High-SNR Analysis of Optimum Multiuser Detection with an Unknown Number of Users

Adrià Tauste Campo  
Department of Engineering  
University of Cambridge  
e-mail: at461@cam.ac.uk

Albert Guillén i Fàbregas  
Department of Engineering  
University of Cambridge  
e-mail: albert.guillen@eng.cam.ac.uk

Ezio Biglieri  
Department of Technology  
Universitat Pompeu Fabra  
e-mail: ezio.biglieri@upf.edu

Abstract—We analyze multiuser detection under the assumption that the number of users accessing the channel is unknown by the receiver. Our main goal is to determine the performance loss caused by the need for estimating the identities of active users, which are not known a priori. To prevent a loss of optimality, we assume that identities and data are estimated jointly, rather than in two separate steps. We examine the performance of multiuser detectors when the number of potential users is large. Statistical-physics methodologies are used to determine the fixed-point equation whose solutions yield the multiuser efficiency of the optimal detector. Special attention is paid to the large signal-to-noise ratio, which yields tight closed-form bounds on the minimum mean-squared error. These bounds analytically illustrate the set of solutions of the fixed-point equation, and their relationship with the maximum system load. By identifying the region of computationally feasible solutions, we study the maximum load that the detector can support for a given SNR and quality of service, specified by the multiuser efficiency.

I. INTRODUCTION

In multiple-access communication, the evolution of user activity plays an important role. From one time instant to the next, some new users may become active and some old users inactive, while parameters of the persisting users, such as power or location, may vary. Most of the available multiuser detection (MUD) theory is based on the assumption that the number of active users is constant, known at the receiver, and equal to the maximum number of users entitled to access the system [1]. If this assumption is not verified, the receiver may exhibit a serious performance loss [2], [3]. In [4], the more realistic scenario in which the number of active users is unknown a priori, and varies with time with known statistics, is the basis of a new approach to detector design. The present work follows in the footsteps of [4], and is devoted to large-system analysis of this new type of detectors for Code Division Multiple Access (CDMA). Our main goal is to determine the performance loss caused by the need for estimating the identities of active users, which are not known a priori. In this paper we restrict our analysis to a worst-case scenario, where detection cannot improve its performance from past experience due to a degeneration of the activity model (for instance, assuming a Markovian evolution of the number of active users) into an independent process [5]. The same analysis applies to systems where the input symbols accounting for data and activity are interleaved before detection. To prevent a loss of optimality, we assume that identities and data are estimated jointly, rather than in two separate steps. Our interest is in randomly spread CDMA systems in terms of multiuser efficiency, whose natural dimensions (number of users $K$, and spreading gain $N$) grow to infinity, while their ratio (the “system load”) is kept fixed. In particular, we consider the optimal maximum a posteriori (MAP) multiuser detector, and use tools recently adopted from statistical physics [6], [7], [8], [9]. Of special relevance in our analysis is the decoupling principle introduced in [7] for randomly spread CDMA.

The results of this paper focus on the degradation of multiuser efficiency when the uncertainty on the activity of the users grows and the SNR is sufficiently large. Here, we go one step beyond the simple application of the large-system fixed-point equation [6], [7] into our model and we provide a new high-SNR analysis on the space of fixed-point solutions showing explicitly its interplay with the system load for a non-uniform ternary and parameter-dependent input distribution. By expanding the minimum mean square error for large SNR, we obtain tight closed-form bounds that describe the large CDMA system as a function of the SNR, the activity factor and the system load and some trade-off results between these quantities are derived. Of special novelty here is the study of the impact of the activity factor in the CDMA performance measures (minimum mean-square error, and multiuser efficiency). Besides, we identify the region of “meaningful” multiuser efficiency solutions and derive consequences for engineering problems of practical interest.

