

# Sequential Estimation of Multipath MIMO-OFDM Channels

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

**Abstract**—Wireless “MIMO” systems, employing multiple transmit and receive antennas, promise a significant increase of channel capacity, while orthogonal frequency-division multiplexing (OFDM) is attracting a good deal of attention due to its robustness to multipath fading. Thus, the combination of both techniques is an attractive proposition for radio transmission. The goal of this paper is the description and analysis of a new and novel pilot-aided estimator of multipath block-fading channels. Typical models leading to estimation algorithms assume the number of multipath components and delays to be constant (and often known), while their amplitudes are allowed to vary with time. Our estimator is focused instead on the more realistic assumption that the number of channel taps is also unknown and varies with time following a known probabilistic model. The estimation problem arising from these assumptions is solved using Random-Set Theory (RST), whereby one regards the multipath-channel response as a single set-valued random entity. Within this framework, Bayesian recursive equations determine the evolution with time of the channel estimator. Due to the lack of a closed form for the solution of Bayesian equations, a (Rao–Blackwellized) particle filter (RBPF) implementation of the channel estimator is advocated. Since the resulting estimator exhibits a complexity which grows exponentially with the number of multipath components, a simplified version is also introduced. Simulation results describing the performance of our channel estimator demonstrate its effectiveness.

**Index Terms**—Bayes theory, channel estimation, MIMO-OFDM, particle filtering, random finite set theory, sequential Monte Carlo.

## I. INTRODUCTION

WIRELESS multiple-input-multiple-output (MIMO) systems, employing several transmit and receive antennas, are known to provide a significant increase of channel capacity [3], [9]. Combination of MIMO with orthogonal-frequency-division multiplexing (OFDM), which has simple implementation and robustness against frequency-selective block-fading channels [22], has been advocated recently [5]. A basic feature of

Manuscript received May 23, 2008; accepted March 02, 2009. First published March 31, 2009; current version published July 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shahram Shahbazpanahi. The work of E. Biglieri was supported by the Spanish Ministry of Education and Science under Project TEC2006-01428/TCM, and by the STREP project No. IST-026905 (MASCOT) within the 6th framework program of the European Commission.

D. Angelosante was with DAEIMI, Università di Cassino, 03043 Cassino (FR), Italy. He is now with the Electrical and Computer Engineering Department, University of Minnesota, Minneapolis, MN USA (e-mail: d.angelosante@unicas.it).

E. Biglieri is with TIC, Universitat Pompeu Fabra, 08018 Barcelona, Spain (e-mail: e.biglieri@iee.org).

M. Lops is with DAEIMI, Università di Cassino, 03043 Cassino (FR), Italy (e-mail: lops@unicas.it).

Digital Object Identifier 10.1109/TSP.2009.2020049

these systems is their reliance upon the knowledge of channel state information (CSI) at the receiver. This can be achieved, for example, by sending a sequence of pilot symbols known to the receiver or by using blind techniques, which do not require the transmission of pilots (for an overview of pilot-aided and blind channel estimators for MIMO-OFDM, see [2], [6], and references therein). Compared with pilot-aided, blind channel estimation generally requires a long data record to extract the information needed, and hence is typically adopted on slowly time-varying channels. In addition, it entails a high complexity, which is to be traded off for the loss of spectral efficiency caused, in pilot-aided techniques, by the transmission of symbols not bearing any information.

This work is devoted to pilot-based sequential estimation of time-varying multipath channels for MIMO-OFDM. Unlike previous studies on the same topic, here we assume that the number of multipath components, their delays, and their amplitudes, may vary with time according to a known evolution model. The availability of a training set makes it possible to define “clairvoyant” detectors, wherein the channel state is first estimated, then employed to detect information symbols. In the time-varying scenario outlined above, channel tracking is a key to achieve satisfactory performance. A classical approach to channel estimation consists of assuming that all paths are always simultaneously active, and tracking their evolution. If this is done, the estimation quality is impaired, because receiver resources are wasted to estimate the amplitudes of paths that are not active. Recently, estimators of channels with unknown order were considered (see, e.g., [16]). However, the channel order is assumed to be constant in time, and all paths active. The new approach we follow here relies instead on modeling the channel response as a finite random set (FRS), which allows us to avoid the assumptions that the exact number of active paths are known and/or constant. Their evolution is tracked by exploiting finite set statistics (FISST) [13], a mathematical tool that has recently found application in multisensor-multi-target tracking [12], multispeaker tracking [11], and multiuser detection [4]. With this technique, path births or deaths, and amplitude variations, are reflected by changes of the elements and/or the cardinality of a properly defined random set, and can be tracked once a model for the channel dynamics is available.

Practical implementation of an FRS estimator relies on *ad hoc* approximations of Bayesian filtering equations. Among them, sequential Monte Carlo (SMC) algorithms [21] and probability hypothesis density (PHD) filters [12], [20] have attracted a good deal of attention. We shall not consider PHD filters here, because their use relies upon a set of assumptions on the dynamic

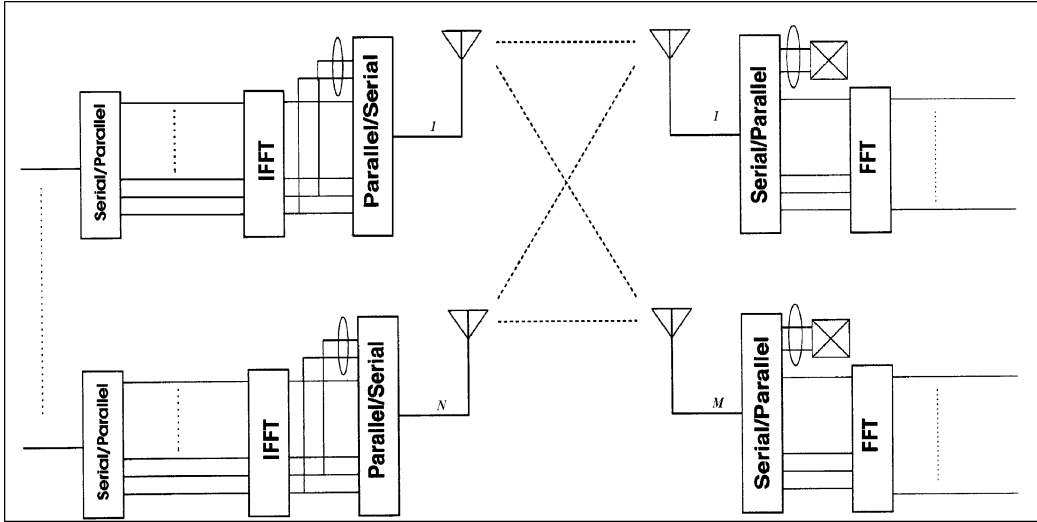


Fig. 1. MIMO-OFDM system.

model and on the available observations which are not met in our problem [13]. We focus instead our attention to SMC algorithms, which approximate Bayesian recursions by numerical integration and are inherently more flexible than PHD filters [1]. Since the accuracy of SMC approximations generally decreases as the number of entities to be estimated increases (for a fixed number of particles), we are faced with the problem of efficiently tracking a densely populated dynamic scenario by using a computationally feasible algorithm [21]. Recently, an efficient multiple-target tracking SMC filter based on the concept of Rao–Blackwellization was advocated [19]. This consists of sampling only some of the random variables involved in SMC, while others are handled analytically through Kalman filtering (KF). The only required assumption is that, upon suitable conditioning, the problem reduces to a linear Gaussian one, viz., the system be conditionally linear Gaussian (CLG) [8].<sup>1</sup> Building partially on the results of [19], we introduce here a Rao–Blackwellized particle filtering (RBPF) algorithm for the joint estimation of the number of active paths and their amplitudes in MIMO-OFDM channels. An approximate (and hence suboptimal) version of this algorithm is also introduced, and its performance is assessed. It achieves a remarkable complexity reduction at the price of a very limited performance loss. We also show that the proposed channel estimation procedure can handle situations where the multipath delays are *not* integer multiples of the inverse of the transmission bandwidth.

The balance of this paper is organized as follows. Section II describes the system model and the problem statement, while Section III is devoted to channel tracking, along with the introduction of FRS estimators and the illustration of their implementation through RBPF. Section IV contains numerical results, while concluding remarks form the object of Section V.

A few words on our notations: Normal-face letters denote scalar values. Lower-case (respectively, upper-case) boldface letters are used for column vectors (respectively, matrices).

<sup>1</sup>Nonlinear evolutions can, in any case, be dealt with through extended Kalman filtering (EKF) or unscented Kalman filtering (UKF).

Upper calligraphic letters are used for random sets. A superscript  $(\cdot)^T$  denotes transposition;  $(\cdot)^H$  denotes Hermitian transposition;  $\mathbf{A}^\dagger$  is the Moore–Penrose pseudoinverse of matrix  $\mathbf{A}$ .  $\mathbb{E}[\cdot]$  denotes expectation. When applied to sets,  $|\mathcal{X}|$  denotes the cardinality of  $\mathcal{X}$ , and when applied to a complex scalar,  $|a|$  denotes the magnitude of  $a$ .  $\mathbf{I}_N$  denotes the  $N \times N$  identity matrix, and  $\mathbf{0}_{M \times N}$  the  $M \times N$  all-zero matrix. If  $\mathbf{x}_t$ ,  $t = 1, \dots, T$ , is a sequence of column vectors of the same dimension then  $\mathbf{x}_{1:T} \triangleq [\mathbf{x}_1, \dots, \mathbf{x}_T]$  and  $\text{diag}(\mathbf{x})$  stands for the diagonal matrix with the column vector  $\mathbf{x}$  on its diagonal.  $\det(\Sigma)$  and  $\text{tr}(\Sigma)$  denote the determinant and the trace of the square matrix  $\Sigma$ , respectively. Finally,  $j \triangleq \sqrt{-1}$  and  $\mathcal{N}_c(\mathbf{x}; \boldsymbol{\mu}, \Sigma) \triangleq (\det(\pi\Sigma))^{-1} \exp\{-\mathbf{x} - \boldsymbol{\mu}^H \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\}$ .

## II. PROBLEM STATEMENT

Consider the system architecture depicted in Fig. 1. This shows a MIMO-OFDM system with  $N$  transmit and  $M$  receive antennas. At each transmit antenna, an OFDM modulator employing  $K$  subcarriers is used. The symbols transmitted by the  $n$ th antenna at discrete time index  $t$  are stacked into the  $K$ -dimensional column vector  $\tilde{\mathbf{d}}_{n;t} \triangleq [d_{1,n;t}, \dots, d_{K,n;t}]^T$ , which undergoes an IFFT and the insertion of a cyclic prefix of length  $\mu \geq L_{\max} - 1$ ,  $L_{\max}$  denoting the maximum length of all multipath channels. At the  $m$ th receive antenna, the cyclic prefix is removed and an FFT is computed, whereby the  $K$ -dimensional column vector of the observations takes the form [2]

$$\tilde{\mathbf{y}}_{m;t} = \sum_{n=1}^N \tilde{\mathbf{D}}_{n;t} \mathbf{T} \mathbf{h}_{m,n;t} + \tilde{\mathbf{z}}_{m;t} \quad (1)$$

where  $\tilde{\mathbf{D}}_{n;t} \triangleq \text{diag}(\tilde{\mathbf{d}}_{n;t})$ , and  $\mathbf{h}_{m,n;t} \triangleq [a_{m,n;t}^{(1)}, \dots, a_{m,n;t}^{(L_{\max})}]^T$  is an  $L_{\max}$ -dimensional column vector representing the channel impulse response from the  $n$ th

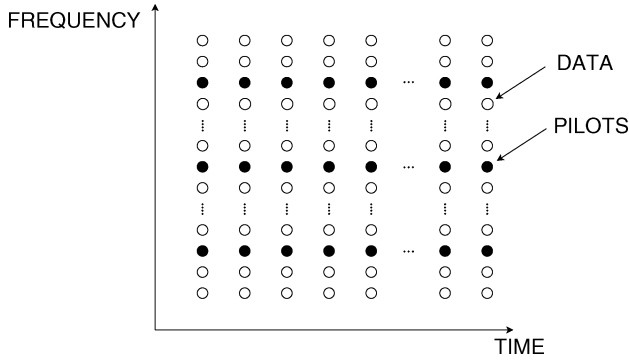


Fig. 2. Symbol pattern of a generic transmission antenna in a MIMO-OFDM system.

transmit to the  $m$ th receive antenna.<sup>2</sup> As to  $\tilde{\mathbf{z}}_{m;t}$ , it represents additive noise, i.e.,  $\tilde{\mathbf{z}}_{m;t} \sim \mathcal{N}_{\mathbb{C}}(\tilde{\mathbf{z}}_{m;t}; \mathbf{0}_{K \times 1}, N_0 \mathbf{I}_K)$ , while  $\mathbf{T}$  is the  $K \times L_{\max}$  Fourier matrix whose entry  $(k, \ell)$  is

$$\{\mathbf{T}\}_{k,\ell} = e^{-j2\pi(\ell-1)\frac{k-1}{K}}, \quad k = 1, \dots, K, \quad \ell = 1, \dots, L_{\max}. \quad (2)$$

Notice the assumption, implicit in the model above, that the delays are integer multiples of the inverse of the transmission bandwidth: the extension to noninteger delays will be dealt with briefly below. A number  $P \geq L_{\max}N$  of pilot symbols are uniformly interleaved with data to enable channel estimation, according to the symbols pattern shown in Fig. 2.

