi\sigma^2\right) + \log\int \mathsf{E}\left\{\exp\left[-\frac{1}{2}\left(y-\sqrt{\beta}\,V_0\right)^2\right]\right.$$
$$\left.\times\prod_{a=1}^{u}\exp\left[-\frac{1}{2\sigma^2}\left(y-\sqrt{\beta}\,V_a\right)^2\right]\middle|\,\boldsymbol{\Gamma},\underline{\boldsymbol{X}}\right\}\frac{\mathrm{d}y}{\sqrt{2\pi}}. \tag{98}$$

Note that given $\boldsymbol{\Gamma}$ and $\underline{\boldsymbol{X}}$, each V_a is a sum of K weighted i.i.d. random chips. Due to a generalization of the central limit theorem, $\boldsymbol{V}$ converges to a zero-mean Gaussian random vector with covariance matrix $\boldsymbol{Q}$ where

$$Q_{ab} = \mathsf{E}\left\{V_a V_b \mid \boldsymbol{\Gamma},\underline{\boldsymbol{X}}\right\} = \frac{1}{K}\sum_{k=1}^{K}\mathsf{snr}_k X_{ak}X_{bk}, \quad a,b = 0,\dots,u. \tag{99}$$

Note that although inexplicit in notation, Q_{ab} is a function of $\{\mathsf{snr}_k, X_{ak}, X_{bk}\}_{k=1}^{K}$. The reader is referred to [24, Appendix B] or [50] for a justification of the asymptotic normality of $\boldsymbol{V}$ through the Edgeworth expansion. As a result,

$$\exp\left[G_K^{(u)}(\boldsymbol{\Gamma},\underline{\boldsymbol{X}})\right] = \exp\left[G^{(u)}(\boldsymbol{Q}) + \mathcal{O}(K^{-1})\right] \tag{100}$$

where the integral of the Gaussian density in (98) can be simplified to obtain (refer to [50] for details)

$$G^{(u)}(\boldsymbol{Q}) = -\frac{1}{2}\log\det(\boldsymbol{I}+\boldsymbol{\Sigma}\boldsymbol{Q}) - \frac{1}{2}\log\left(1+\frac{u}{\sigma^2}\right) - \frac{u}{2}\log\left(2\pi\sigma^2\right) \tag{101}$$

where $\boldsymbol{\Sigma}$ is a $(u+1)\times(u+1)$ matrix:[4]

$$\boldsymbol{\Sigma} = \frac{\beta}{\sigma^2+u}\left[\begin{array}{c|c} u & -\boldsymbol{e}^\top \\ \hline -\boldsymbol{e} & \left(1+\frac{u}{\sigma^2}\right)\boldsymbol{I} - \frac{1}{\sigma^2}\boldsymbol{e}\boldsymbol{e}^\top \end{array}\right] \tag{102}$$

[4]For convenience, the index number of all $(u+1)\times(u+1)$ matrices in this paper starts from 0.

where $\boldsymbol{e}$ is a $u \times 1$ column vector whose entries are all 1. It is clear that $\boldsymbol{\Sigma}$ is invariant if two nonzero indexes are interchanged, i.e., $\boldsymbol{\Sigma}$ is symmetric in the replicas.

By (97) and (100),

$$\begin{aligned} \frac{1}{K} \log \mathsf{E}\left\{ Z^u(\boldsymbol{Y}, \boldsymbol{S}) \right\} &= \frac{1}{K} \log \mathsf{E}\left\{ \exp\left[L \left(G^{(u)}(\boldsymbol{Q}) + \mathcal{O}\left(K^{-1}\right)\right)\right]\right\} && (103) \\ &= \frac{1}{K} \log \int \exp\left[K\, \beta^{-1}\, G^{(u)}(\boldsymbol{Q})\right]\, \mathrm{d}\mu_K^{(u)}(\boldsymbol{Q}) + \mathcal{O}(K^{-1}) && (104) \end{aligned}$$

where the expectation over the replicated symbols is rewritten as an integral over the probability measure of the correlation matrix $\boldsymbol{Q}$, which is expressed as

$$\mu_K^{(u)}(\boldsymbol{Q}) = \mathsf{E}\left\{ \prod_{0 \le a \le b}^{u} \delta\left(\frac{1}{K} \sum_{k=1}^{K} \mathsf{snr}_k X_{ak} X_{bk} - Q_{ab} \right)\right\} \tag{105}$$

where $\delta(\cdot)$ is the Dirac function. Note that the limit in K and the expectation can be exchanged from (103) to (104) by dominated convergence theorem since $\exp\left[G^{(u)}(\boldsymbol{Q})\right]$ is bounded by a function in u independent of $\boldsymbol{Q}$. By Cramér's theorem, the probability measure of the empirical means Q_{ab} defined by (99) satisfies, as $K \to \infty$, the large deviations property with some rate function $I^{(u)}(\boldsymbol{Q})$ [49]. Note the factor K in the exponent in the integral in (104). As $K \to \infty$, the integral is dominated by the maximum of the overall effect of the exponent and the rate of the measure on which the integral takes place. Precisely, by Varadhan's theorem [49],

$$\lim_{K\to\infty} \frac{1}{K} \log \mathsf{E}\left\{ Z^u(\boldsymbol{Y}, \boldsymbol{S}) \right\} = \sup_{\boldsymbol{Q}} \left[\beta^{-1}\, G^{(u)}(\boldsymbol{Q}) - I^{(u)}(\boldsymbol{Q}) \right] \tag{106}$$

where the supremum is over all possible $\boldsymbol{Q}$ that can be obtained from varying X_{ak} in (99).

Let the moment generating function be defined as

$$M^{(u)}(\tilde{\boldsymbol{Q}}) = \mathsf{E}\left\{ \exp\left[\mathsf{snr} \boldsymbol{X}^\top \tilde{\boldsymbol{Q}} \boldsymbol{X} \right]\right\} \tag{107}$$

where $\tilde{\boldsymbol{Q}}$ is a $(u+1) \times (u+1)$ symmetric matrix, $\boldsymbol{X} = [X_0, X_1, \dots, X_u]^\top$, and the expectation in (107) is taken over independent random variables $\mathsf{snr} \sim P_{\mathsf{snr}}$, $X_0 \sim p_X$ and $X_1, \dots, X_u \sim q_X$. The rate of the measure $\mu_K^{(u)}$ is given by the Legendre-Fenchel transform of the cumulant generating function (logarithm of the moment generating function) [49]:

$$I^{(u)}(\boldsymbol{Q}) = \sup_{\tilde{\boldsymbol{Q}}} \left[\mathsf{tr}\left\{ \tilde{\boldsymbol{Q}} \boldsymbol{Q} \right\} - \log M^{(u)}(\tilde{\boldsymbol{Q}}) \right] \tag{108}$$

where the supremum is taken with respect to the symmetric matrix $\tilde{\boldsymbol{Q}}$.

By (106), (108) and (101), one has

$$\begin{aligned} \lim_{K\to\infty} \frac{1}{K} \log \mathsf{E}\left\{ Z^u(\boldsymbol{Y}, \boldsymbol{S}) \right\} &= \sup_{\boldsymbol{Q}} \left\{ \beta^{-1}\, G^{(u)}(\boldsymbol{Q}) - \sup_{\tilde{\boldsymbol{Q}}} \left[\mathsf{tr}\left\{ \tilde{\boldsymbol{Q}} \boldsymbol{Q} \right\} - \log M^{(u)}(\tilde{\boldsymbol{Q}}) \right] \right\} && (109) \\ &= \sup_{\boldsymbol{Q}} \inf_{\tilde{\boldsymbol{Q}}} T^{(u)}(\boldsymbol{Q}, \tilde{\boldsymbol{Q}}) && (110) \end{aligned}$$

where

$$\begin{aligned} T^{(u)}(\boldsymbol{Q}, \tilde{\boldsymbol{Q}}) = &-\frac{1}{2\beta} \log\det(\boldsymbol{I} + \boldsymbol{\Sigma}\boldsymbol{Q}) - \mathsf{tr}\left\{ \tilde{\boldsymbol{Q}} \boldsymbol{Q} \right\} + \log \mathsf{E}\left\{ \exp\left[\mathsf{snr} \boldsymbol{X}^\top \tilde{\boldsymbol{Q}} \boldsymbol{X} \right]\right\} \\ &- \frac{1}{2\beta} \log\left(1 + \frac{u}{\sigma^2}\right) - \frac{u}{2\beta} \log\left(2\pi\sigma^2\right). \end{aligned} \tag{111}$$

For an arbitrary $\boldsymbol{Q}$, we first seek the point of zero gradient with respect to $\tilde{\boldsymbol{Q}}$ and find that for any given $\boldsymbol{Q}$, the extremum in $\tilde{\boldsymbol{Q}}$ satisfies

$$\boldsymbol{Q} = \frac{\mathsf{E}\left\{\mathsf{snr}\boldsymbol{X}\boldsymbol{X}^\top \exp\left[\mathsf{snr}\boldsymbol{X}^\top\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\right\}}{\mathsf{E}\left\{\exp\left[\mathsf{snr}\boldsymbol{X}^\top\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\right\}}. \tag{112}$$

Let $\tilde{\boldsymbol{Q}}^*(\boldsymbol{Q})$ be a sufficiently smooth function that solves (112). We then seek the point of zero gradient of $T^{(u)}\left(\boldsymbol{Q}, \tilde{\boldsymbol{Q}}^*(\boldsymbol{Q})\right)$ with respect to $\boldsymbol{Q}$.[5] By virtue of the relationship (112), one finds that the derivative of $\tilde{\boldsymbol{Q}}^*$ with respect to $\boldsymbol{Q}$ is multiplied by 0 and hence inconsequential. Therefore, the extremum in $\boldsymbol{Q}$ satisfies

$$\tilde{\boldsymbol{Q}} = -\frac{1}{\beta}\left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{Q}\right)^{-1}. \tag{113}$$

It is interesting to note from the resulting joint equations (112)–(113) that the order in which the supremum and infimum are taken in (110) can be exchanged (i.e., minimax is equal to max-min). The solution $\left(\boldsymbol{Q}^*, \tilde{\boldsymbol{Q}}^*\right)$ is in fact a saddle point of $T^{(u)}$. Notice that (112) can also be expressed as

$$\boldsymbol{Q} = \mathsf{E}\left\{\mathsf{snr}\boldsymbol{X}\boldsymbol{X}^\top \,\middle|\, \tilde{\boldsymbol{Q}}\right\} \tag{114}$$

where the expectation is over an appropriately defined Gaussian measure $p_{\boldsymbol{X},\mathsf{snr}|\tilde{\boldsymbol{Q}}}$ dependent on $\tilde{\boldsymbol{Q}}$.

