Multiuser Detection in a Dynamic Environment— Part II: Joint User Identification and Parameter Estimation

Daniele Angelosante, *Student Member, IEEE*, Ezio Biglieri, *Fellow, IEEE*, and Marco Lops, *Senior Member, IEEE*

Abstract—The problem of jointly estimating the number, the identities, and the data of active users in a time-varying multiuser environment was examined in a companion paper (*IEEE Trans. Information Theory*, vol. 53, no. 9, September 2007), at whose core was the use of the theory of finite random sets on countable spaces. Here we extend that theory to encompass the more general problem of estimating unknown *continuous* parameters of the active-user signals. This problem is solved here by applying the theory of random finite sets constructed on hybrid spaces. We do so deriving Bayesian recursions that describe the evolution with time of a posteriori densities of the unknown parameters and data. Unlike in the above cited paper, wherein one could evaluate the exact multiuser set posterior density, here the continuous-parameter Bayesian recursions do not admit closed-form expressions. To circumvent this difficulty, we develop numerical approximations for the receivers that are based on Sequential Monte Carlo (SMC) methods ("particle filtering"). Simulation results, referring to a code-divisin multiple-access (CDMA) system, are presented to illustrate the theory.

Index Terms—Bayesian recursions, multiuser detection, particle filtering, random-set theory.

I. INTRODUCTION

THE estimate of the identities and the parameters of users in a dynamic environment has several applications to communication systems, e.g., user localization in wireless systems, neighbor discovery in *ad hoc* networks, and power-control strategy optimization. Several approaches for tracking single-user evolution have been proposed in the literature under diverse models [18], [23], while in [7] a joint detection–estimation problem is formulated through Bayesian theory and solved by Sequential Monte Carlo (SMC) method, yielding better performance than traditional approaches (most of them based on extended Kalman filtering [17] or sum-of-Gaussian

approximation [1]) in a nonlinear, non-Gaussian environment. The present paper aims at presenting a new approach to multiuser tracking in a scenario whereby users may appear and disappear in a random manner from frame to frame, while the parameters of persisting users are themselves time-varying. In this scenario, tracking involves estimating the number of users, their identities, possibly their data, and a number of channel parameters, typically affected by users' mobility. The object of interest is thus the set of parameters of all users. The cardinality of this set varies with time with a known probabilistic law.

As in the companion paper [6], the mathematical tool we use is random-set theory (RST), which provides a natural, flexible, and mathematically consistent framework for the problem under study. The key concept here is to treat the collection of users as a single set-valued state. After introducing appropriate notions of probability density for random sets, known as Finite Sets Statistics (FISST) (see [13], [19], [26], and [22, Ch. 11]) the multiuser tracking problem can be rigorously formulated as a Bayesian set-valued filtering problem. Unlike [6], where the Bayesian recursions leading to the optimal detector could be evaluated exactly, here we deal with finite random sets defined on hybrid spaces, and hence an exact solution of this problem is not feasible due to the lack, even for linear-Gaussian problems [4], of closed-form expressions for the relevant densities. Among the various techniques aimed at approximating the Bayesian recursion, we have chosen here SMC filtering, which achieves asymptotical optimality [10].

The balance of this paper is organized as follow. Section II defines the signal model and states the problem, while Section III defines optimal set estimators. Section IV describes the SMC approximation of the Bayesian recursive filter. Section V presents an application of our methodology to joint multiuser detection and channel tracking in synchronous direct-sequence code-division multiple-access (DS-CDMA) systems. Some numerical results are also described. Two Appendices are devoted to some technicalities of random-set theory, and to the definition of a model for the evolution of user powers under the effects of motion and fading.

II. CHANNEL MODEL AND STATEMENT OF THE PROBLEM

We assume K users transmitting digital data over a common channel. Let $s(\boldsymbol{x}_t^{(i)})$ denote the signal transmitted by the active user i, $i = 1, \ldots, K$, at discrete time t, $t = 1, 2, \ldots$. Each signal has in it a number of known parameters, included in the deterministic function $s(\cdot)$, and a number of random parameters, summarized by the vector $\boldsymbol{x}_t^{(i)}$. 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 the signals generated by the users active at time t, which are in a random number, and of stationary random noise $\boldsymbol{z}_t$. We write

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

where $\boldsymbol{X}_t$ is a random set, encapsulating what is unknown about the active users. The set $\boldsymbol{X}_t \triangleq \{\boldsymbol{x}_t^{(1)}, \ldots, \boldsymbol{x}_t^{(k)}\}$, whose k elements are random vectors (k is a random integer denoting the number of active users), is a *finite random set defined over a hybrid space* $\mathbb{S}$ (see, e.g., [13], [22], [26] and references therein). As stated in [6], the same problem could be solved by using point-process theory [28]. However, RST, and in particular Mahler's "FISST calculus" [13], [22], is advocated here, because it leads to entities (e.g., the belief density) which have immediate counterparts in standard-probability theory (the probability density function), can be manipulated in the same way, and allow one to exploit the engineering intuition developed under the ordinary probability framework.

A finite random set (FRS) defined over a hybrid space $\mathbb{S}$, is a map between a sample space and a family of subsets of $\mathbb{S}$, where the hybrid space $\mathbb{S}$ is the Cartesian product $\mathbb{S} = \mathbb{U} \times \mathbb{R}^d$ consisting of all pairs $(\boldsymbol{u}, \boldsymbol{d})$, with $\boldsymbol{d} \in \mathbb{R}^d$ and $\boldsymbol{u} \in \mathbb{U}$. $\mathbb{R}^d$ contains the values of the d real parameters of the active users, and $\mathbb{U}$ is a discrete set. In our context, we may have, for example, $\mathbb{U} = \{1, \ldots, K\} \triangleq \mathbb{K}$, the set of the identities of the interferers, or $\mathbb{U} = \mathbb{K} \times \{+1, -1\}$, the set of interferers each transmitting binary antipodal data.

Formally, at time t the random set that concerns us is given by the union of singleton-or-empty sets

$$\boldsymbol{X}_t = \bigcup_{k \in \mathbb{K}} \boldsymbol{X}_t^{(k)}$$

where

$$\boldsymbol{X}_t^{(k)} = \begin{cases} \{\boldsymbol{x}_t^{(k)}\}, & \text{if user } k \text{ is active at time } t \\ \emptyset, & \text{otherwise.} \end{cases}$$

We may consider two distinct situations, namely:

(a) known-data ("trained") systems, where $\mathbb{U} = \mathbb{K}$, $\mathbb{S} = \mathbb{K} \times \mathbb{R}^d$, and $\boldsymbol{x}_t^{(k)} = [k, \boldsymbol{a}_t^{(k)}]^T$, with $\boldsymbol{a}_t^{(k)}$ a d-dimensional random vector;

(b) unknown-data ("untrained") systems, where $\mathbb{U} = \mathbb{K} \times \mathbb{M}$, $\mathbb{S} = \mathbb{K} \times \mathbb{M} \times \mathbb{R}^d$, with $\mathbb{M}$ an M-ary symbol alphabet, and $\boldsymbol{x}_t^{(k)} = (k, d_t^{(k)}, \boldsymbol{a}_t^{(k)})^T$, $d_t^{(k)}$ the transmitted symbol.

For future use, we introduce the notation $\pi(\boldsymbol{G})$ to denote the finite discrete random set containing the identities of the users in $\boldsymbol{G}$. This is the projection of the interferer state $\boldsymbol{G}$ onto $\mathbb{K}$, i.e., the set of indices of the active users. Similarly, $\pi'(\boldsymbol{G})$ is the projection of $\boldsymbol{G}$ onto $\mathbb{R}^d$, so that in a trained system, the elements of set $\boldsymbol{G}$ have the form $(\pi(\boldsymbol{g}), \pi'(\boldsymbol{g}))$, $\boldsymbol{g} \in \boldsymbol{G}$.

A. Measurement Model

With measurement model (1), the receiver observes a superposition of signals. Thus, the random set describing the receiver,
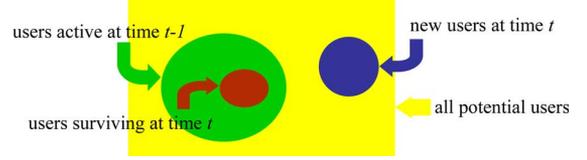

Fig. 1. Dynamic model for users at time t.

denoted $\boldsymbol{Y}_t$ is the singleton $\{\boldsymbol{y}_t\}$, where $\boldsymbol{y}_t$ has conditional probability density function (pdf) given by[1]

$$f_{\boldsymbol{Y}_t|\boldsymbol{X}_t}(\boldsymbol{y}_t|\boldsymbol{X}_t) = f_{\boldsymbol{z}}\left(\boldsymbol{y}_t - \sum_{\boldsymbol{x}_t^{(k)} \in \boldsymbol{X}_t} s(\boldsymbol{x}_t^{(k)})\right) \tag{2}$$

with $f_{\boldsymbol{z}}(\cdot)$ the pdf of the additive noise.