The signal observed in the pilot subcarriers at the  $m$ th receive antenna at time  $t$  is organized into the  $P$ -dimensional column vector  $\mathbf{y}_{m;t} \triangleq [y_{k_1;m;t}, \dots, y_{k_P;m;t}]^T$ . Analogously, the observations from the information subcarriers are organized in the  $(K - P)$ -dimensional column vector  $\mathbf{y}_{m;t}^I = [y_{k_1^I;m;t}, \dots, y_{k_{K-P}^I;m;t}]^T$  where  $k_i^I$ , for  $i = 1, \dots, K - P$ , is the  $i$ th information subcarrier index. With these notations, (1) can be rewritten as

$$\text{Pilot subcarriers:} \quad \mathbf{y}_{m;t} = \sum_{n=1}^N \mathbf{D}_{n;t} \mathbf{F} \mathbf{h}_{m,n;t} + \mathbf{z}_{m;t} \quad (3a)$$

$$\text{Information subcarriers:} \quad \mathbf{y}_{m;t}^I = \sum_{n=1}^N \mathbf{D}_{n;t}^I \mathbf{F}^I \mathbf{h}_{m,n;t} + \mathbf{z}_{m;t}^I \quad (3b)$$

where  $\mathbf{D}_{n;t} \triangleq \text{diag}([d_{k_1,n;t}, \dots, d_{k_P,n;t}]^T)$ ,  $\mathbf{D}_{n;t}^I \triangleq \text{diag}([d_{k_1^I,n;t}, \dots, d_{k_{K-P}^I,n;t}]^T)$ ,  $\mathbf{z}_{m;t} \triangleq [z_{k_1,m;t}, \dots, z_{k_P,m;t}]^T$ ,  $\mathbf{z}_{m;t}^I \triangleq [z_{k_1^I,m;t}, \dots, z_{k_{K-P}^I,m;t}]^T$ , while  $\mathbf{F}$  and  $\mathbf{F}^I$  are sub-matrices of  $\mathbf{T}$ , of order  $P \times L_{\max}$  and  $(K - P) \times L_{\max}$ , respectively.

As anticipated, representation (3) is highly redundant, since it assumes the simultaneous presence of  $L_{\max}$  paths

<sup>2</sup>We hasten to recall here that a “standard” way of proceeding further with this model consists of keeping a constant number  $L_{\max}$  of components in vector  $\mathbf{h}_{m,n;t}$ . Whenever the number of actual paths is lower than  $L_{\max}$ , the corresponding vector components are set to zero. This procedure is far from optimum, as the receiver wastes computational resources to estimate the amplitudes of nonexistent paths. The approach taken in this paper, one which is typical of random-set-theoretical analyses and will be illustrated in detail *infra*, models every component as a singleton-or-empty set, thus avoiding the estimation of absent paths.

for each transmit-receive antenna pair at any epoch  $t$ . To dispense with this assumption, we introduce a random-set model for the channel. Explicitly, let us introduce the singleton-or-empty set  $\mathcal{H}_{m,n;t}^{(\ell)} \triangleq \emptyset$ , if the  $\ell$ th path from the  $n$ th transmit to the  $m$ th receive antenna is not active at time  $t$ , and  $\mathcal{H}_{m,n;t}^{(\ell)} \triangleq \{\mathbf{h}_{m,n;t}^{(\ell)}\} = \{[m, n, \ell, a_{m,n;t}^{(\ell)}]^T\}$ , otherwise. The channel state for the pair  $(m, n)$  at epoch  $t$  is completely described by the disjoint union

$$\mathcal{H}_{m,n;t} = \bigcup_{\ell=1}^{L_{\max}} \mathcal{H}_{m,n;t}^{(\ell)}. \quad (4)$$

Here,  $\mathcal{H}_{m,n;t}$  is a finite random set (FRS) defined on  $\{m\} \times \{n\} \times \{1, \dots, L_{\max}\} \times \mathbb{C}$ . The whole multipath channel observed by the  $m$ th receive antenna at time  $t$ , denoted  $\mathcal{H}_{m;t}$ , can be written as the disjoint union

$$\mathcal{H}_{m;t} = \bigcup_{n=1}^N \mathcal{H}_{m,n;t} \quad (5)$$

where now  $\mathcal{H}_{m;t}$  is an FRS on  $\{m\} \times \{1, \dots, N\} \times \{1, \dots, L_{\max}\} \times \mathbb{C}$ . Notice that  $\mathcal{H}_{m;t}$  is uniquely specified by an FRS on a countable space,  $\pi(\mathcal{H}_{m;t})$ , the projection of  $\mathcal{H}_{m;t}$  onto  $\{m\} \times \{1, \dots, N\} \times \{1, \dots, L_{\max}\}$ , and by an FRS  $\pi'(\mathcal{H}_{m;t}) = \bigcup_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m;t})} \{a_{m,n;t}^{(\ell)}\}$ . On the other hand,  $\pi(\mathcal{H}_{m;t})$  is *totally ordered* by the standard “ $<$ ” relation, and induces a bijection between  $\mathcal{H}_{m;t}$  and  $\mathbb{C}^{NL_{\max}}$ , viz., once  $\pi(\mathcal{H}_{m;t})$  is known,  $\mathcal{H}_{m;t}$  can be associated with an  $NL_{\max}$ -dimensional vector whose nonzeros entries, uniquely identified by  $\pi(\mathcal{H}_{m;t})$ , contain the continuous parts of  $\mathcal{H}_{m;t}$ , namely the elements of  $\pi'(\mathcal{H}_{m;t})$ . Similar considerations apply to  $\mathcal{H}_{m,n;t}$ , which is specified by its projections on  $\{m\} \times \{n\} \times \{1, \dots, L_{\max}\}$  and  $\mathbb{C}$ ,  $\pi(\mathcal{H}_{m,n;t})$  and  $\pi'(\mathcal{H}_{m,n;t})$ , respectively.

The signal received at the  $m$ th receive antenna can thus be written as in (3), with the understanding that the tap weight vectors are now given by the following set functions:

$$[\mathbf{h}_{m,n;t}(\mathcal{H}_{m,n;t})]_{\ell} = \begin{cases} a_{m,n;t}^{(\ell)}, & \text{if } [m, n, \ell]^T \in \pi(\mathcal{H}_{m,n;t}) \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Equation (6) highlights the fact that efficient channel tracking amounts to sequentially estimating the FRSs  $\mathcal{H}_{m,n;t}$ , which, in turn, requires defining a model for their evolution. In what follows, we introduce a Markov model for the channel dynamics, whereby customary Bayesian recursions allow causal channel tracking.

### A. Channel Dynamics

From now on we assume that distinct transmit-receive antenna pairs generate independent channels,<sup>3</sup> and that the coefficients of different paths are themselves independent. This makes it sufficient to characterize only the singleton-or-empty sets  $\mathcal{H}_{m,n;t}^{(\ell)}$ , which, in turn, allows the description of both  $\mathcal{H}_{m,n;t}$  and  $\mathcal{H}_{m;t}$  through the properties of disjoint unions of independent FRS’s. In our model,  $\mathcal{H}_{m,n;t}$  is the union of the

<sup>3</sup>We disregard the consideration of correlated channels, whose analysis is more complicated but does not involve any additional theoretical difficulty in the present context.

path gains surviving from time  $t-1$ , denoted  $\mathcal{S}_{m,n;t}(\mathcal{H}_{m,n;t-1})$ , and of the path gains born at time  $t$ , denoted  $\mathcal{B}_{m,n;t}$ . Hence, we have

$$\begin{aligned}\mathcal{H}_{m,n;t} &= \mathcal{S}_{m,n;t}(\mathcal{H}_{m,n;t-1}) \cup \mathcal{B}_{m,n;t} \\ &= \left[ \bigcup_{\ell=1}^{L_{\max}} \mathcal{S}^{(\ell)}(\mathcal{H}_{m,n;t-1}^{(\ell)}) \right] \cup \mathcal{B}_{m,n;t}\end{aligned}\quad (7a)$$

$$\pi(\mathcal{H}_{m,n;t-1}) \cap \pi(\mathcal{B}_{m,n;t}) = \emptyset \quad (7b)$$

$$\pi(\mathcal{S}_{m,n;t}) \subseteq \pi(\mathcal{H}_{m,n;t-1}). \quad (7c)$$

In (7), we have exploited once again the fact that the FRSs  $\mathcal{S}_{m,n;t}$  and  $\mathcal{B}_{m,n;t}$  can be equivalently described through their projections onto  $\{m\} \times \{n\} \times \{1, \dots, L_{\max}\}$  and  $\mathbb{C}$ , while  $\mathcal{S}^{(\ell)}(\mathcal{H}_{m,n;t-1}^{(\ell)})$  is the singleton-or-empty random set

$$\mathcal{S}^{(\ell)}(\mathcal{H}_{m,n;t-1}^{(\ell)}) = \begin{cases} \emptyset & \text{w.p. } P_{\text{death}} \\ \left[ [m, n, \ell, a_{m,n;t-1}^{(\ell)}] \right]^T & \text{w.p. } 1 - P_{\text{death}}. \end{cases} \quad (8)$$

Here,  $P_{\text{death}}$  is the probability that a path active at  $t-1$  disappears at  $t$ , and  $\mathcal{S}^{(\ell)}(\emptyset) = \emptyset$ . The constraints (7b) and (7c) reflect the facts that no component, among those active at time  $t-1$ , can migrate to the set of new paths, and that the paths surviving at  $t$  are a subset of those active at  $t-1$ . Since

$$\begin{aligned}f(\mathcal{S}_{m,n;t}|\mathcal{H}_{m,n;t-1}) &= f(\pi(\mathcal{S}_{m,n;t})|\pi(\mathcal{H}_{m,n;t-1})) \\ &\times f(\pi'(\mathcal{S}_{m,n;t})|\pi(\mathcal{H}_{m,n;t-1}), \pi(\mathcal{S}_{m,n;t}), \pi'(\mathcal{H}_{m,n;t-1}))\end{aligned} \quad (9)$$

the assumption that the paths may disappear or survive independently of each other yields

$$\begin{aligned}f(\pi(\mathcal{S}_{m,n;t})|\pi(\mathcal{H}_{m,n;t-1})) \\ = P_{\text{death}}^{|\mathcal{H}_{m,n;t-1}| - |\mathcal{S}_{m,n;t}|} (1 - P_{\text{death}})^{|\mathcal{S}_{m,n;t}|}\end{aligned} \quad (10)$$

and

$$\begin{aligned}f(\pi'(\mathcal{S}_{m,n;t})|\pi(\mathcal{H}_{m,n;t-1}), \pi(\mathcal{S}_{m,n;t}), \pi'(\mathcal{H}_{m,n;t-1})) \\ = \prod_{[m,n,\ell]^T \in \pi(\mathcal{S}_{m,n;t})} f(a_{m,n;t}^{(\ell)}|a_{m,n;t-1}^{(\ell)})\end{aligned} \quad (11)$$

where  $f(a_{m,n;t}^{(\ell)}|a_{m,n;t-1}^{(\ell)})$  is the transition density describing the evolution of the gains of the surviving paths. Thus

$$\begin{aligned}f(\mathcal{S}_{m,n;t}|\mathcal{H}_{m,n;t-1}) &= P_{\text{death}}^{|\mathcal{H}_{m,n;t-1}| - |\mathcal{S}_{m,n;t}|} (1 - P_{\text{death}})^{|\mathcal{S}_{m,n;t}|} \\ &\times \prod_{[m,n,\ell]^T \in \pi(\mathcal{S}_{m,n;t})} f(a_{m,n;t}^{(\ell)}|a_{m,n;t-1}^{(\ell)}).\end{aligned} \quad (12)$$

Analogously, defining  $\mathcal{B}_{m,n;t} = \{[m, n, b_1, a_{m,n;t}^{(b_1)}], \dots, [m, n, b_{|\mathcal{B}_{m,n;t}|}, a_{m,n;t}^{(b_{|\mathcal{B}_{m,n;t}|})}]\}$ , the independence assumption for the FRS of the newly born paths yields

$$\begin{aligned}f(\mathcal{B}_{m,n;t}|\mathcal{H}_{m,n;t-1}) &= P_{\text{birth}}^{|\mathcal{B}_{m,n;t}|} \\ &\times (1 - P_{\text{birth}})^{L_{\max} - |\mathcal{H}_{m,n;t-1}| - |\mathcal{B}_{m,n;t}|} \prod_{i=1}^{|\mathcal{B}_{m,n;t}|} f(a_{m,n;t}^{(b_i)})\end{aligned} \quad (13)$$

where  $P_{\text{birth}}$  is the probability that a new path arises, while  $f(a_{m,n;t}^{(\ell)})$  is the marginal probability density function of the  $\ell$ th newly born path gain from the  $n$ th transmit to the  $m$ th receive antenna at epoch  $t$ .