Solving joint equations (112) and (113) directly is prohibitive except in the simplest cases such as q_X being Gaussian. In the general case, because of symmetry in the matrix $\boldsymbol{\Sigma}$ (102), we postulate that the solution to the joint equations satisfies *replica symmetry*, namely, both $\boldsymbol{Q}^*$ and $\tilde{\boldsymbol{Q}}^*$ are invariant if two (nonzero) replica indexes are interchanged. In other words, the extremum can be written as

$$\boldsymbol{Q}^* = \left[\begin{array}{c:c} r & m\,\boldsymbol{e}^\top \\ \hdashline m\,\boldsymbol{e} & (p-q)\boldsymbol{I} + q\,\boldsymbol{e}\boldsymbol{e}^\top \end{array}\right], \qquad \tilde{\boldsymbol{Q}}^* = \left[\begin{array}{c:c} c & d\,\boldsymbol{e}^\top \\ \hdashline d\,\boldsymbol{e} & (g-f)\boldsymbol{I} + f\,\boldsymbol{e}\boldsymbol{e}^\top \end{array}\right] \tag{115}$$

where r, m, p, q, c, d, f, g are some real numbers.

Note that replica symmetry is a reasonable assumption here rather than a proved fact. It can be shown that replica symmetry holds in certain cases, e.g., if the postulated prior q_X is Gaussian. Under equal-power binary input and individually optimal detection, Tanaka showed also that if the system parameters satisfy certain condition, the replica-symmetric solution is stable against replica-symmetry-breaking (RSB), i.e., it is at least a local maximum [24]. In some other cases, replica symmetry can be broken [51]. Unfortunately, there is no known general condition for replica symmetry to hold. The replica-symmetric solution, assumed for analytical tractability in this paper, is consistent with numerical results in the experiments shown in Section 6.

Under replica symmetry, (101) is evaluated to obtain

$$\begin{aligned} G^{(u)}\left(\boldsymbol{Q}^*\right) = &-\frac{u}{2}\log\left(2\pi\sigma^2\right) - \frac{u-1}{2}\log\left[1 + \frac{\beta}{\sigma^2}(p-q)\right] \\ &-\frac{1}{2}\log\left[1 + \frac{\beta}{\sigma^2}(p-q) + \frac{u}{\sigma^2}(1 + \beta(r - 2m + q))\right]. \end{aligned} \tag{116}$$

[5]The following identities are useful:

$$\frac{\partial \log\det \boldsymbol{Q}}{\partial x} = \operatorname{tr}\left\{\boldsymbol{Q}^{-1}\frac{\partial \boldsymbol{Q}}{\partial x}\right\}, \quad \frac{\partial \boldsymbol{Q}^{-1}}{\partial x} = -\boldsymbol{Q}^{-1}\frac{\partial \boldsymbol{Q}}{\partial x}\boldsymbol{Q}^{-1}.$$

The moment generating function (107) is evaluated as

$$
\begin{aligned}
& M^{(u)}(\tilde{\boldsymbol{Q}}^*) \\
= \; & \mathsf{E}\left\{\exp\left[\mathsf{snr}\left(2d\sum_{a=1}^{u} X_0 X_a + 2f\sum_{0<a<b}^{u} X_a X_b + cX_0^2 + g\sum_{a=1}^{u} X_a^2\right)\right]\right\} && (117) \\
= \; & \mathsf{E}\left\{\exp\left[\mathsf{snr}\left(\frac{d}{\sqrt{f}}X_0 + \sqrt{f}\sum_{a=1}^{u} X_a\right)^2 + \left(c - \frac{d^2}{f}\right)\mathsf{snr}X_0^2 + (g-f)\mathsf{snr}\sum_{a=1}^{u} X_a^2\right]\right\}, && (118)
\end{aligned}
$$

where $X_0 \sim p_X$ while $X_a \sim q_X$ are all independent. The expectation (118) with respect to the symbols can be decoupled using a property of Gaussian density, which is also a variant of the so-called Hubbard-Stratonovich transform [52]:

$$
e^{x^2} = \sqrt{\frac{\eta}{2\pi}} \int \exp\left[-\frac{\eta}{2}z^2 + \sqrt{2\eta}\, xz\right] \mathrm{d}z, \quad \forall x, \eta. \tag{119}
$$

Using (119) with $\eta = 2d^2/f$, (118) becomes

$$
\begin{aligned}
M^{(u)}(\tilde{\boldsymbol{Q}}^*) = & \mathsf{E}\left\{\sqrt{\frac{d^2}{f\pi}} \int \exp\left[-\frac{d^2}{f}z^2 + 2\sqrt{\mathsf{snr}}\left(\frac{d^2}{f}X_0 + d\sum_{a=1}^{u} X_a\right) z \right.\right. \\
& \left.\left. + \left(c - \frac{d^2}{f}\right)\mathsf{snr}X_0^2 + (g-f)\mathsf{snr}\sum_{a=1}^{u} X_a^2\right] \mathrm{d}z\right\}.
\end{aligned} \tag{120}
$$

Since $X_0, \dots, X_u$ and snr are independent, the rate of the measure (108) under replica symmetry is obtained from (120) as

$$
\begin{aligned}
I^{(u)}\left(\boldsymbol{Q}^*\right) = & rc + upg + 2umd + u(u-1)qf \\
& - \log \mathsf{E}\Bigg\{\int \sqrt{\frac{d^2}{f\pi}}\, \mathsf{E}\left\{\exp\left[-\frac{d^2}{f}\left(z - \sqrt{\mathsf{snr}}X_0\right)^2 + c\,\mathsf{snr}X_0^2\right] \,\middle|\, \mathsf{snr}\right\} \\
& \times \left[\mathsf{E}_q\left\{\exp\left[2d\sqrt{\mathsf{snr}}Xz + (g-f)\mathsf{snr}X^2\right] \,\middle|\, \mathsf{snr}\right\}\right]^u \mathrm{d}z\Bigg\}.
\end{aligned} \tag{121}
$$

The free energy is then found by (84) and (106):

$$
\mathcal{F} = -\lim_{u\to 0} \frac{\partial}{\partial u}\left[\beta^{-1}G^{(u)}\left(\boldsymbol{Q}^*\right) - I^{(u)}\left(\boldsymbol{Q}^*\right)\right], \tag{122}
$$

where $\boldsymbol{Q}^*$ is the replica-symmetric solution to (112)–(113).

The eight parameters (r, m, p, q, c, d, f, g) that define $\boldsymbol{Q}^*$ and $\tilde{\boldsymbol{Q}}^*$ are the solution to the joint equations (112)–(113) under replica symmetry. It is interesting to note that as functions of u, the derivative of each of the eight parameters with respect to u vanishes as $u \to 0$. Thus for the purpose of the free energy (122), it suffices to find the extremum of $\left[\beta^{-1}G^{(u)} - I^{(u)}\right]$ at $u = 0$. Using (113), it can be shown that at $u = 0$,

$$
\begin{aligned}
c &= 0, && (123a) \\
d &= \frac{1}{2[\sigma^2 + \beta(p-q)]}, && (123b) \\
f &= \frac{1 + \beta(r - 2m + q)}{2[\sigma^2 + \beta(p-q)]^2}, && (123c) \\
g &= f - d. && (123d)
\end{aligned}
$$

The parameters r, m, p, q can be determined from (114) by studying the measure $p_{\boldsymbol{X},\mathsf{snr}|\tilde{\boldsymbol{Q}}}$ under replica symmetry and $u \to 0$. For that purpose, define two useful parameters:

$$\eta = \frac{2d^2}{f} \quad \text{and} \quad \xi = 2d. \tag{124}$$

Noticing that $c = 0$, $g - f = -d$, (120) can be written as

$$\begin{aligned} M^{(u)}(\tilde{\boldsymbol{Q}}^*) =& \mathsf{E}\left\{ \sqrt{\frac{\eta}{2\pi}} \int \exp\left[-\frac{\eta}{2} \left(z - \sqrt{\mathsf{snr}} X_0 \right)^2 \right] \right. \\ & \left. \times \left[\mathsf{E}_q \left\{ \exp\left[-\frac{\xi}{2} z^2 - \frac{\xi}{2} \left(z - \sqrt{\mathsf{snr}} X \right)^2 \right] \,\middle|\, \mathsf{snr} \right\} \right]^u \mathrm{d}z \right\}. \end{aligned} \tag{125}$$

It is clear that the limit of (125) as $u \to 0$ is 1. Hence by (112), as $u \to 0$,

$$Q^*_{ab} = \mathsf{E}\left\{ \mathsf{snr} X_a X_b \mid \tilde{\boldsymbol{Q}}^* \right\} \to \mathsf{E}\left\{ \mathsf{snr} X_a X_b \exp\left[\boldsymbol{X}^\top \tilde{\boldsymbol{Q}}^* \boldsymbol{X} \right] \right\}. \tag{126}$$

We now give a useful representation for the parameters r, m, p, q defined in (115). Note that as $u \to 0$,

$$\begin{aligned} \mathsf{E}\left\{ \mathsf{snr} X_0 X_1 \exp\left[\boldsymbol{X}^\top \tilde{\boldsymbol{Q}}^* \boldsymbol{X} \right] \right\} =& \mathsf{E}\left\{ \mathsf{snr} X_0 \int \sqrt{\frac{\eta}{2\pi}} \exp\left[-\frac{\eta}{2} \left(z - \sqrt{\mathsf{snr}} X_0 \right)^2 \right] \right. \\ & \left. \times \frac{X_1 \sqrt{\frac{\xi}{2\pi}} \exp\left[-\frac{\xi}{2} \left(z - \sqrt{\mathsf{snr}} X_1 \right)^2 \right]}{\mathsf{E}_q\left\{ \sqrt{\frac{\xi}{2\pi}} \exp\left[-\frac{\xi}{2} \left(z - \sqrt{\mathsf{snr}} X_1 \right)^2 \right] \,\middle|\, \mathsf{snr} \right\}} \mathrm{d}z \right\}. \end{aligned} \tag{127}$$