II. SYSTEM MODEL

We consider a CDMA system with an unknown number of users [4], and examine the optimum user-and-data detector. In particular, we study randomly spread direct-sequence (DS) CDMA with a maximum of $K$ active users, where the received signal at time $t$ is

$$y_t = SA b_t + z_t$$  \hspace{1cm} (1)

where $y_t \in \mathbb{R}^N$, with $N$ the length of the spreading sequences, $S \in \mathbb{R}^{N \times K}$ the matrix of the sequences, $A = \text{diag}(a_1, \ldots, a_K) \in \mathbb{R}^{K \times K}$ is the diagonal matrix of the users’ signal amplitudes, $b_t = (b_1^t, \ldots, b^K_t) \in \mathbb{R}^K$ is the users’ data vector, and $z_t$ is an additive white Gaussian noise vector with i.i.d. entries $\sim \mathcal{N}(0, 1)$. We assume that $a_k = \sqrt{\gamma}$, where
\(\gamma\) is the average received signal-to-noise ratio (SNR)\(^1\). We define the system’s activity rate as \(\alpha \triangleq \Pr\{\text{user } k \text{ is active}\}, 1 \leq k \leq K\). Active users employ binary phase-shift keying (BPSK) with equal probabilities. This scheme is equivalent to one where each user transmits a ternary constellation (BPSK) with equal probabilities. This scheme is equivalent to one where each user transmits a ternary constellation \(X \triangleq \{-1, 0, +1\}\) with probabilities \(\Pr\{b_k^* = -1\} = \Pr\{b_k^* = +1\} = \frac{1}{3}\) and \(\Pr\{b_k^* = 0\} = 1 - \alpha\). We define the maximum system load as \(\beta \triangleq \frac{K}{N}\).

In a static channel model, the detector operation remains invariant along a data frame, indexed by \(t\), but we often omit this time index for the sake of simplicity. Assuming that the receiver knows \(S\) and \(A\), the a posteriori probability (APP) of the transmitted data has the form
\[
p(b|y, S, A) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\|y - SAb\|^2}{2}} \frac{p(b)}{p(y|S, A)}.
\]
Hence, the maximum a posteriori (MAP) joint activity-and-data multiuser detector solves
\[
\hat{b} = \arg \max_{b \in X^K} p(b|y, S, A).
\]
Similarly, optimum detection of single-user data and activity is obtained by marginalizing over the undesired users as follows:
\[
\hat{b}^k = \arg \max_{b \in \{0, 1\}} \sum_{b \neq b^k} p(b|y, S, A).
\]

A. The large-system decoupling principle

In a communications scheme such as the one modeled by (2), the goal of the multiuser detector is to infer the information-bearing symbols given the received signal \(y\) and the knowledge about the channel state. This leads naturally to defining the corresponding free energy \[6\], normalized to the number of users and with the choice \(T = 1\), as
\[
\mathcal{F}_K \triangleq -\frac{1}{K} \ln p(y | S)
\]
To calculate this expression we make the self-averaging assumption, which states that the randomness of (5) vanishes as \(K \to \infty\). This is tantamount to saying that the free energy per user \(\mathcal{F}_K\) converges in probability to its expected value over the distribution of the random variables \(y\) and \(S\), denoted by
\[
\mathcal{F} \triangleq \lim_{K \to \infty} \mathbb{E}\left\{-\frac{1}{K} \ln p(y | S)\right\}.
\]
To compute the free energy through replica analysis, one of the cornerstones in large deviation theory, the Varadhan’s theorem \[10\], is invoked to transform the calculation of the limiting free energy into a simplified optimization problem, whose solution is assumed to exhibit symmetry among its replicas. More specifically, in the case of a MAP individually optimum detector, the optimization yields a fixed-point equation, whose unknown is a single operational macroscopic parameter, which is claimed to be the multiuser efficiency (reflecting the degradation factor of SNR due to interference) of an equivalent Gaussian channel \[7\]. Due to the structure of the optimization problem, the multiuser efficiency must minimize the free energy. The above is tantamount to formulating the decoupling principle:

**Claim 2.1:** \[7\], \[9\] Given a multiuser channel, the distribution of the output \(\hat{b}^k\) of the individually optimum (IO) detector, conditioned on \(b^k = b \in X\) being transmitted with constant amplitude \(a = \sqrt{\gamma}\), converges to the distribution of the posterior mean estimate of the single-user Gaussian channel
\[
y = \sqrt{\gamma}b^k + \frac{1}{\sqrt{\eta}} z
\]
where \(z \sim \mathcal{N}(0, 1)\), and \(\eta\), the multiuser efficiency \[1\], is the solution of the following fixed-point equation:
\[
\eta^{-1} = 1 + \beta \gamma \text{MMSE}(\gamma, \eta, \alpha).
\]
If (8) admits more than one solution, we must choose the one minimizing the free energy function
\[
\mathcal{F} = -\mathbb{E}\left[\int p(y | b^k) \ln p(y | b^k) dy\right] - \frac{1}{2} \ln 2\pi e + \frac{1}{2\beta} \left(\eta \ln \eta + \frac{2\pi e}{\eta}\right).
\]
In (8)-(9), \(p(y | b^k)\) is the transition probability of the large-system equivalent single-user Gaussian channel described by (7), and
\[
\text{MMSE}(\gamma, \eta, \alpha) \triangleq \mathbb{E}\left[(b^k - \hat{b}^k)^2\right]
\]
 denotes the minimum mean-square error in estimating \(b^k\) in Gaussian noise with amplitude equal to \(\sqrt{\gamma}\), where \(\hat{b}^k = \mathbb{E}[b^k|y]\) is the posterior mean estimate, known to minimize the MMSE \[11\].

III. LARGE-SYSTEM MULTIUSER EFFICIENCY

We start by shaping our problem into the statistical-physics framework \[6\], \[7\] and invoke the decoupling principle (Claim 2.1) in the multiuser system (1), so as to use its single-user characterization. By doing this, the system is converted into the scalar Gaussian channel (7), where the input distribution for an arbitrary user \(k\) takes values \(X = \{-1, 0, +1\}\) with probabilities \(\alpha = \frac{1}{2}\) and \(\alpha = \frac{1}{2}\), respectively, the signal amplitudes from matrix \(A\) are \(a = \sqrt{\gamma}\), where \(\gamma\) is the SNR per active user (we shall call it SNR), and the inverse noise variance is equal to the multiuser efficiency \(\eta\). Hence, \(\eta\) is the solution of the fixed-point equation (8) that minimizes (9), where the MMSE is given by (10). By applying Claim 2.1 \[7\], the fixed-point equation of the user-and-data detector can be written as
\[
\eta = \frac{1}{1 + \beta \gamma \left(\alpha - \alpha^2 \frac{\alpha^2 \sinh(\pi (\gamma - \gamma \sqrt{\eta})/(1 - \alpha \gamma) + (1 - \alpha) e^{\gamma/2})}{\alpha \cosh(\pi (\gamma - \gamma \sqrt{\eta})/(1 - \alpha \gamma) + (1 - \alpha) e^{\gamma/2})}\right)}
\]
Numerical solutions of (11), showing \(\eta\) vs. SNR at a load \(\beta = 3/7\) are shown in Fig. 1. For low SNR, noise dominates, and the performance of the MMSE and the multiuser efficiency

\(^1\)The analysis presented in this paper can be easily extended to different statistics of the \(a_k\) coefficients, like for example those induced by Rayleigh fading.
is degraded as $\alpha$ grows, since the presence of more active users adds more noise to the system. On the other hand, for high SNR the MMSE strongly depends on the minimum distance between the transmitted symbols, and the activity rate here plays a secondary role. Hence, the gap between the multiuser efficiencies with $\alpha = 1$ and $\alpha < 1$ for larger SNR is due to the fact that the former constellation has twice the minimum distance of the latter.

For the whole range of activity rates, i.e., $\alpha \in [0, 1]$, we can derive lower and upper bounds illustrating analytically the transition between the classical assumption ($\alpha = 1$) and the introduction of activity detection ($\alpha < 1$) for large SNR. Our calculations bring about a new analytical framework for large-system analysis, as we will see in next section. The most general result is stated as follows:

**Theorem 3.1.** The MMSE of joint user identification and data detection in a large system with an unknown number of users has the following behavior, valid for large values of the product $\eta \gamma$:

$$
\kappa_1(\gamma, \eta, \alpha) \leq \text{MMSE}(\gamma, \eta, \alpha) \leq \kappa_2(\gamma, \eta, \alpha)
$$

where $\kappa_1(\gamma, \eta, \alpha) \triangleq 2 \sqrt{\frac{\alpha(1-\alpha)}{\pi \eta \gamma}} e^{-\eta \gamma/8}$ and $\kappa_2(\gamma, \eta, \alpha) \triangleq 2\alpha e^{-\eta \gamma/2} + \sqrt{\frac{\pi \alpha(1-\alpha)}{\eta \gamma}} e^{-\eta \gamma/8}$.