The transition density of  $\mathcal{H}_{m,n;t}$ , i.e.,

$$\begin{aligned}f(\mathcal{H}_{m,n;t}|\mathcal{H}_{m,n;t-1}) &= f(\pi(\mathcal{H}_{m,n;t})|\pi(\mathcal{H}_{m,n;t-1})) \\ &\times f(\pi'(\mathcal{H}_{m,n;t})|\pi(\mathcal{H}_{m,n;t-1:t}), \pi'(\mathcal{H}_{m,n;t-1:t}))\end{aligned} \quad (14)$$

can now be determined through the generalized-convolution formula [4] as shown in (15) shown at the bottom of the page, and

$$\begin{aligned}f(\pi'(\mathcal{H}_{m,n;t})|\pi(\mathcal{H}_{m,n;t-1:t}), \pi'(\mathcal{H}_{m,n;t-1:t})) \\ = \prod_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m,n;t}) \cap \pi(\mathcal{H}_{m,n;t-1})} f(a_{m,n;t}^{(\ell)}|a_{m,n;t-1}^{(\ell)}) \\ \times \prod_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m,n;t}) \setminus (\pi(\mathcal{H}_{m,n;t}) \cap \pi(\mathcal{H}_{m,n;t-1}))} f(a_{m,n;t}^{(\ell)}).\end{aligned} \quad (16)$$

Under the assumption that  $\mathcal{H}_{m,n;t}$  are independent across  $n$  and  $m$ , we have

$$f(\mathcal{H}_{m;t}|\mathcal{H}_{m;t-1}) = \prod_{n=1}^N f(\mathcal{H}_{m,n;t}|\mathcal{H}_{m,n;t-1}). \quad (17)$$

$$\begin{aligned}f(\pi(\mathcal{H}_{m,n;t})|\pi(\mathcal{H}_{m,n;t-1})) \\ = f(\pi(\mathcal{H}_{m,n;t}) \cap \pi(\mathcal{H}_{m,n;t-1})|\pi(\mathcal{H}_{m,n;t-1})) \\ \times f(\pi(\mathcal{H}_{m,n;t}) \setminus \pi(\mathcal{H}_{m,n;t}) \cap \pi(\mathcal{H}_{m,n;t-1})|\pi(\mathcal{H}_{m,n;t-1})) \\ = P_{\text{death}}^{|\mathcal{H}_{m,n;t-1}| - |\pi(\mathcal{H}_{m,n;t}) \cap \pi(\mathcal{H}_{m,n;t-1})|} \\ \times (1 - P_{\text{death}})^{|\pi(\mathcal{H}_{m,n;t}) \cap \pi(\mathcal{H}_{m,n;t-1})|} \\ \times P_{\text{birth}}^{|\pi(\mathcal{H}_{m,n;t}) \setminus \pi(\mathcal{H}_{m,n;t}) \cap \pi(\mathcal{H}_{m,n;t-1})|} \\ \times (1 - P_{\text{birth}})^{L_{\max} - |\mathcal{H}_{m,n;t-1}| - |\pi(\mathcal{H}_{m,n;t}) \setminus \pi(\mathcal{H}_{m,n;t}) \cap \pi(\mathcal{H}_{m,n;t-1})|}\end{aligned} \quad (15)$$

Equation (17) represents the transition model for the multipath channel experienced by the  $m$ th receive antenna.

### B. Clairvoyant Decoding

The availability of a set of pilot signals allows the implementation of a clairvoyant<sup>4</sup> data detector, wherein the observations (3a) are used to obtain estimates of the channel state  $\mathcal{H}_{m;t}$ , and then, in the conditional likelihood of the observations (3b), to detect information symbols. Notice indeed that

$$f(\mathbf{y}_{m;t}^I | \mathbf{D}_{1;t}^I, \dots, \mathbf{D}_{N;t}^I; \mathcal{H}_{m;t}) = \mathcal{N}_c \left( \mathbf{y}_{m;t}^I; \sum_{n=1}^N \mathbf{D}_{n;t}^I \mathbf{F}^I \mathbf{h}_{m,n;t}(\mathcal{H}_{m,n;t}), N_0 \mathbf{I}_{K-P} \right) \quad (18)$$

whereby a clairvoyant detector is obtained as

$$\begin{aligned} & \left[ \widehat{\mathbf{D}}_{1;t}^I, \dots, \widehat{\mathbf{D}}_{N;t}^I \right] \\ & = \arg \min_{\mathbf{D}_{1;t}^I, \dots, \mathbf{D}_{N;t}^I} \sum_{m=1}^M \left\| \mathbf{y}_{m;t}^I - \sum_{n=1}^N \mathbf{D}_{n;t}^I \mathbf{F}^I \mathbf{h}_{m,n;t}(\widehat{\mathcal{H}}_{m,n;t}) \right\|^2 \end{aligned} \quad (19)$$

with  $\widehat{\mathcal{H}}_{m,n;t}$  is an estimate of  $\mathcal{H}_{m,n;t}$  obtained through the pilot observations (3a). Availability of  $\widehat{\mathcal{H}}_{m,n;t}$  allows one to split the minimization in (19) into  $K-P$  disjoint minimizations, as one can easily verify by noticing that

$$\begin{aligned} & \sum_{m=1}^M \left\| \mathbf{y}_{m;t}^I - \sum_{n=1}^N \mathbf{D}_{n;t}^I \mathbf{F}^I \mathbf{h}_{m,n;t}(\widehat{\mathcal{H}}_{m,n;t}) \right\|^2 \\ & = \sum_{m=1}^M \left( \sum_{i=1}^{K-P} \left| y_{k_i^I, m;t} - \sum_{n=1}^N d_{k_i^I, n;t} \widehat{H}_{k_i^I, m, n;t} \right|^2 \right) \\ & = \sum_{i=1}^{K-P} \left( \sum_{m=1}^M \left| y_{k_i^I, m;t} - \sum_{n=1}^N d_{k_i^I, n;t} \widehat{H}_{k_i^I, m, n;t} \right|^2 \right) \\ & = \sum_{i=1}^{K-P} \|\mathbf{r}_{i;t} - \mathbf{C}_{i;t} \mathbf{s}_{i;t}\|^2 \end{aligned} \quad (20)$$

where we have defined

$$\mathbf{r}_{i;t} \triangleq [y_{k_i^I, 1;t}, \dots, y_{k_i^I, M;t}]^T \quad (21a)$$

$$\mathbf{s}_{i;t} \triangleq [d_{k_i^I, 1;t}, \dots, d_{k_i^I, N;t}]^T \quad (21b)$$

$$\mathbf{C}_{i;t} \triangleq \begin{bmatrix} \widehat{H}_{k_i^I, 1, 1;t} & \cdots & \widehat{H}_{k_i^I, 1, N;t} \\ \vdots & \ddots & \vdots \\ \widehat{H}_{k_i^I, M, 1;t} & \cdots & \widehat{H}_{k_i^I, M, N;t} \end{bmatrix} \quad (21c)$$

$$\widehat{H}_{k_i^I, m, n;t} \triangleq \sum_{[m, n, \ell] \in \pi(\widehat{\mathcal{H}}_{m,n;t})} \widehat{a}_{m, n; t}^{(\ell)} e^{-j2\pi(k_i^I - 1) \frac{\ell-1}{K}}. \quad (21d)$$

Here,  $\widehat{H}_{k_i^I, m, n;t}$  models the frequency-flat fading experienced by the  $i$ th information subcarrier from the  $n$ th transmit to the  $m$ th receive antenna at time  $t$ . Notice that the complexity of the

above receiver grows linearly with the number of information subcarriers, and exponentially with the number of transmit antennas.<sup>5</sup>

The final step to be taken, one which forms the object of next sections, can be formulated as the following problem. Given the observation of the pilot subcarriers at the generic  $m$ th receive antenna up to time  $t$  (i.e.,  $\mathbf{y}_{m;\tau}$  for  $\tau = 1, \dots, t$ ), the pilot symbols sent by the transmit antennas (i.e.,  $\mathbf{D}_{n;\tau}$  for  $n = 1, \dots, N$  and  $\tau = 1, \dots, t$ ), and the channel model of Section II-A, derive a causal estimator for  $\mathcal{H}_{m;t} = \bigcup_{n=1}^N \mathcal{H}_{m,n;t}$ , so as to enable real-time implementation of the clairvoyant detector (19).

### III. RECURSIVE FILTERING

The evolution with time of the predicted density of the FRS  $\mathcal{H}_{m;t}$  is described by the Chapman–Kolmogorov equation

$$f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t-1}) = \int f(\mathcal{H}_{m;t} | \mathcal{H}_{m;t-1}) \times f(\mathcal{H}_{m;t-1} | \mathbf{y}_{m;1:t-1}) \delta \mathcal{H}_{m;t-1} \quad (22)$$

where the symbol  $\delta \mathcal{H}_{m;t-1}$  emphasizes the fact that the integral in (22) is a set integral [13]. At time  $t$ , the observation  $\mathbf{y}_{m;t}$  is made available, and hence the filtering distribution can be updated by using Bayes' rule

$$f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t}) = \frac{f(\mathbf{y}_{m;t} | \mathcal{H}_{m;t}) f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t-1})}{f(\mathbf{y}_{m;t} | \mathbf{y}_{m;1:t-1})} \propto f(\mathbf{y}_{m;t} | \mathcal{H}_{m;t}) f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t-1}). \quad (23)$$

Equations (22) and (23), the ‘‘Bayesian recursions’’ (BR), define the so-called Bayesian filter. A closed-form solution of the above equations is, in general, unavailable, and hence approximate techniques are needed. We defer to Section III-B a description of the approximations to the BR. Before doing that, we discuss how an estimate of the random set  $\mathcal{H}_{m;t}$  can be derived from  $f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t})$ .

#### A. Defining FRS estimators

Once the *a posteriori* density  $f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t})$  is computed, a ‘‘Bayesian’’ estimate of  $\mathcal{H}_{m;t}$  can be obtained, as outlined for example in [13], by minimizing the expected value of a suitable cost function. Two popular such estimators are known as GMAP-I (also, ‘‘Marginal Multitarget Estimator’’) and GMAP-II (also, ‘‘Joint Multitarget Estimator’’). GMAP-I is a two-stage estimator, wherein the set cardinality is estimated first. Defining

$$P(|\mathcal{H}_{m;t}| = n_t | \mathbf{y}_{m;1:t}) \triangleq \int_{\mathcal{H}_{m;t}: |\mathcal{H}_{m;t}| = n_t} f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t}) \delta \mathcal{H}_{m;t} \quad (24)$$

the probability that the cardinality of  $\mathcal{H}_{m;t}$  equals  $n_t$ , we have

$$\text{GMAP-I: } \begin{cases} \widehat{n}_t = \arg \max_{n_t \in \{1, \dots, N L_{\max}\}} P(|\mathcal{H}_{m;t}| = n_t | \mathbf{y}_{m;1:t}) \\ \widehat{\mathcal{H}}_{m;t} = \arg \max_{\mathcal{H}_{m;t}: |\mathcal{H}_{m;t}| = \widehat{n}_t} f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t}) \end{cases} \quad (25)$$

<sup>5</sup>If needed, linearity in the number of transmit antennas can always be granted by resorting to suboptimal strategies borrowed from multiuser-detection theory [17].

<sup>4</sup>The word ‘‘clairvoyant’’ is used in the same sense as in [18, ch. 14].

GMAP-II performs the estimation in a single step:

$$\text{GMAP - II} : \widehat{\mathcal{H}}_{m;t} = \arg \max_{\mathcal{H}_{m;t}} f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t}) \frac{c^{|\mathcal{H}_{m;t}|}}{|\mathcal{H}_{m;t}|!} \quad (26)$$

where  $c$  is a small constant determined by the cost function that this estimator minimizes [13]. GMAP-I minimizes first the expected value of the cost function  $C_0$ , involving only the cardinality of the set [13, p. 192]:

$$C_0(\mathcal{H}_{m;t}, \widehat{\mathcal{H}}_{m;t}) \triangleq \begin{cases} 0, & |\mathcal{H}_{m;t}| = |\widehat{\mathcal{H}}_{m;t}| \\ 1, & |\mathcal{H}_{m;t}| \neq |\widehat{\mathcal{H}}_{m;t}| \end{cases} \quad (27)$$

and, successively, performs a MAP estimation given the estimated cardinality. Indeed, GMAP-II minimizes the expected value of the cost function  $C_2 = C_0 + C_1$ , where  $C_1$  is defined as

$$C_1(\mathcal{H}_{m;t}, \widehat{\mathcal{H}}_{m;t}) \triangleq \begin{cases} 0, & \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) \text{ and} \\ & \pi'(\mathcal{H}_{m;t}), \pi'(\widehat{\mathcal{H}}_{m;t}) \in \mathcal{A} \\ 1, & \text{otherwise} \end{cases} \quad (28)$$

where  $\mathcal{A}$  is a closed ball in  $\mathbb{C}^{|\mathcal{H}_{m;t}|}$ . The choice of this estimator implies that no cost is incurred whenever (a) the number of active paths and their location are estimated correctly, and (b) the true and estimated values of the continuous components are both contained in a closed ball  $\mathcal{A}$  whose size is reflected by the choice of the constant  $c$  in (26). If either (a) or (b) is not satisfied, then the cost is 1.

For the specific problem studied in this paper, we advocate a novel estimation rule, which turns out to be preferable to GMAP-I and GMAP-II. The reason for this is that GMAP-I and GMAP-II were originally defined in the context of a multisensor-multitarget tracking framework, wherein the discrete elements typically describe the target type and/or its tactical importance. In our scenario, we have  $|\mathcal{H}_{m;t}| = |\pi(\mathcal{H}_{m;t})|$ , which is generally not true in situations where the discrete components of the objects of interest are allowed to coincide, as in multisensor-multitarget tracking. In our specific context, we look instead for a cost function associating proper weights to an error in the estimate of the countable part, i.e.,  $\pi(\mathcal{H}_{m;t})$ , and to that in the estimate of the continuous part, i.e.,  $\pi'(\mathcal{H}_{m;t})$ . Based on this rationale, the cost function we advocate is

$$C_3(\mathcal{H}_{m;t}, \widehat{\mathcal{H}}_{m;t}) \triangleq \begin{cases} Q, & \pi(\mathcal{H}_{m;t}) \neq \pi(\widehat{\mathcal{H}}_{m;t}) \\ \sum_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m;t})} g(a_{m,n;t}^{(\ell)}, \tilde{a}_{m,n;t}^{(\ell)}), & \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}). \end{cases} \quad (29)$$

This assigns cost  $Q$  to any wrong estimate of the set of active paths or of their location. If this estimate is correct, the cost is a function  $g$  of the discrepancies between estimated and true continuous parameters. The actual selection of  $Q$  and of  $g$  reflects the relative weights assigned to the errors in the discrete and the continuous parameters. Define

$$\mathbb{P}(\pi(\mathcal{H}_{m;t}) = \mathcal{P} | \mathbf{y}_{m;1:t}) \triangleq \int_{\mathcal{H}_{m;t}: \pi(\mathcal{H}_{m;t}) = \mathcal{P}} f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t}) \delta \mathcal{H}_{m;t} \quad (30)$$

the probability mass function of  $\pi(\mathcal{H}_{m;t})$ . The causal optimum estimator corresponding to (29) minimizes, at each epoch  $t$ , the conditional risk

$$\begin{aligned} R(\widehat{\mathcal{H}}_{m;t} | \mathbf{y}_{m;1:t}) &\triangleq \int C_3(\mathcal{H}_{m;t}, \widehat{\mathcal{H}}_{m;t}) f(\mathcal{H}_{m;t} | \mathbf{y}_{m;1:t}) \delta \mathcal{H}_{m;t} \\ &= Q \mathbb{P}(\pi(\mathcal{H}_{m;t}) \neq \pi(\widehat{\mathcal{H}}_{m;t}) | \mathbf{y}_{m;1:t}) \\ &\quad + \mathbb{P}(\pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) | \mathbf{y}_{m;1:t}) \\ &\quad \times \mathbb{E} \left[ \sum_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m;t})} g(a_{m,n;t}^{(\ell)}, \tilde{a}_{m,n;t}^{(\ell)}) \right. \\ &\quad \left. | \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t} \right]. \end{aligned} \quad (31)$$

Not unexpectedly, the estimates of  $\pi(\mathcal{H}_{m;t})$  and  $\pi'(\mathcal{H}_{m;t})$  are inherently coupled. However, for large  $Q$  (i.e., as errors in the discrete part, containing the object identities, carry much more weight than the errors in the continuous parts), the two estimates tend to decouple, which suggests a two-step estimator minimizing separately the two summands in (31). This is what we call GMAP-III. If  $g$  is a quadratic function, the second minimization yields the conditional expectation, and hence the GMAP-III estimator is the following:

$$\begin{aligned} \widehat{\mathcal{H}}_{m;t} &= \arg \max_{\mathcal{P}} \mathbb{P}(\pi(\mathcal{H}_{m;t}) = \mathcal{P} | \mathbf{y}_{m;1:t}), \\ \tilde{a}_{m,n;t}^{(\ell)} &= \mathbb{E} [a_{m,n;t}^{(\ell)} | \mathbf{y}_{m;1:t}] \text{ if } [m,n,\ell]^T \in \pi(\widehat{\mathcal{H}}_{m;t}). \end{aligned} \quad (32)$$

We now review the SMC methods enabling the implementation of the above estimation rules.