Let two single-user Gaussian channels be defined as in Section 2.4, i.e., $p_{Z|X,\mathsf{snr};\eta}$ is given by (17) and $q_{Z|X,\mathsf{snr};\xi}$ by (18). Assuming that the input distribution to the channel $q_{Z|X,\mathsf{snr};\xi}$ is q_X, a posterior probability distribution $q_{X|Z,\mathsf{snr};\xi}$ is induced, which defines a retrochannel. Let X_0 be the input to the channel $p_{Z|X,\mathsf{snr};\eta}$ and X be the output of the retrochannel $q_{X|Z,\mathsf{snr};\xi}$. The posterior mean with respect to the measure q, denoted by $\langle X \rangle_q$ is given by (19). The Gaussian channel $p_{Z|X,\mathsf{snr};\eta}$, the retrochannel $q_{X|Z,\mathsf{snr};\xi}$ and the PME, all in the single-user setting, are depicted in Figure 5(b). Then, (127) can be understood as an expectation over X_0, X and Z to obtain

$$\begin{aligned} Q^*_{01} &= \mathsf{E}\left\{ \mathsf{snr} X_0 X_1 \exp\left[\boldsymbol{X}^\top \tilde{\boldsymbol{Q}}^* \boldsymbol{X} \right] \right\} && (128) \\ &= \mathsf{E}\left\{ \mathsf{snr} X_0 \int p_{Z|X,\mathsf{snr};\eta}(z|X_0, \mathsf{snr}; \eta)\, \mathsf{E}_q\left\{ X \mid Z = z, \mathsf{snr}; \xi \right\} \mathrm{d}z \right\} && (129) \\ &= \mathsf{E}\left\{ \mathsf{snr}\, X_0 \langle X \rangle_q \right\}. && (130) \end{aligned}$$

Similarly, (126) can be evaluated for all indexes (a, b) yielding together with (115):

$$\begin{aligned} r &= Q^*_{00} = \mathsf{E}\left\{ \mathsf{snr}\, X_0^2 \right\} = \mathsf{E}\left\{ \mathsf{snr} \right\}, && (131a) \\ m &= Q^*_{01} = \mathsf{E}\left\{ \mathsf{snr}\, X_0 \langle X \rangle_q \right\}, && (131b) \\ p &= Q^*_{11} = \mathsf{E}\left\{ \mathsf{snr}\, X^2 \right\}, && (131c) \\ q &= Q^*_{12} = \mathsf{E}\left\{ \mathsf{snr}\, (\langle X \rangle_q)^2 \right\}. && (131d) \end{aligned}$$

In summary, under replica symmetry, the parameters c, d, f, g are given by (123) as functions of r, m, p, q, which are in turn determined by the statistics of the two channels (17) and (18) parameterized by $\eta = 2d^2/f$ and $\xi = 2d$ respectively. It is not difficult to see that

$$
\begin{aligned}
r - 2m + q &= \mathsf{E}\left\{ \mathsf{snr} \left(X_0 - \langle X \rangle_q \right)^2 \right\}, && \text{(132a)} \\
p - q &= \mathsf{E}\left\{ \mathsf{snr} \left(X - \langle X \rangle_q \right)^2 \right\}. && \text{(132b)}
\end{aligned}
$$

Using (123) and (124), it can be checked that

$$
\begin{aligned}
p - q &= \frac{1}{\beta}\left(\frac{1}{\xi} - \sigma^2 \right), && \text{(133a)} \\
r - 2m + q &= \frac{1}{\beta}\left(\frac{1}{\eta} - 1 \right). && \text{(133b)}
\end{aligned}
$$

Thus $G^{(u)}$ and $I^{(u)}$ given by (116) and (121) can be expressed in η and ξ. Using (122) and (133), the free energy is found as (23), where (η, ξ) satisfies fixed-point equations

$$
\begin{aligned}
\eta^{-1} &= 1 + \beta\, \mathsf{E}\left\{ \mathsf{snr} \left(X_0 - \langle X \rangle_q \right)^2 \right\}, && \text{(134a)} \\
\xi^{-1} &= \sigma^2 + \beta\, \mathsf{E}\left\{ \mathsf{snr} \left(X - \langle X \rangle_q \right)^2 \right\}. && \text{(134b)}
\end{aligned}
$$

In case of multiple solutions to (134), (η, ξ) is chosen as the solution that gives the minimum free energy $\mathcal{F}$. By defining $\mathcal{E}(\mathsf{snr}; \eta, \xi)$ and $\mathcal{V}(\mathsf{snr}; \eta, \xi)$ as in (20) and (21), the coupled equations (123) and (131) can be summarized to establish the key fixed-point equations (22). It will be shown in Section 4.2 that, from a single user's stance, the multiuser PME and the multiuser retrochannel, parameterized by arbitrary (q_X, σ), have an equivalence as a single-user PME and a single-user retrochannel.

Here, for the purpose of the total spectral efficiency, we set the postulated measure q to be identical to the actual measure p (i.e., $q_X = p_X$ and $\sigma = 1$). The inverse noise variances (η, ξ) satisfy joint equations but we choose the replica-symmetric solution $\eta = \xi$ as argued in Section 2.4.2. Using (80), the total spectral efficiency is

$$
\mathsf{C}_{\text{joint}} = -\beta\, \mathsf{E}\left\{ \int p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta) \log p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta)\, \mathrm{d}\, z \right\} - \frac{\beta}{2} \log \frac{2\pi e}{\eta} + \frac{1}{2}(\eta - 1 - \log \eta), \quad (135)
$$

where η satisfies

$$
\eta + \eta\, \beta\, \mathsf{E}\left\{ \mathsf{snr} \left[1 - \int \frac{[p_1(z, \mathsf{snr}; \eta)]^2}{p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta)} \mathrm{d}\, z \right] \right\} = 1. \quad (136)
$$

The optimal spectral efficiency of the multiuser channel is thus found.

4.2 Joint Moments

Consider again the Gaussian channel, the PME and the retrochannel in the multiuser setting depicted in Figure 5(a). The joint moments (82) are of interest here. For simplicity, we first study joint moments of the input symbol and the retrochannel output, which can be obtained as expectations under the replicated system:

$$
\mathsf{E}\left\{ X_{0k}^i X_k^j \right\} = \mathsf{E}\left\{ X_{0k}^i X_{mk}^j \right\}, \quad m = 1, \dots, u. \quad (137)
$$

It is then straightforward to calculate (82) by following the same procedure.

The following lemma allows us to determine the expected value of a function of the symbols and their replicas by considering a modified partition function akin to (86).

Lemma 1 *Given an arbitrary function* $f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)$*, define*

$$Z^{(u)}(\boldsymbol{y}, \boldsymbol{S}, \boldsymbol{x}_0; h) = \mathsf{E}_q \left\{ \exp\left[h\, f(\boldsymbol{x}_0, \underline{\boldsymbol{X}}_a)\right] \prod_{a=1}^{u} q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{X}_a, \boldsymbol{S}) \,\middle|\, \boldsymbol{S} \right\}. \tag{138}$$

If $\mathsf{E}\{ f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a) \mid \boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0\}$ *is not dependent on* u*, then*

$$\mathsf{E}\{f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)\} = \lim_{u\to 0} \frac{\partial}{\partial h} \log \mathsf{E}\left\{ Z^{(u)}(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0; h)\right\}\Big|_{h=0}. \tag{139}$$

Proof: It is easy to see that

$$Z^{(u)}(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0; h)\Big|_{h=0} = Z^u(\boldsymbol{Y}, \boldsymbol{S}). \tag{140}$$

By taking the derivative and letting $h = 0$, the right hand side of (139) is

$$\frac{1}{K} \lim_{u\to 0} \mathsf{E}\left\{ \mathsf{E}_q \left\{ f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}'_a) \prod_{a=1}^{u} q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{Y}|\boldsymbol{X}'_a, \boldsymbol{S}) \,\middle|\, \boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0 \right\}\right\}, \tag{141}$$

where $\underline{\boldsymbol{X}}'_a$ has the same statistics as $\underline{\boldsymbol{X}}_a$ (i.e., contains i.i.d. entries with distribution q_X) but independent of $(\boldsymbol{X}_0, \boldsymbol{Y}, \boldsymbol{S})$. Also note that

$$q_{\underline{\boldsymbol{X}}_a|\boldsymbol{Y},\boldsymbol{S}}(\underline{\boldsymbol{X}}_a \mid \boldsymbol{Y}, \boldsymbol{S}) = Z^{-u}(\boldsymbol{Y}, \boldsymbol{S})\, q_{\underline{\boldsymbol{X}}_a}(\underline{\boldsymbol{X}}_a) \prod_{a=1}^{u} q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{Y} \mid \boldsymbol{X}_a, \boldsymbol{S}). \tag{142}$$

One can change the expectation over the replicas $\underline{\boldsymbol{X}}'_a$ independent of $(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0)$ to an expectation over $\underline{\boldsymbol{X}}_a$ conditioned on $(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0)$. Hence (141) can be further written as

$$\begin{aligned} \frac{1}{K} \lim_{u\to 0} \mathsf{E}\left\{\mathsf{E}\left\{f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)|\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0\right\}\, Z^u(\boldsymbol{Y}, \boldsymbol{S})\right\} &= \frac{1}{K} \mathsf{E}\left\{\mathsf{E}\left\{f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)|\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{X}_0\right\}\right\} && (143)\\ &= \frac{1}{K} \mathsf{E}\left\{f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)\right\} && (144) \end{aligned}$$

where $Z^u(\boldsymbol{Y}, \boldsymbol{S})$ can be dropped as $u \to 0$ in (143) since the conditional expectation is not dependent on u by the assumption in the lemma. ∎

For the function $f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)$ to have influence on the free energy, it must grow at least linearly with K. Assume that $f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)$ involves users 1 through $K_1 = \alpha_1 K$ where $0 < \alpha_1 < 1$ is fixed as $K \to \infty$:

$$f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a) = \sum_{k=1}^{K_1} X_{0k}^{i} X_{mk}^{j} \tag{145}$$

where m is an arbitrary replica number in $\{1, \ldots, u\}$. Without loss of generality, we calculate (137) for a user $\kappa \in \{1, \ldots, K_1\}$. It is also assumed that user 1 through K_1 take the same signal-to-noise ratio snr. We will finally take the limit $\alpha_1 \to 0$ so that the equal-power constraint for the first K_1 users becomes superfluous.