Both bounds describe explicitly, in the high-SNR region, the relationship among the MMSE, the users’ activity rate, and the SNR. Furthermore, we can observe that, for sufficiently large SNR, the behavior vs. $\alpha$ of the optimal detector is symmetric with respect to $\alpha = 1/2$, which corresponds to the maximum uncertainty on the activity rate. Fig. 2 illustrates the bounds and the aforementioned fact for SNR = 20 dB. The MMSE essentially depends on the minimum distance between the inactivity “symbol” $\{0\}$ and the data symbols $\{-1, 1\}$, and thus users’ identification prevails over data detection. Summarizing, the dependence of the MMSE must be symmetrical with respect to $\alpha = 1/2$, since it reflects the impact of prior knowledge on the user’s activity into the estimation.

**IV. SYSTEM LOAD AND RELATED CONSIDERATIONS**

Recall the definition of maximum system load $\beta = \frac{K}{N}$, where $K$ is the maximum number of users accessing the multiuser channel. When the number of active users is unknown, and there is a priori knowledge of the activity rate, the actual system load is $\beta' = \alpha \beta$. In this section, we focus on $\beta$ and study some of its properties. Notice that, given an activity rate, results on the actual system load follow trivially.

**A. Solutions of the large-system fixed-point equation**

We characterize the behavior of the system load subject to quality-of-service constraints. Doing this helps shedding light into the nature of the solutions of the fixed-point equation (11). In particular, there might be cases where (11) has multiple solutions. These solutions correspond to those appearing in any simple mathematical model of magnetism based on the evaluation of the free energy with the fixed-point method [8]. In the case of a multiuser detection problem, the multiuser efficiencies for the IO detector might vary significantly depending on the value of the system load and SNR. More specifically, as the SNR grows large, stable solutions may switch between two regions: one where the multiuser efficiency tends to 1, and thus, the single-user performance is approached, and another one, where the multiuser efficiency tends to 0. Following previous literature [6], we shall call the former solutions good and the latter bad. Thus, in case of multiplicity, the solution that minimizes the free-energy might belong to the good or to the bad region. When the solution is unique, the multiuser efficiency is trivially a globally stable solution in any of these two regions.

**B. System load and the space of fixed-point solutions**

Even in the case of good solutions, the multiuser efficiency can be greatly degraded by the joint effect of the activity rate and the system load. In order to analyze (11) from a different perspective and shed light into the interplay between these parameters, we express the maximum system load as the
following function, finding $\beta$ in (8):

$$\Upsilon_{\beta}(\gamma, \eta, \alpha) \triangleq \frac{(1 - \eta)}{\eta \gamma \text{MMSE}(\gamma, \eta, \alpha)}$$

(13)

It is easy to observe that, for small values of $\eta$, $\Upsilon_{\beta}(\gamma, \ldots, \alpha)$ grows to infinity, whereas in the high-$\eta$ region, which is of interest here, it decays to 0. Before analyzing the behavior of (13), we introduce a few definitions that help describe the boundaries between the regions with and without coexistence (in the statistical-physics literature, these boundaries are called spinodal lines [6]). We also define approximately the regions of potentially stable solutions as introduced before.

**Definition 4.1:** The critical system load $\beta^*(\gamma, \alpha)$ is the maximum load up to which a stable good solution of (11) exists.

**Definition 4.2:** The transition system load $\beta_*(\gamma, \alpha)$ is the load at which a single good solution of (11), $\eta_*$, starts to coexist with other solutions $\eta'$.

**Definition 4.3:** The good-solution region corresponds to the domain of (13) formed by the maximum $\eta$ in every set of pre-images of $\Upsilon_{\beta}$ below the critical system load:

$$\mathcal{R}_g = \left\{ \eta \in [0, 1], \eta = \max\{\Upsilon_{\beta}^{-1}(\beta)\}, \forall \beta \in [0, \beta^*) \right\}$$

(14)

Similarly, the bad-solution region corresponds to the domain of (13) formed by the minimum $\eta$ in every set of pre-images of $\Upsilon_{\beta}$ above the transition system load:

$$\mathcal{R}_b = \left\{ \eta \in [0, 1], \eta = \min\{\Upsilon_{\beta}^{-1}(\beta)\}, \forall \beta \in [\beta_*, +\infty) \right\}$$