---

#### Algorithm 1: Generic RBPF

---

- 1: Given  $\{w_{t-1}^{(r)}, \pi(\mathcal{H}_{m;1:t-1})^{(r)}\}_{r=1}^R$   
**for**  $r = 1$  to  $R$ 
    - 2: Sample  $\pi(\mathcal{H}_{m;t})^{(r)} \sim q(\cdot | \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \mathbf{y}_{m;1:t})$
    - 3: Set  $\pi(\mathcal{H}_{m;1:t})^{(r)} \triangleq \{\pi(\mathcal{H}_{m;1:t-1})^{(r)}, \pi(\mathcal{H}_{m;t})^{(r)}\}$
    - 4: Update the weight  
 $\tilde{w}_t^{(r)} = \frac{w_{t-1}^{(r)} (f(\mathbf{y}_{m;t} | \pi(\mathcal{H}_{m;1:t})^{(r)}, \mathbf{y}_{m;1:t-1}) f(\pi(\mathcal{H}_{m;t})^{(r)} | \pi(\mathcal{H}_{m;t-1})^{(r)}))}{q(\pi(\mathcal{H}_{m;t})^{(r)} | \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \mathbf{y}_{m;1:t})}$
    - end**
  - Normalization  
**for**  $r = 1$  to  $R$ 
    - 5:  $w_t^{(r)} = \tilde{w}_t^{(r)} / \sum_{r'=1}^R \tilde{w}_t^{(r')}$
    - end**
  - 6: Resampling with replacement
  - 7: OUTPUT  $\{w_t^{(r)}, \pi(\mathcal{H}_{m;1:t})^{(r)}\}_{r=1}^R$
- 

#### B. Tracking the Posterior Density

Whenever a closed-form expression for the Bayesian recursions (22), (23) is missing (which occurs in almost all

cases, a notable exception being the linear Gaussian problem), approximate and efficient filtering techniques are typically employed to track *a posteriori* densities. In particular, SMC techniques, which rely upon a sample-based construction of the densities involved in (22) and (23), have established themselves as an invaluable tool for the implementation of random-set tracking systems [21]. Unfortunately, SMC efficiency decreases as the dimension of the state space increases, hence possibly requiring an inordinately large number of particles.<sup>6</sup> To reduce complexity, a clever choice consists of resorting to “Rao–Blackwellization,” a technique that improves efficiency by analytically marginalizing some of the random variables (those being Gaussian and appearing linearly) from the joint filtering density. Thus, only the nonlinear part of the system model is estimated through SMC, while its linear part is estimated through conventional Kalman filtering. Rao-Blackwellized particle filtering can result in a decrease of the variance of the SMC estimates and of their complexity, at the cost of its limited applicability—the system under consideration must be conditionally linear Gaussian (CLG). In our framework, the dynamic channel model of Section II-A, under the assumption that the densities  $f_{a_{m,n;t}|a_{m,n;t-1}}^{(\ell)}$  and  $f_{a_{m,n;t}}^{(\ell)}$  in (12) and (13), respectively, are Gaussian, yields in fact a CLG system. Thus, RBPF can be applied outright through the two following basic steps:

- 1) sample  $\pi(\mathcal{H}_{m;1:t})$  using efficient SMC methods;
- 2) evaluate  $f(\pi'(\mathcal{H}_{m;1:t})|\pi(\mathcal{H}_{m;1:t}), \mathbf{y}_{m;1:t})$  through Kalman filtering for each sample of  $\pi(\mathcal{H}_{m;1:t})$ ;

where we have employed the customary notation  $\mathcal{H}_{m;1:t}$  to denote the sequence of the FRSS  $\mathcal{H}_{m;1}, \dots, \mathcal{H}_{m;t}$ . Recall that, while general Bayesian filtering aims at propagating the filtering distribution  $f(\mathcal{H}_{m;t}|\mathbf{y}_{m;1:t})$ , RBPF relies on a recursion for the so-called *joint a posteriori density*  $f(\mathcal{H}_{m;1:t}|\mathbf{y}_{m;1:t})$ , from which the filtering density can often be easily obtained. Under the Markov assumption for  $\mathcal{H}_{m;t}$ , the joint *a posteriori* density satisfies the recursion

$$f(\mathcal{H}_{m;1:t}|\mathbf{y}_{m;1:t}) = \frac{f(\mathbf{y}_{m;t}|\mathcal{H}_{m;t})f(\mathcal{H}_{m;t}|\mathcal{H}_{m;t-1})}{f(\mathbf{y}_{m;t}|\mathbf{y}_{m;1:t-1}) \times f(\mathcal{H}_{m;1:t-1}|\mathbf{y}_{m;1:t-1})} \quad (33)$$

Resorting to the specification of  $\mathcal{H}_{m;t}$  through  $\pi(\mathcal{H}_{m;t})$  and  $\pi'(\mathcal{H}_{m;t})$ , we have, from (14)

$$\begin{aligned} f(\mathcal{H}_{m;t}|\mathcal{H}_{m;t-1}) &= f(\pi(\mathcal{H}_{m;t}), \pi'(\mathcal{H}_{m;t})|\mathcal{H}_{m;t-1}) \\ &= f(\pi(\mathcal{H}_{m;t})|\pi(\mathcal{H}_{m;t-1})) \\ &\quad \times f(\pi'(\mathcal{H}_{m;t})|\pi(\mathcal{H}_{m;t-1:t}), \pi'(\mathcal{H}_{m;t-1})). \end{aligned} \quad (34)$$

Notice that  $\pi(\mathcal{H}_{m;t})$  is a random set defined on a countable space, while  $\pi'(\mathcal{H}_{m;t})$ , conditioned on  $\pi(\mathcal{H}_{m;t})$ , is a random set whose cardinality equals  $|\pi(\mathcal{H}_{m;t})|$ . Due to the CLG hypothesis, the conditional *a posteriori* density  $f(\pi'(\mathcal{H}_{m;1:t})|\pi(\mathcal{H}_{m;1:t}), \mathbf{y}_{m;1:t})$  is analytically tractable through Kalman filtering. Hence, we can easily marginalize

<sup>6</sup>For a thorough discussion of SMC in random-set tracking, the interested reader is deferred to the now abundant literature on this issue, such as, e.g., [13].

out  $\pi'(\mathcal{H}_{m;1:t})$  from the joint *a posteriori* density, and look for an estimator of  $f(\pi(\mathcal{H}_{m;1:t})|\mathbf{y}_{m;1:t})$ , which lies in a space of reduced dimension (viz., a countable space). Formally, we have

$$f(\pi(\mathcal{H}_{m;1:t}), \pi'(\mathcal{H}_{m;1:t})|\mathbf{y}_{m;1:t}) = f(\pi(\mathcal{H}_{m;1:t})|\mathbf{y}_{m;1:t}) \times f(\pi'(\mathcal{H}_{m;1:t})|\pi(\mathcal{H}_{m;1:t}), \mathbf{y}_{m;1:t}) \quad (35)$$

where the density  $f(\pi(\mathcal{H}_{m;1:t})|\mathbf{y}_{m;1:t})$  follows the recursion

$$\begin{aligned} f(\pi(\mathcal{H}_{m;1:t})|\mathbf{y}_{m;1:t}) &= \frac{f(\mathbf{y}_{m;t}|\pi(\mathcal{H}_{m;1:t}), \mathbf{y}_{m;1:t-1}) f(\pi(\mathcal{H}_{m;t})|\pi(\mathcal{H}_{m;t-1}))}{f(\mathbf{y}_{m;t}|\mathbf{y}_{m;1:t-1})} \\ &\quad \times f(\pi(\mathcal{H}_{m;1:t-1})|\mathbf{y}_{m;1:t-1}). \end{aligned} \quad (36)$$

Assume that we have an approximation of the posterior density at time  $t-1$  in terms of particles:

$$f(\pi(\mathcal{H}_{m;1:t-1})|\mathbf{y}_{m;1:t-1}) = \sum_{r=1}^R w_{t-1}^{(r)} m_{\pi(\mathcal{H}_{m;1:t-1})^{(r)}}(\pi(\mathcal{H}_{m;1:t-1})) \quad (37)$$

where  $m_{\mathcal{Y}}(\mathcal{X})$  is the “0–1” measure

$$\int_{\mathcal{C}} m_{\mathcal{Y}}(\mathcal{X}) \delta \mathcal{X} = \begin{cases} 1, & \text{if } \mathcal{Y} \subseteq \mathcal{C} \\ 0, & \text{otherwise.} \end{cases} \quad (38)$$

When the measurement  $\mathbf{y}_{m;t}$  is available, the RBPF performs the operations described in Algorithm 1 in order to determine an approximation in terms of particles of  $f(\pi(\mathcal{H}_{m;1:t})|\mathbf{y}_{m;1:t})$ . Note that the quantity  $q(\pi(\mathcal{H}_{m;t})|\pi(\mathcal{H}_{m;1:t-1})^{(r)}, \mathbf{y}_{m;1:t})$  in Algorithm 1 is the “importance distribution,” which plays a key role in SMC algorithms and whose choice will be discussed later on in this subsection. From the particle approximation of the joint *a posteriori* density

$$f(\pi(\mathcal{H}_{m;1:t})|\mathbf{y}_{m;1:t}) \approx \sum_{r=1}^R w_t^{(r)} m_{\pi(\mathcal{H}_{m;1:t})^{(r)}}(\pi(\mathcal{H}_{m;1:t})) \quad (39)$$

it is straightforward to obtain the filtering density required for causal channel estimation

$$P(\pi(\mathcal{H}_{m;t}) = \mathcal{P}|\mathbf{y}_{m;1:t}) \approx \sum_{r=1}^R w_t^{(r)} m_{\pi(\mathcal{H}_{m;t})^{(r)}}(\mathcal{P}). \quad (40)$$

Moreover, since the density  $f(\pi'(\mathcal{H}_{m;1:t})|\pi(\mathcal{H}_{m;1:t}), \mathbf{y}_{m;1:t})$  is analytically tractable, we can obtain a set of particles for the elements of  $\pi'(\mathcal{H}_{m;1:t})$  as

$$\mathbf{h}_{m,n;t}^{(r)} = \mathbb{E}[\mathbf{h}_{m,n;t}|\pi(\mathcal{H}_{m;1:t}), \mathbf{y}_{m;1:t}] \text{ for } n=1, \dots, N \quad (41)$$

for  $r=1, \dots, R$  (this follows from the mapping in (6)). We remark incidentally that, in the CLG scenario we are assuming, (41) represents the conditional minimum-mean-square error (MMSE) and the conditional maximum *a posteriori* (MAP) estimator of the elements of  $\pi'(\mathcal{H}_{m;t})$ . The full joint *a posteriori* density can finally be written

$$\begin{aligned} f(\pi(\mathcal{H}_{m;1:t}), \pi'(\mathcal{H}_{m;1:t})|\mathbf{y}_{m;1:t}) &= \sum_{r=1}^R w_t^{(r)} m_{\pi(\mathcal{H}_{m;1:t})^{(r)}, \pi'(\mathcal{H}_{m;1:t})^{(r)}}(\pi(\mathcal{H}_{m;1:t}), \pi'(\mathcal{H}_{m;1:t})). \end{aligned} \quad (42)$$

Online estimation of the paths active at epoch  $t$  can be thus performed by maximizing the filtering density of  $\pi(\mathcal{H}_{m;t})$ :

$$\pi(\widehat{\mathcal{H}}_{m;t}) = \arg \max_{\mathcal{P}} \mathbb{P}(\pi(\mathcal{H}_{m;t}) = \mathcal{P} | \mathbf{y}_{m;1:t}) \quad (43)$$

while the elements of  $\pi'(\mathcal{H}_{m;t})$  are estimated as

$$\mathbf{h}_{m,n;t} = \mathbb{E} \left[ \mathbf{h}_{m,n;t} | \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t} \right] \text{ for } n = 1, \dots, N \quad (44)$$

a task that can be accomplished by KF. Indeed, from the law of total expectation

$$\begin{aligned} \mathbf{h}_{m,n;t} &= \mathbb{E} \left[ \mathbb{E} \left[ \mathbf{h}_{m,n;t} | \pi(\mathcal{H}_{m;1:t-1}), \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t} \right] \right] \\ &\approx \frac{1}{\sum_{r:\pi(\mathcal{H}_{m;t})^{(r)}=\pi(\widehat{\mathcal{H}}_{m;t})} w_t^{(r)}} \\ &\quad \times \sum_{r:\pi(\mathcal{H}_{m;t})^{(r)}=\pi(\widehat{\mathcal{H}}_{m;t})} w_t^{(r)} \widehat{\mathbf{h}}_{m,n;t}^{(r)}. \end{aligned} \quad (45)$$

From (43) and (44), we may notice that RBPF is akin to the family of the GMAP-III estimators. Actually, we prove in the Appendix that the RBPF-based estimator is asymptotically (in signal-to-noise ratio) Bayes-optimal for the cost family (29) when we choose  $g$  to be quadratic:  $g(a_{m,n;t}^{(\ell)}, \tilde{a}_{m,n;t}^{(\ell)}) = |a_{m,n;t}^{(\ell)} - \tilde{a}_{m,n;t}^{(\ell)}|^2$ .