Clearly, the moments (137) for user κ can be rewritten as

$$\mathsf{E}\left\{X_{0\kappa}^{i}X_{m\kappa}^{j}\right\} = \frac{1}{K_1}\sum_{k=1}^{K_1}\mathsf{E}\left\{X_{0k}^{i}\,X_{mk}^{j}\right\} \tag{146}$$

$$= \frac{1}{K_1}\,\mathsf{E}\left\{f(\boldsymbol{X}_0,\underline{\boldsymbol{X}}_a)\right\}. \tag{147}$$

Note that

$$\mathsf{E}\left\{f(\boldsymbol{X}_0,\underline{\boldsymbol{X}}_a)\mid \boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0\right\} = \mathsf{E}\left\{\sum_{k=1}^{K_1}X_{0k}^{i}\,X_{k}^{j}\,\middle|\,\boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0\right\} \tag{148}$$

is not dependent on u. By Lemma 1, the moments (147) can be obtained as

$$\lim_{u\to 0}\frac{\partial}{\partial h}\frac{1}{\alpha_1 K}\log\mathsf{E}\left\{Z^{(u)}(\boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0;h)\right\}\Big|_{h=0} \tag{149}$$

where

$$Z^{(u)}(\boldsymbol{y},\boldsymbol{S},\boldsymbol{x}_0;h) = \left(2\pi\sigma^2\right)^{-\frac{uL}{2}}\mathsf{E}_q\left\{\exp\left[h\sum_{k=1}^{K_1}x_{0k}^{j}X_{mk}^{i}\right]\prod_{a=1}^{u}\exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{S}\boldsymbol{X}_a\|^2\right]\,\middle|\,\boldsymbol{S}\right\}. \tag{150}$$

Regarding (150) as a partition function for some random system allows the same techniques in Section 4.1 to be used to write

$$\lim_{K\to\infty}\frac{1}{K}\log\mathsf{E}\left\{Z^{(u)}(\boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0;h)\right\} = \sup_{\boldsymbol{Q}}\left\{\beta^{-1}G^{(u)}(\boldsymbol{Q}) - I^{(u)}(\boldsymbol{Q};h)\right\} \tag{151}$$

where $G^{(u)}(\boldsymbol{Q})$ is given by (101) and $I^{(u)}(\boldsymbol{Q};h)$ is the rate of the following measure

$$\mu_K^{(u)}(\boldsymbol{Q};h) = \mathsf{E}\left\{\exp\left[h\sum_{k=1}^{K_1}X_{0k}^{i}X_{mk}^{j}\right]\prod_{0\le a\le b}^{u}\delta\left(\sum_{k=1}^{K}\mathsf{snr}_k X_{ak}X_{bk} - KQ_{ab}\right)\right\}. \tag{152}$$

By the large deviations property, one finds the rate

$$I^{(u)}(\boldsymbol{Q};h) = \sup_{\tilde{\boldsymbol{Q}}}\left[\mathsf{tr}\left\{\tilde{\boldsymbol{Q}}\boldsymbol{Q}\right\} - \log M^{(u)}(\tilde{\boldsymbol{Q}}) - \alpha_1\left(\log M^{(u)}(\tilde{\boldsymbol{Q}},\mathsf{snr};h) - \log M^{(u)}(\tilde{\boldsymbol{Q}},\mathsf{snr};0)\right)\right] \tag{153}$$

where $M^{(u)}(\tilde{\boldsymbol{Q}})$ is defined in (107), and

$$M^{(u)}(\tilde{\boldsymbol{Q}},\mathsf{snr};h) = \mathsf{E}\left\{\exp\left[h\,X_0^{i}X_m^{j}\right]\exp\left[\mathsf{snr}\boldsymbol{X}^{\top}\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\,\middle|\,\mathsf{snr}\right\}. \tag{154}$$

¿From (151) and (153), taking the derivative in (149) with respect to h at $h=0$ leaves only one term

$$\frac{\partial}{\partial h}\log M^{(u)}(\tilde{\boldsymbol{Q}},\mathsf{snr};h)\Big|_{h=0} = \frac{\mathsf{E}\left\{X_0^{i}X_m^{j}\exp\left[\mathsf{snr}\boldsymbol{X}^{\top}\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\right\}}{\mathsf{E}\left\{\exp\left[\mathsf{snr}\boldsymbol{X}^{\top}\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\right\}}. \tag{155}$$

Since

$$Z^{(u)}(\boldsymbol{Y},\boldsymbol{S},\boldsymbol{X}_0;h)\Big|_{h=0} = Z^{u}(\boldsymbol{Y},\boldsymbol{S}), \tag{156}$$

the $\tilde{\boldsymbol{Q}}$ in (155) that give the supremum in (153) at $h\to 0$ is exactly the $\tilde{\boldsymbol{Q}}$ that gives the supremum of (108), which is replica-symmetric by assumption. By introducing the parameters (η,ξ) the same as in Section 4.1, and by definition of q_i and p_i in (25) and (28) respectively, (155) can be further evaluated as

$$\frac{\int \left(\sqrt{\frac{2\pi}{\xi}} e^{\frac{\xi z^2}{2}}\right)^u p_i(z,\mathsf{snr};\eta)\, q_0^{u-1}(z,\mathsf{snr};\xi)\, q_j(z,\mathsf{snr};\xi)\,\mathrm{d}z}{\int \left(\sqrt{\frac{2\pi}{\xi}} e^{\frac{\xi z^2}{2}}\right)^u p_0(z,\mathsf{snr};\eta) q_0^u(z,\mathsf{snr};\xi)\,\mathrm{d}z} \tag{157}$$

Taking the limit $u \to 0$, one has from (147)–(157) that as $K \to \infty$,

$$\frac{1}{K_1}\sum_{k=1}^{K_1} \mathsf{E}\left\{X_{0k}^i\, X_{mk}^j\right\} \to \int p_i(z,\mathsf{snr};\eta)\, \frac{q_j(z,\mathsf{snr};\xi)}{q_0(z,\mathsf{snr};\xi)}\,\mathrm{d}z. \tag{158}$$

Let $X_0 \sim p_X$ be the input to the single-user Gaussian channel $p_{Z|X,\mathsf{snr};\eta}$ and Z be its output (see Figure 5(b)). Let X be the corresponding output of the companion retrochannel with Z as its input. Then X_0–Z–X is a Markov chain. By definition of p_i and q_i, the right hand side of (158) is

$$\int p_0(z,\mathsf{snr};\eta)\, \frac{p_i(z,\mathsf{snr};\xi)}{p_0(z,\mathsf{snr};\xi)}\, \frac{q_j(z,\mathsf{snr};\xi)}{q_0(z,\mathsf{snr};\xi)}\,\mathrm{d}z = \mathsf{E}\left\{\mathsf{E}\left\{ X_0^i \mid Z\right\} \mathsf{E}\left\{ X^j \mid Z\right\}\right\}. \tag{159}$$

Letting $K_1 \to 1$ (thus $\alpha_1 \to 0$) so that the requirement that the first K_1 users take the same SNR becomes unnecessary, we have proved by (137), (146), (158) and (159) that for every SNR distribution and every every user $k \in \{1,\dots,K\}$

$$\mathsf{E}\left\{X_{0k}^i\, X_k^j\right\} \to \mathsf{E}\left\{X_0^i X^j\right\} \quad \text{as } K \to \infty. \tag{160}$$

Since the moments (160) are uniformly bounded, the distribution is thus uniquely determined by the moments due to Carleman's Theorem [53, p. 227]. Therefore, for every user k, the joint distribution of the input X_{0k} to the multiuser channel and the output X_k of the multiuser retrochannel converges to the joint distribution of the input X_0 to the single-user Gaussian channel $p_{Z|X,\mathsf{snr};\eta}$ and the output X of the single-user retrochannel $q_{X|Z,\mathsf{snr};\xi}$.

Applying the same methodology as developed thus far in this subsection, one can also calculate the joint moments (82) by letting

$$f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a) = \sum_{k=1}^{K_1} X_{0k}^i X_{mk}^j \prod_{a=1}^{l} X_{ak} \tag{161}$$

where it is assumed that $m > l$. The rationale is that $\boldsymbol{X}_0$–$(\boldsymbol{Y},\boldsymbol{S})$–$\boldsymbol{X}_a$ is a Markov chain and X_a's are i.i.d. conditioned on $(\boldsymbol{Y},\boldsymbol{S})$; hence (82) can be calculated as expectations under the replicated system:

$$\begin{aligned}\mathsf{E}\left\{X_{0k}^i\, X_k^j\, \langle X_k\rangle_q^l\right\} &= \mathsf{E}\left\{X_{0k}^i\, X_{mk}^j \prod_{a=1}^{l} \mathsf{E}\left\{X_{ak} \mid \boldsymbol{Y},\boldsymbol{S}\right\}\right\} &(162)\\ &= \mathsf{E}\left\{f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)\right\}. &(163)\end{aligned}$$

It is straightforward by Lemma 1 to calculate (163) and obtain that, as $K \to \infty$,

$$\mathsf{E}\left\{f(\boldsymbol{X}_0, \underline{\boldsymbol{X}}_a)\right\} \to \int p_i(z,\mathsf{snr};\eta)\, \frac{q_j(z,\mathsf{snr};\xi)}{q_0(z,\mathsf{snr};\xi)} \left(\frac{q_1(z,\mathsf{snr};\xi)}{q_0(z,\mathsf{snr};\xi)}\right)^l \mathrm{d}z. \tag{164}$$

Let $\langle X\rangle_q$ be the single-user PME output as seen in Figure 5(b), which is a function of the Gaussian channel output Z. Then the right hand side of (164) represents a joint moment and thus

$$\mathsf{E}\left\{X_{0k}^i\, X_k^j\, \langle X_k\rangle_q^l\right\} \to \mathsf{E}\left\{X_0^i X^j \langle X\rangle_q^l\right\}. \tag{165}$$

Again, by Carleman's Theorem, the joint distributions of $(X_{0k}, X_k, \langle X_k \rangle_q)$ converge to that of $(X_0, X, \langle X \rangle_q)$. Indeed, from the viewpoint of user k, the multiuser setting is equivalent to the single-user setting in which the SNR suffers a degradation η (compare Figures 5(b) and 5(a)). Hence we have proved the decoupling principle and Claim 1.

