EZIO BIGLIERI (M'73 – SM'82 – F'89) was born in Aosta (Italy). He studied Electrical Engineering at Politecnico di Torino (Italy), where he received his Dr. Engr. degree in 1967. He is presently a Professor at Universitat Pompeu Fabra, Barcelona, Spain, and an Adjunct Professor of Electrical Engineering at University of California Los Angeles (UCLA). Previously he was a Professor at the University of Napoli (Italy), at Politecnico di Torino, and at UCLA.

He has held visiting positions with the Department of System Science, UCLA, the Mathematical Research Center, Bell Laboratories, Murray Hill, NJ, the Bell Laboratories, Holmdel, NJ, the Department of Electrical Engineering, UCLA, the Telecommunication Department of The Ecole Nationale Supérieure des Télécommunications, Paris, France, the University of Sydney, Australia, the Yokohama National University, Japan, the Electrical Engineering Department of Princeton University, the University of South Australia, Adelaide, the University of Melbourne, the Institute for Communications Engineering, Munich Institute of Technology, Germany, and the Institute for Infocomm Research, National University of Singapore.

He was elected three times to the Board of Governors of the IEEE Information Theory Society, and he served as its President in 1999. He is now the Editor-in-Chief of the *IEEE Transactions on Information Theory* and of the *Journal of Communications and Netowrks*.

Among other honors, in 2000 he received the IEEE Third-Millennium Medal and the IEEE Donald G. Fink Prize Paper Award, in 2001 the IEEE Communications Society Edwin Howard Armstrong Achievement Award and a Best Paper Award from WPMC'01, Aalborg, Denmark, and in 2004 the Journal of Communications and Networks Best Paper Award.

MARCO LOPS was born and educated in Napoli (Italy). He got his "Laurea" and his PhD degrees, both in Electronic and Computer Engineering, from the "Federico II" University. After a short experience at Selenia (currently Selex) as a Radar Engineer, he joined the Department of Electronic and Telecommunications Engineering at "Federico II", first as an assistant professor, then as an associate professor. Since 2000 he has been a professor at DAEIMI (University of Cassino). Marco Lops spent sabbatical periods at University of Connecticut, Rice University and Princeton University. His research interests are in communications, in radar detection theory and, more recently, in tracking.

Multiuser detection in a dynamic environment Part I: User identification and data detection

Ezio Biglieri, *Fellow, IEEE,* and Marco Lops, *Senior Member, IEEE*

***Abstract*— In random-access communication systems, the number of active users varies with time, and has considerable bearing on receiver's performance. Thus, techniques aimed at identifying not only the information transmitted, but also that number, play a central role in those systems. An example of application of these techniques can be found in multiuser detection (MUD). In typical MUD analyses, receivers are based on the assumption that the number of active users is constant and known at the receiver, and coincides with the maximum number of users entitled to access the system. This assumption is often overly pessimistic, since many users might be inactive at any given time, and detection under the assumption of a number of users larger than the real one may impair performance.**

The main goal of this paper is to introduce a general approach to the problem of identifying active users and estimating their parameters and data in a random-access system where users are continuously entering and leaving the system. The tool whose use we advocate is Random-Set Theory: applying this, we derive optimum receivers in an environment where the set of transmitters comprises an unknown number of elements. In addition, we can derive Bayesian-filter equations which describe the evolution with time of the a posteriori probability density of the unknown user parameters, and use this density to derive optimum detectors. In this paper we restrict ourselves to interferer identification and data detection, while in a companion paper we shall examine the more complex problem of estimating users' parameters.

***Index Terms*— Multiuser detection, Random set theory, Bayesian recursions.**

I. INTRODUCTION

In typical random-access communication systems, the number of active users, their location, as well as the parameters that characterize their channel state, vary with time. Thus, techniques aimed at identifying not only the data transmitted, but also the user parameters, play a central role in analysis and design of those systems. Examples of application of these techniques can be found in multiuser detection (MUD), spatial multiplex schemes, and ad hoc networks.

In MUD, it has long been recognized that one of the important issues is that the set of active users at any time may not be known to the receiver. For conventional (matched-filter) receivers, this dearth of information does not affect performance, while for other receivers the simplistic assumption that all users are active will cause significant degradation. [1] In addition, certain detectors based on interference cancellation must know the strongest active users in order to perform satisfactorily [7]. Moreover, as observed in [25], "identification of active users will help the system to promptly process requests and efficiently allocate channels. In such a way, system capacity can be increased."

The problem of detecting active users in a multiuser system has been addressed by several authors in a CDMA context [4], [7], [16]–[18], [24]. Typically, the resulting multiuser receiver is a combination of two separate modules, namely, the active-user identifier and the multiuser detector. The treatment in [7], [16], [17] focuses on the problem of detecting a single user entering or leaving the system. Ref. [24] advocates a subspace-based method (MUSIC algorithm) for identifying the active users, under the assumption that the receiver knows the pool of all possible spreading codes that may be used in transmission. In [10], the active-user-identification algorithm is subspace-based, as in [24]; the receiver is not interested in decoding all active users, but only those transmitting a message to it. Ref. [25] addresses the problem of estimating the number of active users when synchronous and asynchronous users co-exist in the system.

Further performance improvement can be expected if the receiver can exploit a priori information about users entering and exiting the system. The knowledge of a traffic model is exploited in [5], whose authors use it to improve the detection of active users. They model bursty traffic for an individual source as a two-state Markov chain.

Other systems in which user identification is necessary include spatial multiplexing schemes, where the total system throughput can be optimized by properly selecting a subset of users to which the power is allocated [11], [26]. Thus, optimum power-control strategy requires the identification of this best set of users. In ad hoc networks, optimal transmission strategies require the identification and localization of active nodes in the neighborhood of the transmitter.

A. *In this paper...*

In this paper, and in its companion [2], we examine the general problem of detecting data and parameters of a set of users whose number is unknown. Unlike previous work done in this area, which advocates a two-stage receiver, we focus on

Manuscript received XX, XXXX; revised YY, YYYY.

Ezio Biglieri is with Universitat Pompeu Fabra, Barcelona, Spain. email: ezio.biglieri@upf.edu This work was supported by the STREP project No. IST-026905 (MASCOT) within the 6th framework program of the European Commission, and by the Spanish Ministry of Education and Science under Project TEC2006-01428/TCM.

Marco Lops is with DAEIMI, Università di Cassino, Italy. email: lops@unicas.it

[1] See, e.g., [5], [16], [24]. As an example, if a decorrelator detector [21, Chapter 5] does more nulling than needed, its performance is impaired. Ref. [8] describes a case where a multiuser detector suffers from catastrophic error if a new user becomes active.

optimal design of a receiver which estimates simultaneously the number of users and their parameters and data. The receiver performance can be enhanced if one uses the a priori information offered by a traffic model, i.e., a dynamic model for the number of active users and for their parameters. In fact, information from the past history of the parameters may bring a considerable amount of extra information if their changes are not overly abrupt (for example, if the number of active users does not change considerably from frame to frame).

Having to deal with random sets, i.e., with sets comprising a random number of random vectors (those including what is unknown about each user), a tool that can be used is *Random Set Theory* (RST: see the Appendix and the references therein). RST, recently applied in the context of multitarget tracking and identification (see, e.g., [6], [12]–[15], [22], [23]) is based on a probability theory of finite sets that exhibit randomness not only in each element, but also in the number of elements. RST (and in particular its formulation, referred to as Finite Set Statistics, or FISST [12]–[14], specifically tailored to problems in whose class lie those we are considering in this paper) develops concepts which are not part of conventional probability theory. In fact, a central point in FISST is the generation of "densities" which are not the usual Radon-Nikodým derivatives of probability measures, but rather "set derivatives" of nonadditive "belief functions." On the other hand, these densities, which capture what is known about measurement state space, users state space, and users dynamics, can be derived in a rather straightforward way from the system model by using the FISST toolbox. RST is a tool that has considerable generality and flexibility, is consistent with engineering intuition, and is easy to use.[2]

To illustrate application of RST to random-access communication, we focus on MUD problems, and derive Bayesian-filter equations describing the evolution with time of the a posteriori probability density of the unknown user parameters and data. Specifically, here we restrict ourselves to interferer identification and data detection, while in a companion paper [2] we examine the problem of estimating users' parameters as well. We hasten to claim that the applications considered here do not exhaust the potential of RST for the analysis of random-access system: thus, many of the simplifying assumptions are not made because more realistic models cannot be dealt with using our theory, but rather because we do not want to muddle the intrinsic simplicity of the RST tool with marginal details. This paper is organized as follows. Section II describes the channel model, while Section III states the MUD problem in the context of RST. Section IV describes an application to CDMA, while Section V provides some numerical results illustrating the theory.

[2]Advocating the use of RST to solve the problem at hand, we do not imply that it is the only tool that can be used. Actually, standard probability theory could be applied to achieve the same results, although, we argue, in a less elegant and concise way. The companion paper [2], which deals with estimates of random sets of continuous parameters, should be even more convincing about the usefulness of RST.

II. Channel model and statement of the problem

We assume $K+1$ users transmitting digital data over a common random-access channel. Let $s(\mathbf{x}_t^{(0)})$ denote the signal transmitted by the reference user at discrete time t, $t=1,2,\ldots$, and $s(\mathbf{x}_t^{(i)})$, $i=1,\ldots,K$, the signals that may be transmitted at the same time by K interferers. Each signal has in it a number of known parameters, reflected by the deterministic function $s(\,\cdot\,)$, and a number of random parameters, summarized by $\mathbf{x}_t^{(i)}$. The index i reflects the identity of the user, and is typically associated with its signature. The observed signal at time t is a sum of $s(\mathbf{x}_t^{(0)})$, of the signals generated by the users active at time t, which are in a random number, and of stationary random noise $\mathbf{z}_t$. We write

$$\mathbf{y}_t = s(\mathbf{x}_t^{(0)}) + \sum_{\mathbf{x}_t^{(i)}\in\mathbf{X}_t} s(\mathbf{x}_t^{(i)}) + \mathbf{z}_t \tag{1}$$

where $\mathbf{X}_t$ is a random set, encapsulating what is unknown about the active users. The notations of (1) implicitly assume that user 0 is active with probability 1 and its parameters (but not its data) are known (this restriction can be easily removed).