As anticipated earlier, the ‘‘importance distribution’’ represents a degree of freedom whose choice has a great impact on the overall algorithm efficiency. The optimal importance distribution, i.e., the one minimizing the variance of the particle weights, is [13], [19]

$$\begin{aligned} q(\pi(\mathcal{H}_{m;t}) | \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \mathbf{y}_{m;1:t}) \\ &= f(\pi(\mathcal{H}_{m;t}) | \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \mathbf{y}_{m;1:t}) \\ &= \frac{f(\mathbf{y}_{m;t} | \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \pi(\mathcal{H}_{m;t}), \mathbf{y}_{m;1:t-1})}{f(\mathbf{y}_{m;t} | \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \mathbf{y}_{m;1:t-1})} \\ &\quad \times f(\pi(\mathcal{H}_{m;t}) | \pi(\mathcal{H}_{m;t-1})^{(r)}) \end{aligned} \quad (46)$$

corresponding to the weight update law

$$\tilde{w}_t^{(r)} = w_{t-1}^{(r)} f(\mathbf{y}_{m;t} | \pi(\mathcal{H}_{1:t-1})^{(r)}, \mathbf{y}_{m;1:t-1}). \quad (47)$$

Notice that the computation of the above distribution requires an additional integration stage, viz.,

$$\begin{aligned} &f(\mathbf{y}_{m;t} | \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \mathbf{y}_{m;1:t-1}) \\ &= \int f(\mathbf{y}_{m;t} | \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \pi(\mathcal{H}_{m;t}), \mathbf{y}_{m;1:t-1}) \\ &\quad \times f(\pi(\mathcal{H}_{m;t}) | \pi(\mathcal{H}_{m;t-1})^{(r)}) \delta\pi(\mathcal{H}_{m;t}). \end{aligned} \quad (48)$$

Since  $\pi(\mathcal{H}_{m;t})$  is an FRS on a countable space, the integral above becomes a summation, whose first term can be evaluated through a KF for each value of  $\pi(\mathcal{H}_{m;t})$ .

### C. Low-Complexity Approximation

The complexity of the channel estimator introduced in Section III grows linearly with  $R$ , but the number of particles to approximate the optimal inference grows exponentially with  $L_{\max}$ . Thus, if no simplification is introduced, the applicability of the estimator is limited to situations where  $L_{\max}$  is small. The main reason for its complexity is that, given the realization  $\mathcal{H}_{m,n;t-1}$  of the channel at time  $t-1$ , the range of  $\pi(\mathcal{H}_{m,n;t})$  is the whole power set of  $\{1, \dots, L_{\max}\}$ . To be able to apply the algorithm to channels with a large  $L_{\max}$ , this range should be reduced. A reasonable simplification consists of assuming that the channel variations between two adjacent discrete-time instants are so slow that at most one multipath component is added (with probability  $\beta_1$ ) or subtracted (with probability  $\beta_2$ ). This is tantamount to assuming that at most  $L_{\max} + 1$  transitions may take place at each epoch, whereby the model transition density is the one in (49), shown at the bottom of the page. Should the multipath channel vary so rapidly that the above assumption becomes invalid, one can expect that the simplified RFS algorithm takes some discrete-time instants to adjust to large channel variations. However, as we shall describe below, simulation results show that the loss in performance to be traded off for complexity reduction (the number of particles to perform this suboptimal inference grows linearly with  $L_{\max}$ ) is acceptable.

### D. Extension to Noninteger Delays

The generalization of the channel estimation procedure to the case of noninteger delays is fairly straightforward, and requires only a few changes in the discussion above. Let  $T_s$  denote the

$$f(\mathcal{H}_{m,n;t} | \mathcal{H}_{m,n;t-1}) = \begin{cases} (1 - \beta_1 - \beta_2) \\ \quad \times \prod_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m,n;t})} f(a_{m,n;t}^{(\ell)} | a_{m,n;t-1}^{(\ell)}), & \text{if } \pi(\mathcal{H}_{m,n;t}) = \pi(\mathcal{H}_{m,n;t-1}) \\ \frac{\beta_1}{L_{\max} - |\mathcal{H}_{m,n;t-1}|} f(a_{m,n;t}^{(b)}) \\ \quad \times \prod_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m,n;t-1})} f(a_{m,n;t}^{(\ell)} | a_{m,n;t-1}^{(\ell)}), & \text{if } \pi(\mathcal{H}_{m,n;t}) = \pi(\mathcal{H}_{m,n;t-1}) \cup [m,n,b]^T \\ \frac{\beta_2}{|\mathcal{H}_{m,n;t-1}|} \\ \quad \times \prod_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m,n;t})} f(a_{m,n;t}^{(\ell)} | a_{m,n;t-1}^{(\ell)}), & \text{if } \pi(\mathcal{H}_{m,n;t}) = \pi(\mathcal{H}_{m,n;t-1}) \setminus [m,n,d]^T \\ 0, & \text{otherwise} \end{cases} \quad (49)$$



sampling interval of the OFDM system (so that the total bandwidth of the waveform is  $1/T_s$ ). Then, the singleton-or-empty set  $\mathcal{H}_{m,n;t}^{(\ell)}$  can be rewritten as in (50), shown at the bottom of the page, where  $m \in \{1, \dots, M\}$ ,  $n \in \{1, \dots, N\}$ ,  $\ell \in \{1, \dots, L_{\max}\}$ ,  $a_{m,n;t}^{(\ell)} \in \mathbb{C}$  and  $\tau_{m,n;t}^{(\ell)} \in [(\ell-1)T_s, \ell T_s]$  represents the delay of the  $\ell$ th path from the  $n$ th transmit antenna to the  $m$ th receive antenna at time  $t$ . A reasonable statistical model for the delays, which we assume from now on, is the following:

- 1)  $f_{\tau_{m,n;t}^{(\ell)}}(\tau_{m,n;t}^{(\ell)}) \sim U[(\ell-1)T_s, \ell T_s]$ ;
- 2)  $f_{\tau_{m,n;t}^{(\ell)} | \tau_{m,n;t-1}^{(\ell)}}(\tau_{m,n;t}^{(\ell)} | \tau_{m,n;t-1}^{(\ell)}) = U[\max(\tau_{m,n;t-1}^{(\ell)} - \Delta, (\ell-1)T_s), \min(\tau_{m,n;t-1}^{(\ell)} + \Delta, \ell T_s)]$  (a Markov process with reflective boundary);

where  $U(a, b)$  denotes the uniform distribution in the interval  $(a, b)$ . Defining  $\mathcal{H}_{m,n;t}$  and  $\mathcal{H}_{m;t}$  as in (4) and (5), respectively, the FRS  $\mathcal{S}^{(\ell)}(\mathcal{H}_{m,n;t-1}^{(\ell)})$  introduced in (7) becomes as in (51).

$$\begin{aligned} \mathcal{S}^{(\ell)}(\mathcal{H}_{m,n;t-1}^{(\ell)}) &= \begin{cases} \emptyset, & \text{w.p. } P_{\text{death}} \\ \left\{ \left[ m, n, \ell, \tau_{m,n;t}^{(\ell)}, a_{m,n;t}^{(\ell)} \right]^T \right\}, & \text{w.p. } 1 - P_{\text{death}}. \end{cases} \quad (51) \end{aligned}$$

Likewise, we may define the set of newly born paths in the  $(m, n)$  link as in

$$\mathcal{B}_{m,n;t} = \left\{ \left[ m, n, b_1, \tau_{m,n;t}^{(b_1)}, a_{m,n;t}^{(b_1)} \right], \dots, \left[ m, n, b_{|\mathcal{B}_{m,n;t}|}, \tau_{m,n;t}^{(b_{|\mathcal{B}_{m,n;t}|})}, a_{m,n;t}^{(b_{|\mathcal{B}_{m,n;t}|})} \right] \right\}. \quad (52)$$

Next, we generalize the measurement model of (1), by extending the one examined in [10] to the case of MIMO systems. Adopting the same notation as in Section II, the measurement model with noninteger delays turns out to be very similar to (1), the major difference being in the matrix  $\mathbf{T}$ , which becomes now *channel-dependent*. We have

$$\{\mathbf{T}\}_{k,\ell} = \{\mathbf{T}(\mathcal{H}_{m,n;t})\}_{k,\ell} = e^{-j2\pi k \tau_{m,n;t}^{(\ell)} / KT} \quad (53)$$

and the observations are written as

$$\tilde{\mathbf{y}}_{m;t} = \sum_{n=1}^N \tilde{\mathbf{D}}_{n;t} \mathbf{T}(\mathcal{H}_{m,n;t}) \mathbf{h}_{m,n;t} + \tilde{\mathbf{z}}_{m;t} \quad (54)$$

showing that the channel delays enter nonlinearly in the measurement model.

Generalization of the Rao–Blackwellized SMC algorithm for the noninteger case is also rather straightforward. Indeed,  $\mathcal{H}_{m;t}$  is completely specified by the three FRSs.

- 1)  $\pi(\mathcal{H}_{m;t})$ , the projection of  $\mathcal{H}_{m;t}$  onto  $\{m\} \times \{1, \dots, N\} \times \{1, \dots, L_{\max}\}$ .
- 2)  $\pi'(\mathcal{H}_{m;t}) = \bigcup_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m;t})} \{a_{m,n;t}^{(\ell)}\}$ .
- 3)  $\tau(\mathcal{H}_{m;t}) = \bigcup_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m;t})} \{\tau_{m,n;t}^{(\ell)}\}$ .

Once again, the system is CLG given  $\pi(\mathcal{H}_{m;t})$  and  $\tau(\mathcal{H}_{m;t})$ , whereby the FRS  $\pi'(\mathcal{H}_{m;t})$  can be integrated out. The only difference with respect to the case of integer delays is that we have to draw samples from  $\{\pi(\mathcal{H}_{m;t}), \tau(\mathcal{H}_{m;t})\}$  rather than from  $\pi(\mathcal{H}_{m;t})$ . Since the space where  $\{\pi(\mathcal{H}_{m;t}), \tau(\mathcal{H}_{m;t})\}$  lies is not countable any longer (it is indeed a *hybrid* one), a larger number of particles is required, but asymptotic optimality (as  $R \rightarrow \infty$ ) is not undone. An alternative approach (which we shall not pursue here) consists of resorting to EKF (or UKF) to locally linearize the model.

The Rao–Blackwellized SMC for noninteger delays is summarized in Algorithm 2.

---

**Algorithm 2: Generic RBPF Noninteger Delays**

---

- 1: Given  $\{w_{t-1}^{(r)}, \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \tau(\mathcal{H}_{m;1:t-1})^{(r)}\}_{r=1}^R$
  - 2: **for**  $r = 1$  to  $R$
  - 3: Sample  $\{\pi(\mathcal{H}_{m;t})^{(r)}, \tau(\mathcal{H}_{m;t})^{(r)}\} \sim q(\cdot | \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \tau(\mathcal{H}_{m;1:t-1})^{(r)}, \mathbf{y}_{m;1:t})$
  - 4: Set  $\pi(\mathcal{H}_{m;1:t})^{(r)} \triangleq \{\pi(\mathcal{H}_{m;1:t-1})^{(r)}, \pi(\mathcal{H}_{m;t})^{(r)}\}$
  - 5: Set  $\tau(\mathcal{H}_{m;1:t})^{(r)} \triangleq \{\tau(\mathcal{H}_{m;1:t-1})^{(r)}, \tau(\mathcal{H}_{m;t})^{(r)}\}$
  - 6: Update the weight [see (55), shown at the bottom of the page]
  - 7: **end**
  - 8: Normalization
  - 9: **for**  $r = 1$  to  $R$
  - 10:  $w_t^{(r)} = \tilde{w}_t^{(r)} / \sum_{r'=1}^R \tilde{w}_t^{(r')}$
  - 11: **end**
  - 12: Resampling with replacement
  - 13: OUTPUT  $\{w_t^{(r)}, \pi(\mathcal{H}_{m;1:t})^{(r)}, \tau(\mathcal{H}_{m;1:t})^{(r)}\}_{r=1}^R$
- 

$$\mathcal{H}_{m,n;t}^{(\ell)} \triangleq \begin{cases} \emptyset, & \text{if the } \ell\text{th path from the } n\text{th transmit to the } m\text{th receive antenna is not active at time } t \\ \left\{ \left[ m, n, \ell, \tau_{m,n;t}^{(\ell)}, a_{m,n;t}^{(\ell)} \right]^T \right\}, & \text{otherwise} \end{cases} \quad (50)$$

$$\tilde{w}_t^{(r)} = w_{t-1}^{(r)} \frac{f(\mathbf{y}_{m;t} | \pi(\mathcal{H}_{m;1:t})^{(r)}, \tau(\mathcal{H}_{m;1:t})^{(r)}, \mathbf{y}_{m;1:t-1}) f(\pi(\mathcal{H}_{m;t})^{(r)}, \tau(\mathcal{H}_{m;t})^{(r)} | \pi(\mathcal{H}_{m;t-1})^{(r)}, \tau(\mathcal{H}_{m;t-1})^{(r)})}{q(\pi(\mathcal{H}_{m;t})^{(r)}, \tau(\mathcal{H}_{m;t})^{(r)} | \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \tau(\mathcal{H}_{m;1:t-1})^{(r)}, \mathbf{y}_{m;1:t})} \quad (55)$$