B. Dynamic Model

The multiuser set can be expressed in the general form

$$\boldsymbol{X}_t = \boldsymbol{S}_t \cup \boldsymbol{N}_t \tag{3}$$

with $\boldsymbol{S}_t$ the set of users surviving from time $t-1$ into time t, and $\boldsymbol{N}_t$ the set of newly born users. These two sets are related by

$$\pi(\boldsymbol{X}_{t-1}) \cap \pi(\boldsymbol{N}_t) = \emptyset \quad \text{and} \quad \pi(\boldsymbol{S}_t) \subseteq \pi(\boldsymbol{X}_{t-1}).$$

The first condition expresses the fact that no user being active at $t-1$ can migrate to the set of new users; the second states that the surviving users at t are a subset of those active at $t-1$ (see Fig. 1). Since $\boldsymbol{X}_t = \boldsymbol{S}_t \cup \boldsymbol{N}_t$, and since $\boldsymbol{S}_t$ and $\boldsymbol{N}_t$ are conditionally independent given $\boldsymbol{X}_{t-1}$, for the conditional belief functions (Appendix A) we have

$$\begin{aligned} \beta_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1}}(\boldsymbol{Z} \mid \boldsymbol{B}) &\triangleq \mathbb{P}(\boldsymbol{X}_t \subseteq \boldsymbol{Z} \mid \boldsymbol{X}_{t-1} = \boldsymbol{B}) \\ &= \beta_{\boldsymbol{S}_t|\boldsymbol{X}_{t-1}}(\boldsymbol{Z} \mid \boldsymbol{B})\beta_{\boldsymbol{N}_t|\boldsymbol{X}_{t-1}}(\boldsymbol{Z} \mid \boldsymbol{B}). \end{aligned} \tag{4}$$

We make the assumption that $\{\boldsymbol{X}_t\}_{t=1}^{\infty}$ forms a Markov set sequence, i.e., that $\boldsymbol{X}_t$ depends on its past only through $\boldsymbol{X}_{t-1}$, and that the death-and-birth process is governed by a binomial law. More precisely, denote by α the probability that a user inactive at time $t-1$ becomes active at time t, and by μ the probability that a user active at epoch $t-1$ survives into epoch t. We assume a trained system first.

The conditional belief functions (given $\boldsymbol{X}_{t-1} = \boldsymbol{B}$) of surviving and newly born user sets can be written as

$$\begin{aligned} &\beta_{\boldsymbol{S}_t|\boldsymbol{X}_{t-1}}(\boldsymbol{Z}|\boldsymbol{B}) \\ &= \sum_{j=0}^{|\pi(\boldsymbol{Z})|} \sum_{\pi(\boldsymbol{G}) \subseteq \pi(\boldsymbol{Z}):|\pi(\boldsymbol{G})|=j} f_{\pi(\boldsymbol{S}_t)|\boldsymbol{X}_{t-1}}\big(\pi(\boldsymbol{G})|\boldsymbol{B}\big) \\ &\quad \times \mathbb{P}\{\boldsymbol{S}_t \subseteq \boldsymbol{G}|\boldsymbol{X}_{t-1} = \boldsymbol{B}, \pi(\boldsymbol{S}_t) = \pi(\boldsymbol{G})\} \end{aligned} \tag{5}$$

$$\begin{aligned} &\beta_{\boldsymbol{N}_t|\boldsymbol{X}_{t-1}}(\boldsymbol{Z}|\boldsymbol{B}) \\ &= \sum_{j=0}^{|\pi(\boldsymbol{Z})|} \sum_{\pi(\boldsymbol{G}) \subseteq \pi(\boldsymbol{Z}):|\pi(\boldsymbol{G})|=j} s f_{\pi(\boldsymbol{N}_t)|\boldsymbol{X}_{t-1}}\big(\pi(\boldsymbol{G})|\boldsymbol{B}\big) \\ &\quad \times \mathbb{P}\{\boldsymbol{N}_t \subseteq \boldsymbol{G}|\boldsymbol{X}_{t-1} = \boldsymbol{B}, \pi(\boldsymbol{N}_t) = \pi(\boldsymbol{G})\} \end{aligned} \tag{6}$$

---

[1]This measurement model is not the most general one. For more complex models, including provisions for sensor failure, missed measures, multiple observers, etc., see [13].

with [6]

$$
\begin{aligned}
&f_{\pi(\boldsymbol{S}_t)|\boldsymbol{X}_{t-1}}\big(\pi(\boldsymbol{G})|\boldsymbol{B}\big)\\
&\triangleq \mathbb{P}\left\{\pi(\boldsymbol{S}_t)=\pi(\boldsymbol{G})|\boldsymbol{X}_{t-1}=\boldsymbol{B}\right\}\\
&=\begin{cases}\mu^{|\pi(\boldsymbol{G})|}(1-\mu)^{|\boldsymbol{B}|-|\pi(\boldsymbol{G})|}, & \text{if } \pi(\boldsymbol{G})\subseteq\pi(\boldsymbol{B})\\ 0, & \text{otherwise}\end{cases}
\end{aligned} \tag{7}
$$

and

$$
\begin{aligned}
&f_{\pi(\boldsymbol{N}_t)|\boldsymbol{X}_{t-1}}\big(\pi(\boldsymbol{G})|\boldsymbol{B}\big)\triangleq\mathbb{P}\left\{\pi(\boldsymbol{N}_t)=\pi(\boldsymbol{G})|\boldsymbol{X}_{t-1}=\boldsymbol{B}\right\}\\
&=\begin{cases}\alpha^{|\pi(\boldsymbol{G})|}(1-\alpha)^{K-|\boldsymbol{B}|-|\pi(\boldsymbol{G})|}, & \text{if } \pi(\boldsymbol{G})\cap\pi(\boldsymbol{B})=\emptyset\\ 0, & \text{otherwise.}\end{cases}
\end{aligned} \tag{8}
$$

Likewise, assuming that $\boldsymbol{a}_t^{(k)}$ and $\boldsymbol{a}_t^{(m)}$ are independent for $k\neq m$, we have

$$
\begin{aligned}
&\mathbb{P}\left\{\boldsymbol{S}_t\subseteq\boldsymbol{G}\mid\boldsymbol{X}_{t-1}=\boldsymbol{B},\pi(\boldsymbol{S}_t)=\pi(\boldsymbol{G})\right\}\\
&\qquad=\int_{\pi'(\boldsymbol{G})}\prod_{k\in\pi(\boldsymbol{G})}f_{\boldsymbol{a}_t^{(k)}|\boldsymbol{a}_{t-1}^{(k)}}(\boldsymbol{a}_t^{(k)}\mid\boldsymbol{b}^{(k)})d\boldsymbol{a}_t^{(k)}
\end{aligned} \tag{9}
$$

$$
\begin{aligned}
&\mathbb{P}\left\{\boldsymbol{N}_t\subseteq\boldsymbol{G}\mid\boldsymbol{X}_{t-1}=\boldsymbol{B},\pi(\boldsymbol{N}_t)=\pi(\boldsymbol{G})\right\}\\
&\qquad=\int_{\pi'(\boldsymbol{G})}\prod_{k\in\pi(\boldsymbol{G})}f_{\boldsymbol{a}_t^{(k)}}(\boldsymbol{a}_t^{(k)})\,d\boldsymbol{a}_t^{(k)}.
\end{aligned} \tag{10}
$$

Computing the set derivative of the above (see Appendix A) we obtain, with some notational abuse

$$
\begin{aligned}
f_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1}}(\boldsymbol{C}\mid\boldsymbol{B})=&\,f_{\pi(\boldsymbol{S}_t)|\boldsymbol{X}_{t-1}}(\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})\mid\boldsymbol{B})\\
&\times\prod_{k\in\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})}f_{\boldsymbol{a}_t^{(k)}|\boldsymbol{a}_{t-1}^{(k)}}(\boldsymbol{a}_t^{(k)}\mid\boldsymbol{b}^{(k)})\\
&\times f_{\pi(\boldsymbol{N}_t)|\boldsymbol{X}_{t-1}}(\pi(\boldsymbol{C})\backslash\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})\mid\boldsymbol{B})\\
&\times\prod_{k\in\pi(\boldsymbol{C})\backslash\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})}f_{\boldsymbol{a}_t^{(k)}}(\boldsymbol{a}_t^{(k)})\\
=&\,\mu^{|\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})|}(1-\mu)^{|\boldsymbol{B}|-|\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})|}\\
&\times\alpha^{|\pi(\boldsymbol{C})\backslash\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})|}\\
&\times(1-\alpha)^{K-|\boldsymbol{B}|-|\pi(\boldsymbol{C})\backslash\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})|}\\
&\times\prod_{k\in\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})}f_{\boldsymbol{a}_t^{(k)}|\boldsymbol{a}_{t-1}^{(k)}}(\boldsymbol{a}_t^{(k)}\mid\boldsymbol{b}^{(k)})\\
&\times\prod_{k\in\pi(\boldsymbol{C})\backslash\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})}f_{\boldsymbol{a}_t^{(k)}}(\boldsymbol{a}_t^{(k)}).
\end{aligned} \tag{11,12}
$$

The companion case of untrained systems may be dealt with in a very similar way. Indeed, if we denote by $d_t^{(k)}$ the data transmitted by the kth user, the Markov assumption on $\boldsymbol{X}_t$ is not violated under either one of the following conditions:

(a) the source data are memoryless;
(b) the source data form a first-order Markov chain; in this case, the densities $f_{d_t^{(k)}}(\cdot)$ and $f_{d_t^{(k)}|d_{t-1}^{(k)}}(\cdot\mid\cdot)$, are assumed known.