In the large-system limit, the transformation from the input X_{0k} to the multiuser detection output $\langle X_k \rangle_q$ is nothing but a single-user Gaussian channel $p_{Z|X,\mathsf{snr};\eta}$ concatenated with a decision function (24). The decision function can be ignored from both detection- and information-theoretic viewpoints due to its monotonicity:

Proposition 1 *The decision function (24) is strictly monotone increasing in z for all snr and ξ.*

Proof: Let $(\cdot)'$ denote derivative with respect to z. One can show that

$$q_i'(z, \mathsf{snr}; \xi) = \xi \sqrt{\mathsf{snr}}\, q_{i+1}(z, \mathsf{snr}; \xi) - \xi z\, q_i(z, \mathsf{snr}; \xi) \quad i = 0, 1, \dots \tag{166}$$

Clearly,

$$\left[\frac{q_1(z, \mathsf{snr}; \xi)}{q_0(z, \mathsf{snr}; \xi)}\right]' = \xi \sqrt{\mathsf{snr}}\, \frac{q_2(z, \mathsf{snr}; \xi) q_0(z, \mathsf{snr}; \xi) - [q_1(z, \mathsf{snr}; \xi)]^2}{[q_0(z, \mathsf{snr}; \xi)]^2}. \tag{167}$$

The numerator in (167) is positive (≥ 0) by the Cauchy-Schwartz inequality. For the numerator in (167) to be 0, X must be a constant, which contradicts the assumption that X has zero mean and unit variance. Therefore, (24) is strictly increasing. ∎

Collecting relevant results in the above, the equivalent single-user channel is then an additive Gaussian noise channel with input signal-to-noise ratio snr and noise variance η^{-1} as depicted in Figure 5(b). Corollaries 1 and 2 are thus proved. In the special case that the postulated measure q is identical to the actual measure p, Claim 1 reduces to Claim 2.

The single-user mutual information is now simply that of a Gaussian channel with input distribution p_X,

$$I(\eta\, \mathsf{snr}) = -\frac{1}{2} \log \frac{2\pi e}{\eta} - \int p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr}; \eta) \log p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr}; \eta)\, \mathrm{d}z. \tag{168}$$

The overall spectral efficiency under separate decoding is therefore

$$\mathsf{C}_{\text{sep}} = \beta\, \mathsf{E}\left\{ I(\eta\, \mathsf{snr}) \right\}. \tag{169}$$

Hence the proof of (27). Claim 3 is proved by comparing (169) to (135).

5 Complex-valued Channels

Until now the discussion is based on a real-valued setting of the multiuser system, namely, both the inputs X_k and the spreading chips S_{nk} take real values. In practice, particularly in carrier-modulated communications where spectral efficiency is a major concern, transmission in the complex domain must be addressed. Either the input symbols or the spreading chips or both can take values in the complex number set. In the complex-valued setting, the channel model (4) is equivalent to the following real-valued one:

$$\begin{bmatrix} \boldsymbol{Y}^{(\mathrm{r})} \\ \boldsymbol{Y}^{(\mathrm{i})} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}^{(\mathrm{r})} & -\boldsymbol{S}^{(\mathrm{i})} \\ \boldsymbol{S}^{(\mathrm{i})} & \boldsymbol{S}^{(\mathrm{r})} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}^{(\mathrm{r})} \\ \boldsymbol{X}^{(\mathrm{i})} \end{bmatrix} + \begin{bmatrix} \boldsymbol{N}^{(\mathrm{r})} \\ \boldsymbol{N}^{(\mathrm{i})} \end{bmatrix}, \tag{170}$$

where the superscripts (r) and (i) denote real and imaginary components respectively. Note that the previous analysis does not apply to (170) since the channel state matrix does not contain i.i.d. entries in this case.

If the inputs take complex values but the spreading is real-valued ($\boldsymbol{S}^{(\mathrm{i})} = 0$), the channel can be considered as two uses of the real-valued channel $\boldsymbol{S} = \boldsymbol{S}^{(\mathrm{r})}$, where the inputs $\boldsymbol{X}^{(\mathrm{r})}$ and $\boldsymbol{X}^{(\mathrm{i})}$ to the two channels may be dependent. Since independent inputs maximize the channel capacity, there is little reason to transmit dependent signals in the two subchannels. Thus the analysis of the real-valued channel in previous sections also applies to the case of independent in-phase and quadrature components, while the only change is that the spectral efficiency is the sum of that of the two subchannels.

We can also compare the real-valued and the complex-valued channel assuming the same real-valued input distribution. Under the complex-valued channel,

$$\begin{bmatrix} \boldsymbol{Y}^{(\mathrm{r})} \\ \boldsymbol{Y}^{(\mathrm{i})} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}^{(\mathrm{r})} \\ \boldsymbol{S}^{(\mathrm{i})} \end{bmatrix} \boldsymbol{X} + \begin{bmatrix} \boldsymbol{N}^{(\mathrm{r})} \\ \boldsymbol{N}^{(\mathrm{i})} \end{bmatrix}, \tag{171}$$

which is equivalent to transmitting the same $\boldsymbol{X}$ twice over two uses of real-valued channels. This is equivalent to having a real-valued channel with the load β halved.

If both the symbols and the spreading chips are complex-valued, the analysis in the previous sections can be modified to take this into account. For convenience it is assumed that the real and imaginary components of spreading chips, $S_{nk}^{(\mathrm{r})}$, $S_{nk}^{(\mathrm{i})}$ are i.i.d. with zero mean and unit variance. The noise vector has i.i.d. circularly symmetric Gaussian entries, i.e., $\mathsf{E}\{\boldsymbol{N}\boldsymbol{N}^{\mathrm{H}}\} = 2\boldsymbol{I}$. Thus the conditional probability density function of the actual multiuser channel is

$$p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S}) = (2\pi)^{-L} \exp\left[-\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right], \tag{172}$$

whereas that of the postulated channel is

$$q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{S}) = \left(2\pi\sigma^2\right)^{-L} \exp\left[-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|^2\right]. \tag{173}$$

Also, the actual and the postulated input distributions p_X and q_X have both zero-mean and unit variance, $\mathsf{E}\left\{|X|^2\right\} = \mathsf{E}_q\left\{|X|^2\right\} = 1$. Note that the in-phase and the quadrature components are intertwined due to complex spreading.

The replica analysis can be carried out in parallel as that in Section 4. In the following we highlight the major differences. Given $(\boldsymbol{\Gamma}, \underline{\boldsymbol{X}})$, the variables V_a defined in (96) have asymptotically independent real and imaginary components. Thus, $G_K^{(u)}$ can be evaluated to be 2 times that under real-valued channels with

$$Q_{ab} = \frac{1}{K}\sum_{k=1}^{K} \mathsf{snr}_k \mathrm{Re}\left\{X_{ak}X_{bk}^*\right\}, \quad a, b = 0, \ldots, u. \tag{174}$$

The rate $I^{(u)}$ of the measure $\mu_K^{(u)}$ of $\boldsymbol{Q}$ is obtained as

$$I^{(u)}(\boldsymbol{Q}) = \sup_{\tilde{\boldsymbol{Q}}} \left[\mathsf{tr}\left\{\tilde{\boldsymbol{Q}}\boldsymbol{Q}\right\} - \log \mathsf{E}\left\{\exp\left[\mathsf{snr}\boldsymbol{X}^{\mathrm{H}}\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\right\} / \log e\right]. \tag{175}$$

As a result, the fixed-point joint equations for $\boldsymbol{Q}$ and $\tilde{\boldsymbol{Q}}$ are

$$\tilde{\boldsymbol{Q}} = -\frac{2}{\beta}\left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{Q}\right)^{-1}, \tag{176a}$$

$$\boldsymbol{Q} = \frac{\mathsf{E}\left\{\mathsf{snr}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}} \exp\left[\mathsf{snr}\boldsymbol{X}^{\mathrm{H}}\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\right\}}{\mathsf{E}\left\{\exp\left[\mathsf{snr}\boldsymbol{X}^{\mathrm{H}}\tilde{\boldsymbol{Q}}\boldsymbol{X}\right]\right\}}. \tag{176b}$$

Under replica symmetry (115), the parameters (c, d, f, g) are found to be 2 times the corresponding values given in (123), and (r, m, p, q) are found the same as in (131) except that all squares are replaced by squared norms. By defining two parameters (differ from (124) by a factor of 2):

$$\eta = \frac{d^2}{f} \quad \text{and} \quad \xi = d, \tag{177}$$

we have the following result.

Claim 5 *Let the multiuser posterior mean estimate of the complex-valued multiple-access channel (172) with complex-valued spreading be $\langle \boldsymbol{X} \rangle_q$ parameterized by a postulated input distribution q_X and noise level σ. Then, in the large-system limit, the distribution of the multiuser detection output $\langle X_k \rangle_q$ conditioned on $X_k = x$ being transmitted with signal-to-noise ratio snr_k is identical to the distribution of the estimate $\langle X \rangle_q$ of a single-user complex Gaussian channel*

$$Z = \sqrt{\mathsf{snr}}\, X + \frac{1}{\sqrt{\eta}} N \tag{178}$$

conditioned on $X = x$ being transmitted with $\mathsf{snr} = \mathsf{snr}_k$, where N is circularly symmetric Gaussian with unit variance, $\mathsf{E}\left\{|N|^2\right\} = 1$. The multiuser efficiency η and the inverse noise variance ξ of the postulated single-user channel (173) satisfy the coupled equations (22), where the mean-square error $\mathcal{E}(\mathsf{snr}; \eta, \xi)$ of the posterior mean estimate and the variance $\mathcal{V}(\mathsf{snr}; \eta, \xi)$ of the retrochannel are defined similarly as that of the real-valued channel, with the squares in (20) and (21) replaced by squared norms. In case of multiple solutions to (22), (η, ξ) are chosen to minimize the free energy:

$$\begin{aligned} \mathcal{F} = & -\mathsf{E}\left\{ \int p_{Z|\mathsf{snr};\eta}(z|\mathsf{snr};\eta) \log q_{Z|\mathsf{snr};\xi}(z|\mathsf{snr};\xi)\,\mathrm{d}\,z \right\} \\ & + \log\frac{\xi}{\pi} - \frac{\xi}{\eta}\log e + \frac{\sigma^2\xi(\eta-\xi)}{\beta\eta}\log e + \frac{1}{\beta}[(\xi-1)\log e - \log\xi] + \frac{1}{\beta}\log(2\pi) + \frac{\xi}{\beta\eta}\log e. \end{aligned} \tag{179}$$

Corollary 3 *For the complex-valued channel (172), the mutual information of the single-user channel seen at the multiuser posterior mean estimator output for a user with signal-to-noise ratio snr is*

$$I(\eta\, \mathsf{snr}) = \mathsf{D}\left(p_{Z|X,\mathsf{snr};\eta} \| p_{Z|\mathsf{snr};\eta} | p_X\right). \tag{180}$$

where η is the multiuser efficiency given by Claim 5 and $p_{Z|\mathsf{snr};\eta}$ is the marginal probability distribution of the output of channel (178). The overall spectral efficiency under suboptimal separate decoding is

$$\mathsf{C}_{\text{sep}}(\beta) = \beta\,\mathsf{E}\left\{I(\eta\,\mathsf{snr})\right\}. \tag{181}$$

Claim 6 *The optimal spectral efficiency under joint decoding is*

$$\mathsf{C}_{\text{joint}}(\beta) = \beta\,\mathsf{E}\left\{I(\eta\,\mathsf{snr})\right\} + (\eta - 1)\log e - \log\eta, \tag{182}$$

where η is the optimal multiuser efficiency determined by Claim 5 by postulating a measure q that is identical to p.