To motivate the development presented in this paper, and in particular our use of RST, we proceed to formulate the general problem through three intermediate steps. Specifically, we examine three scenarios of increasing complexity, under the assumptions that the users' parameters are all known, that the number of interferers is random and unknown to the receiver, and that we are interested in detecting the data transmitted by the reference user:

① *The receiver has no information about the a priori probabilities that the individual interferers are active.* Two options we may consider here are maximum-likelihood (ML) detection of the reference user's data under the assumption that all potential interferers are active, or joint ML detection of the number of active interferers and of the reference user's data. Consider binary transmission for simplicity. In the first case, detection implies choosing among 2×2^K hypotheses. In the second case, the choice is among 2×3^K hypotheses, as every interferer may transmit one between two binary symbols, or be inactive. The difference in performance between the two situations is illustrated in Fig. 1, which compares the two detectors described above. The ordinate shows the bit error probability of the reference user in a multiuser system with 2 independent interferers transmitting binary antipodal signals over an additive white Gaussian noise (AWGN) channel with the same a priori probability of activity α, spreading-sequences consisting of Kasami sequences with length 15 [9, p. 240], and perfect power control. The single-user bound is also shown as a reference. It is seen that RST yields a detector much more robust than classic MUD to variations in the users activity factor. We also observe that classic MUD can outperform RST for high values of α, as this situation corresponds to its having reliable side information about the number of active users.

② *The receiver knows the a priori probabilities that the individual interferers are active.* System performance can

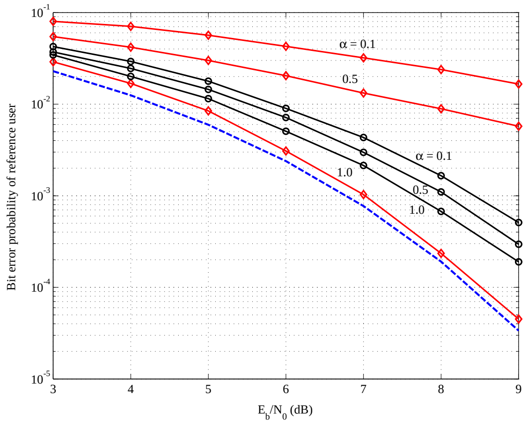

Fig. 1. *Bit error probability of the reference user in a multiuser system with 2 interferers, independently active with probability α. Lines with diamond markers: Classic multiuser ML detection, assuming that all users are active. Lines with circle markers: ML detection of data and interferers number. Dashed curve: Single-user bound.*

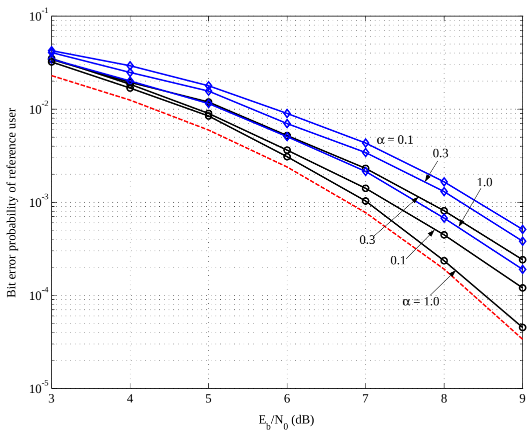

Fig. 2. *Bit error probability of the reference user in a multiuser system with 2 interferers, independently active with probability α. Lines with diamond markers: ML detection of active users and data. Line with circle markers: MAP detection using a priori knowledge of the value of α. Dashed curve: Single-user bound.*

be further improved if the receiver is able to exploit additional side information in the form of a priori probabilities of user activity. By assuming that the activity factor α is known, maximum a posteriori (MAP) detection yields the results shown in Fig. 2.

③ *The receiver has a dynamic model of users' activity.* The receiver performance can be further improved by using additional information about the interferers, in the form of a model of their dynamic behavior. This information can be generated once a model of users' mobility is available.

Observe again that the information carried by the interferers is contained in the set

$$\mathbf{X}_t = \{\mathbf{x}_t^{(1)}, \ldots, \mathbf{x}_t^{(k)}\}$$

whose elements are random vectors, and k is itself a random integer. RST develops a probability theory on random sets of this form, which are modeled as single entities. Roughly speaking, a random set is a map $\mathbf{X}$ between a sample space and a family of subsets of a space $\mathbb{S}$. This is the space of the unknown data and parameters of the active interferers. For example, if everything about the interferers is known, except for their number and identity, then $\mathbf{X}$ takes values in the power set $2^{\mathbb{K}}$, where $\mathbb{K} \triangleq \{1, \ldots, K\}$. We may also consider a situation in which one or more parameters (the interferer power, etc.) are also unknown in addition to the interferers' number and identities, while the transmitted data are known (for example, in a training phase). In mathematical terms, $\mathbb{S}$ is generally a *hybrid space* $\mathbb{S} \triangleq \mathbb{R}^d \times U$, with U a finite discrete set, and d the number of parameters to be estimated for each user. In the remainder of this paper we shall restrict ourselves to the case $d = 0$, and leave to a companion paper [2] the discussion of the case $d \neq 0$.

With channel model (1), the receiver detects only a superposition of interfering signals. Thus, the random set describing the receiver, denoted $\mathbf{Y}_t$, is the singleton $\{\mathbf{y}_t\}$, where $\mathbf{y}_t$ has conditional probability density function

$$f_{\mathbf{Y}_t|\mathbf{X}_t}(\mathbf{y}_t \mid \mathbf{B}) = f_{\mathbf{z}}(\mathbf{y}_t - \sigma(\mathbf{B})) \tag{2}$$

where $\mathbf{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ is a realization of $\mathbf{X}_t$, that is, a realization of a random set of users and their parameters/data. Moreover, $f_{\mathbf{z}}(\cdot)$ is the probability density function (pdf) of the additive noise, and

$$\sigma(\mathbf{B}) \triangleq \sum_{\mathbf{b}_i \in \mathbf{B}} s(\mathbf{b}_i) \tag{3}$$

A. *Defining estimators*

Development of estimators with our model must take into account the peculiarities of RST. For example, expectations cannot be defined, because there is no notion of set addition, and hence estimators based on a posteriori expectations do not exist (this point is discussed thoroughly and eloquently in [12]). A possible estimator maximizes the a posteriori probability (APP) of $\mathbf{X}_t$ given $\mathbf{y}_{1:T}$, the latter denoting the whole set of observations corresponding to a data frame transmitted from $t = 1$ to $t = T$. Another possibility is to restrict oneself to a *causal* estimator, which searches for the maximum probability of $\mathbf{X}_t$ given $\mathbf{y}_{1:t}$. In a delay-constrained system, one may estimate $\mathbf{X}_t$ on the basis of the observations $\mathbf{y}_{t-\Delta:t+\Delta}$, with Δ a fixed interval duration (sliding-window estimator).

B. *Consideration of a dynamic environment*

Since $\{\mathbf{X}_t\}_{t=1}^{\infty}$ forms a random set sequence, the statistical characterization of $\mathbf{X}_t$ is needed for all discrete time instants t. If a dynamic model of the transmission system is available (which is what we assume in this paper), then the APPs can be updated recursively, thus allowing one to take advantage of the information gathered from the past evolution of the system. We observe in passing that the concept of an adaptive receiver was examined previously by several authors (see, e.g., [20] and references therein), while the effects on analysis of a dynamic model were touched upon, among others, by the authors of [3], [7], [16], [17].

We make the assumption that $\{\mathbf{X}_t\}_{t=1}^{\infty}$ forms a Markov set sequence, i.e., that $\mathbf{X}_t$ depends on its past only through $\mathbf{X}_{t-1}$.

This allows us to use *Bayesian-filter* recursions for $\widehat{\mathbf{X}}_t$ [13]:

$$f_{\mathbf{X}_{t+1}|\mathbf{Y}_{1:t}}(\mathbf{B} \mid \mathbf{y}_{1:t}) = \int f_{\mathbf{X}_{t+1}|\mathbf{X}_t}(\mathbf{B} \mid \mathbf{C}) f_{\mathbf{X}_t|\mathbf{Y}_{1:t}}(\mathbf{C} \mid \mathbf{y}_{1:t})\, \delta\mathbf{C} \quad (4)$$

$$f_{\mathbf{X}_{t+1}|\mathbf{Y}_{1:t+1}}(\mathbf{B} \mid \mathbf{y}_{1:t+1}) \propto f_{\mathbf{Y}_{t+1}|\mathbf{X}_{t+1}}(\mathbf{y}_{t+1} \mid \mathbf{B}) f_{\mathbf{X}_{t+1}|\mathbf{Y}_{1:t}}(\mathbf{B} \mid \mathbf{y}_{1:t}) \quad (5)$$

The integrals appearing in the equations are *set integrals*, defined in the sense of RST (see the Appendix). The notation δ for the differential reflects this definition.