Under condition (b), which trivially subsumes (a), (11) generalizes to

$$
\begin{aligned}
&f_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1}}(\boldsymbol{C}\mid\boldsymbol{B})\\
&\quad=\mu^{|\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})|}(1-\mu)^{|\boldsymbol{B}|-|\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})|}\\
&\qquad\times\alpha^{|\pi(\boldsymbol{C})\backslash\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})|}
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
&\times(1-\alpha)^{K-|\boldsymbol{B}|-|\pi(\boldsymbol{C})\backslash\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})|}\\
&\times\prod_{k\in\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})}f_{d_t^{(k)}|d_{t-1}^{(k)}}(d_t^{(k)}|d_{t-1}^{(k)})\\
&\times f_{\boldsymbol{a}_t^{(k)}|\boldsymbol{a}_{t-1}^{(k)}}(\boldsymbol{a}_t^{(k)}\mid\boldsymbol{b}^{(k)})\\
&\times\prod_{k\in\pi(\boldsymbol{C})\backslash\pi(\boldsymbol{C})\cap\pi(\boldsymbol{B})}f_{d_t^{(k)}}(d_t^{(k)})f_{\boldsymbol{a}_t^{(k)}}(\boldsymbol{a}_t^{(k)}).
\end{aligned} \tag{14}
$$

III. OPTIMAL BAYESIAN SET ESTIMATORS

In causal dynamic set estimation, a random set at time t is described by its *a posteriori* belief density given the observations up to time t, which we denote $f_{\boldsymbol{X}_t|\boldsymbol{Y}_{1:t}}(\boldsymbol{X}_t\mid\boldsymbol{y}_{1:t})$. A common method for the evaluation of the above density relies on Bayesian recursions, which in RST take the form:[2]

$$
\begin{aligned}
&f_{\boldsymbol{X}_t|\boldsymbol{Y}_{1:t-1}}(\boldsymbol{X}_t\mid\boldsymbol{y}_{1:t-1})\\
&=\int f_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1}}(\boldsymbol{X}_t\mid\boldsymbol{X}_{t-1})f_{\boldsymbol{X}_{t-1}|\boldsymbol{Y}_{1:t-1}}(\boldsymbol{X}_{t-1}\mid\boldsymbol{y}_{1:t-1})\delta\boldsymbol{X}_{t-1}
\end{aligned}
$$

$$
f_{\boldsymbol{X}_t|\boldsymbol{Y}_{1:t}}(\boldsymbol{X}_t\mid\boldsymbol{y}_{1:t})\propto f_{\boldsymbol{Y}_t|\boldsymbol{X}_t}(\boldsymbol{y}_t\mid\boldsymbol{X}_t)f_{\boldsymbol{X}_t|\boldsymbol{Y}_{1:t-1}}(\boldsymbol{X}_t\mid\boldsymbol{y}_{1:t-1})
$$

for $t=1,2,\ldots$, with $f_{\boldsymbol{X}_1|\boldsymbol{y}_0}(\boldsymbol{X}_1\mid\boldsymbol{y}_0)\triangleq f_{\boldsymbol{X}_1}(\boldsymbol{X}_1)$, a known initial prior density (the notation $\delta\boldsymbol{X}_{t-1}$ indicates that the integral is a "set integral" in the sense of random-set theory).

Once $f_{\boldsymbol{X}_t|\boldsymbol{Y}_{1:t}}(\boldsymbol{X}_t|\boldsymbol{y}_{1:t})$ is made available, we are faced with the problem of defining a suitable set estimator. In general, if we assume that an observation $\boldsymbol{y}$ is given, and the *a posteriori* probabilities $f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{X}\mid\boldsymbol{y})$ have been computed, they can be used to generate a Bayesian estimator, i.e., a function $\widehat{\boldsymbol{X}}(\boldsymbol{y})$ that minimizes the risk $\mathcal{R}\triangleq\mathbb{E}[C(\boldsymbol{X},\widehat{\boldsymbol{X}})]$ associated with a given cost function C (see, e.g., [25, p. 54 ff.] –[22, p. 63 ff.]), where the expectation is taken with respect to the joint pdf of $\boldsymbol{X}$ and $\boldsymbol{y}$. Within RST, the choice of a cost function may not be an obvious task. In fact, while with discrete sets a natural choice is the minimization of error probability (i.e., the probability of choosing an erroneous set of active users and their data), with hybrid sets one must balance the cost of choosing the wrong set of users with the cost of a discrepancy between the estimated and the correct set of continuous parameters. In addition, the cost function chosen should result in an estimator with tractable complexity.

Recall that the hybrid random set $\boldsymbol{X}$ has elements $(b_i,\boldsymbol{x}_i)$, with $\boldsymbol{x}_i$ a random vector, b_i a random variable taking on a finite number of values, and $i=1,\ldots,|\boldsymbol{X}|$. We have defined $\pi(\boldsymbol{X})$ as the discrete part of $\boldsymbol{X}$, i.e., the set $\{b_1,\ldots,b_{|\boldsymbol{X}|}\}$. Also, we denote with a hat ˆ the estimated quantities. Two common cost functions related to hybrid random sets can be found in RST literature. The cost C_0 involves only the number of elements in $\boldsymbol{X}$ [13, p. 192]

$$
C_0(\boldsymbol{X},\widehat{\boldsymbol{X}})\triangleq\begin{cases}0, & |\widehat{\boldsymbol{X}}|=|\boldsymbol{X}|\\ 1, & |\widehat{\boldsymbol{X}}|\neq|\boldsymbol{X}|.\end{cases} \tag{15}
$$

In our context, this cost function implies that only the choice of the correct number of active users has relevance. The cost

---

[2]We could also seek for a fixed-lag smoothing estimation. However, in this paper we neglect this attractive alternative.

function C_1 weights the error in the estimate of the continuous part of $\boldsymbol{X}$ [13, p. 192]

$$C_1(\boldsymbol{X}, \widehat{\boldsymbol{X}}) \triangleq \begin{cases} 0, & \begin{cases} \pi(\widehat{\boldsymbol{X}}) = \pi(\boldsymbol{X}) \quad \text{and} \\ \hat{b}_i = b_i, \ i = 1, \ldots, |\boldsymbol{X}|, \ \text{and} \\ \pi'(\boldsymbol{X}), \pi'(\widehat{\boldsymbol{X}}) \in \mathcal{A} \end{cases} \\ 1, & \text{otherwise} \end{cases} \quad (16)$$

where $\mathcal{A}$ is a closed ball in $\mathbb{R}^{d|\boldsymbol{X}|}$. The choice of this function implies that no cost is incurred whenever the number of active users, their identities, and their discrete data are estimated correctly, and in addition there exists an index permutation such that the two $d|\boldsymbol{X}|$-dimensional vectors $(\widehat{\boldsymbol{x}}_1, \ldots, \widehat{\boldsymbol{x}}_{|\boldsymbol{X}|})$ and $(\boldsymbol{x}_{\tau 1}, \ldots, \boldsymbol{x}_{\tau|\boldsymbol{X}|})$ are close enough (the size of $\mathcal{A}$ reflects the definition of "closeness," i.e., the amount of tolerable difference between the two vectors). If any of these conditions is not satisfied, then the cost is 1. Before commenting on the suitability of C_1 in our context, let us observe the structure of the estimators derived from the cost functions above. Minimization of the cost function C_0 yields the estimator (whose consistency has not been proved) known as *GMAP-I* [13, p. 191], also called a *Marginal Multitarget Estimator* in [22, p. 497 ff.]

$$\widehat{n} \triangleq \arg\max_n f_{n|\boldsymbol{y}}(n \mid \boldsymbol{y}) \quad (17)$$

$$\widehat{\boldsymbol{X}}_I \triangleq \arg\max_{\substack{\boldsymbol{X} \\ |\boldsymbol{X}|=\widehat{n}}} f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{X} \mid \boldsymbol{y}) \quad (18)$$

where

$$f_{n|\boldsymbol{y}}(n \mid \boldsymbol{y}) = \int_{|\boldsymbol{X}|=n} f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{X} \mid \boldsymbol{y}) \, \delta\boldsymbol{X}. \quad (19)$$

This is a two-step estimator, which first evaluates the number of active users, next their parameters under the assumption that the number of active users is known and coincides with $\widehat{n}$. Minimization of the cost function $C_0 + C_1$ yields the consistent estimator known as *GMAP-II* [13, p. 191], [19], also called a *Joint Multitarget Estimator* in [22, p. 497 ff.]