It is interesting to compare the performance of the real-valued channel and that of the complex-valued channel. We assume the in-phase and quadrature components of the input symbols are independent with identical distribution p'_X which has a variance of $\frac{1}{2}$. By Claim 5, the equivalent

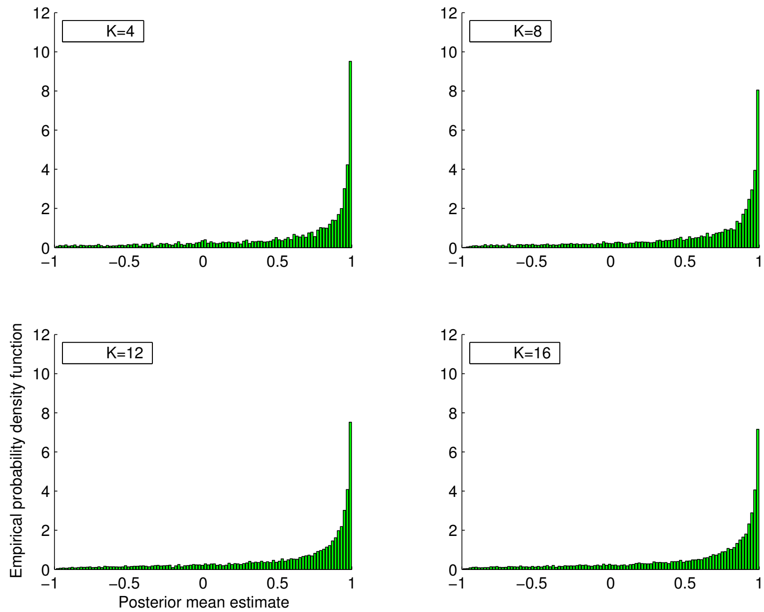


Figure 7: Simulated probability density function of the posterior mean estimates under binary input conditioned on "+1" being transmitted. Systems with 4, 8, 12 and 16 equal-power users are simulated with $\beta = 2/3$. The SNR is 2 dB.

single-user channel (178) can also be regarded as two independent subchannels. The mean-square error and the variance in (22) are the sum of those of the subchannels. It can be checked that the performance of each subchannel is identical to that of the real-valued channel with input distribution p'_X normalized to unit variance. Note, however, that the total transmit energy in case of complex spreading take twice the energy of their real counterparts. In all, the error performance under complex-valued spreading is the exactly the same as those under real-valued spreading. This result simplifies the analysis of complex-valued channels such as those arise in multiantenna systems. On the other hand, if we have control over the channel state matrix, as in CDMA systems, complex-valued spreading should be avoided due to higher complexity with no performance gain.

6 Numerical Results

In Figures 7–8 we plot the simulated distribution of the posterior mean estimate and its corresponding "hidden" Gaussian statistic. Equal-power users with binary input are considered. We simulate CDMA systems of 4, 8, 12 and 16 users respectively. The load is fixed to $\beta = 2/3$ and the SNR is 2 dB. Let $X_k = 1$ be transmitted. We collect the output decision statistics of the posterior mean estimator (i.e., the soft output of the individually optimal detector, $\langle X_k \rangle$) out of several thousand experiments. A histogram of the statistic is obtained and then scaled to plot an estimate of the probability density function in Figure 7. We also apply the inverse nonlinear decision function to recover the "hidden" Gaussian decision statistic (normalized so that its conditional mean is equal to $X_k = 1$), which in this case is

$$\tilde{Z}_k = \frac{\tanh^{-1}(\langle X_k \rangle)}{\eta \, \mathsf{snr}_k}. \tag{183}$$

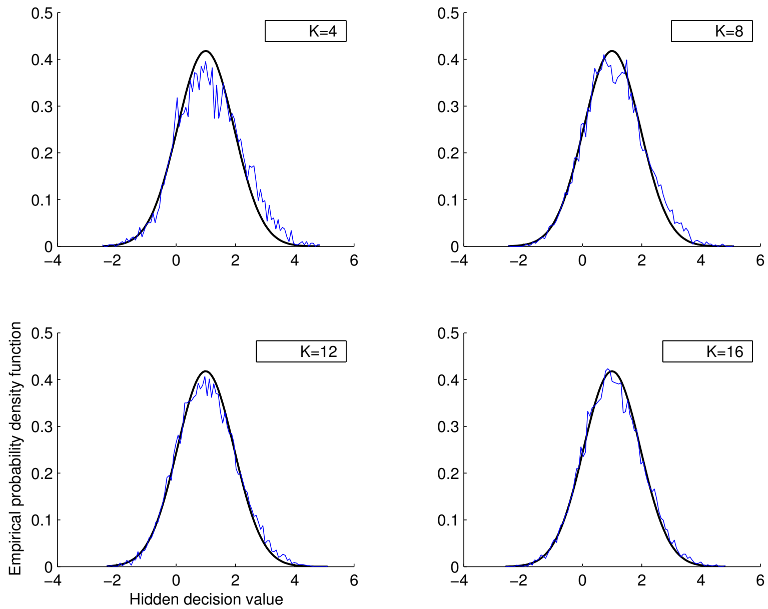


Figure 8: Simulated probability density function of the "hidden" Gaussian statistic recovered from the posterior mean estimates under binary input conditioned on "+1" being transmitted. Systems with 4, 8, 12 and 16 equal-power users are simulated with $\beta = 2/3$. The SNR is 2 dB. The asymptotic Gaussian distribution predicted by our theory is also plotted for comparison.

The probability density function of $\tilde{Z}_k$ estimated from its histogram is then compared to the theoretically predicted Gaussian density function in Figure 8. It is clear that even though the PME output $\langle X_k \rangle$ takes a non-Gaussian distribution, the equivalent statistic $\tilde{Z}_k$ converges to a Gaussian distribution centered at X_k as K becomes large. This result is particularly powerful considering that the "fit" to the Gaussian distribution is quite good even for a system with merely 8 users.

In Figures 9–10, the multiuser efficiency and the spectral efficiency are plotted as functions of the average SNR. We consider three input distributions, namely, real-valued Gaussian and binary inputs, and (complex-valued) 8PSK inputs. Under Gaussian and binary inputs, where real-valued spreading is considered, the multiuser efficiencies are given by (54) and (62) respectively, and the spectral efficiencies are given by (58)–(59), (63) and (64) respectively. Under 8PSK, where complex-valued spreading is assumed, the multiuser efficiency and the spectral efficiency are given by Claim 5 and Corollary 3 respectively. We also consider two SNR distributions: 1) identical SNRs for all users (perfect power control), and 2) two groups of users of equal population with a power difference of 10 dB. We first assume a system load of $\beta = 1$ and then redo the experiments with $\beta = 3$.

In Figure 9(a), the multiuser efficiency under Gaussian inputs and linear MMSE detection is plotted as a function of the average SNR. The load is $\beta = 1$. We find the multiuser efficiencies decrease from 1 to 0 as the SNR increases. The monotonicity can be easily verified by inspecting the Tse-Hanly equation (54). Transmission with unbalanced power improves the multiuser efficiency. The corresponding spectral efficiencies of the system are plotted in Figure 9(b). Both joint decoding and separate decoding are considered. The gain in the spectral efficiency due to joint decoding is little under low SNR but significant under high SNR. Unbalanced SNR reduces the spectral efficiency, where under separate decoding the loss is almost negligible.