Thus, the causal maximum-a-posteriori estimate of $\mathbf{X}_t$ is obtained by maximizing, over $\mathbf{B}$, the APP $f_{\mathbf{X}_t|\mathbf{Y}_{1:t}}(\mathbf{B} \mid \mathbf{y}_{1:t})$, which is tantamount to minimizing

$$m(\mathbf{B}) \triangleq (\mathbf{y}_t - \sigma(\mathbf{B}))^2 - \varepsilon(\mathbf{B})$$

where $\varepsilon(\mathbf{B}) \triangleq N_0 \ln f_{\mathbf{X}_t|\mathbf{Y}_{1:t-1}}(\mathbf{B} \mid \mathbf{y}_{1:t-1})$. The first term in the RHS of definition above is the Euclidean distance between the observation and the sum of the interfering signals at time t. Its minimization yields ML estimates of $\mathbf{X}_t$. The second term in the RHS, generated by the uppermost step of iterations, reflects the influence on $\mathbf{X}_t$ of its past history, and its consideration yields MAP estimates.

The Bayesian-filter recipe (4)-(5) requires two ingredients. The first one is the channel model, through the pdf $f_{\mathbf{z}}(\mathbf{y}_t \mid \mathbf{X}_t)$. For example, assuming real signals, and the noise to be Gaussian with mean 0 and known variance $N_0/2$, we have

$$f_{\mathbf{z}}(\mathbf{y}_t \mid \mathbf{X}_t) \propto \exp\{-(\mathbf{y}_t - \sigma(\mathbf{X}_t))^2/N_0\}$$

The second ingredient is the dynamic model of the random set sequence $\mathbf{X}_t$, described by the function $f_{\mathbf{X}_{t+1}|\mathbf{X}_t}(\cdot \mid \cdot)$ that describes the time evolution of data and parameters of the system. Examples of this modeling procedure are available for the problem of tracking multiple targets [13], [22].

From now on we restrict ourselves to the detection of the number and identity of active interferers, and of the data they carry, under the assumption that the remaining parameters, which were previously estimated by the receiver in a training phase, do not change in any appreciable way during the tracking phase. Estimation of these parameters using RST is described in the companion paper [2].

III. Detection

A. Active users

We assume first that we are only interested in detecting which interferers, out of a universe of K potential system users, are present at time t. This information may be used for example to do decorrelation detection, under the assumption that the signatures of all users are known at the receiver. In our theory, $\mathbf{X}_t$ takes values in $2^{\mathbb{K}}$. Since this set is finite, a probability measure for $\mathbf{X}_t$ can be defined by assigning all probabilities $\mathbb{P}(\mathbf{A})$, $\mathbf{A} \in 2^{\mathbb{K}}$.

1) Static model: At any fixed time t, suppose that the probability of interferer $\mathbf{x}_t^{(i)}$ to be active is α, independent of t and i. In this case the probability of the interferer set $\mathbf{X}_t$ depends only on its cardinality $|\mathbf{X}_t|$, and we can write

$$f_{\mathbf{X}_t}(\mathbf{B}) = \alpha^{|\mathbf{B}|}(1-\alpha)^{K-|\mathbf{B}|} \quad (6)$$

To derive this result, we use RST by first computing the belief function

$$\begin{aligned} \beta_{\mathbf{X}}(\mathbf{S}) &\triangleq \mathbb{P}(\mathbf{X} \subseteq \mathbf{S}) \\ &= \sum_{j=0}^{|\mathbf{S}|} \sum_{\mathbf{B}:\mathbf{B}\subseteq\mathbf{S}\ \&\ |\mathbf{B}|=j} \mathbb{P}(\mathbf{X} = \mathbf{B}) \\ &= \sum_{j=0}^{|\mathbf{S}|} \binom{|\mathbf{S}|}{j} \alpha^j (1-\alpha)^{K-j} \end{aligned} \quad (7)$$

and subsequently computing its set derivative, which, in the discrete case, becomes the Möbius inversion formula (50).

2) Dynamic model: Consider now the evolution of $\mathbf{X}_t$. We assume that from $t-1$ to t some new users become active, while some old users become inactive. We write

$$\mathbf{X}_t = \mathbf{S}_t \cup \mathbf{N}_t \quad (8)$$

where $\mathbf{S}_t$ is the set of *surviving* users still active from $t-1$, and $\mathbf{N}_t$ is the set of *new* users becoming active at t. The condition $\mathbf{N}_t \cap \mathbf{X}_{t-1} = \emptyset$ is forced, because a user ceasing transmission at time $t-1$ cannot reënter the set of active users at time t. We proceed by constructing separate dynamic models for $\mathbf{S}_t$ and $\mathbf{N}_t$, which will be eventually combined to yield a model for $\mathbf{X}_t$.

Consider first $\mathbf{S}_t$. Suppose that there are k active users at $t-1$, the elements of the random set $\mathbf{X}_{t-1} = \{\mathbf{x}_{t-1}^{(1)}, \dots, \mathbf{x}_{t-1}^{(k)}\}$. Then we may write, for the set of surviving users,

$$\mathbf{S}_t = \bigcup_{i=1}^{k} \mathbf{X}_t^{(i)} \quad (9)$$

where $\mathbf{X}_t^{(i)}$ denotes either an empty set (if user i has become inactive) or the singleton $\{\mathbf{x}_t^{(i)}\}$ (user i is still active). Let μ denote the "persistence" probability, i.e., the probability that a user survives from $t-1$ to t. We obtain, for the conditional probability of $\mathbf{S}_t$ given that $\mathbf{X}_{t-1} = \mathbf{B}$:

$$f_{\mathbf{S}_t|\mathbf{X}_{t-1}}(\mathbf{C} \mid \mathbf{B}) = \begin{cases} \mu^{|\mathbf{C}|}(1-\mu)^{|\mathbf{B}|-|\mathbf{C}|}, & \mathbf{C} \subseteq \mathbf{B} \\ 0, & \mathbf{C} \nsubseteq \mathbf{B} \end{cases} \quad (10)$$

For new users, a reasonable model has

$$f_{\mathbf{N}_t|\mathbf{X}_{t-1}}(\mathbf{C} \mid \mathbf{B}) = \begin{cases} \alpha^{|\mathbf{C}|}(1-\alpha)^{K-|\mathbf{B}|-|\mathbf{C}|}, & \mathbf{C} \cap \mathbf{B} = \emptyset \\ 0, & \mathbf{C} \cap \mathbf{B} \neq \emptyset \end{cases} \quad (11)$$

Finally, by assuming that births and deaths of users are conditionally independent given $\mathbf{X}_{t-1} = \mathbf{B}$, the conditional pdf of the union of the random sets $\mathbf{S}_t$ and $\mathbf{N}_t$ is obtained from the *generalized convolution* [6]

$$\begin{aligned} &f_{\mathbf{X}_t|\mathbf{X}_{t-1}}(\mathbf{C} \mid \mathbf{B}) \\ &\quad = \sum_{\mathbf{W}\subseteq\mathbf{C}} f_{\mathbf{S}_t|\mathbf{X}_{t-1}}(\mathbf{W} \mid \mathbf{B})\, f_{\mathbf{N}_t|\mathbf{X}_{t-1}}(\mathbf{C} \setminus \mathbf{W} \mid \mathbf{B}) \\ &\quad = f_{\mathbf{S}_t|\mathbf{X}_{t-1}}(\mathbf{C} \cap \mathbf{B}) f_{\mathbf{N}_t|\mathbf{X}_{t-1}}(\mathbf{C} \setminus (\mathbf{C} \cap \mathbf{B})) \end{aligned} \quad (12)$$

3) Bayesian-filter recursions: In our context, recursions (4)-(5) can be implemented as follows. Determine first:

① The a priori probability distribution of $\mathbf{X}_0$ at the beginning of the detection process. Description of this distribution consists of assigning probabilities to all the elements of $2^{\mathbb{K}}$. This can be done for example by assuming independent users with the same stationary activity factor.

② The set of observations $\mathbf{y}_t$, $t = 1, \dots, T$.

③ The conditional pdf's $f_{\mathbf{Y}_t|\mathbf{X}_t}$, depending on the channel model.

④ The "evolution" pdf's $f_{\mathbf{X}_{t+1}|\mathbf{X}_t}$, depending on the dynamic model.

The recursion goes as follows: omitting the subscripts for notational simplicity here, and identifying random sets with their realizations, we have

$$f(\mathbf{X}_1) = \int f(\mathbf{X}_1 \mid \mathbf{X}_0) f(\mathbf{X}_0)\, \delta\mathbf{X}_0$$

With this, we can compute

$$f(\mathbf{X}_1 \mid \mathbf{y}_1) \propto f(\mathbf{y}_1 \mid \mathbf{X}_1)\, f(\mathbf{X}_1)$$

which allows the calculation of the causal MAP estimate $\widehat{\mathbf{X}}_1$. Next, we compute

$$f(\mathbf{X}_2 \mid \mathbf{y}_1) = \int f(\mathbf{X}_2 \mid \mathbf{X}_1)\, f(\mathbf{X}_1 \mid \mathbf{y}_1)\, \delta\mathbf{X}_1$$

and hence

$$f(\mathbf{X}_2 \mid \mathbf{y}_{1:2}) \propto f(\mathbf{y}_2 \mid \mathbf{X}_2)\, f(\mathbf{X}_2 \mid \mathbf{y}_1)$$

which allows the calculation of $\widehat{\mathbf{X}}_2$. The general recursion has, for $t = 2, \dots$:

$$\begin{aligned} f(\mathbf{X}_{t+1} \mid \mathbf{y}_{1:t}) &= \int f(\mathbf{X}_{t+1} \mid \mathbf{X}_t)\, f(\mathbf{X}_t \mid \mathbf{y}_{1:t})\, \delta\mathbf{X}_t \\ f(\mathbf{X}_{t+1} \mid \mathbf{y}_{1:t+1}) &\propto f(\mathbf{y}_{t+1} \mid \mathbf{X}_{t+1})\, f(\mathbf{X}_{t+1} \mid \mathbf{y}_{1:t}) \end{aligned}$$

and, in the case examined in this section,

$$f(\mathbf{X}_{t+1} \mid \mathbf{y}_{1:t}) = \sum_{\mathbf{X}_t \in 2^{\mathbb{K}}} f(\mathbf{X}_{t+1} \mid \mathbf{X}_t)\, f(\mathbf{X}_t \mid \mathbf{y}_{1:t}) \tag{13}$$

$$f(\mathbf{X}_{t+1} \mid \mathbf{y}_{1:t+1}) \propto f_{\mathbf{z}}(\mathbf{y}_{t+1} - \sigma(\mathbf{X}_{t+1}))\, f(\mathbf{X}_{t+1} \mid \mathbf{y}_{1:t}) \tag{14}$$