$$\widehat{\boldsymbol{X}}_{II} \triangleq \arg\max_{\boldsymbol{X}} f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{X} \mid \boldsymbol{y}) \frac{c^{|\boldsymbol{X}|}}{|\boldsymbol{X}|!} \quad (20)$$

where c is a small constant reflecting the size of $\mathcal{A}$, and hence the accuracy to which the estimate need be performed [22, p. 500]. Unlike GMAP-I, this estimator is based on a single operation, although it can also be realized in two steps as follows [22, p. 500]. For all n, compute

$$\widehat{\boldsymbol{X}}_n \triangleq \arg\max_{|\boldsymbol{X}|=n} f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{X} \mid \boldsymbol{y})$$

next

$$\widehat{\boldsymbol{X}}_{II} \triangleq \widehat{\boldsymbol{X}}_{\hat{n}}$$

where

$$\hat{n} \triangleq \arg\max_n f_{\boldsymbol{X}|\boldsymbol{y}}(\widehat{\boldsymbol{X}}_n \mid \boldsymbol{y}) \frac{c^n}{n!}.$$

We observe now that neither of the above estimators, in spite of their relative simplicity and intuitively satisfactory structure,

is perfectly suited to the problem considered in this paper. In fact, GMAP-I does not weight the errors in the continuous parameters, while GMAP-II cost function is not sensitive to large errors in the continuous parameters. A simple family of cost functions appropriate to our problem is

$$C(\boldsymbol{X}, \widehat{\boldsymbol{X}}) \triangleq \begin{cases} Q, & \pi(\widehat{\boldsymbol{X}}) \neq \pi(\boldsymbol{X}) \\ \sum_{i \in \pi(\boldsymbol{X})} g(\boldsymbol{x}_i, \widehat{\boldsymbol{x}}_i), & \pi(\widehat{\boldsymbol{X}}) = \pi(\boldsymbol{X}). \end{cases} \quad (21)$$

This assigns cost Q to any wrong estimate of the set of active users or of their data. If this estimate is correct, the cost depends on a function g of the discrepancies between estimated and true continuous parameters. The actual selection of Q and of g reflects the relative weight of an error in the discrete and the continuous parameters. Straightforward modifications of the arguments in [13, pp. 192–193] lead to the risk associated with the cost function C

$$\begin{aligned} \mathcal{R} &\triangleq \mathbb{E}[C(\boldsymbol{X}, \widehat{\boldsymbol{X}})] \\ &= Q\, \mathbb{P}\big(\pi(\widehat{\boldsymbol{X}}) \neq \pi(\boldsymbol{X})\big) + \mathbb{P}\big(\pi(\widehat{\boldsymbol{X}}) = \pi(\boldsymbol{X})\big) \\ &\quad \times \mathbb{E}\left[\sum_{i \in \pi(\boldsymbol{X})} g(\boldsymbol{x}_i, \widehat{\boldsymbol{x}}_i) \,\middle|\, \pi(\widehat{\boldsymbol{X}}) = \pi(\boldsymbol{X}) \right]. \end{aligned} \quad (22)$$

The resulting Bayesian estimator requires the minimization of the sum (22). Besides the necessity of selecting Q and g (which may not be an obvious task), computation of the minimum may involve an unacceptable complexity. For this reason, we define a suboptimum, simpler version of the estimator, one that minimizes *separately* the two terms, and is realized with a two-step procedure. The first step minimizes the probability $\mathbb{P}\big(\pi(\widehat{\boldsymbol{X}}) \neq \pi(\boldsymbol{X}) \mid \boldsymbol{y}\big)$. This is obtained through a maximum *a posteriori* probability (MAP) estimator of $\pi(\boldsymbol{X})$, the set of active users and their data, which maximizes the conditional pdf of $\boldsymbol{X}$ by assuming the continuous parameters of $\boldsymbol{X}$ as nuisance. In the second step, $\sum_{i \in \pi(\widehat{\boldsymbol{X}})} \mathbb{E}[g(\boldsymbol{x}_i, \widehat{\boldsymbol{x}}_i)]$ is minimized. If g is a quadratic function, this second minimization yields the conditional expectation

$$\widehat{\boldsymbol{x}}_i = \mathbb{E}[\boldsymbol{x}_i \mid \boldsymbol{y}], \quad i \in \pi(\widehat{\boldsymbol{X}}). \quad (23)$$

We call GMAP-III the resulting estimator. [3]

We finally observe that, although the above discussion refers for simplicity to a single observation, we may easily extend it to the estimation of a whole sequence $\boldsymbol{X}_{1:T}$. In this case, if we constrain the estimator to be causal, that is, we have $\widehat{\boldsymbol{X}}_t = \widehat{\boldsymbol{X}}_t(\boldsymbol{y}_{1:t})$, the derivations in this section apply to the cumulated risk

$$\mathcal{R} = \sum_{t=1}^{T} \mathbb{E}\left[C(\boldsymbol{X}_t, \widehat{\boldsymbol{X}}_t) \right]$$

and result into the approximate estimators

$$\pi(\widehat{\boldsymbol{X}}_t) = \arg\max f\big(\pi(\boldsymbol{X}_t) \mid \boldsymbol{y}_{1:t}\big)$$

$$\widehat{\boldsymbol{x}}_t^{(i)} = \mathbb{E}\left[\boldsymbol{x}_t^{(i)} \mid \boldsymbol{y}_{1:t} \right], \quad i \in \pi(\widehat{\boldsymbol{X}}_t).$$

[3]Observe that, unlike GMAP-I and GMAP-II, estimator GMAP-III relies upon the condition $|\boldsymbol{X}| = |\pi(\boldsymbol{X})|$. This is the requirement that the discrete elements of the random set $\boldsymbol{X}$ be distinct. Now, this condition is obviously met in the multiuser communication scenario of this paper, where each user is uniquely identified, but may not be in multitarget applications [22], where the discrete elements could identify type or threat level of the target. In the latter applications, since different targets may share the same type or threat level, we have in general $|\boldsymbol{X}| \geq |\pi(\boldsymbol{X})|$. See also [22, p. 505 ff.].

These equations, along with their counterparts obtained for different cost functions, emphasize the relevance of Bayesian recursions.

We hasten to observe that the computational complexity involved by the exact implementation of the proposed estimation rules may be intolerable if the number of potential users is large. As in [6], joint multiuser detection and parameters estimation yields a complexity of the order $\mathcal{O}(2^K)$ for trained systems and $\mathcal{O}((|\mathbb{M}| + 1)^K)$ for untrained systems, which shows how the complexity of the optimal detector scales exponentially with the number of users. Techniques aimed at reducing the computational burden, and yet retaining reasonable performance at the price of a moderate loss of optimality, are available: see, e.g., [2].

IV. NUMERICAL APPROXIMATIONS OF BAYESIAN RECURSIONS

We have seen how Bayesian recursive filtering provides a framework for optimal set estimation enabling sequential evaluation of the multiuser set *a posteriori* density. Now, the main problem with Bayesian recursive filtering for random sets defined over hybrid spaces is the lack of a closed form to implement the inference. In fact, for hybrid stochastic systems, not even the linear-Gaussian case admits a closed form [4],[4] which calls for approximate methods. The simplest among these use grid-based filters, in which the continuous parameters are discretized into a fixed grid. The main difficulty with this approach is that, for reasonable accuracy, a very dense grid is needed, which may easily result in unmanageable computational burden. A more efficient approach relies on SMC, or "particle filtering," methods (see, e.g., [9] and references therein). Suppose we have a set of random samples $\{\boldsymbol{X}_{t-1,i}\}_{i=1}^N$, each associated with a weight $\{w_{t-1,i}\}_{i=1}^N$, such that the sample–weight pairs ("particle") represent the density $f_{\boldsymbol{X}_{t-1}|\boldsymbol{Y}_{1:t-1}}$ as

$$f_{\boldsymbol{X}_{t-1}|\boldsymbol{Y}_{1:t-1}}(\boldsymbol{X}_{t-1} \mid \boldsymbol{y}_{1:t-1}) \approx \sum_{i=1}^N w_{t-1,i} m_{\boldsymbol{X}_{t-1}}(\boldsymbol{X}_{t-1,i}) \tag{24}$$

where $m_{\boldsymbol{X}_t}(\boldsymbol{A})$ is the "0–1" measure defined by

$$\int_{\boldsymbol{C}} m_{\boldsymbol{X}_t}(\boldsymbol{A}) \, \delta \boldsymbol{X}_t = \begin{cases} 1, & \text{if } \boldsymbol{A} \subseteq \boldsymbol{C} \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

SMC methods provide a rule for propagating and updating these samples and weights to obtain a set of new particles $\{\boldsymbol{X}_{t,i}, w_{t,i}\}_{i=1}^N$ representing $f_{\boldsymbol{X}_t|\boldsymbol{Y}_{1:t}}$ as

$$f_{\boldsymbol{X}_t|\boldsymbol{Y}_{1:t}}(\boldsymbol{X}_t|\boldsymbol{y}_{1:t}) \approx \sum_{i=1}^N w_{t,i} m_{\boldsymbol{X}_t}(\boldsymbol{X}_{t,i}). \tag{26}$$

Among the various SMC methods, the *bootstrap filter* [14] has received considerable attention for its simplicity. With this, particles are recursively updated as follows:

$$\boldsymbol{X}_{t,i} \sim f_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1,i}), \quad i = 1, 2, \ldots, N \tag{27}$$

$$\hat{w}_{t,i} = w_{t-1,i} f_{\boldsymbol{Y}_t|\boldsymbol{X}_t}(\boldsymbol{y}_t|\boldsymbol{X}_{t,i}) \tag{28}$$

[4]The only case of hybrid systems that admits an optimum closed expression seems to be the jump-Markov linear-Gaussian system [11]. Yet, this expression is not a recursive one, and the optimum rule has a complexity which grows exponentially with the frame length.

$$w_{t,i} = \frac{\hat{w}_{t,i}}{\sum_{j=1}^N \hat{w}_{t,j}}. \tag{29}$$

The feature that makes the bootstrap filter especially attractive for approximation of the Bayesian recursive filter is its asymptotic optimality [10]. Indeed, as N grows to infinity, under weak technical assumptions, [8] proves almost sure convergence to the true ones of the empirical distributions generated by particle filtering.