(15)

![Figure 3. System load function in the multiuser efficiency domain for $\alpha = 0.5$ and $\gamma = 18$dB.](image)

Fig. 3 illustrates $\Upsilon_{\beta}$ (for fixed SNR and activity rate) and shows the regions defined by the aforementioned parameters. It is important to remark that both system loads defined above delimit the regions from which there is phase coexistence ($\beta_* \leq \beta \leq \beta^*$) from the areas where there is one solution ($\beta > \beta^*$ or $\beta < \beta_*$). Additionally, Fig. 3 illustrates the set of solutions that satisfy conditions (14) and (15).

We are interested in describing the domains where good and bad solutions of (11) can be found and show their relationship with the system parameters ($\beta$, $\gamma$ and $\alpha$). In case of multiple solutions the multiuser detector is not guaranteed to achieve the good solution [6]. Hence, our last goal is to give an explicit analytical representation of the single-good-solution (also called computationally feasible) region.

We focus here only on $\alpha \in (0, 1)$. In order to analyze the conditions on the system load, SNR, and activity rate, under which we can find a good solution, we use the asymptotic results on the MMSE, yielding lower and upper bounds

$$L(\ldots) \leq \Upsilon_{\beta}(\gamma, \eta, \alpha) \leq U(\ldots),$$

where

$$L(\gamma, \eta, \alpha) \triangleq \frac{(1 - \eta)}{\eta \gamma^{(1-\eta)} e^{\eta^2/8}}$$

and

$$U(\gamma, \eta, \alpha) \triangleq \frac{(1 - \eta)}{2\sqrt{\frac{\pi}{\eta \gamma(1-\eta)}} e^{\eta^2/8}}.$$

Although not exact for low SNR, the dependence on $\eta$ of the upper and lower bound provides a good approximation for the dependence of $\Upsilon_{\beta}(\ldots)$ for large SNR and given $\alpha$. Hence, using $U(\ldots)$ and $L(\ldots)$, we obtain necessary and sufficient conditions that determine the regions of stable solutions and provide analytical expressions for the transition and critical system loads. The main result for $\alpha \in (0, 1)$ follows:

**Theorem 4.4:** Given the range of activity rates $\alpha \in (0, 1)$, a necessary condition for phase coexistence is

$$\gamma \geq 4(3 + 2\sqrt{2}), \quad (\gamma \geq 13.67 \text{ dB})$$

(16)

Moreover, for large enough $\gamma$, the condition is met and the transition system load is bounded by

$$L(\gamma, \eta_m, \alpha) < \beta_*(\gamma, \alpha) < U(\gamma, \eta_m, \alpha)$$

(17)

while the critical system load is bounded by

$$L(\gamma, \eta_M, \alpha) < \beta^*(\gamma, \alpha) < U(\gamma, \eta_M, \alpha)$$

(18)

and $\eta_m$, $\eta_M$ are given by $\eta_m \triangleq (\gamma/2 - 2 - 4\Delta(\gamma))/\gamma$, $\eta_M \triangleq (\gamma/2 - 2 + 4\Delta(\gamma))/\gamma$, where $\Delta(\gamma) = \sqrt{(\gamma/8)^2 - 3\gamma/8 + 1/4}$. Hence, the bad-solution region is given by $\mathcal{R}_b = [0, \eta_m]$, whereas the good-solution region is $\mathcal{R}_g = [\eta_M, 1]$. Similarly, the (computationally feasible) subregions of single bad solutions, that we shall denote $\mathcal{R}_{bc} = [\eta_{bc}, \eta_m] \subset \mathcal{R}_b$, and single good solutions, denoted by $\mathcal{R}_{gc} = [\eta_{gc}, 1] \subset \mathcal{R}_g$, satisfy the following bounds:

$$\eta_{bc} = \min\{\Upsilon_{\beta}^{-1}(\beta^*)\} > \eta^*_bc,$$

$$\eta_{gc} = \max\{\Upsilon_{\beta}^{-1}(\beta_*)\} < \eta^*_gc,$$

where $\eta^*_bc \triangleq \min\{U^{-1}(\beta^*)\}$, and $\eta^*_gc \triangleq \max\{L^{-1}(\beta^*)\}$ are obtained from the bounds.