B. Active users and their data

Assume binary information data, independent from time to time and across users, and a discrete-time unit such that from t to $t+1$ each user transmits N binary symbols. In this case $\mathbf{X}_t$ takes values in a set with

$$\sum_{k=0}^{K} \binom{K}{k} 2^{kN} = (1 + 2^N)^K$$

elements, that we denote $(1 + 2^N)^{\mathbb{K}}$. Eq. (6) becomes

$$f_{\mathbf{X}_t}(\mathbf{B}) = 2^{-N|\mathbf{B}|} \alpha^{|\mathbf{B}|} (1-\alpha)^{K-|\mathbf{B}|} \tag{15}$$

where the new factor $2^{-N|\mathbf{B}|}$ accounts for the fact that there are $N|\mathbf{B}|$ equally likely binary symbols transmitted at time t by $|\mathbf{B}|$ interferers.

Similarly, (10) is transformed into

$$f_{\mathbf{S}_t|\mathbf{X}_{t-1}}(\mathbf{C} \mid \mathbf{B}) = \begin{cases} 2^{-N|\mathbf{C}|} \mu^{|\mathbf{C}|} (1-\mu)^{|\mathbf{B}|-|\mathbf{C}|}, & \mathbf{C} \subseteq \mathbf{B} \\ 0, & \mathbf{C} \nsubseteq \mathbf{B} \end{cases} \tag{16}$$

and (11) into

$$f_{\mathbf{N}_t|\mathbf{X}_{t-1}}(\mathbf{C} \mid \mathbf{B}) = \begin{cases} 2^{-N|\mathbf{C}|} \alpha^{|\mathbf{C}|} (1-\alpha)^{K-|\mathbf{B}|-|\mathbf{C}|}, & \mathbf{C} \cap \mathbf{B} = \emptyset \\ 0, & \mathbf{C} \cap \mathbf{B} \neq \emptyset \end{cases} \tag{17}$$

C. Possible scenarios

We recall that throughout this paper we assume that the only unknown signal quantities may be the identities of the users and their data. Specifically, we may distinguish four cases in our context:

① *Static channel, unknown identities, known data.* This corresponds to a training phase intended at identifying users, and assumes that the user identities do not change during transmission. In this case we write $\mathbf{X}$ in lieu of $\mathbf{X}_t$.

② *Static channel, unknown identities, unknown data.* This may correspond to a tracking phase following ① above. We write again $\mathbf{X}$ in lieu of $\mathbf{X}_t$, and assume that $\mathbf{X}$ contains the whole transmitted data sequence.

③ *Dynamic channel, unknown identities, known data.* This corresponds to identification of users preliminary to data detection (which, for example, may be based on decorrelation).

④ *Dynamic channel, unknown identities, unknown data.* This corresponds to simultaneous user identification and data detection in a time-varying environment.

IV. An Example of Application

Assume now the specific situation of a DS-CDMA system with signature sequences of length L and additive white Gaussian noise. At discrete time t, we may write, for the sufficient statistics of the received signal,

$$\mathbf{y}_t = \mathbf{R}\mathbf{A}\mathbf{b}_t(\mathbf{X}_t) + \mathbf{z}_t, \qquad t = 1, \dots, T \tag{18}$$

where $\mathbf{X}_t$ is now the random set of all active users, $\mathbf{R}$ is the $L \times L$ correlation matrix of the signature sequences (assumed to have unit norm), $\mathbf{A}$ is the diagonal matrix of the users' signal amplitudes, the vector $\mathbf{b}_t(\mathbf{X}_t)$ has nonzero entries in the locations corresponding to the active-user identities described by the components of $\mathbf{X}_t$, and $\mathbf{z}_t \sim \mathcal{N}(0, (N_0/2)\mathbf{R})$ is the noise vector, with $N_0/2$ the power spectral density of the received noise. We further assume $N = 1$, i.e., that at every discrete time instant only one binary antipodal symbol is transmitted.

A. Static channel

The a posteriori probability of $\mathbf{X}$, given the whole received sequence (we omit the time subscript for simplicity), is

$$
\begin{aligned}
&f(\mathbf{X} \mid \mathbf{y}_1, \dots, \mathbf{y}_T) \\
&\quad \propto \quad f_{\mathbf{X}}(\mathbf{X})\, f(\mathbf{y}_1, \dots, \mathbf{y}_T \mid \mathbf{X}) \qquad (19)\\
&\quad = \exp\left\{-\frac{1}{N_0}\sum_{t=1}^{T}\left(\mathbf{y}_t - \mathbf{R}\mathbf{A}\mathbf{b}_t(\mathbf{X})\right)' \mathbf{R}^{-1}\left(\mathbf{y}_t - \mathbf{R}\mathbf{A}\mathbf{b}_t(\mathbf{X})\right)\right\} \\
&\qquad \times f_{\mathbf{X}}(\mathbf{X})
\end{aligned}
$$

Thus, the MAP estimator of users' identities is

$$
\widehat{\mathbf{X}} = \arg\max_{\mathbf{X}\in 2^{\mathbb{K}}} f(\mathbf{X} \mid \mathbf{y}_{1:T}) \tag{20}
$$

where, as usual, $\mathbf{y}_{1:T} \triangleq \mathbf{y}_1, \dots, \mathbf{y}_T$. The MAP estimator of users' identities and data is

$$
\widehat{\mathbf{X}} = \arg\max_{\mathbf{X}} f(\mathbf{X} \mid \mathbf{y}_{1:T}) \tag{21}
$$

where the set of possible realizations of $\mathbf{X}$ includes $(1+2^T)^K$ elements: in fact, in T time interval the number of transmitted binary symbols is $2^{|\mathbf{X}|T}$, and

$$
\sum_{|\mathbf{X}|=0}^{K}\binom{K}{|\mathbf{X}|}2^{|\mathbf{X}|T} = (1+2^T)^K
$$

The expression above can be rewritten in such a way that the presence of the sequence of transmitted data is made more explicit. Specifically, we may write, in lieu of $\mathbf{X}$, the sequence $(\mathbf{X}, \mathbf{b}_1(\mathbf{X}), \dots, \mathbf{b}_T(\mathbf{X}))$. Doing so, we may express the MAP estimator of users' identities and data in the more explicit form

$$
\begin{aligned}
&(\widehat{\mathbf{X}}, \widehat{\mathbf{b}}_1(\mathbf{X}), \dots, \widehat{\mathbf{b}}_T(\mathbf{X})) \\
&\quad = \quad \arg\max f(\mathbf{X}, \mathbf{b}_1(\mathbf{X}), \dots, \mathbf{b}_T(\mathbf{X}) \mid \mathbf{y}_{1:T}) \qquad (22)
\end{aligned}
$$

where the maximum has to be taken with respect to $\mathbf{X}$ and $\mathbf{b}_t(\mathbf{X})$, $t = 1, \dots, t$. The introduction of this "fine-grain" notation for the random set suggests that the MAP detector may be implemented in the form of a sequential detector, thus simplifying its operation (more on this *infra*).

Before moving on, it may be worth pointing out that, for the case considered here, the same results could be obtained through ordinary probability theory by introducing a K-dimensional vector sequence, taking on binary or ternary values according to whether a training sequence is used or not: as already anticipated, though, such equivalence holds only for the discrete case, and, unlike the RST formulation, does not suggest an immediate extension accounting for more general channel models [2].

B. Dynamic channel

Consider now a dynamic channel, and examine first the case of known data. We have, accounting for the Markov property of our channel model,

$$
\begin{aligned}
&f(\mathbf{X}_1, \dots, \mathbf{X}_T \mid \mathbf{y}_{1:T}) \qquad (23)\\
&\quad \propto \quad f(\mathbf{y}_{1:T} \mid \mathbf{X}_1, \dots, \mathbf{X}_T) \times f(\mathbf{X}_1)\prod_{t=2}^{T} f(\mathbf{X}_t \mid \mathbf{X}_{t-1})
\end{aligned}
$$

with $f(\mathbf{X}_1)$ a density whose assignment is based upon prior knowledge of the channel state at the beginning of the transmission. The MAP estimator here maximizes the RHS of the above (or its logarithm) with respect to the values taken on by the sequence $(\mathbf{X}_1, \dots, \mathbf{X}_T)$. Even in this case we may think of a sequential detector, which searches for the maximum-APP path traversing a trellis having T stages and a number of states at stage i equal to the number of realizations of the random set $\mathbf{X}_i$.

a) Implementing a sequential detector.: Implementation of the sequential detector through a version of Viterbi algorithm leads to the following consequences:

① The decision on the whole sequence of users' identities and their data should be taken only after the whole sequence of observations $\mathbf{y}_1, \dots, \mathbf{y}_T$ has been recorded.
② The decision on the users' identities and their data at time t depends not only on the past observations, but also on observations that have not been recorded yet at time t.
③ A suboptimum version of the optimum sequential algorithm, the *sliding-window Viterbi* algorithm (see, e.g., [1, p. 133 ff.]) can be implemented. This consists of forcing a decision on $\mathbf{X}_t$, $\mathbf{b}_t(\mathbf{X}_t)$ based on a sliding window of observations that includes $\mathbf{y}_t$, but whose length is smaller than T.