The multiuser efficiency under binary inputs and nonlinear MMSE (individually optimal) detection is plotted in Figure 9(c). The multiuser efficiency is not monotone. The multiuser efficiency

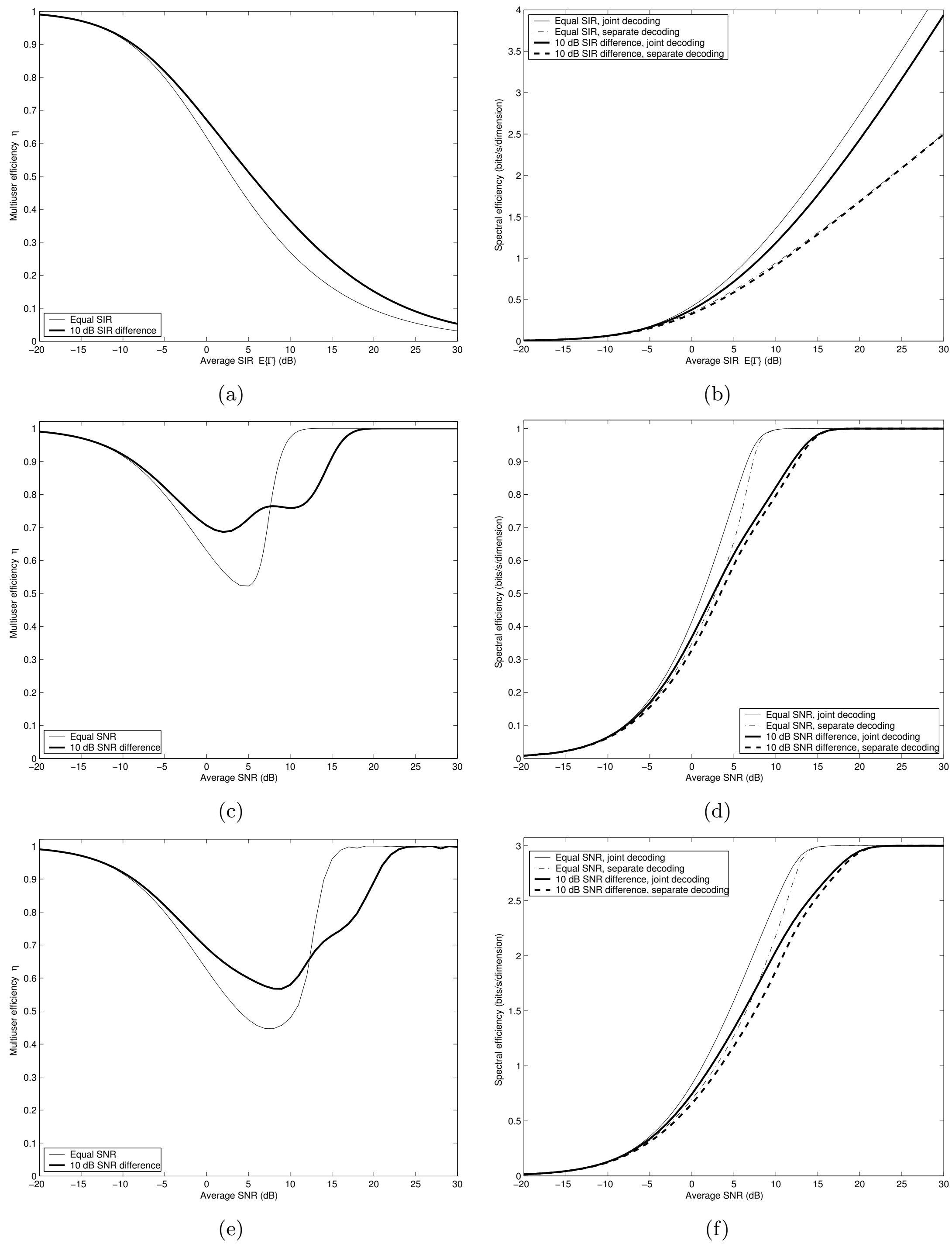


Figure 9: Plots of the multiuser efficiency and spectral efficiency as functions of the SNR. The load is $\beta = 1$. (a) The multiuser efficiency, Gaussian inputs. (b) The spectral efficiency, Gaussian inputs. (c) The multiuser efficiency, binary inputs. (d) The spectral efficiency, binary inputs. (e) The multiuser efficiency, 8PSK inputs. (f) The spectral efficiency, 8PSK inputs.

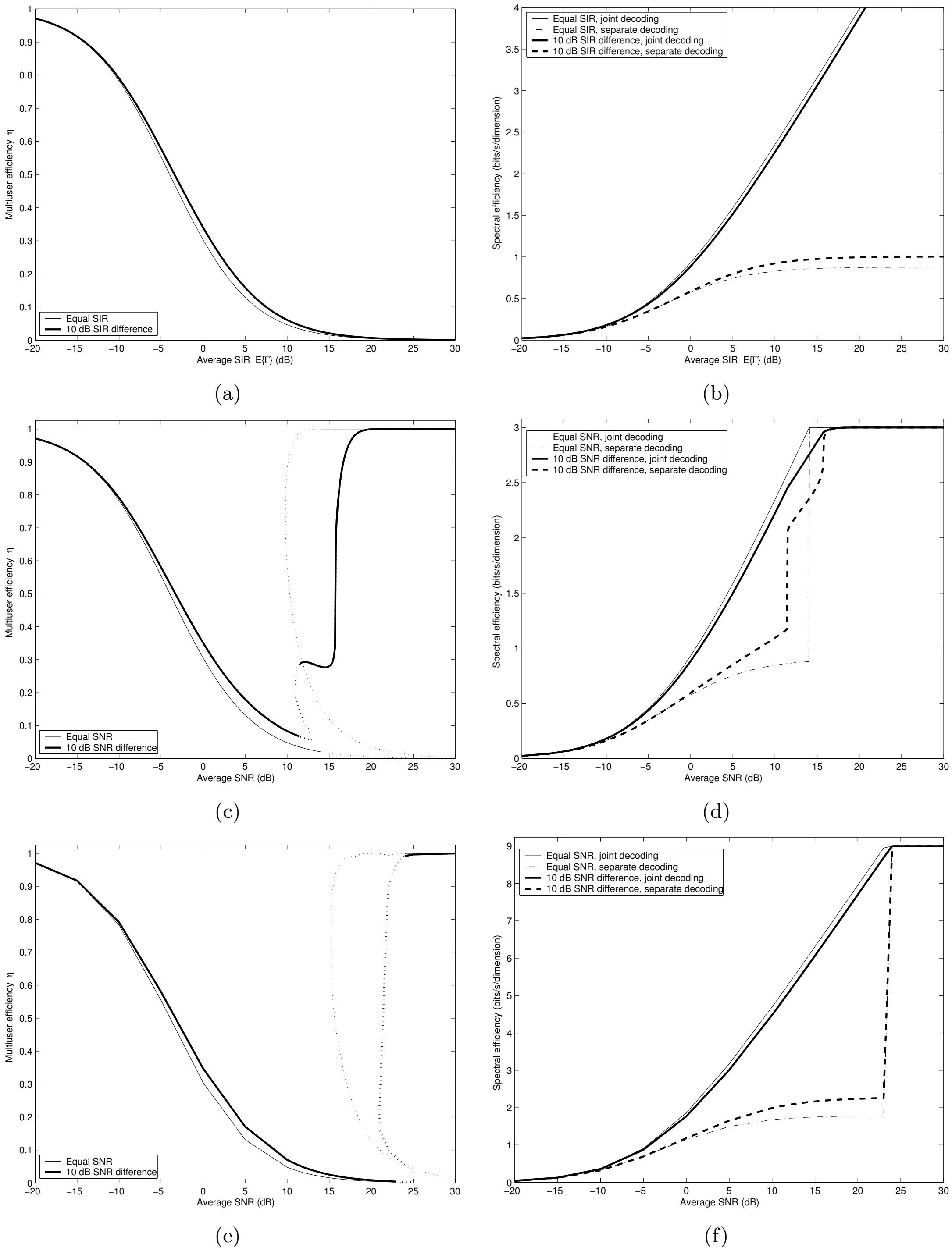


Figure 10: Plots of the multiuser efficiency and spectral efficiency as functions of the SNR. The load is $\beta = 3$. (a) The multiuser efficiency, Gaussian inputs. (b) The spectral efficiency, Gaussian inputs. (c) The multiuser efficiency, binary inputs. (d) The spectral efficiency, binary inputs. (e) The multiuser efficiency, 8PSK inputs. (f) The spectral efficiency, 8PSK inputs.

converges to 1 for both diminishing SNR and infinite SNR. While for diminishing SNR this follows directly from the definition of multiuser efficiency, the convergence to unity as the SNR goes to infinity was shown in [54] for the case of binary inputs. A single dip is observed for the case of identical SNRs while two dips are observed in the case of two SNRs of equal population with 10 dB difference in SNR (the gap is about 10 dB). The corresponding spectral efficiencies are plotted in Figure 9(d). The spectral efficiencies saturate to 1 bit/s/dimension at high SNR. The difference between joint decoding and separate decoding is quite small for both very low and very high SNRs while it can be 25% at around 8 dB.

The multiuser efficiency under 8PSK inputs and nonlinear MMSE detection is plotted in Figure 9(e). The multiuser efficiency curve is slightly better than that for binary inputs. The corresponding spectral efficiencies are plotted in Figure 9(f). The spectral efficiencies saturate to 3 bit/s/dimension at hight SNR.

In Figure 10, we redo the previous experiments only with a different system load $\beta = 3$. The results are to be compared with those in Figure 9.

Under Gaussian inputs, the multiuser efficiency curves in Figure 10(a) take a similar shape as in Figure 9(a), but are significantly lower due to higher load. The corresponding spectral efficiencies are shown in Figure 10(b). It is clear that higher load results in higher spectrum usage under joint decoding. Separate decoding, however, is interference limited and saturates under high SNR (cf. [11, Figure 1]).

In Figure 10(c), we plot the multiuser efficiency under binary inputs. All solutions to the fixed-point equation (35) of the multiuser efficiency are shown. Under equal SNR, multiple solutions coexist for an average SNR of 10 dB or higher. If two groups of users with 10 dB difference in SNR, multiple solutions are seen in between 11 to 13 dB. The solution that minimizes the free energy is valid and is shown in solid lines, while invalid solutions are plotted using dotted lines. An almost 0 to 1 jump is observed under equal SNR and a much smaller jump is seen under unbalanced SNRs. This is known as phase transition in statistical physics. The asymptotics under equal SNR can be shown by taking the limit $\mathsf{snr} \to \infty$ in (62). Essentially, if $\eta\mathsf{snr} \to \infty$, then $\eta \to 1$; while if $\eta\mathsf{snr} \to \tau$ where τ is the solution to

$$\tau \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \left[1 - \tanh(\tau - z\sqrt{\tau})\right] \mathrm{d}z = \frac{1}{\beta}, \tag{184}$$

then $\eta \to 0$. If $\beta > 2.085$, there exists a solution to (184) so that two solutions coexist for large SNR.

The spectral efficiency under binary inputs and $\beta = 3$ is shown in Figure 10(d). As a result of phase transition, one observes a jump to saturation in the spectral efficiency under equal-power binary inputs. The gain due to joint decoding can be significant in moderate SNRs. In case of two groups of users with 10 dB difference in SNR, the spectral efficiency curve also shows one jump and the loss due to separate decoding is reduced significantly for a small window of SNRs around the areas of phase transition (11–13 dB). Therefore, perfect power control may not be the best strategy in such cases.