A common problem with (27)–(29) stems from the *degeneracy* phenomenon, which causes all but one particle to have negligible weights after a few iterations [10]. A brute-force approach to reducing its effect consists of using a very large N, thus increasing the computational effort. To reduce the degeneracy phenomenon with low complexity one could use *adaptive resampling*. This consists of eliminating particles that have small weights, and concentrating on particles with large weights whenever the degeneracy becomes relevant. An indicator of this relevance is the *effective sample size*, $N_{\text{eff}} \triangleq 1/\sum_{i=1}^N w_{t,i}^2$. Here, $N_{\text{eff}} \in [1, N]$, with a small N_{eff} (in practice, smaller than a threshold N_{th} typically equal to $N/4$ or $N/2$ in most applications) suggesting degeneracy. The resampling step involves generating a new set of particles $\{\boldsymbol{X}_{t,i}^*, w_{t,i}^*\}_{i=1}^N$ by resampling (with replacement) N times from the approximate posterior density in (26), computed by (27)–(29), in such a way that

$$P(\boldsymbol{X}_{t,i}^* = \boldsymbol{X}_{t,i}) = w_{t,i}. \tag{30}$$

The new weights are set to be uniform, i.e., $w_{t,i}^* = 1/N$. Finally, if resampling is performed, the *a posteriori* density is set to be

$$f_{\boldsymbol{X}_t|\boldsymbol{Y}_{1:t}}(\boldsymbol{X}_t \mid \boldsymbol{Y}_{1:t}) \approx \frac{1}{N} \sum_{i=1}^N m_{\boldsymbol{X}_t}(\boldsymbol{X}_{t,i}^*). \tag{31}$$

With a representation of $f_{\boldsymbol{X}_t|\boldsymbol{Y}_{1:t}}(\boldsymbol{X}_t \mid \boldsymbol{y}_{1:t})$ in term of particles, the estimators described in Section III can be implemented in a rather straightforward manner.

We would like to point out that the bootstrap filter, in spite of its versatility and simplicity, is not the most efficient way of approximating the Bayesian recursive filter. Indeed, if the system is conditionally linear and Gaussian (a condition that can be granted in our case by assuming that the users amplitude evolves according to a Gauss–Markov model), then a Rao–Blackwellized particle filter (RBPF) can be used. In a nutshell, RBPFs adopt an SMC algorithm to track births and deaths of users, while, conditioned on the particles realizations, their amplitudes are exactly estimated through Kalman filtering [10]. Since the RBPF draws particles from a discrete space, the number of particles required by a the latter is on average much smaller than the number of particles required by the bootstrap filter. RBPF has been previously applied in RST for multitarget tracking [27] and for estimating abruptly time-varying multipath channels in multiple-input, multiple-output (MIMO) systems based on orthogonal frequency-division multiplexing (OFDM) [3]. Since we are mainly concerned with a proof-of-concept rather than with implementation details, here it suffices to add that, since in most of the cases dealt with in this paper linearity

and Gaussianity cannot be granted, the bootstrap filter enables the evaluation of the limiting performance.

V. AN APPLICATION: JOINT MULTIUSER DETECTION AND AMPLITUDE ESTIMATION IN CDMA

Following in the footpath of [6], in this section we apply the methodologies described above to the scenario of an uncoded, synchronous, single-rate DS-CDMA system with a maximum number K of users, processing gain L, and additive white Gaussian noise. Let us assume coherent detection, while the amplitudes of transmitted signals are unknown. The chip-wise matched-filtered received signal at time t has the form

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{v}(\boldsymbol{X}_t) + \boldsymbol{z}_t, \quad t = 1, 2, \ldots, T \tag{32}$$

where
(a) $\boldsymbol{S}$ is an $L \times K$ matrix whose columns contain the spreading codes of the users.
(b) $\boldsymbol{v}(\boldsymbol{X}_t)$ is a K-dimensional vector with nonzero entries in the locations dictated by $\pi(\boldsymbol{X}_t)$. In particular, we have

$$v_m(\boldsymbol{X}_t) = \begin{cases} d_t^{(m)} a_t^{(m)}, & \text{if } m \in \pi(\boldsymbol{X}_t) \\ 0, & \text{otherwise} \end{cases} \tag{33}$$

where, as before, $d_t^{(m)}$ denote the datum of user m at time t, while $a_t^{(m)}$ is the corresponding received amplitude.
(c) $\boldsymbol{z}_t \sim \mathcal{N}(0, N_0/2\boldsymbol{I}_L)$ is an L-dimensional white Gaussian noise vector, where $N_0/2$ is the power spectral density of the received noise, and $\boldsymbol{I}_L$ denotes the $L \times L$ identity matrix.

As a consequence, the set $\boldsymbol{X}_t$ consists of a random number of elements, each element containing the active-user identity, the transmitted data, and the received power (in random-set theoretical parlance, the hybrid space is now $\mathbb{S} = \mathbb{K} \times \mathbb{M} \times \mathbb{R}^+$, with $\mathbb{M}$ the symbol alphabet). Thus, the measurement model is

$$f_{\boldsymbol{Y}_t|\boldsymbol{X}_t}(\boldsymbol{y}_t|\boldsymbol{X}_t) = f_{\boldsymbol{z}}(\boldsymbol{y}_t - \boldsymbol{S}\boldsymbol{v}(\boldsymbol{X}_t)). \tag{34}$$

We consider a binomial birth-and-death process with parameters α and μ, as in (7)–(8). Assuming independent, equally-likely symbols, and memoryless modulation, we have, defining $M \triangleq |\mathbb{M}|$

$$f(d_t^{(k)} \mid d_{t-1}^{(k)}) = f(d_t^{(k)}) = \frac{1}{M}, \quad \forall k \in \pi(\boldsymbol{X}_t). \tag{35}$$

Denote now by $P_t^{(k)} \triangleq \left(a_t^{(k)}\right)^2$ the power received from the kth user at time t. Its level may vary with time, due to both user mobility and fading. Using the model and the notations of Appendix A, we define the random variable $\eta_t^{(k)}$ such that

$$P_t^{(k)} = \eta_t^{(k)} P_{t-1}^{(k)}. \tag{36}$$

Thus

$$f_{P_t^{(k)}|P_{t-1}^{(k)}}(P_t^{(k)} \mid P_{t-1}^{(k)}) = \frac{1}{P_{t-1}^{(k)}} f_{\eta_t^{(k)}}\left(\frac{P_t^{(k)}}{P_{t-1}^{(k)}}\right) \tag{37}$$

whereby the amplitude-transition model can be obtained as

$$a_t^{(k)} = \sqrt{\eta_t^{(k)}} a_{t-1}^{(k)} \tag{38}$$

yielding

$$f_{a_t^{(k)}|a_{t-1}^{(k)}}(a_t^{(k)} \mid a_{t-1}^{(k)}) = \frac{1}{a_{t-1}^{(k)}} f_{\sqrt{\eta_t^{(k)}}}\left(\frac{a_t^{(k)}}{a_{t-1}^{(k)}}\right) \tag{39}$$

and

$$f_{\sqrt{\eta_t^{(k)}}}(w) = \begin{cases} 2w f_{\eta_t^{(k)}}(w^2), & \text{for } w > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{40}$$

With this, the set densities in (11) and (13) are fully specified.