The above result provides the general boundaries of the space of solutions of our problem. It is important to note that $\eta_m$ and $\eta_M$ are very good approximations for high SNR of the positions of the minimum and maximum observed in Fig. 3, which determine transition and critical system loads. Similarly, $\eta^*_bc$ and $\eta^*_gc$ are tight bounds of the boundaries of the single-solution regions as $U(\ldots)$ and $L(\ldots)$ are of $\Upsilon_{\beta}(\ldots)$.

Remark also that the activity rate affects the transition and
critical system load, and the subregions $\mathcal{R}_{ge}$ and $\mathcal{R}_{bc}$ in the same symmetrical manner as it does the MMSE (i.e., the worst case here also corresponds to $\alpha = 0.5$) but has no impact on the whole regions of good $\mathcal{R}_g$ and bad solutions $\mathcal{R}_b$, that are only reduced in size by increasing SNR. In particular, these regions are characterized, in the limit of high SNR, as follows:

Corollary 4.5: In the limit of high SNR, $\mathcal{R}_g \rightarrow \{1\}$, and $\mathcal{R}_b \rightarrow \{0\}$ and consequently $\mathcal{R}_{ge} \rightarrow \{1\}$, and $\mathcal{R}_{bc} \rightarrow \{0\}$.

Note that, given a system load $\beta$ with $\beta < \beta^*$, for sufficiently large SNR the unique true (large-system) solution is $\eta = 1$, which corroborates the main result in [12]. Moreover, the description of the computationally feasible solutions by analytical means allows the derivation of a sufficient condition on the system load to guarantee a given (single-good) multiuser efficiency in practical implementations. More specifically, we use the aforementioned lower bound on $\gamma$ to state that any system load below $L(\ldots)$ guarantees that the requirement is met.

Corollary 4.6: The maximum system load, $\beta_{\alpha, \eta}$, for a given activity rate and multiuser efficiency requirement in the computationally feasible region, $\eta = 1 - \epsilon$ (0 < $\epsilon$ < 1) such that $\eta \geq \eta_{ge}$, is lower-bounded in the high-SNR region by:

$$\beta_{\alpha, \eta} \geq \frac{\epsilon}{\sqrt{\pi \gamma \alpha (1 - \alpha)}} e^{(1-\epsilon)/\gamma/8} \quad (19)$$

In Fig. 4 we show the transition and the critical system load as function of the SNR in a ($\gamma, \beta$) space. We also use the asymptotic expansion to derive upper and lower bounds, respectively. The curves plotted are the spinodal lines, and mark the boundary between the regions with and without solution coexistence. The $\beta_\star$ (lower branch) separates the region where the good solution disappears, whereas $\beta^\star$ (upper branch) contains the bifurcation points at which the good solution disappears. The intersection point between both branches corresponds to the SNR threshold (16), which provides the necessary condition for solution coexistence.

V. CONCLUSIONS

We have analyzed the worst-case performance of a MAP multiuser detector for CDMA, where the evolution of active users accessing the network degenerates into an independent process and activity and data must be jointly estimated at each symbol time. By applying the large-system fixed-point equation to a ternary and activity-dependent constellation, we derive explicit bounds on the MMSE for high SNR. This leads to a novel high-SNR analysis of the space of multiuser efficiency solutions where the critical system loads as well as the feasible-solution regions can be well approximated by analytical means. Our results show that for the ternary constellation assumed in the model, the boundaries of the space of solutions and more remarkably, the computationally feasible subregions, do depend on the activity rate, whereas the general regions are only affected by the SNR. In applications with high quality-of-service demands, we are interested in a critical system load as large as possible, keeping the optimal detector in the good subregion of computationally feasible multiuser efficiencies, so that a wider range of potential users can access successfully the channel with a given rate. The above analysis can be easily extended to more general non-uniform and parameter-dependent constellations.

VI. ACKNOWLEDGMENTS

The work of Ezio Biglieri was supported by the Project CONSOLIDER-INGENIO 2010 CSD2008-00010 "COMONSENS".

REFERENCES


Figure 4. Upper and lower bounds on the numerical spinodal lines (thicker line) for $\alpha = 0.5$. 