A significant difference with the previous case is that no explicit form of the optimal importance distribution is available. For this reason, we resort here to the so-called *bootstrap filter*, i.e., we adopt the prior distribution as importance distribution

$$\begin{aligned} & q\left(\pi(\mathcal{H}_{m;t}), \tau(\mathcal{H}_{m;t}) \mid \pi(\mathcal{H}_{m;1:t-1})^{(r)}, \tau(\mathcal{H}_{m;1:t-1})^{(r)}, \mathbf{y}_{m;1:t}\right) \\ &= \prod_{n=1}^N f\left(\pi(\mathcal{H}_{m,n;t}) \mid \pi(\mathcal{H}_{m,n;t-1})^{(r)}\right) \\ & \quad \times f\left(\tau(\mathcal{H}_{m,n;t}) \mid \tau(\mathcal{H}_{m,n;t-1})^{(r)}, \pi(\mathcal{H}_{m,n;t-1})^{(r)}\right). \end{aligned} \quad (56)$$

Paralleling the arguments leading to (16), the transition density of the FRS of the set of the delays is written as

$$\begin{aligned} & f\left(\tau(\mathcal{H}_{m,n;t}) \mid \pi(\mathcal{H}_{m,n;t-1:t}), \tau(\mathcal{H}_{m,n;t-1})\right) \\ &= \prod_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m,n;t}) \cap \pi(\mathcal{H}_{m,n;t-1})} f\left(\tau_{m,n;t}^{(\ell)} \mid \tau_{m,n;t-1}^{(\ell)}\right) \\ & \quad \times \prod_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m,n;t}) \setminus (\pi(\mathcal{H}_{m,n;t}) \cap \pi(\mathcal{H}_{m,n;t-1}))} f\left(\tau_{m,n;t}^{(\ell)}\right) \end{aligned} \quad (57)$$

and the weight update is

$$\tilde{w}_t^{(r)} = w_{t-1}^{(r)} f\left(\mathbf{y}_{m;t} \mid \pi(\mathcal{H}_{m;1:t})^{(r)}, \tau(\mathcal{H}_{m;1:t})^{(r)}, \mathbf{y}_{m;1:t-1}\right) \quad (58)$$

where  $f(\mathbf{y}_{m;t} \mid \pi(\mathcal{H}_{m;1:t})^{(r)}, \tau(\mathcal{H}_{m;1:t})^{(r)}, \mathbf{y}_{m;1:t-1})$  can be evaluated by a Kalman filter.

#### IV. NUMERICAL RESULTS

To illustrate our algorithm, we choose a MIMO-OFDM system with  $K = 64$  subcarriers,  $N = 2$ ,  $M = 3$ ,  $P = 8$ , and frequency spacing  $D_f = 8$  between pilot subcarriers. Pilot subcarriers are located at

$$k_p = \frac{P}{2} + (p-1)D_f, \quad p = 1, \dots, P. \quad (59)$$

The average energy per symbol,  $\sigma_s^2 \triangleq \mathbb{E}[|d_{m,n;t}|^2]$ , is uniform (so that the pilot symbols are equally spaced and equally powered). Moreover, the pilot symbols are assumed to be phase-shift orthogonal (see [2]), meaning that the pilot sequence from each antenna must not only be orthogonal to the pilot sequences from other antennas, but also to be circularly shifted versions thereof. The modulation scheme is 4-QAM. As to the channel, we assume a uniform multipath delay profile, a multipath spread smaller than the guard time, and uncorrelated path gains. The overall channel energy is normalized to one.

Consider first the transmission of  $T = 100$  OFDM symbols through a channel with  $L_{\max} = 4$ ,  $P_{\text{birth}} = 0.05$ , and  $P_{\text{death}} = 0.05$ . The path gains follow the Gauss-Markov model

$$f\left(a_{m,n;1}^{(\ell)}\right) = \mathcal{N}_c\left(a_{m,n;1}^{(\ell)}; 0, \sigma_h^2\right) \quad (60)$$

$$f\left(a_{m,n;t}^{(\ell)} \mid a_{m,n;t-1}^{(\ell)}\right) = \mathcal{N}_c\left(a_{m,n;t}^{(\ell)}; \lambda a_{m,n;t-1}^{(\ell)}, (1-\lambda^2)\sigma_h^2\right) \quad (61)$$

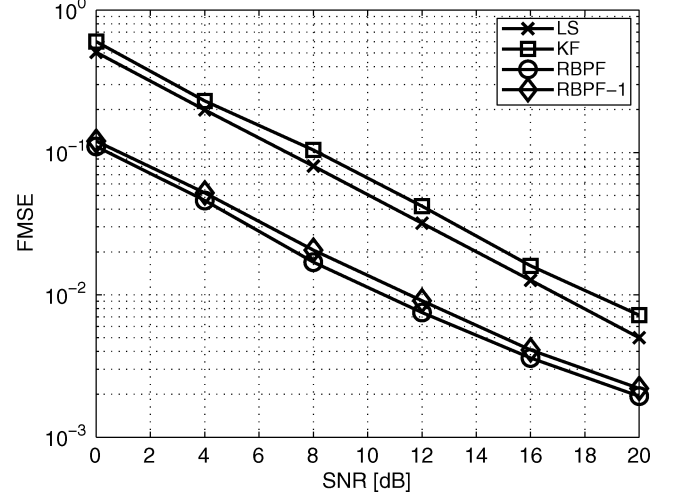


Fig. 3. FMSE versus SNR.

with  $\sigma_h^2$  the average energy of one path, and  $\lambda = 0.999$ . We have simulated an RBPF with optimal importance distribution and  $R = 50$ , the approximate version of the RBPF presented in Section III-C (referred as RBPF-1), and compared it to the least-square (LS) channel estimator of [2], which assumes that all paths are active, neglects the dynamic model of the channel, and performs the ML channel estimate in the form

$$\hat{\mathbf{h}}_{m;t} = \arg \max_{\mathbf{h}_{m;t}} f(\mathbf{y}_{m;t} \mid \mathbf{h}_{m;t}). \quad (62)$$

We also examine the behavior of a “standard” channel estimator, viz., one assuming that all paths are active and performing Kalman filtering to track their variations.

The figure of merit represented in Fig. 3 is the frequency-domain mean-square error (FMSE), defined as

$$\text{FMSE} \triangleq \frac{\sum_{i=1}^{K-P} \sum_{m=1}^M \sum_{n=1}^N \sum_{t=1}^T |H_{k_i^I, m, n; t} - \hat{H}_{k_i^I, m, n; t}|^2}{MNT(K-P)}$$

while the abscissa is labeled by the signal-to-noise ratio (SNR), defined as  $\text{SNR} \triangleq \sigma_s^2/N_0$ . This figure shows that the RBPF outperforms the standard LS estimator by about 6 dB at a FMSE of  $10^{-2}$ , while the RBPF-1 ends up in a small loss with respect to the original RBPF.

Fig. 4 shows the bit-error rate (BER) achieved by the clairvoyant detector (19) based on the channel estimates of the LS, KF, RBPF, and RBPF-1 estimators. For comparison purposes, the performance of an ML detector assuming complete channel state information (CCSI) is also shown. These results show that, at  $\text{BER} = 10^{-4}$ , RBPF outperforms LS by about 3 dB, and closely follows the ML-CCSI ideal performance for all SNRs. Notice also that the LS estimator outperforms the KF. As expected, RBPF-1 performance is very close to RBPF.

Figs. 5 and 6 show the FMSE and the BER, respectively, for a MISO (multi-input single-output) system with same parameters of the previous scenario except for  $M = 1$ ,  $L_{\max} = 6$ ,  $P = 12$ . From Fig. 5 it is clear that, augmenting the number of pilot symbols, the quality of the channel estimator does not

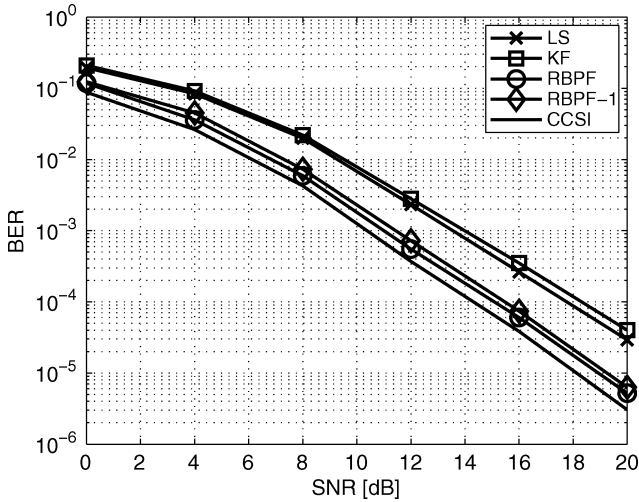


Fig. 4. BER versus SNR.

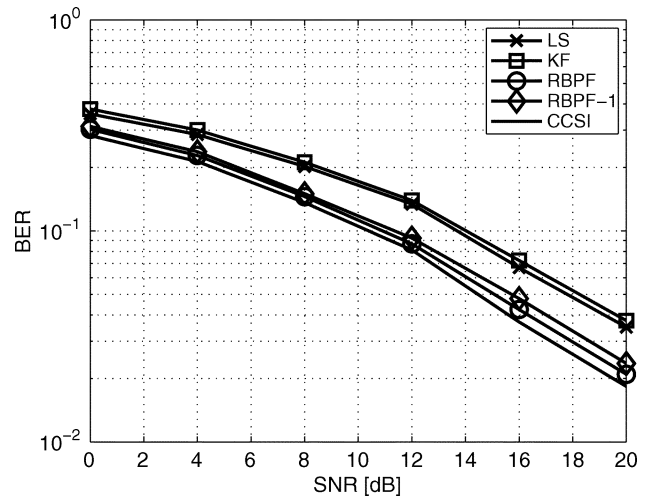


Fig. 6. BER versus SNR.

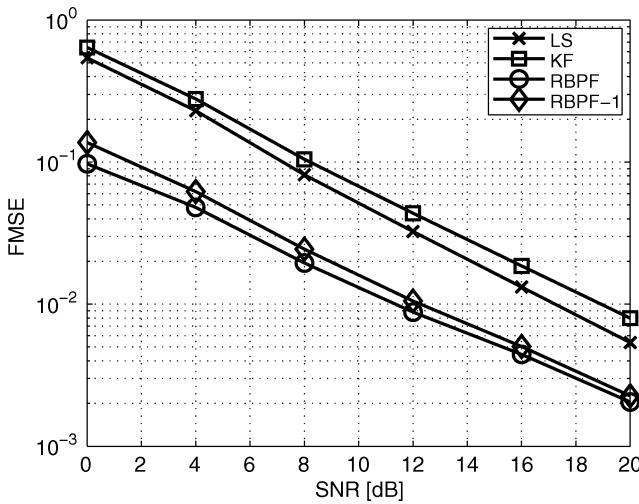


Fig. 5. FMSE versus SNR.

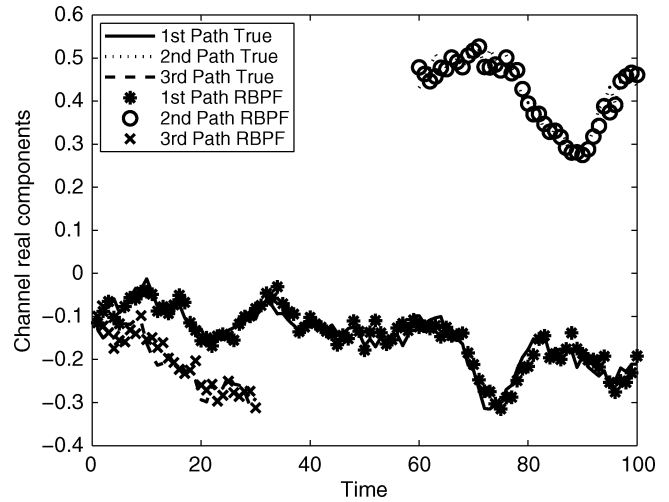


Fig. 7. Snapshot of the RBPF estimates.

degrade and the proposed algorithms outperforms the competitors. As expected, with a single receiving antenna the BER is far larger than the previous scenario wherein spatial diversity was exploited.

Figs. 7–9 show a snapshot of the real components of the estimated channel at SNR = 20 dB from the first transmit antenna to the first receive antenna, i.e.,  $\mathcal{H}_{1,1;t}$  (the parameters are the same as for Figs. 3 and 4). The first path is assumed active from  $t = 1$  to  $t = 100$ , the second path from  $t = 60$  to  $t = 100$ , and the third path from  $t = 1$  to  $t = 30$ . The fourth path is never active. Fig. 7 refers to an RBPF with optimal importance function and  $R = 50$ , Fig. 8 to an LS. The degradation of the LS estimates, with respect to the RBPF, is apparent. Indeed, the RBPF-based channel estimator estimates the active paths more accurately, immediately detecting births and deaths, while LS suffers from the lack of suitable exploitation of the channel dynamics. The advantage of the random-set-theoretic framework is especially apparent from Fig. 9, which represents the estimated real components of a conventional KF. Notice that path deaths and births start transient phases in the KF, which ultimately impair the overall estimation accuracy. This is the pri-

mary reason of the weak performance of the KF in terms of BER and FMSE: in fact, as the figure clearly shows, as long as a path remains active, the KF is able to track closely the evolution of its amplitude.