C. PEP analysis

We now evaluate the performance of the detectors described above. We assume $N = 1$ for simplicity, and derive bounds and approximations to error probabilities using the pairwise error probability (PEP) $P(\mathbf{X}_t \to \widehat{\mathbf{X}}_t)$. This is the probability that, when $\mathbf{X}_t$ is the true value of the random set to be detected, the receiver assigns a higher APP to $\widehat{\mathbf{X}}_t \neq \mathbf{X}_t$ (see, e.g., [1, p. 43]) [3]

1) Static channel: Defining

$$
\begin{aligned}
&\mathbf{S}_t(\mathbf{X}, \widehat{\mathbf{X}}) \qquad (24)\\
&\quad \triangleq \mathbf{R}^{-1}\Big[\big(\mathbf{y}_t - \mathbf{R}\mathbf{A}\mathbf{b}_t(\widehat{\mathbf{X}})\big)\big(\mathbf{y}_t - \mathbf{R}\mathbf{A}\mathbf{b}_t(\widehat{\mathbf{X}})\big)'\Big]\Big|_{\mathbf{y}_t = \mathbf{R}\mathbf{A}\mathbf{b}_t(\mathbf{X}) + \mathbf{z}_t}
\end{aligned}
$$

we have

$$
\mathbf{S}_t(\mathbf{X}, \mathbf{X}) = \mathbf{R}^{-1}\mathbf{z}_t\mathbf{z}_t' \tag{25}
$$

and

$$
\mathbf{S}_t(\mathbf{X}, \widehat{\mathbf{X}}) = \mathbf{R}^{-1}\left[\big(\mathbf{R}\mathbf{A}\mathbf{d}_t(\mathbf{X}, \widehat{\mathbf{X}}) + \mathbf{z}_t\big)\big(\mathbf{R}\mathbf{A}\mathbf{d}_t(\mathbf{X}, \widehat{\mathbf{X}}) + \mathbf{z}_t\big)'\right] \tag{26}
$$

where $\mathbf{d}_t(\mathbf{X}, \widehat{\mathbf{X}}) \triangleq \mathbf{b}_t(\mathbf{X}) - \mathbf{b}_t(\widehat{\mathbf{X}})$. Based on the above, the PEP with ML detection of unknown user identities can be written as

$$
P(\mathbf{X} \to \widehat{\mathbf{X}}) = \mathbb{P}\left\{\mathrm{tr}\left[\sum_{t=1}^{T}\mathbf{S}_t(\mathbf{X}, \widehat{\mathbf{X}}) - \mathbf{S}_t(\mathbf{X}, \mathbf{X})\right] < 0\right\} \tag{27}
$$

[3] It might be worth observing here that, contrary to a fairly widespread misconception, $P(\mathbf{X}_t \to \widehat{\mathbf{X}}_t)$ is *not* the probability of mistaking $\widehat{\mathbf{X}}_t$ for $\mathbf{X}_t$, unless $\mathbf{X}_t$ and $\widehat{\mathbf{X}}_t$ are the only possible alternatives.

Now, observe that

$$\begin{aligned}
&\operatorname{tr}\left[\mathbf{S}_t(\mathbf{X},\widehat{\mathbf{X}})-\mathbf{S}_t(\mathbf{X},\mathbf{X})\right]\\
&\quad=\ \operatorname{tr}\left[\mathbf{R}^{-1}\left(\mathbf{RAd}_t\mathbf{d}_t'\mathbf{AR}+\mathbf{z}_t\mathbf{d}_t'\mathbf{AR}+\mathbf{RAd}_t\mathbf{z}_t'\right)\right]\\
&\quad=\ \operatorname{tr}\left[\mathbf{Ad}_t\mathbf{d}_t'\mathbf{AR}+\mathbf{R}^{-1}\mathbf{z}_t\mathbf{d}_t'\mathbf{AR}+\mathbf{Ad}_t\mathbf{z}_t'\right]\\
&\quad=\ \operatorname{tr}\left[\mathbf{Ad}_t\mathbf{d}_t'\mathbf{AR}+\mathbf{z}_t\mathbf{d}_t'\mathbf{A}+\mathbf{Ad}_t\mathbf{z}_t'\right]\\
&\quad=\ \operatorname{tr}\left[\mathbf{RAd}_t\mathbf{d}_t'\mathbf{A}\right]+2\operatorname{tr}\left[\mathbf{Ad}_t\mathbf{z}_t'\right]
\end{aligned}$$

Denoting by $a(i)$ the ith diagonal element of matrix $\mathbf{A}$, by $d_t(i)$ the ith entry of $\mathbf{d}_t$, by $r_{j,k}$ the entry in row j and column k of $\mathbf{R}$, and by $z_t(i)$ the ith entry of vector $\mathbf{z}_t$, we have

$$\operatorname{tr}\left[\mathbf{RAd}_t\mathbf{d}_t'\mathbf{A}\right] = \sum_{i=1}^{K}\sum_{j=1}^{K} a(i)a(j)d_t(i)d_t(j)r_{i,j} \tag{28}$$

$$\operatorname{tr}\left[\mathbf{Ad}_t\mathbf{z}_t'\right] = \sum_{i=1}^{K} a(i)d_t(i)z_t(i) \tag{29}$$

Finally, since we are assuming $\mathbf{z}_t \sim \mathcal{N}(\mathbf{0},(N_0/2)\mathbf{R})$, we have

$$\operatorname{tr}\left[\sum_{t=1}^{T}\left(\mathbf{S}_t(\mathbf{X},\widehat{\mathbf{X}})-\mathbf{S}_t(\mathbf{X},\mathbf{X})\right)\right] \sim \mathcal{N}(\xi_T, 2N_0\xi_T) \tag{30}$$

where

$$\xi_T \triangleq \sum_{t=1}^{T}\sum_{i=1}^{K}\sum_{j=1}^{K} a(i)a(j)d_t(i)d_t(j)r_{i,j} \tag{31}$$

In conclusion, we obtain

$$P(\mathbf{X}\to\widehat{\mathbf{X}}) = Q\left(\sqrt{\frac{\xi_T}{2N_0}}\right) \tag{32}$$

where $Q(\,\cdot\,)$ denotes the Gaussian tail function.

Before proceeding further, we comment briefly on the structure of vectors $\mathbf{d}_t$. They have the following nonzero entries:

① The $|\mathbf{X}\cap\widehat{\mathbf{X}}|$ terms corresponding to users present in both sets: these terms may take on values in $\{0,\pm 2\}$.

② The $|\widehat{\mathbf{X}}\setminus\mathbf{X}\cap\widehat{\mathbf{X}}|$ terms corresponding to users present in $\widehat{\mathbf{X}}$ only: these terms may take on values in $\{\pm 1\}$.

③ The $|\mathbf{X}\setminus\mathbf{X}\cap\widehat{\mathbf{X}}|$ terms corresponding to users present in $\mathbf{X}$ only: these terms may take on values in $\{\pm 1\}$.

Similarly, the PEP with MAP detection has the form

$$P(\mathbf{X}\to\widehat{\mathbf{X}}) = Q\left(\frac{\xi_T-\eta}{\sqrt{2N_0\xi_T}}\right) \tag{33}$$

where

$$\eta \triangleq N_0 \ln\left[\frac{f_{\mathbf{X}}(\widehat{\mathbf{X}})}{f_{\mathbf{X}}(\mathbf{X})}\right]$$

Observe here that, with $\mathbf{R}$ a positive definite matrix, we have

$$(\mathbf{Ad}_t)'\mathbf{R}(\mathbf{Ad}_t) > 0 \quad \text{for } \mathbf{Ad}_t \neq \mathbf{0}$$

which entails

$$\lim_{T\to\infty} P(\mathbf{X}\to\widehat{\mathbf{X}}) = 0$$

Notice also that, for given signal-to-noise ratio, (33) also suggests a minimum length of the data frame in the form of the "open-eye" condition $T \geq T_{\min}$, where

$$T_{\min} \triangleq \inf\left\{T : \min_{\mathbf{X},\widehat{\mathbf{X}}}\left[\xi_T(\mathbf{X},\widehat{\mathbf{X}})-\eta(\mathbf{X},\widehat{\mathbf{X}})\right] > 0\right\}$$

The PEP for the case of detection of user identities and data can be dealt with with similar techniques, and we shall not delve in this issue any further here.