A. Trained CDMA

A synchronous, uncoded, single-rate DS-CDMA system with processing gain $L = 7$ is assumed in this computer simulation. The maximum number of active users is $K = 3$. Each user is assigned an m-sequence of length 7, and the transmission has duration $T = 10$. Equations (11) and (34) describe the measurement model and the dynamic model, respectively. The birth rate and the persistence probability take values $\alpha = 0.2$ and $\mu = 0.8$, respectively. The parameters of the power-evolution model are $\rho^{(k)} = 0.9999$, $\varepsilon^{(k)} = 1.2$, and $\beta^{(k)} = 0.5$ for all k. We assume that the users transmit known symbols (trained system), whereby only the identities and the transmitted powers of active users are to be estimated. In approximating Bayesian recursions, a bootstrap filter is used with $N = 1000$ particles and threshold $N_{\text{th}} = N/4$. Define the *Discrete Set Error Probability* (DSEP) as $\mathbb{P}(\pi(\widehat{\boldsymbol{X}}_t) \neq \pi(\boldsymbol{X}_t))$. Since this macro-parameter only measures the system ability to identify the active users, a further performance measure is needed to assess the quality of the power estimates. To this end, we use the standard root-mean square error (RMSE)

$$\text{RMSE} \triangleq T^{-1} \sum_{t=1}^{T} \frac{1}{|\pi(\widehat{\boldsymbol{X}}_t) \cap \pi(\boldsymbol{X}_t)|}$$
$$\times \left(\sum_{k \in \pi(\widehat{\boldsymbol{X}}_t) \cap \pi(\boldsymbol{X}_t)} \left| \hat{a}_t^{(k)} - a_t^{(k)} \right|^2 \right)^{1/2} \tag{41}$$

which makes sense as long as $\pi(\widehat{\boldsymbol{X}}_t) \cap \pi(\boldsymbol{X}_t) \neq \emptyset$.

Fig. 2 shows the DSEP versus the signal-to-noise ratio (SNR) for the GMAP-I, GMAP-II, and GMAP-III estimators. As a baseline for performance comparisons, we also report the DSEP corresponding to complete channel–state information (labeled CCSI), i.e., perfect knowledge of the channel parameters (in this case, the power). Fig. 3 shows the distance (41) for the GMAP-I, GMAP-II, and GMAP-III estimators.

Notice from Fig. 2 that GMAP-I and GMAP-III perform equivalently in terms of DSEP, both being slightly better than GMAP-II. This behavior can be justified by noticing that both GMAP-I and GMAP-III rely upon a marginalization of the conditional belief density with respect to the continuous parameters of the random sets. Moreover, since GMAP-I
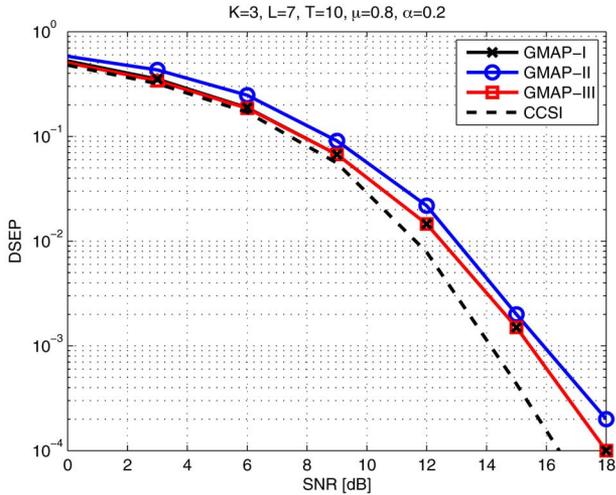
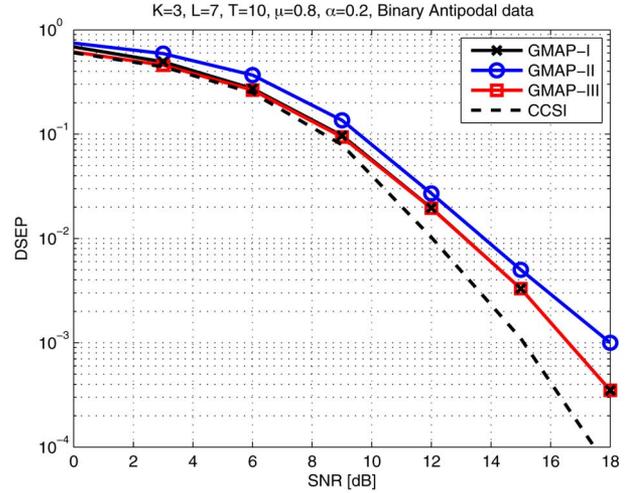Fig. 2. DSEP for various receivers (trained CDMA).

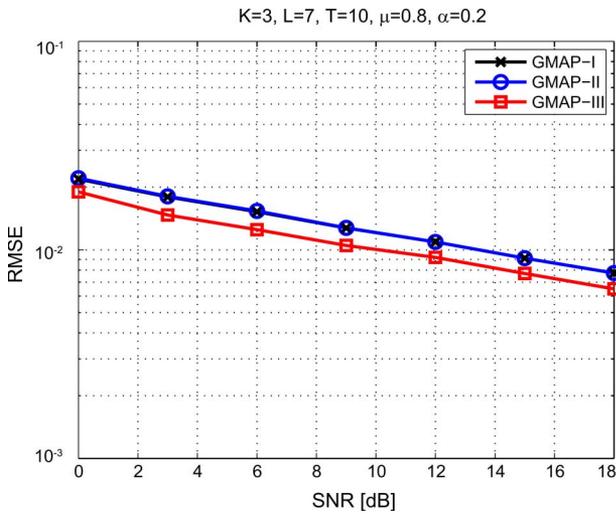Fig. 4. DSEP for various receivers (untrained CDMA).

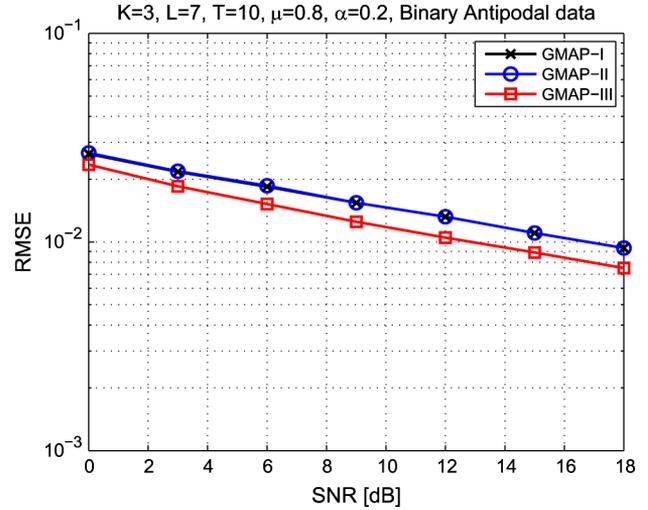Fig. 3. RMSE for various receivers (trained CDMA).

Fig. 5. RMSE error for various receivers (untrained CDMA).

estimates only the cardinality of the set of active users, while GMAP-III also estimates their identities, and since an error on $|\pi(\boldsymbol{X}_t)|$ necessarily implies an error on $\pi(\boldsymbol{X}_t)$ (but not *vice versa*), we argue that the most critical aspect of active-users identification is the estimate of the cardinality of their set. This statement is consistent with observations in [6], where it has been shown that erroneous set estimation typically occurs due to errors in set cardinality. As far as estimation of continuous parameters is concerned, the advantage of GMAP-III over GMAP-I and GMAP-II is apparent from Fig. 3, which shows the RMSE for the three estimators examined here. This result is also consistent with standard estimation theory. In fact, the Expected *A Posteriori* (EAP) estimator, i.e., the second step of the GMAP-III estimator, is the optimal estimator as far as mean-squared error (MSE) is the performance measure.

B. Untrained CDMA

Here the system parameters are the same as in Section V-A, and binary antipodal transmission is assumed ($\mathbb{M} = \{\pm 1\}$). DSEP is now defined as the probability that the estimated set containing the identities *and* the data of active users differs from

the true set in either its cardinality, or its elements, or both. Fig. 4 shows such a DSEP for GMAP-I, GMAP-II, GMAP-III, and CCSI receivers, while Fig. 5 shows the set distance for GMAP-I, GMAP-II, and GMAP-III.

From Fig. 4, we observe that trained and untrained CDMA perform very similarly as far as DSEP is concerned, and that, as in trained CDMA, GMAP-I and GMAP-III are practically equivalent. This is again consistent with observations in [6], where it was noticed how errors in joint user identification and data detection are mainly caused by erroneous estimation of the cardinality of the set of active users. Fig. 5 confirms the superiority of GMAP-III in the estimation of the continuous parameters of the active users.

VI. Conclusion

We have extended the theory presented in [6] to encompass the study of multiuser detection in an environment where not only the number and the identities of active users, but also their continuous parameters, are unknown and must be estimated. This problem is solved by applying the theory of random finite sets constructed on hybrid spaces. In particular, we have derived

Bayesian recursions describing the evolution with time of *a posteriori* densities of the unknown parameters and data. Unlike in [6], where the optimal estimator can be expressed in a closed form yielding the limiting performance of multiuser detection with an unknown and time-varying number of users, here, since we deal with joint multiuser detection and continuous-parameter estimation, Bayesian recursions cannot be solved exactly. Among the various approximation techniques, numerical approximations based on SMC methods ("particle filtering") were used, yielding asymptotically optimum set estimators and enabling numerical evaluation of the limiting performances.

APPENDIX
COMPUTING BELIEF DENSITIES

In this Appendix we show how belief densities can be computed in the context of the problem described in this paper. For simplicity, we consider a static environment, and the set of random variables modeling the continuous unknown parameters of the K potential users, assumed to be independent and equally distributed. We denote by $P(\cdot)$ the probability measure of each set, and by $p(\cdot)$ the corresponding pdf. Moreover, α denote the probability that a user is active.