The effectiveness of the RBPF estimator can be justified analytically by considering the conditional MSE at time  $t$ , given the configuration  $\pi(\mathcal{H}_{m,1:t})$  of the paths active up to epoch  $t$ :

$$\text{MSE}_t = \mathbb{E} \left[ \sum_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m;t})} \left| a_{m,n;t}^{(\ell)} - \hat{a}_{m,n;t}^{(\ell)} \right|^2 \middle| \mathbf{y}_{1:t}, \pi(\mathcal{H}_{m;1:t}) \right]. \quad (63)$$

Since, conditioned upon  $\pi(\mathcal{H}_{m,1:t})$ , the system is linear Gaussian, the KF represents the optimum unbiased estimator at any given time  $t$ , i.e., the one minimizing (63) once  $\pi(\mathcal{H}_{m,1:t})$  is known. On the other hand, the conditional posterior Cramér Rao lower bound (PCRLB) [15] is given by the trace of the matrix  $\mathbf{M}_{m;t|t} = \mathbb{E}[(\mathbf{h}_{m;t} - \hat{\mathbf{h}}_{m;t})(\mathbf{h}_{m;t} - \hat{\mathbf{h}}_{m;t})^H | \pi(\mathcal{H}_{m;1:t}), \mathbf{y}_{m;1:t}]$  which, in turn, can be evaluated through KF. Since RBPF does not assume any prior knowledge as to the set sequence

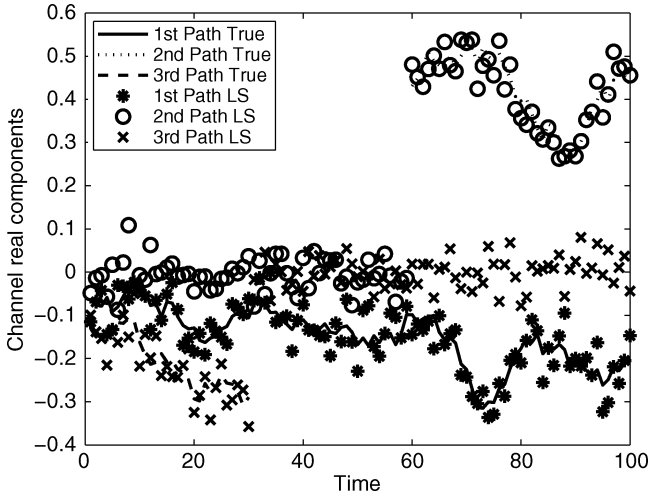


Fig. 8. Snapshot of the LS estimates.

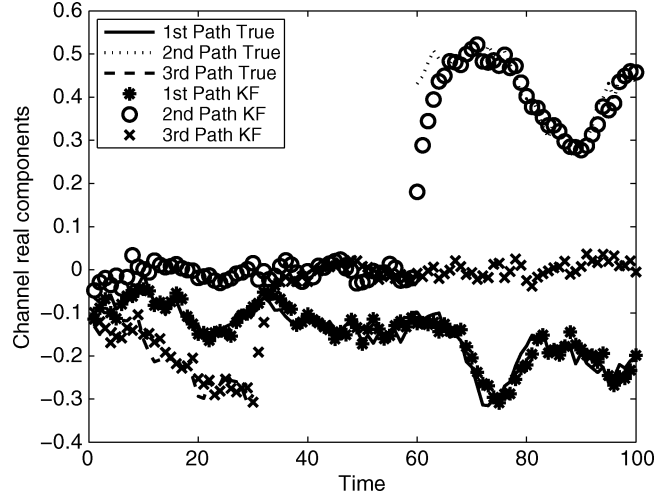


Fig. 9. Snapshot of the KF estimates.

$\pi(\mathcal{H}_{m,1:t})$ , its conditional MSE is obtained by averaging over  $\widehat{\mathcal{H}}_{m;t}$  the quantity

$$\begin{aligned} \text{MSE}_t^{\text{RBPF}} = & \underbrace{\sum_{[m,n,\ell]^T \in \pi(\widehat{\mathcal{H}}_{m;t}) \cap \pi(\mathcal{H}_{m;t})} |a_{m,n;t}^{(\ell)} - \hat{a}_{m,n;t}^{(\ell)}|^2}_{\text{Detected paths}} \\ & + \underbrace{\sum_{[m,n,\ell]^T \in \pi(\widehat{\mathcal{H}}_{m;t}) \setminus \pi(\mathcal{H}_{m;t}) \cap \pi(\widehat{\mathcal{H}}_{m;t})} |\hat{a}_{m,n;t}^{(\ell)}|^2}_{\text{False paths}} \\ & + \underbrace{\sum_{[m,n,\ell]^T \in \pi(\mathcal{H}_{m;t}) \setminus \pi(\widehat{\mathcal{H}}_{m;t})} |a_{m,n;t}^{(\ell)}|^2}_{\text{Missed paths}} \end{aligned} \quad (64)$$

showing that  $\text{tr}(\mathbf{M}_{m;t|t})$  represents a lower bound to the conditional MSE attainable by an RBPF estimator. Fig. 10 compares the conditional PCRLB to the conditional MSE of an RBPF-based channel estimator. These results refer to a scenario similar to that of Figs. 3 and 4, and consider the channel estimated by the first receive antenna, i.e.,  $\widehat{\mathcal{H}}_{1;t}$ . Indeed, the multipath channel coefficients from the first transmit antenna to the first receive antenna, i.e.,  $\widehat{\mathcal{H}}_{1,1;t}$ , are characterized as in Figs. 7–9, while  $\widehat{\mathcal{H}}_{1,2;t}$  has its first path active from  $t = 1$  to  $t = 100$ , its second path active from  $t = 1$  to  $t = 100$ , its third path active from  $t = 1$  to  $t = 45$ , and its fourth path active from  $t = 50$  to  $t = 100$ . The performance has been evaluated through 100 independent realizations of the channel amplitudes. The closeness of the conditional generalized MSE of an RBPF estimator to its lower bound confirms its suitability for channel tracking. Fig. 11 shows the FMSE of the RBPF and of the LS estimator for the case of noninteger delays. Here  $T_s = 0.05 \mu\text{s}$ ,  $T_g = 0.8 \mu\text{s}$  (the guard interval),  $\Delta = 0.01 \mu\text{s}$  and the same parameters as for Fig. 3, with  $R = 500$  (as already stated, when considering noninteger delays a larger number of particles is required). Once again, RBPF distinctively outperforms LS whose

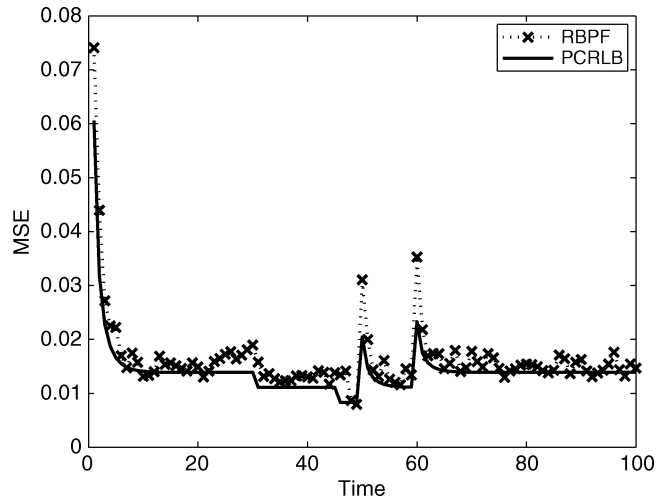


Fig. 10. MSE of the RBPF versus conditional PCRLB.

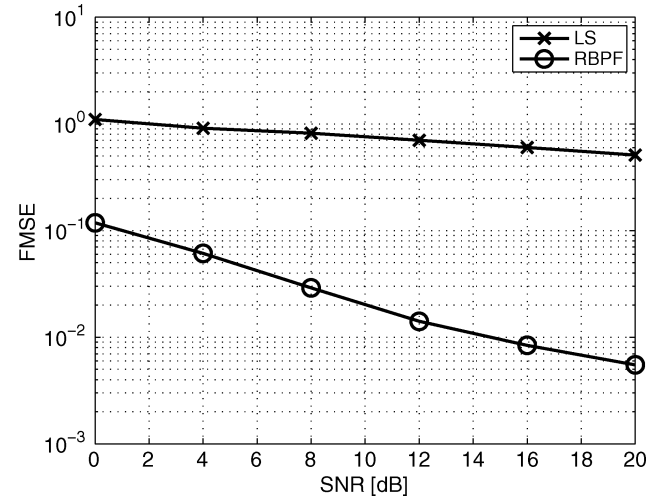


Fig. 11. MSE of the RBPF for noninteger delays.

performances are totally disappointing due to the energy leakage phenomenon [7].

## V. CONCLUSION

We have examined the estimate of multipath MIMO-OFDM channels with an unknown and time-varying number of paths. By using Random-Set Theory, we have derived recursive equations for the evolution with time of the *a posteriori* density of the channel responses, and introduced a new Bayesian channel estimator. Due to the lack of a closed-form solution of the recursive equations, sequential Monte Carlo (“particle-filtering”) methods have been applied. Rao–Blackwellized particle filtering, as applied to a conditionally linear Gaussian model, has been shown to provide an estimator which outperforms all those previously proposed. The problem of the complexity arising from long channel responses has been solved by introducing a simplification of the estimator that guarantees a linear growth of the computational burden with the channel length. A generalization to channel responses not constrained to integer delays has finally been described.

## APPENDIX

In this Appendix, we prove that the GMAP-III estimator in (43) and (44) attains the global minimum of (31) for high SNR, and hence, is asymptotically Bayes-optimal for the cost function  $C_3$  in (29). Denote the estimator  $\widehat{\mathcal{H}}_{m;t} \triangleq \widehat{\mathcal{H}}_{m;t}(\mathbf{y}_{m;1:t})$ . The Bayes-optimal estimator minimizes the conditional risk

$$\begin{aligned}
 R(\widehat{\mathcal{H}}_{m;t}|\mathbf{y}_{m;1:t}) &\triangleq \int C_3(\mathcal{H}_{m;t}, \widehat{\mathcal{H}}_{m;t}) f(\mathcal{H}_{m;t}|\mathbf{y}_{m;1:t}) \delta\mathcal{H}_{m;t} \\
 &= \int_{\pi(\mathcal{H}_{m;t}) \neq \pi(\widehat{\mathcal{H}}_{m;t})} C_3(\mathcal{H}_{m;t}, \widehat{\mathcal{H}}_{m;t}) f(\mathcal{H}_{m;t}|\mathbf{y}_{m;1:t}) \delta\mathcal{H}_{m;t} \\
 &\quad + \int_{\pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t})} C_3(\mathcal{H}_{m;t}, \widehat{\mathcal{H}}_{m;t}) f(\mathcal{H}_{m;t}|\mathbf{y}_{m;1:t}) \delta\mathcal{H}_{m;t} \\
 &= Q \left[ 1 - \mathbb{P} \left( \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) | \mathbf{y}_{m;1:t} \right) \right] \\
 &\quad + \mathbb{P} \left( \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) | \mathbf{y}_{m;1:t} \right) \\
 &\quad \times \mathbb{E} \left[ \sum_{[m,n,\ell] \in \pi(\mathcal{H}_{m;t})} g \left( a_{m,n;t}^{(\ell)}, \widehat{a}_{m,n;t}^{(\ell)} \right) \right. \\
 &\quad \left. | \mathbf{y}_{m;1:t}, \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) \right] \\
 &= Q - \mathbb{P} \left( \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) | \mathbf{y}_{m;1:t} \right) \\
 &\quad \times \left\{ Q - \mathbb{E} \left[ \sum_{[m,n,\ell] \in \pi(\mathcal{H}_{m;t})} g \left( a_{m,n;t}^{(\ell)}, \widehat{a}_{m,n;t}^{(\ell)} \right) \right. \right. \\
 &\quad \left. \left. | \mathbf{y}_{m;1:t}, \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) \right] \right\}. \quad (65)
 \end{aligned}$$

Consider  $g(a_{m,n;t}^{(\ell)}, \widehat{a}_{m,n;t}^{(\ell)}) \triangleq |a_{m,n;t}^{(\ell)} - \widehat{a}_{m,n;t}^{(\ell)}|^2$ . Then,  $\mathbb{E}[\sum_{[m,n,\ell] \in \pi(\mathcal{H}_{m;t})} |a_{m,n;t}^{(\ell)} - \widehat{a}_{m,n;t}^{(\ell)}|^2 | \mathbf{y}_{m;1:t}, \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t})]$  corresponds to the total mean-squared error in the estimation of the continuous parameters, under the assumption that the

discrete parameter has been correctly estimated. The CLG system modeling our MIMO-OFDM system can be described as

$$\mathbf{h}_{m;t} = \mathbf{G}_t(\pi(\mathcal{H}_{m;t-1:t})) \mathbf{h}_{m;t-1} + \mathbf{N}_t(\pi(\mathcal{H}_{m;t-1:t})) \mathbf{n}_{m;t} \quad (66a)$$

$$\mathbf{y}_{m;t} = \mathbf{A}_t \mathbf{h}_{m;t} + \mathbf{z}_{m;t} \quad (66b)$$

where  $\mathbf{G}_t(\pi(\mathcal{H}_{m;t-1:t}))$  is a  $NL_{\max} \times NL_{\max}$  matrix whose  $(n_1\ell_1, n_2\ell_2)$  entry is given in (67)

$$\begin{aligned}
 &\mathbf{G}_t(\pi(\mathcal{H}_{m;t-1:t}))_{n_1\ell_1, n_2\ell_2} \\
 &= \begin{cases} \lambda, & \text{if } \ell_1 = \ell_2, n_1 = n_2, [m, n_1, \ell_1] \in \pi(\mathcal{H}_{m;t}) \\ & \text{and } [m, n_1, \ell_1] \in \pi(\mathcal{H}_{m;t-1}) \\ 0, & \text{otherwise} \end{cases} \quad (67)
 \end{aligned}$$

and  $\mathbf{N}_t(\pi(\mathcal{H}_{m;t-1:t}))$  is a  $NL_{\max} \times NL_{\max}$  matrix whose  $(n_1\ell_1, n_2\ell_2)$  entry is given in<sup>7</sup>

$$\begin{aligned}
 &\{\mathbf{N}_t(\pi(\mathcal{H}_{m;t-1:t}))\}_{n_1\ell_1, n_2\ell_2} \\
 &= \begin{cases} \sqrt{1 - \lambda^2} \sigma_h, & \text{if } \ell_1 = \ell_2, n_1 = n_2, [m, n_1, \ell_1] \in \pi(\mathcal{H}_{m;t}) \\ & \text{and } [m, n_1, \ell_1] \in \pi(\mathcal{H}_{m;t-1}) \\ \sigma_h, & \text{if } \ell_1 = \ell_2, n_1 = n_2, [m, n_1, \ell_1] \in \pi(\mathcal{H}_{m;t}) \\ & \text{and } [m, n_1, \ell_1] \notin \pi(\mathcal{H}_{m;t-1}) \\ 0, & \text{otherwise.} \end{cases} \quad (68)
 \end{aligned}$$