2) Dynamic channel: In this case, defining the true state sequence $\underline{\mathbf{X}} \triangleq (\mathbf{X}_t)_{t=1}^{T}$ and the competing state sequence $\underline{\widehat{\mathbf{X}}} \triangleq (\widehat{\mathbf{X}}_t)_{t=1}^{T}$, we obtain the PEP for the MAP detection of unknown identities:

$$P(\underline{\mathbf{X}}\to\underline{\widehat{\mathbf{X}}}) = Q\left(\frac{\xi_T(\underline{\mathbf{X}},\underline{\widehat{\mathbf{X}}})-\eta_T(\underline{\mathbf{X}},\underline{\widehat{\mathbf{X}}})}{\sqrt{2N_0\xi_T(\underline{\mathbf{X}},\underline{\widehat{\mathbf{X}}})}}\right) \tag{34}$$

and the PEP for the MAP detection of unknown identities and data:

$$\begin{aligned}
&P\big(\underline{\mathbf{X}},(\mathbf{b}_t(\mathbf{X}_t))_{t=1}^{T}\to\underline{\widehat{\mathbf{X}}},(\mathbf{b}_t(\widehat{\mathbf{X}}_t))_{t=1}^{T}\big)\\
&\quad= Q\left(\frac{\xi_T(\underline{\mathbf{X}},\underline{\widehat{\mathbf{X}}})-\tilde{\eta}_T(\underline{\mathbf{X}},\underline{\widehat{\mathbf{X}}})}{\sqrt{2N_0\xi_T(\underline{\mathbf{X}},\underline{\widehat{\mathbf{X}}})}}\right)
\end{aligned} \tag{35}$$

where

$$\xi_T(\underline{\mathbf{X}},\underline{\widehat{\mathbf{X}}}) \triangleq \sum_{t=1}^{T}\mathbf{d}_t'\mathbf{ARAd}_t \tag{36}$$

$$\eta_T(\underline{\mathbf{X}},\underline{\widehat{\mathbf{X}}}) \triangleq N_0\sum_{t=1}^{T}\ln\left[\frac{f_{\mathbf{X}_t|\mathbf{X}_{t-1}}(\widehat{\mathbf{X}}_t\mid\widehat{\mathbf{X}}_{t-1})}{f_{\mathbf{X}_t|\mathbf{X}_{t-1}}(\mathbf{X}_t\mid\mathbf{X}_{t-1})}\right] \tag{37}$$

$$\begin{aligned}
\tilde{\eta}_T(\underline{\mathbf{X}},\underline{\widehat{\mathbf{X}}}) &\triangleq \eta_T(\underline{\mathbf{X}},\underline{\widehat{\mathbf{X}}})\\
&\quad+N_0\sum_{t=1}^{T}\left[|\mathbf{X}_t|-|\widehat{\mathbf{X}}_t|\right]\ln 2
\end{aligned} \tag{38}$$

Similar arguments, which we omit here for brevity's sake, apply to ML detection.

D. Error probabilities

Several approximations to error probabilities are possible, based on the union bound (see, e.g., [1]), on the PEP derivations outlined *supra*, and on assumptions on user statistics, spreading codes, and users' amplitudes. We obtain, for the union bound to the probability of mistaking the set of active users,

$$P(e) \leq \sum_{i=1}^{2^K} f(\mathbf{X}_i)\sum_{j\neq i} P(\mathbf{X}_i\to\mathbf{X}_j) \tag{39}$$

which, under maximum prior uncertainty as to the channel occupancy, becomes:

$$P(e) \leq \frac{1}{2^K}\sum_{i=1}^{2^K}\sum_{j\neq i} P(\mathbf{X}_i\to\mathbf{X}_j) \tag{40}$$

where $\mathbf{X}_i,\mathbf{X}_j \in 2^{\mathbb{K}}$. This union bound can be simplified by restricting the inner summation to those pairs of realizations of the random sets that are most likely to contribute significantly to error probability. For example, if we restrict it to the pairs that differ in at most n entries, we obtain an approximation depending on n:

$$P(e) \approx P^{(n)}(e) \triangleq \frac{1}{2^K} \sum_{i=1}^{2^K} \sum_{j:\delta_{i,j} \leq n} P(\mathbf{X}_i \to \mathbf{X}_j) \tag{41}$$

where

$$\delta_{i,j} \triangleq |\mathbf{X}_i \setminus \mathbf{X}_i \cap \mathbf{X}_j| + |\mathbf{X}_j \setminus \mathbf{X}_i \cap \mathbf{X}_j| \leq n$$

Likewise, for a dynamic scenario, we have that the union bound for user identification is written as:

$$P(e) \leq \sum_{\underline{\mathbf{X}} \in (2^{\mathbb{K}})^T} f(\underline{\mathbf{X}}) \sum_{\underline{\mathbf{X}} \neq \widehat{\underline{\mathbf{X}}}} P(\underline{\mathbf{X}} \to \widehat{\underline{\mathbf{X}}}) \tag{42}$$

where $f(\underline{\mathbf{X}})$ can be easily determined by applying the chain rule to the Markov set sequence $\underline{\mathbf{X}}$. Approximations similar to (41) can be developed; likewise, the case of joint user identification and data detection can be handled by noticing that the configurations of $\underline{\mathbf{X}}$ become now 3^{KT}, and the joint density $f(\underline{\mathbf{X}})$ is written as:

$$f(\underline{\mathbf{X}}) = 2^{-|\mathbf{X}_1|} f(\mathbf{X}_1) \prod_{t=2}^{T} 2^{-|\mathbf{X}_t|} f(\mathbf{X}_t \mid \mathbf{X}_{t-1}) \tag{43}$$

The above relationships also suggest a semi-analytical method to evaluate the approximations without summing up an exponential number of terms: indeed, since an average over the joint density (43) is to be performed, this can be efficiently evaluated through Monte-Carlo counting by generating a substantially smaller number of independent set sequence patterns obeying the Markov law (43).

V. Numerical results

In this section we show some numerical examples that illustrate the theory developed before.

Fig. 3 shows how the knowledge of the channel dynamics can improve the performance of a multiuser detector.

Fig. 4 refers again to a static channel and to the case that the active users transmit a known sequence of bits in order to be identified: we assume here that all users (including the reference user) are active with probability $\alpha = 0.5$. Now $K = 6$, the transmitted signals are binary antipodal, spreading is done through m-sequences with length 7, the power control is perfect (hence, $\mathbf{A}$ is a scalar matrix) and the data-frame length varies from $T = 1$ to $T = 3$. Here we evaluate the accuracy of the union bound to the probability of an error in the identification of active users (set-error probability, or SEP), and of its approximation $P^{(1)}(e)$ (obtained by assuming that the errors can only be generated by the event, denoted $\mathcal{E}(1)$, of mistaking an active-user set by another differing by only one of its elements). Simulation results are also shown for reference's sake. It can be observed how, especially for large values of signal-to-noise ratio, the error probability is dominated by the event $\mathcal{E}(1)$.

Fig 5 refers to a system with the same configuration as in Fig. 4, but on a dynamic channel with $K = 3$, $\alpha = 0.2$ and $\mu = 0.8$. The system dynamics are tracked over an

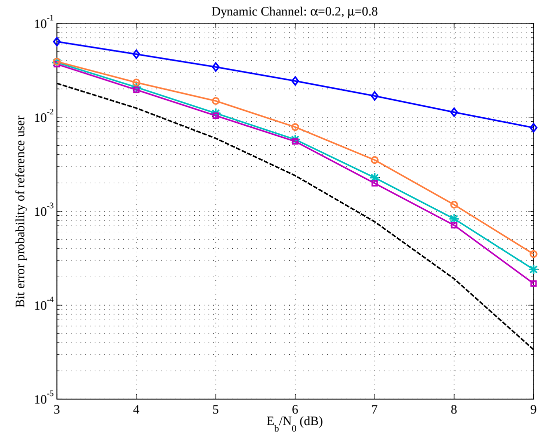

Fig. 3. *Bit error probability of the reference user in a multiuser system with* 2 *interferers, following a dynamic model described above with* $\alpha = 2$ *and* $\mu = 0.8$. *Line with diamond markers: Classic multiuser ML detection, assuming that all users are active. Line with circle markers: MAP detection based on the knowledge of* α *alone. Line with star markers: causal RST detector, based on Bayes recursions. Line with square markers: Viterbi RST detector. Dashed curve: Single-user bound.*

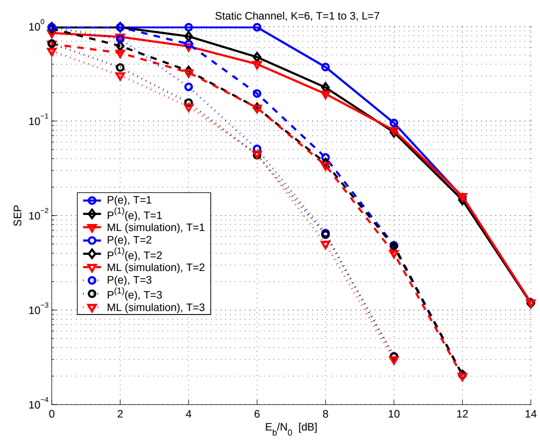

Fig. 4. *Set-error probability with ML detection based on RST with* $K = 6$, $L = 7$. *Comparison among "exact" probability (obtained by simulation), union-bound to it (denoted* $P(e)$*), and approximation (41) to the union bound.*

interval with length $T = 10$. The ordinate shows the set error probability (SEP), i.e., the probability of an erroneous estimate of the active-user sequence. Here a comparison is made between a non-causal Viterbi set estimate and a causal estimate, obtained through Bayesian-filter recursions. To elicit the impact of the causality constraint, we represent the SEP for the set $\mathbf{X}_1$, where the causality constraint prevents sequence detection, and for the set $\mathbf{X}_{10}$, where such a constraint has no effect: as expected, the performances of the Viterbi algorithm and of the Bayesian recursions coincide when estimating $\mathbf{X}_{10}$, while the causality constraint has a perceivable effect on the performance when $\mathbf{X}_1$ is estimated.