According to RST, the belief density of the random set $\boldsymbol{X}$, denoted $f_{\boldsymbol{X}}(\cdot)$, is defined through the set derivative of the belief function of $\boldsymbol{X}$. Thus, we first need the belief function

$$\beta_{\boldsymbol{X}}(\boldsymbol{S}) \triangleq \mathbb{P}(\boldsymbol{X} \subseteq \boldsymbol{S}) \qquad (42)$$

of a subset $\boldsymbol{S}$ of interferers. In general

$$\beta_{\boldsymbol{X}}(\boldsymbol{S}) = \sum_{j=0}^{|\boldsymbol{S}|} \alpha^j (1-\alpha)^{K-j} \sum_{\boldsymbol{T}_j(\boldsymbol{S})} \prod_{i=1}^{j} P(\{t_i\}) \qquad (43)$$

where $\boldsymbol{T}_j(\boldsymbol{S}) \triangleq \{t_1, \ldots, t_j\}$ denotes one of the $\binom{|\boldsymbol{S}|}{j}$ subsets of $\boldsymbol{S}$ with j elements.[5]

Next, the belief density is obtained by computing the set derivative of (43) at $\boldsymbol{S} = \emptyset$

$$f_{\boldsymbol{X}}(\boldsymbol{Z}) = \left. \frac{\delta \beta(\boldsymbol{S})}{\delta \boldsymbol{Z}} \right|_{\boldsymbol{S}=\emptyset}. \qquad (44)$$

As illustrated for example in [13, p. 163], the value $f_{\boldsymbol{X}}(\boldsymbol{Z})$ of the belief density specifies the likelihood with which the random set $\boldsymbol{X}$ takes the set $\boldsymbol{Z}$ as its specific realization.

We observe the following.
(1) The belief function of a *random vector* is just its ordinary probability measure. Thus, if $\boldsymbol{X} = \{\boldsymbol{x}\}$, then

$$\beta_{\boldsymbol{X}}(\boldsymbol{S}) = \mathbb{P}(\boldsymbol{X} \subseteq \boldsymbol{S}) = \mathbb{P}(\boldsymbol{x} \in \boldsymbol{S}) = P(\boldsymbol{S})$$

and the belief density corresponds to the ordinary pdf $f(\boldsymbol{s})$, with $\boldsymbol{S} = \{\boldsymbol{s}\}$: more precisely, we have

$$f_{\boldsymbol{X}}(\boldsymbol{S}) = \begin{cases} f(\boldsymbol{s}), & \boldsymbol{S} = \{\boldsymbol{s}\} \\ 0, & |\boldsymbol{S}| \neq 1. \end{cases}$$

[5]Observe how the belief function has the general form examined in [26, Theorem 5.17, p. 52].

(2) If a is a constant, then [26, p. 56]

$$\frac{\delta}{\delta \boldsymbol{Z}} a = 0, \quad \text{if } \boldsymbol{Z} \neq \emptyset.$$

(3) The set derivative of $P(\{\boldsymbol{s}_1\}) \times P(\{\boldsymbol{s}_2\})$ is obtained by using the following property [26, Definition 5.22, p. 55]:

$$\frac{\delta \Phi}{\delta \boldsymbol{Z}}(\boldsymbol{T}) = \begin{cases} \frac{\delta^k}{\delta \boldsymbol{s}_1 \cdots \delta \boldsymbol{s}_k} \Phi(\boldsymbol{T}), & \boldsymbol{Z} \neq \emptyset \\ \Phi(\boldsymbol{T}), & \boldsymbol{Z} = \emptyset \end{cases}$$

where Φ is a set function, and $\boldsymbol{Z} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k\}$. We obtain

$$\frac{\delta}{\delta \boldsymbol{Z}} P(\{\boldsymbol{s}_1\}) \times P(\{\boldsymbol{s}_2\})$$

$$= \begin{cases} P(\{\boldsymbol{s}_1\}) \times P(\{\boldsymbol{s}_2\}), & \boldsymbol{Z} = \emptyset \\ p(\boldsymbol{s}_1) \times P(\{\boldsymbol{s}_2\}), & \boldsymbol{Z} = \{\boldsymbol{s}_1\} \\ P(\{\boldsymbol{s}_1\}) \times p(\boldsymbol{s}_2), & \boldsymbol{Z} = \{\boldsymbol{s}_2\} \\ p(\boldsymbol{s}_1) \times p(\boldsymbol{s}_2), & \boldsymbol{Z} = \{\boldsymbol{s}_1, \boldsymbol{s}_2\} \\ 0, & |\boldsymbol{Z}| > 2. \end{cases}$$

Applying the above to the belief function (43), we obtain

$$f_{\boldsymbol{X}}(\boldsymbol{Z})$$
$$= \begin{cases} (1-\alpha)^K, & \boldsymbol{Z} = \emptyset \\ \alpha^j (1-\alpha)^{K-j} \prod_{\ell=1}^{j} p(\boldsymbol{s}_{i_\ell}), & \boldsymbol{Z} = \{\boldsymbol{s}_{i_1}, \ldots, \boldsymbol{s}_{i_j}\} \subseteq \boldsymbol{S} \\ 0, & |\boldsymbol{Z}| > K \end{cases}$$

Here we model the effects of user's mobility and of fading on the power received from user k at discrete time t, denoted $P_t^{(k)}$. The superposition of these two effects yields a model for the dynamics of signal-amplitude variations.

We denote by $x_t^{(k)}$ the noiseless part of the signal received from user k at time t. This is a deterministic function of the random power $P_t^{(k)}$, so we can write

$$x_t^{(k)} = x(P_t^{(k)}) \qquad (45)$$

and hence

$$p(x_{t+1}^{(k)} \mid x_t^{(k)}) = p(x(P_{t+1}^{(k)}) \mid P_t^{(k)}). \qquad (46)$$

For notational simplicity, in the balance of this Appendix we omit the superscript k associate with the user.

A. Effect of Fading

Denoting by P the reference power, and by R_t the fading amplitude at time t, we may write

$$P_t = R_t^2 P \qquad (47)$$

and hence

$$P_{t+1} = \frac{R_{t+1}^2}{R_t^2} P_t = \gamma_t P_t \qquad (48)$$

where γ_t is the random variable defined as

$$\gamma_t \triangleq \frac{R_{t+1}^2}{R_t^2} \qquad (49)$$

B. Effect of Motion

We consider two classes of users: one includes stationary users, the other includes users in motion. The latter class experiences a power variation from time t to time $t+1$ equal to $\varepsilon^{\pm 1}$. Thus, we can write, for the effect of motion

$$P_{t+1} = \delta_t P_t \tag{50}$$

where δ_t is a random variable taking values

$$\delta_t = \begin{cases} 1, & \text{with probability } 1 - \lambda \\ \varepsilon, & \text{with probability } \lambda/2 \\ \varepsilon^{-1}, & \text{with probability } \lambda/2 \end{cases} \tag{51}$$

where $\varepsilon > 1$ depends on the user's velocity, and λ may be called the *mobility factor* of the user.

C. Joint Effects of Fading and Motion

We have

$$P_{t+1} = \gamma_t \delta_t P_t = \eta_t P_t \tag{52}$$

where the random variable

$$\eta_t \triangleq \gamma_t \delta_t \tag{53}$$

models the joint effects of fading and motion. Hence

$$p\big(x(P_{t+1}) \mid x(P_t)\big) = p\big(x(\eta_t P_t) \mid P_t\big) \tag{54}$$

and the dynamic modeling problem is reduced to the problem of modeling statistically the random process η_n. Our basic assumptions here are that fast fading and motion are independent, and that the fading process is first-order Markov (see [29]).

D. Modeling η_n With Rayleigh Fading

1) R_n, R_{n+1} Independent Rayleigh: If R_n and R_{n+1} are *independent* Rayleigh random variables, with normalized pdf

$$p_R(r) = 2re^{-r^2}, \quad r > 0 \tag{55}$$

then the pdf of $S \triangleq R^2$ is exponential

$$p_S(s) = e^{-s}, \quad s > 0 \tag{56}$$

and hence the ratio $\gamma_{n+1} = R_{n+1}^2/R_n^2$ has pdf

$$p_\gamma(r) = \int_0^\infty y e^{-ry} e^{-y} \, dy = \frac{1}{(1+r)^2}, \quad r > 0. \tag{57}$$

Consequently, with the assumptions above we obtain the following pdf for the random variable η_n:

$$p_\eta(y) = (1-\lambda)\frac{1}{(1+y)^2} + \frac{\lambda}{2}\frac{\varepsilon}{(\varepsilon+y)^2} + \frac{\lambda}{2}\frac{\varepsilon^{-1}}{(\varepsilon^{-1}+y)^2} \tag{58}$$

valid for $y > 0$.