Moreover,  $\mathbf{A}_t = [\mathbf{D}_{1;t} \mathbf{F} \cdots \mathbf{D}_{N;t} \mathbf{F}]$  is a  $P \times L_{\max} N$  matrix and  $\mathbf{h}_{m;t} = [\mathbf{h}_{m,1;t}^T \cdots \mathbf{h}_{m,N;t}^T]^T$  is a  $L_{\max} N$ -dimensional column vector.  $\mathbf{n}_t$  is a  $NL_{\max}$ -dimensional column vector such that  $\mathbf{n}_t \sim \mathcal{N}_c(\mathbf{n}_t; \mathbf{0}_{NL_{\max} \times 1}, \mathbf{I}_{NL_{\max}})$ . Conditioned on  $\pi(\mathcal{H}_{m;1:t})$ , the matrix sequences  $\mathbf{G}_t$  and  $\mathbf{N}_t$ <sup>8</sup> are known, thus Kalman filtering can be applied to generate estimates of  $\mathbf{h}_{m;t}$ . Let us compute the total mean squared error, i.e.,  $\mathbb{E}[\sum_{[m,n,\ell] \in \pi(\mathcal{H}_{m;t})} |a_{m,n;t}^{(\ell)} - \widehat{a}_{m,n;t}^{(\ell)}|^2 | \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t}]$ . From the law of total expectation, we have

$$\begin{aligned}
 &\mathbb{E} \left[ \sum_{[m,n,\ell] \in \pi(\mathcal{H}_{m;t})} |a_{m,n;t}^{(\ell)} - \widehat{a}_{m,n;t}^{(\ell)}|^2 \right. \\
 &\quad \left. | \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t} \right] \\
 &= \mathbb{E} \left[ \mathbb{E} \left[ \sum_{[m,n,\ell] \in \pi(\mathcal{H}_{m;t})} |a_{m,n;t}^{(\ell)} - \widehat{a}_{m,n;t}^{(\ell)}|^2 \right. \right. \\
 &\quad \left. \left. | \pi(\mathcal{H}_{m;1:t-1}), \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t} \right] \right] \\
 &= \sum_{\pi(\mathcal{H}_{m;1:t-1})} f \left( \pi(\mathcal{H}_{m;1:t-1}) | \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t} \right) \\
 &\quad \times \mathbb{E} \left[ \sum_{[m,n,\ell] \in \pi(\mathcal{H}_{m;t})} |a_{m,n;t}^{(\ell)} - \widehat{a}_{m,n;t}^{(\ell)}|^2 \right. \\
 &\quad \left. | \pi(\mathcal{H}_{m;1:t-1}), \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t} \right]. \quad (69)
 \end{aligned}$$

<sup>7</sup>This definition of  $\mathbf{G}_t$  and  $\mathbf{N}_t$  are consistent with the channel model developed *supra* for the linear and Gaussian dynamics of (60) and (61).

<sup>8</sup>In the following we will not indicate explicitly the dependence of  $\mathbf{G}_t$  and  $\mathbf{N}_t$  on  $\pi(\mathcal{H}_{m;t-1:t})$  when no confusion arises.

Let us define the *a posteriori* mean-squared error matrix as

$$\mathbf{M}_{m;t|t} = \mathbb{E} \left[ (\mathbf{h}_{m;t} - \widehat{\mathbf{h}}_{m;t})(\mathbf{h}_{m;t} - \widehat{\mathbf{h}}_{m;t})^H \right. \\ \left. \left| \pi(\mathcal{H}_{m;1:t-1}), \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t} \right] \right]. \quad (70)$$

For each realization of  $\pi(\mathcal{H}_{m;1:t-1})$ , the mean-squared error matrix and the Kalman Filter estimate satisfy the recursions

$$\widehat{\mathbf{h}}_{m;t|t-1} = \mathbf{G}_t \widehat{\mathbf{h}}_{m;t-1} \quad (71a)$$

$$\mathbf{M}_{m;t|t-1} = \mathbf{G}_t \mathbf{M}_{m;t-1} \mathbf{G}_t^H + \mathbf{N}_t \mathbf{N}_t^H \quad (71b)$$

$$\mathbf{K}_t = \mathbf{M}_{m;t|t-1} \mathbf{A}_t^H (\sigma_z^2 \mathbf{I} + \mathbf{A}_t \mathbf{M}_{m;t|t-1} \mathbf{A}_t^H)^{-1} \quad (71c)$$

$$\widehat{\mathbf{h}}_{m;t} = \widehat{\mathbf{h}}_{m;t|t-1} + \mathbf{K}_t (\mathbf{y}_{m;t} - \mathbf{A}_t \widehat{\mathbf{h}}_{m;t|t-1}) \quad (71d)$$

$$\mathbf{M}_{m;t} = (\mathbf{I} - \mathbf{K}_t \mathbf{A}_t) \mathbf{M}_{m;t|t-1} \quad (71e)$$

where  $\widehat{\mathbf{h}}_{m;t|t-1}$  is the one-step prediction

$$\widehat{\mathbf{h}}_{m;t|t-1} = \mathbb{E} \left[ \mathbf{h}_{m;t} \left| \pi(\mathcal{H}_{m;1:t-1}), \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t-1} \right] \right] \quad (72)$$

and  $\mathbf{M}_{m;t|t-1}$  is the predicted mean-squared error matrix

$$\mathbf{M}_{m;t|t-1} = \mathbb{E} \left[ (\mathbf{h}_{m;t} - \widehat{\mathbf{h}}_{m;t|t-1})(\mathbf{h}_{m;t} - \widehat{\mathbf{h}}_{m;t|t-1})^H \right. \\ \left. \left| \pi(\mathcal{H}_{m;1:t-1}), \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}), \mathbf{y}_{m;1:t-1} \right] \right]. \quad (73)$$

Assume that  $P \geq NL_{\max}$ ,<sup>9</sup> so that  $\mathbf{A}_t$  is a tall matrix. In the literature,  $\mathbf{K}_t$  is referred to as the *Kalman gain* matrix. We study the asymptotic behavior of the Kalman gain as  $\sigma_z^2 \rightarrow 0$ , i.e., as  $\text{SNR} \rightarrow \infty$ . Bearing in mind that the matrix  $\mathbf{A}_t \mathbf{M}_{m;t|t-1} \mathbf{A}_t^H$  is not full-rank, and replacing the inverse in (71c) by the Moore–Penrose pseudoinverse, after straightforward calculations it can be proven that

$$\lim_{\sigma_z^2 \rightarrow 0} \mathbf{K}_t = \mathbf{A}_t^\dagger. \quad (74)$$

Thus, using a conditional Kalman filter we obtain

$$\lim_{\sigma_z^2 \rightarrow 0} \mathbf{M}_{m;t} = \mathbf{0}_{NL_{\max} \times NL_{\max}} \quad (75)$$

for each  $\pi(\mathcal{H}_{m;1:t-1})$ . This means that, provided that  $\pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t})$ , by using the expected *a posteriori* mean (i.e., the Kalman filter), for large SNR, the mean squared error tends to zero, irrespective of  $\pi(\mathcal{H}_{m;1:t-1})$ . Since each term of (69) can be seen as the trace of a mean-squared error matrix obtained by Kalman filtering, then

$$\lim_{\sigma_z^2 \rightarrow 0} \mathbb{E} \left[ \sum_{[m,n,\ell] \in \pi(\mathcal{H}_{m;t})} \left| a_{m,n;t}^{(\ell)} - \widehat{a}_{m,n;t}^{(\ell)} \right|^2 \right. \\ \left. \left| \mathbf{y}_{m;1:t}, \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) \right] \right] = 0. \quad (76)$$

If conditional Kalman filtering is used, in the high-SNR region the conditional risk in (65) is

$$\lim_{\sigma_z^2 \rightarrow 0} R(\widehat{\mathcal{H}}_{m;t} | \mathbf{y}_{m;1:t}) \\ = Q \left[ 1 - P \left( \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) | \mathbf{y}_{m;1:t} \right) \right] \quad (77)$$

<sup>9</sup>This condition is necessary in this analysis to guarantee the asymptotical optimality of GMAP-III in the SNR, but in general has to be fulfilled according to the sampling theorem, meaning that the number of pilot tones must increase when the multipath delay spread and the number of transmit antennas increase [14].

which is minimized when  $\mathbb{P}(\pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) | \mathbf{y}_{m;1:t})$  is maximized. Thus, GMAP-III is asymptotically (in SNR) Bayes-optimal. In fact, it attains the global minimum of the unconditional risk, since it first maximizes  $\mathbb{P}(\pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) | \mathbf{y}_{m;1:t})$  and then, for high SNR, forces  $\mathbb{E} \left[ \sum_{[m,n,\ell] \in \pi(\mathcal{H}_{m;t})} \left| a_{m,n;t}^{(\ell)} - \widehat{a}_{m,n;t}^{(\ell)} \right|^2 \right] \pi(\mathcal{H}_{m;t}) = \pi(\widehat{\mathcal{H}}_{m;t}) | \mathbf{y}_{m;1:t}$  to approach zero.

## REFERENCES

- [1] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
- [2] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 5, pp. 1615–1624, Jun. 2003.
- [3] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Transmission Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
- [4] E. Biglieri and M. Lops, "Multiuser detection in a dynamic environment. Part I: User identification and data detection," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3158–3170, Sep. 2007.
- [5] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.
- [6] H. Bölcskei, R. W. Heath, Jr., and A. J. Paulraj, "Blind channel identification and equalization in OFDM-based multi-antenna systems," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 96–109, Jan. 2002.
- [7] R. Chen and K. B. Letaief, "Channel estimation for space time coded OFDM systems in non-sample-spaced multipath channels," in *Proc. IEEE Wireless Communications Networking Conf. (WCNC)*, Mar. 2002, vol. 1, pp. 61–66.
- [8] A. Doucet, N. de Freitas, K. Murphy, and S. Russell, "Rao-Blackwellised particle filtering for dynamic Bayesian networks," in *Proc. 16th Annu. Conf. Uncertainty in Artificial Intelligence*, San Francisco, CA, Morgan Kaufmann Publishers, 2000, pp. 176–183.
- [9] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.
- [10] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.
- [11] W.-K. Ma, B.-N. Vo, S. S. Singh, and A. Baddeley, "Tracking an unknown time-varying number of speakers using TDOA measurements: A random finite set approach," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3291–3304, Sep. 2006.
- [12] R. Mahler, "Multi-target Bayes filtering via first-order multi-target moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, pp. 1152–1178, 2003.
- [13] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Boston, MA: Artech House, 2007.
- [14] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3065–3073, Dec. 2001.
- [15] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1386–1396, May 1998.
- [16] M. A. Vazquez and J. Miguez, "Maximum likelihood sequence detection in time frequency selective MIMO channels with unknown order," *IEEE Trans. Veh. Technol.*, to be published.
- [17] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [18] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.
- [19] M. Vihola, "Rao-Blackwellised particle filtering in random set multi-target tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 2, pp. 689–705, Apr. 2007.
- [20] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 409–4104, Nov. 2006.

- [21] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for Bayesian multi-target filtering with random finite sets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 4, pp. 1224–1245, Oct. 2005.
- [22] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Trans. Commun.*, vol. 19, pp. 628–634, Oct. 1971.



**Daniele Angelosante** (S'07–M'08) was born in Frosinone, Italy, on June 1, 1981. He received the B.Sc., Laurea Magistrale, and Ph.D. degrees in telecommunication engineering from the University of Cassino, Italy, in 2003, 2005, and 2009, respectively. He received the M.Sc. degree in electrical engineering with specialization in signal and information processing for communication from the University of 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, MN. He is currently a Postdoctoral Associate with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. His research interests lie in the areas of statistical signal processing, with emphasis on wireless communications, tracking, and compressed sensing 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, Italy, where he received the Dr. Engr. degree in 1967.

He is presently an Adjunct Professor of electrical engineering at the 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, Italy, at the 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, Australia; the Institute for Communications Engineering, Munich Institute of Technology, Germany; the Institute for Infocomm Research, National University of Singapore; the National Taiwan University, Taipei, China; the University of Cambridge, U.K., and ETH Zurich, Switzerland.

Dr. 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, 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, 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 and, during 1991, 1998, and 2000, he was on sabbatical leave at University of Connecticut, Rice University, and Princeton University, respectively. During 2008 he was a visiting professor with University of Minnesota and Columbia University.