In order to provide global figures of merit of both trained and untrained systems in a dynamic environment, we use the "Set Sequence Error Probability"(SSEP). For trained systems, this is the probability that for some t, $1 \leq t \leq T$, the estimated set $\widehat{\mathbf{X}}_t$ differs from the true set $\mathbf{X}_t$ either in its cardinality or in its elements. For untrained systems, it is the probability that

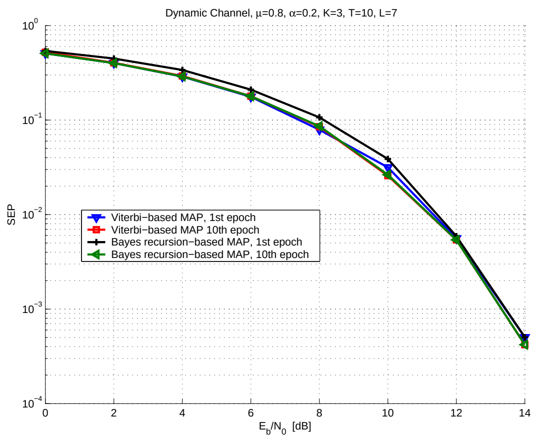

Fig. 5. *Trained acquisition of the set of active users through the Viterbi Algorithm and Bayes recursion: effect of the causality constraint*

at some t the estimated and the true set differ either in the cardinality and/or in the identities of the active users and/or in the transmitted data. Plots of the SSEP are shown in Fig. 6 for a trained system with $K = 6$ maximum number of active users: also shown in the figure is the curve obtained through the semi-analytical approximation suggested in the previous section, which apparently follows the numerical results quite closely.

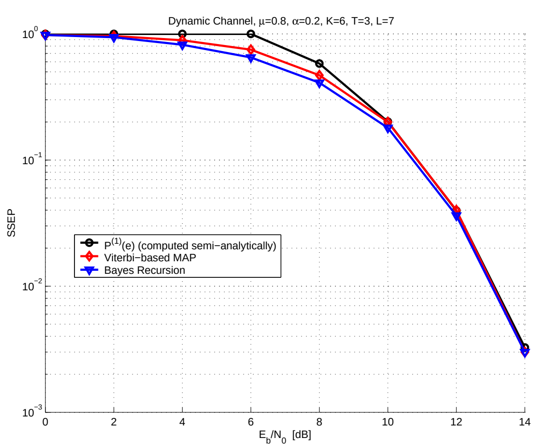

Fig. 6. *Set sequence error probability with $K = 6$, $L = 7$. Also, the curve corresponding to a semi-analytical performance evaluation under trained acquisition of the set of the active users.*

The case that not only the identities, but also the data of the active users are to be estimated is shown in Fig. 7, assuming a maximum of $K = 3$ active users and, again, $L = 7$; the data-frame length is $T = 10$. Here we compare a Viterbi-algorithm receiver with one based on Bayesian recursions for estimating the set of interferers and the transmitted bits. The ordinate shows the bit-sequence error probability, at time $t = 1$ and at time $t = T = 10$, defined as the probability that the estimated and the true set do not coincide: the term ”bit sequence error probability” is tied here to the fact that an error in estimating the identities of the active users automatically implies an error in estimating the stream $\mathbf{b}(\mathbf{X}_t)$, while the converse is not true. Once again, the effect of the causality constraint on the performance is elicited, and the results are in accordance with the intuition as well as with the curves of Fig. 5.

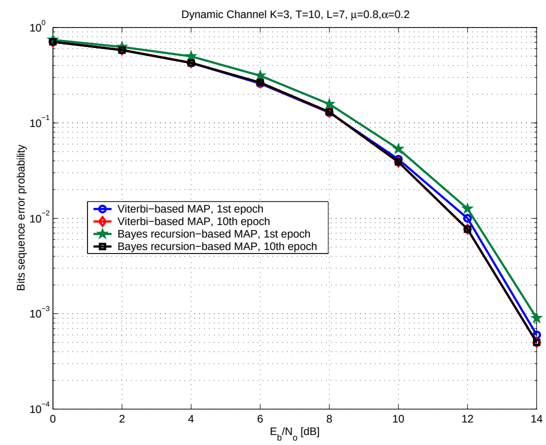

Fig. 7. *Bit-sequence error probability of the reference user. Data estimated with Viterbi algorithm and Bayesian recursions.*

VI. Conclusions

We have described a technique for estimating the received-signal parameters and data in a multiuser transmission system. Since the number of active interferers is itself a random variable, the set of parameters to be estimated has a random number of random elements. A dynamic model for the evolution of this random set accounts for new interferers appearing and old interferers disappearing in each measurement interval. Random-set theory can be used to develop a multiuser detection scheme in this context. This is done by developing Bayesian-filtering equations that describe the evolution of the MAP multiuser detector in a dynamic environment.

One question that may naturally arise at this point concerns the need for RST, and in particular the question if the results obtained with RST can also be obtained in a more straightforward way. Although other ways of obtaining the same results with conventional probability techniques may be available (see,e.g., [23], where the connections between RST and point processes are explored), we argue that not only the rigor and the generality of RST, but especially its simplicity, make it a tool of choice for the study of random-access systems. The results of this paper are also meant to support this claim.

Appendix

This appendix describes, in a rather qualitative fashion, the fundamentals of Random-Set Theory. For a rigorous approach and additional details, see [6], [22], [23] and the references therein.

Given a sample space Ω (the space of all the outcomes of a random experiment), a probability measure can be defined on it. If a random variable (i.e., a mapping from Ω to another space $\mathbb{S}$) is defined, it is convenient to generate a probability measure directly on $\mathbb{S}$. This can be given in terms of a density function, once certain mathematical operations, such as integration, are defined on $\mathbb{S}$. Random sets can be viewed as a generalization of the concept of a random variable. A *finite random set* is a mapping $\mathbf{X} : \Omega \to \mathcal{F}(\mathbb{S})$ from the sample space Ω to the collection of closed sets of the space $\mathbb{S}$, with $|\mathbf{X}(\omega)| < \infty$ for all $\omega \in \Omega$. For our purposes, the space $\mathbb{S}$ of finite random sets is assumed to be the *hybrid space* $\mathbb{S} = \mathbb{R}^d \times U$, the direct product of the d-dimensional Euclidean space $\mathbb{R}^d$ and a finite discrete space U. The elements of $\mathbb{S}$ characterize the users' parameters, some of which continuous (d real numbers) and some discrete (for example, the users' signatures and their transmitted data). An element of $\mathbb{S}$ is the pair $(\mathbf{v}, u)$, $\mathbf{v}$ a d-dimensional real vector, and $u \in U$.

A. Belief mass functions

A fairly natural probability law for $\mathbf{X}$ is the probability distribution $P_{\mathbf{X}}$, defined for any (Borel) subset $\mathcal{T}$ of $\mathcal{F}(\mathbb{S})$ by

$$P_{\mathbf{X}}(\mathcal{T}) \triangleq \mathbb{P}(\mathbf{X} \in \mathcal{T})$$

However, RST is based on a probability law given differently. Specifically, the *belief mass function* of a finite random set $\mathbf{X}$ is defined as

$$\beta_{\mathbf{X}}(\mathbf{C}) \triangleq \mathbb{P}(\mathbf{X} \subseteq \mathbf{C}) \tag{44}$$

where $\mathbf{C}$ is a closed subset of $\mathbb{S}$. As observed in [22, p. 42], the belief function corresponds to the cumulative distribution function of a real random variable, and differs from it because subsets are only partially ordered by the inclusion relation $\subseteq$. The belief function characterizes the probability distribution of a random finite set $\mathbf{X}$, and allows the construction of a density function of $\mathbf{X}$ through the definition of a *set integral* and a *set derivative*. Specifically, the *belief density*, i.e., the *set derivative* of the belief function, plays the role of a probability density function in ordinary probability calculus (for this reason, in this paper we refer to it simply as a *density*). The belief density is obtained as

$$f_{\mathbf{X}}(\mathbf{Z}) = \left.\frac{\delta\beta(\mathbf{S})}{\delta\mathbf{Z}}\right|_{\mathbf{S}=\emptyset} \tag{45}$$

where δ denotes the set derivative, to be defined below. As observed in [6, p. 163], the value $\beta_{\mathbf{X}}(\mathbf{Z})$ of the belief density specifies the likelihood with which the random set $\mathbf{X}$ takes the set $\mathbf{Z}$ as a specific realization.

Notice how, in the special case of a random set consisting of a singleton, $\mathbf{X} = \{\mathbf{x}\}$, $\mathbf{x}$ a random vector, we have

$$\beta_{\mathbf{X}}(\mathbf{C}) \triangleq \mathbb{P}(\mathbf{X} \subseteq \mathbf{C}) = \mathbb{P}(\mathbf{x} \in \mathbf{C}) = P(\mathbf{C})$$

with $P(\,\cdot\,)$ the ordinary probability measure of $\mathbf{x}$.