2) R_n, R_{n+1} Correlated Rayleigh: In this case, we assume the general expression provided in [24, eq. (121)], for the bivariate pdf of correlated exponentials $(s_1, s_2) \triangleq (R_1^2, R_2^2)$ with correlation ρ

$$p_{s_1, s_2}(s_1, s_2)$$
$$= \frac{1}{1-\rho^2} \exp\left(-\frac{1}{1-\rho^2}(s_1 + s_2)\right) I_0\left(\frac{2|\rho|}{1-\rho^2}\sqrt{s_1 s_2}\right)$$
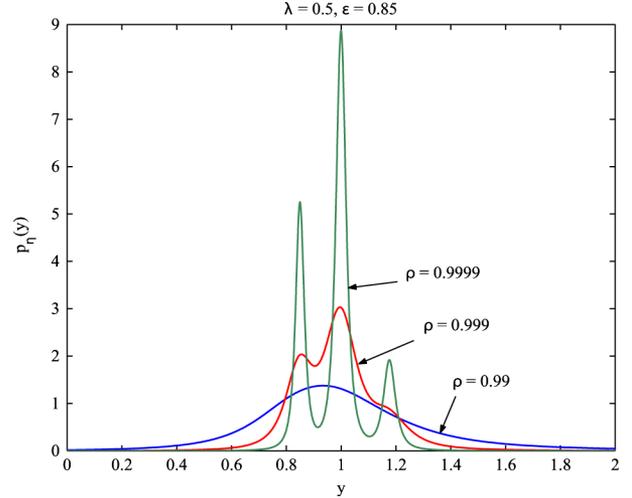

Fig. 6. Probability density function of the random variable η defined in (53).

where $s_1, s_2 > 0$. The pdf of the ratio $\gamma \triangleq R_2^2/R_1^2$ is obtained from the general formula

$$p_\gamma(z) = \int_0^\infty x p_{s_1, s_2}(zx, x) \, dx$$
$$= \frac{1}{1-\rho^2} \int_0^\infty x \exp\left(-\frac{z+1}{1-\rho^2}x\right) I_0\left(\frac{2|\rho|\sqrt{z}}{1-\rho^2}x\right) dx$$
$$= (1-\rho^2)\frac{1+z}{((1+z)^2 - 4\rho^2 z)^{3/2}}$$

which yields

$$p_\eta(y) = (1-\lambda)(1-\rho^2)\frac{(1+y)}{((1+y)^2 - 4\rho^2 y)^{3/2}}$$
$$+ \frac{\lambda}{2}\frac{\varepsilon(1-\rho^2)(1+\varepsilon y)}{((1+\varepsilon y)^2 - 4\rho^2 \varepsilon y)^{3/2}}$$
$$+ \frac{\lambda}{2\varepsilon}\frac{(1-\rho^2)(1+\frac{y}{\varepsilon})}{((1+\frac{y}{\varepsilon})^2 - 4\rho^2\frac{y}{\varepsilon})^{3/2}}.$$

Fig. 6 illustrates the behavior of $p_\eta(y)$ for some values of its parameters.

REFERENCES

[1] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Autom. Contr.*, vol. 17, pp. 439–448, Aug. 1972.

[2] D. Angelosante, E. Biglieri, and M. Lops, "Low-complexity receivers for multiuser detection with an unknown number of active users," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 3481–3484.

[3] D. Angelosante, E. Biglieri, and M. Lops, "Sequential estimation of time-varying multipath channel for MIMO-OFDM systems," in *Proc. IIEEE Int. Symp. Information Theory (ISIT)*, Toronto, ON, Canada, Jul. 2008, pp. 589–593.

[4] Y. Bar-Shalom, S. Challa, and H. A. P. Blom, "IMM estimator versus optimal estimator for hybrid systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, pp. 986–991, Jul. 2005.

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

[6] E. Biglieri and M. Lops, "Multiuser detection in a dynamic environment. Part I: User identification and data detection," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3158–3170, Sep. 2007.

[7] R. Chen, X. Wang, and J. S. Liu, "Adaptive joint detection and decoding in flat-fading channels via mixture Kalman filtering," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2079–2094, Sep. 2000.

[8] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 736–746, Mar. 2002.

[9] P. M. Djurić *et al.*, "Particle filtering," *IEEE Signal Process. Mag.*, pp. 19–38, Sep. 2003.

[10] A. Doucet, On Sequential Monte Carlo Methods for Bayesian Filtering Dept. Eng., Univ. Cambridge, U.K., Tech. Rep., 1998.

[11] A. Doucet, N. J. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump Markov linear systems," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 613–624, Mar. 2001.

[12] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

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

[14] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Proc. IEE—Radar and Signal Process.*, vol. 140, pp. 107–113, Apr. 1993.

[15] J. R. Hoffman and R. P. S. Mahler, "Multitarget miss distance via optimal assignment," *IEEE Trans. Systems, Man Cybern.*, vol. 34, pp. 327–336, May 2004.

[16] V. P. Ipatov, *Spread Spectrum and CDMA*. Chichester, U.K.: Wiley, 2005.

[17] A. H. Jazwinski, "Adaptive filtering," *Automatica*, vol. 5, pp. 475–575, 1969.

[18] J. H. Lodge and M. L. Moher, "Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 38, no. 6, pp. 787–794, Jun. 1990.

[19] R. Mahler, *An Introduction to Multisource-Multitarget Statistics and its Applications*. Eagan, MD: Lockheed Martin, 2000, Technical Monograph.

[20] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.

[21] R. Mahler, "Random sets: Unification and computation for information fusion—A retrospective assessment," in *Proc. 7th Int. Conf. Information Fusion*, Stockholm, Sweden, Jun. 2004, pp. 1–20.

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

[23] D. Makrakis, P. T. Mathiopoulos, and D. P. Bouras, "Optimal decoding of coded PSK and QAM signals in correlated fast fading channels and AWGN: A combined envelop, multiple differential and coherent detection approach," *IEEE Trans. Commun.*, vol. 42, no. 1, pp. 63–75, Jan. 1994.

[24] R. K. Mallik, "On multivariate Rayleigh and exponential distributions," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1499–1515, Jun. 2003.

[25] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. New York: Wiley, 1968.

[26] M. Vihola, "*Random Sets for Multitarget Tracking and Data Fusion*," Licentiate thesis, Tampere University of Technology, Tampere, Finland, 2004.

[27] M. Vihola, "Rao-Blackwellised particle filtering in random sets for multitarget tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 2, pp. 689–705, Apr. 2007.

[28] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multi-target filtering with random finite sets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 4, pp. 1224–1245, Oct. 2005.

[29] H. S. Wang and P.-C. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 353–357, May 1996.

Daniele Angelosante was born in Frosinone, Italy, on June 1, 1981. He received the B.Sc. and *Laurea Magistrale* degrees in telecommunication engineering from the University of Cassino, Cassino, Italy, in 2003 and 2005, respectively. He received the M.Sc. degree in electrical engineering with specialization in signal and information processing for communication from the University of Aalborg, Aalborg, Denmark, in 2005.

From April to June 2007, he worked at the University Pompeu Fabra, Spain. In 2008, he was a Visiting Scholar at the University of Minnesota, Minneapolis, MN. He is currently working towards the Ph.D. degree at the University of Cassino. His research interests lie in the areas of statistical signal processing, with emphasis on communications and tracking, and statistical learning theory.

Ezio Biglieri (M'73–SM'82–F'89) was born in Aosta, Italy. He received his formal training in Electrical Engineering at Politecnico di Torino, Torino, Italy, where he received the Dr. Engr. degree in 1967.

He is presently an Adjunct Professor of Electrical Engineering at University of California Los Angeles (UCLA) and an honorary Professor at Universitat Pompeu Fabra, Barcelona, Spain. Previously he was a Professor at the University of Napoli, 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 École Nationale Supérieure des Télécommunications, Paris, France, the University of Sydney, Australia, the Yokohama National University, Japan, the Electrical Engineering Department of Princeton University, Princeton, NJ, the University of South Australia, Adelaide, the University of Melbourne, Melbourne, Australia, the Institute for Communications Engineering, Munich Institute of Technology, Munich, Germany, the Institute for Infocomm Research, National University of Singapore, the National Taiwan University, Taipei, Republic of China, the University of Cambridge, Cambridge, U.K., and ETH Zurich, Switzerland.

Prof. Biglieri 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 currently the Editor-in-Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY and of the *Journal of Communications and Networks*. 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 (M'96–SM'01) was born in Napoli, Italy, on March 16, 1961. He received the Dr. Eng. degree in electronic engineering from the University of Naples, Naples, Italy, in 1986.

In 1987, he joined the Department of Electronic and Telecommunications Engineering of the University of Naples as a Ph.D. student in Electronic Engineering. From 1991 to 2000, he was an Associate Professor of Radar Theory and Digital Transmission Theory at the University of Naples, while, since March 2000, he has been a Full Professor at the University of Cassino, Cassino, Italy, engaged in research in the field of statistical signal processing, with emphasis on radar processing and spread spectrum multiuser communications. He also held teaching positions at the University of Lecce, Italy, and, during 1991, 1998, and 2000, he was on sabbatical leave at the University of Connecticut, Storrs; Rice University, Houston, TX; and Princeton University, Princeton, NJ; respectively. During 2008 he was a Visiting Professor with University of Minnesota, Minneapolis, and Columbia University, New York City.