Under 8PSK inputs, the multiuser efficiency and spectral efficiency curves in Figure 10(e) and 10(f) take similar shape as the curves under binary inputs in Figure 10(c) and 10(d). Phase transition causes jumps in both the multiuser efficiency and the spectral efficiency. In Figures 10(e)–10(d) a sharp bend upward is observed at the point of phase transition. This is known as "spinodal" in statistical physics.

A comparison of Figures 10(b), 10(d) and 10(f) shows that under separate decoding, the spectral efficiency under Gaussian inputs saturates well below that of binary and 8PSK inputs. This implies that the nonlinear MMSE detector with separate decoding is not efficient in case of dense constellation.

7 Summary

The main contribution of this paper is a simple characterization of the performance of CDMA multiuser detection under arbitrary input distribution and SNR (and/or flat fading) in the large-system limit. A broad family of multiuser detectors is studied under the name of posterior mean estimators, which includes well-known detectors such as the matched filter, decorrelator, linear MMSE detector, maximum likelihood (jointly optimal) detector, and the individually optimal detector.

A key conclusion is the decoupling of a Gaussian multiuser channel concatenated with a generic multiuser detector front end. It is found that the multiuser detection output is a deterministic function of a hidden Gaussian statistic centered at the transmitted symbol. Hence the single-user channel seen at the multiuser detection output is equivalent to a Gaussian channel in which the overall effect of multiple-access interference is a degradation factor in the effective signal-to-interference ratio. This degradation factor, known as the multiuser efficiency, is the solution to a pair of coupled fixed-point equations, and can be easily computed numerically if not analytically.

Another set of results, tightly related to the decoupling principle, are some general formulas for the large-system spectral efficiency of multiuser channels expressed in terms of the multiuser efficiency, both under joint and separate decoding. It is found that the decomposition of optimum spectral efficiency as a sum of single-user efficiencies and a joint decoding gain applies under more general conditions than shown in [12], thereby validating Müller's conjecture [26]. A relationship between the spectral efficiencies under joint and separate decoding is an outcome of an intriguing formula that links the mutual information and MMSE [37].

From a practical viewpoint, this paper presents new results on the efficiency of CDMA communication under arbitrary input signaling such as m-PSK and m-QAM and arbitrary power profile. More importantly, the results in this paper allow the performance of multiuser detection to be characterized by a single parameter, the multiuser efficiency. The efficiency of spectrum usage is also easily quantified by means of this parameter. Thus, the results offer valuable insights in the design and analysis of CDMA systems, e.g., in power control [55].

The linear system in our study also models multiple-input multiple-output channels under various circumstances. The results can thus be used to evaluate the output SNR or spectral efficiency of high-dimensional MIMO channels (such as multiple-antenna systems) with arbitrary signaling and various detection techniques. Some of the results in this paper have been generalized to MIMO channels with spatial correlation at both transmitter and receiver sides [56].

8 Acknowledgements

The authors are grateful to the anonymous referees for their helpful reviews. Thanks also to Dr. Ralf Müller, Dr. Toshiyuki Tanaka and Mr. Chih-Chun Wang for interesting discussions.

References

[1] S. Verdú, "Computational complexity of optimum multiuser detection," *Algorithmica*, vol. 4, no. 3, pp. 303–312, 1989.

[2] S. Verdú, "Minimum probability of error for asynchronous multiple access communication systems," in *Proceedings of IEEE Military Communications Conference*, vol. 1, pp. 213–219, Nov. 1983.

[3] S. Verdú, *Optimum multiuser signal detection.* PhD thesis, Dept. of Electrical Engineering, University of Illinois at Urbana-Champaign, Aug. 1984. Technical Report T-151, Coordinated Science Laboratory.

[4] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 32, pp. 85–96, Jan. 1986.

[5] S. Verdú, "Optimum multiuser asymptotic efficiency," *IEEE Trans. Inform. Theory*, vol. 34, pp. 890–897, Sept. 1986.

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

[7] 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, March 1999.

[8] Y. C. Eldar and A. M. Chan, "On the asymptotic performance of the decorrelator," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2309–2313, Sept. 2003.

[9] D. Guo and S. Verdú, "Multiuser detection and statistical mechanics," in *Communications, Information and Network Security* (V. Bhargava, H. V. Poor, V. Tarokh, and S. Yoon, eds.), ch. 13, pp. 229–277, Kluwer Academic Publishers, 2002.

[10] 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.

[11] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, March 1999.

[12] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1302–1327, May 2001.

[13] 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.

[14] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices, a review," *Statistica Sinica*, vol. 9, pp. 611–677, July 1999.

[15] S. Verdú, "Capacity region of Gaussian CDMA channels: The symbol-synchronous case," in *Proceedings 24th Allerton Conference on Communication, Control and Computing*, pp. 1025–1034, Oct. 1986.

[16] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2832–2836, Nov. 1998.

[17] P. B. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output channel," *IEEE Trans. Commun.*, vol. 48, pp. 1245–1248, Aug. 2000.

[18] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[19] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, Nov./Dec. 1999.

[20] G. J. Rodgers, "Comment: Spectral analysis of random matrices using the replica method," *Statistica Sinica*, vol. 9, pp. 662–667, 1999.

[21] T. Tanaka, "Analysis of bit error probability of direct-sequence CDMA multiuser demodulators," in *Advances in Neural Information Processing Systems* (T. K. Leen et al., ed.), vol. 13, pp. 315–321, The MIT Press, 2001.

[22] T. Tanaka, "Average-case analysis of multiuser detectors," in *Proceedings 2001 IEEE International Symposium on Information Theory*, p. 287, Washington, D.C., June 2001.

[23] T. Tanaka, "Statistical mechanics of CDMA multiuser demodulation," *Europhysics Letters*, vol. 54, no. 4, pp. 540–546, 2001.

[24] 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.

[25] R. R. Müller and W. H. Gerstacker, "On the capacity loss due to separation of detection and decoding," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1769–1778, Aug. 2004.

[26] R. R. Müller, "On channel capacity, uncoded error probability, ML-detection and spin glasses," in *Proceedings Workshop on Concepts in Information Theory*, pp. 79–81, Breisach, Germany, June 2002.

[27] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, pp. 965–974, May 1975.

[28] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Number 111 in International Series of Monographs on Physics, Oxford University Press, 2001.

[29] N. Sourlas, "Spin-glass models as error-correcting codes," *Nature*, vol. 339, pp. 693–695, June 1989.

[30] Y. Kabashima and D. Saad, "Statistical mechanics of error-correcting codes," *Europhysics Letters*, vol. 45, no. 1, pp. 97–103, 1999.

[31] A. Montanari and N. Sourlas, "The statistical mechanics of turbo codes," *European Physical Journal B*, vol. 18, no. 1, pp. 107–119, 2000.

[32] Y. Kabashima and D. Saad, "Statistical mechanics of low-density parity-check codes," *Journal of Physics A: Mathematical and General*, vol. 37, pp. R1–R43, 2004.

[33] M. Talagrand, "Rigorous results for mean field models for spin glasses," *Theoretical Computer Science*, vol. 265, pp. 69–77, Aug. 2001.

[34] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*. World Scientific, 1987.

[35] V. Dotsenko, *The Theory of Spin Glasses and Neural Networks*. World Scientific, 1994.

[36] D. Guo, S. Shamai, and S. Verdú, "Mutual information and MMSE in Gaussian channels," in *Proceedings 2004 IEEE International Symposium on Information Theory*, p. 347, Chicago, IL, USA, June 2004.

[37] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *submitted to IEEE Trans. on Inform. Theory*, 2004.

[38] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, pp. 1405–1409, Monterey, CA, Nov. 1997.

[39] P. Rapajic, M. Honig, and G. Woodward, "Multiuser decision-feedback detection: Performance bounds and adaptive algorithms," in *Proceedings of 1998 IEEE International Symposium on Information Theory*, p. 34, Cambridge, MA USA, Aug. 1998.

[40] S. L. Ariyavisitakul, "Turbo space-time processing to improve wireless channel capacity," *IEEE Trans. Commun.*, vol. 48, pp. 1347–1359, Aug. 2000.

[41] R. R. Müller, "Multiuser receivers for randomly spread signals: Fundamental limits with and without decision-feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 268–283, Jan. 2001.

[42] T. Guess and M. K. Varanasi, "An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels," *IEEE Trans. Inform. Theory*, 2004. To appear.

[43] F. Comets, "The martingale method for mean-field disordered systems at high temperature," in *Mathematical Aspects of Spin Glasses and Neural Networks* (A. Bovier and P. Picco, eds.), pp. 91–113, Birkhäuser, 1998.

[44] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York : Wiley, 1991.

[45] J. Yedidia, W. Freeman, and Y. Weiss, "Generalized belief propagation," in *Advances in Neural Information Processing Systems 13*, pp. 689–695, MIT Press, 2001.

[46] J. van Mourik, D. Saad, and Y. Kabashima, "Critical noise levels for LPDC decoding," *Physics Review E*, vol. 66, no. 2, pp. 026705:1–8, 2002.

[47] M. Chiang and S. Boyd, "Geometric programming duals of channel capacity and rate distortion," *IEEE Trans. Inform. Theory*, vol. 50, pp. 245–258, Feb. 2004.

[48] R. Müller, "Random matrices, free probability, and the replica method," in *Proceedings of European Signal Processing Conference*, (Vienna, Austria), Sept. 2004.

[49] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, vol. 271 of *A series of comprehensive studies in mathematics.* Springer-Verlag, 1985.

[50] D. Guo, *Gaussian Channels: Information, Estimation and Multiuser Detection.* PhD thesis, Princeton University, 2004.

[51] 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.

[52] J. Hubbard, "Calculation of partition functions," *Physics Review Letters*, vol. 3, pp. 77–78, July 1959.

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

[54] D. N. C. Tse and S. Verdú, "Optimum asymptotic multiuser efficiency for randomly spread CDMA," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2718–2722, Nov. 2000.

[55] F. Meshkati, D. Guo, H. V. Poor, S. Schwartz, and N. B. Mandayam, "A unified approach to power control for multiuser detectors," in *Proceedings 2nd International Workshop on Signal Processing for Wireless Communications*, London, England, June 2004.

[56] C.-K. Wen and K.-K. Wong, "Asymptotic spectral efficiency of spatially correlated MIMO multiple-access channels with arbitrary signaling inputs for joint and separate decoding," in *Proceedings 2004 IEEE International Symposium on Information Theory*, p. 282, Chicago, IL, USA, June 2004.