B. Set derivative

Let $\mathcal{C}(\mathbb{S})$ denote the collection of closed subsets of $\mathbb{S}$. If F is a set function defined on $\mathcal{C}(\mathbb{S})$, then its set derivative at $\mathbf{x}$ is defined as the set function

$$\frac{\delta F(\mathbf{S})}{\delta\mathbf{x}} = \lim_{j\to\infty}\lim_{i\to\infty}\frac{F\left[(\mathbf{S}\setminus B_x(1/j))\bigcup\overline{B_x}(1/i)\right] - F(\mathbf{S})}{\overline{m}\left(\overline{B_x}(1/i)\right)}$$

where $B_x(1/j)$, $\overline{B_x}(1/i)$ are an open ball of radius $1/j$ and a closed ball of radius $1/i$, respectively, both centered at $\mathbf{x}$, and $\overline{m}(\cdot)$ denotes the hybrid Lebesgue measure, i.e., the product of the ordinary measure in $\mathbb{R}^d$ and of the counting measure. The set derivative of F at a finite set $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is defined by the recursion

$$\frac{\delta F(\mathbf{S})}{\delta\mathbf{X}} \triangleq \frac{\delta}{\delta\mathbf{x}_n}\left(\frac{\delta F(\mathbf{S})}{\delta\{\mathbf{x}_1, \dots, \mathbf{x}_{n-1}\}}\right)$$

In particular, the belief density of the random set $\mathbf{X}$ is given by

$$f_{\mathbf{X}}(\mathbf{B}) = \left.\frac{\delta\beta_{\mathbf{X}}(\mathbf{S})}{\delta\mathbf{B}}\right|_{\mathbf{S}=\emptyset} \tag{46}$$

Two useful rules of derivation are the following (see also [15, p. 386 ff.]). Let F, G be set functions, and $a, b \in \mathbb{R}$. Then

$$\frac{\delta(aF(\mathbf{S}) + bG(\mathbf{S}))}{\delta\mathbf{B}} = a\frac{\delta F(\mathbf{S})}{\delta\mathbf{B}} + b\frac{\delta G(\mathbf{S})}{\delta\mathbf{B}} \tag{47}$$

and

$$\frac{\delta F(\mathbf{S})G(\mathbf{S})}{\delta\mathbf{B}} = \sum_{\mathbf{C}\subseteq\mathbf{B}}\frac{\delta F(\mathbf{S})}{\delta\mathbf{B}}\frac{\delta G(\mathbf{S})}{\delta(\mathbf{B}\setminus\mathbf{C})} \tag{48}$$

C. Set integral

Let f denote a set function defined by

$$f(\mathbf{X}) = \left.\frac{\delta F(\mathbf{S})}{\delta\mathbf{X}}\right|_{\mathbf{S}=\emptyset}$$

The set integral of f over the closed subset $\mathbf{S} \subseteq \mathbb{S}$ is given by

$$\begin{aligned}&\int_{\mathbf{S}} f(X)\,\delta X \qquad (49)\\ &= \quad f(\{\emptyset\}) + \sum_{k=1}^{\infty}\frac{1}{k!}\int_{\mathbf{S}^k} f(\{\mathbf{x}_1, \dots, \mathbf{x}_k\})d\bar{m}(\mathbf{x}_1)\cdots d\bar{m}(\mathbf{x}_k)\end{aligned}$$

where $f(\{\mathbf{x}_1, \dots, \mathbf{x}_k\}) = 0$ if $\mathbf{x}_1, \dots, \mathbf{x}_k$ are not distinct. Since we are dealing with *finite* random sets, the summation above contains only a finite number of terms.

D. Special case: $d = 0$

The special case $d = 0$ (which corresponds to making $\mathbb{S}$ a discrete finite set) reduces the set derivative to the *Möbius inversion formula* [22, p. 43]:

$$f_{\mathbf{X}}(\mathbf{A}) = \sum_{\mathbf{B}\subseteq\mathbf{A}}(-1)^{|\mathbf{A}\setminus\mathbf{B}|}\beta_{\mathbf{X}}(\mathbf{B}) \tag{50}$$

and the set integral to

$$f(\{\emptyset\}) + \sum_{k=1}^{\infty}\frac{1}{k!}\sum_{\mathbf{x}_1\neq\ldots\neq\mathbf{x}_k} f(\{\mathbf{x}_1, \dots, \mathbf{x}_k\}) \tag{51}$$

where the summation is extended to all possible combinations of k distinct elements $\mathbf{x}_k \in \mathbf{S}$ (in this case, the hybrid Lebesgue measure reduces to the counting measure, and hence the Lebesgue integrals in (49) become summations).

E. Generalized fundamental theorem of calculus

Set derivatives and set integrals turn out to be the inverse of each other: we have

$$f(\mathbf{X}) = \left.\frac{\delta F(\mathbf{S})}{\delta \mathbf{X}}\right|_{\mathbf{S}=\emptyset} \Longleftrightarrow F(\mathbf{S}) = \int_{\mathbf{S}} f(\mathbf{X})\, \delta\mathbf{X} \qquad (52)$$

By using the above result, belief functions and belief densities can be derived from one another.

ACKNOWLEDGMENTS

A preliminary version of this paper was presented at *WPMC 2006*, San Diego, CA, September 17–20, 2006. Ezio Biglieri was exposed to Random-Set Theory during his visit to the Institute for Infocomm Research, Singapore, in 2005. Discussions he had there with Professor Jiankang Wang are gratefully acknowledged. He also benefited from discussions with Professor Kung Yao, whose competence in target-tracking problems helped him to put in the right perspective the topics discussed in this paper. The authors are grateful to Shlomo Shamai (Shitz), who called their attention to [19], which contains an information-theoretical analysis of communication systems with a a random number of active users, and to Daniele Angelosante for his help with the derivation of numerical results.

The constructive comments of the reviewers, which helped to improve the focus and clarity of the manuscript, are gratefully acknowledged.

REFERENCES

[1] E. Biglieri, *Coding for Wireless Channels*. New York: Springer, 2005.

[2] D. Angelosante, E. Biglieri and M. Lops, "Multiuser detection in a dynamic environment. Part II: Parameter estimation," *in preparation.* A preliminary version of this work was presented at the *2007 IEEE International Symposium on Information Theory*, Nice, France, June 24–28, 2007.

[3] T. N. Bui, V. Krishnamurthy, and H. V. Poor, "Online Bayesian activity detection in DS/CDMA networks," *IEEE Trans. Signal Processing*, Vol. 53, No. 1, pp. 371– 375, January 2005.

[4] S. Buzzi, A. De Maio, and M. Lops, "Code-aided blind adaptive new user detection in DS/CDMA systems with fading time-dispersive channels," *IEEE Trans. Signal Processing*, Vol. 51, No. 10, pp. 2637–2649, October 2003.

[5] B. Chen and L. Tong, "Traffic-aided multiuser detection for random-access CDMA networks," *IEEE Trans. Signal Processing*, Vol. 49, No. 7, pp. 1570–1580, July 2001.

[6] I. R. Goodman, R. P. S. Mahler, and H. T. Nguyen, *Mathematics of Data Fusion*. Dordrecht, The Netherlands: Kluwer, 1997.

[7] K. W. Halford and M. Brandt-Pearce, "New-user identification in a CDMA system," *IEEE Trans. Commun.*, Vol. 46, No. 1, pp. 144–155, January 1998.

[8] M. L. Honig and H. V. Poor, "Adaptive interference suppression in wireless communication systems," in *Wireless Communications: Signal Processing Perspectives*, H. V. Poor and G. W. Wornell, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[9] V. P. Ipatov, *Spread Spectrum and CDMA*. Chichester, UK: J. Wiley & Sons, 2005.

[10] D. D. Lin and T. J. Lim, "Subspace-based active user identification for a collision-free slotted ad hoc network," *IEEE Trans. Commun.*, Vol. 52, No. 4, pp. 612–621, April 2004.

[11] M. A. Maddah-Ali, M. Ansari, and A. K. Khandani, "An efficient algorithm for user selection and signalling over MIMO multiuser systems," Technical Report UW-E&CE 2004-19, University of Waterloo, Waterloo, ON, Augustt 3, 2004.

[12] R. Mahler, *An Introduction to Multisource–Multitarget Statistics and its Applications*. Lockheed Martin Technical Monograph, Eagan, MD, March 15, 2000.

[13] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 39, No. 4, pp. 1152–1178, October 2003.

[14] R. Mahler, "Random sets: Unification and computation for information fusion—A retrospective assessment," *7th International Conference on Information Fusion*, Stockholm, Sweden, June 2004.

[15] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Boston, MA: Artech House, 2007.

[16] U. Mitra and H. V. Poor, "Activity detection in a multi-user environment," *Wireless Personal Communications*, Vol. 3, No. 1–2, pp. 149–174, January 1996.

[17] U. Mitra and H. V. Poor, "Analysis of an adaptive decorrelating detector for synchronous CDMA channels," *IEEE Trans. Commun.* , Vol. 44, pp. 257-268, 1996.

[18] T. Oskiper and H. V. Poor, "Online activity detection in a multiuser environment using the matrix CUSUM algorithm," *IEEE Trans. Inform. Theory*, Vol. 48, No. 2, pp. 477–493, February 2002.

[19] E. Plotnik, "On the capacity region of the random-multiple access channel," *The 16th Conference of Electrical and Electronics Engineers in Israel*, pp. 1.1.6.1–4, Tel Aviv, Israel, March 7–9, 1989.

[20] P. B. Rapajic and D. K. Bora, "Adaptive MMSE maximum likelihood CDMA multiuser detection," *IEEE J. Select. areas Commun.*, Vol. 17, No. 12, pp. 2110–2122, December 1999.

[21] S. Verdú, *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.

[22] M. Vihola, *Random Sets for Multitarget Tracking and Data Fusion*. Licentiate Thesis, Tampere University of Technology, 2004.

[23] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multi-target filtering with random finite sets," *IEEE Trans. Aerospace and Electronic Systems*, Vol. 41, No. 4, pp. 1224–1245, October 2005.

[24] W.-C. Wu and K.-C. Chen, "Identification of active users in synchronous CDMA multiuser detection," *IEEE J. Select. Areas Commun.*, Vol. 16, No. 9, pp. 1723–1735, December 1998.

[25] Z. Xu, "Blind identification of co-existing synchronous and asynchronous users for CDMA systems," *IEEE Signal Processing Letters*, Vol. 8, No. 7, pp. 212–214, July 2001.

[26] W. Yu and W. Rhee, "Degrees of freedom in multi-user spatial multiplex systems with multiple antennas," *IEEE Trans. on Commun.*, Vol. 54, No. 10, pp. 1747–1753, October 2006.