A Blind Sequential Monte Carlo Detector for OFDM Systems in the Presence of Phase Noise, Multipath fading, and Channel Order Uncertainty

Derek Yee, James P. Reilly*, and T. Kirubarajan
Department of Electrical and Computer Engineering,
McMaster University, 1280 Main St. W.,
Hamilton, Ontario, Canada L8S 4K1

Abstract

In this paper, an efficient detector is developed to address the blind detection problem for an orthogonal frequency division multiplexing (OFDM) system in the presence of phase noise and unknown multipath fading, even with channel order that is possibly not known and time–varying.

The proposed maximum *a posteriori* detector is a combination of the sequential Monte Carlo (SMC) method and the variance reduction strategy known as Rao–Blackwellization. Being blind, the developed detector, does not rely on pilot tones for the detection of the transmitted data. However, as in most work found in the literature, the aforementioned detector, which we call the RB–SMC detector, invokes the assumption of a fixed and known channel order which may be a limitation in a number of scenarios. Therefore, to relax this assumption, we model channel order uncertainty via a first order Markov process, and subsequently introduce appropriate extensions to the RB–SMC detector, thereby, proposing a novel algorithm called the E–RB–SMC detector.

The performance of the novel SMC–based detectors are validated through computer simulations. It is shown that the proposed SMC–based detectors achieve near bound performance. In terms of convergence speed, the proposed E–RB–SMC detector also shows the smallest acquisition time amongst the considered algorithms.

Index Terms

OFDM, phase noise, blind detection, sequential Monte Carlo, particle filtering, frequency selective fading, channel order.

J. Reilly, corresponding author: ph: 905 525 9140 x22895, fax: 905 521 2922, email: reillyj@mcmaster.ca

I. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) is a multi–carrier transmission technique that has quickly become the preferred modulation scheme for high rate communication over wireless frequency selective fading channels. Some applications include the digital audio broadcasting (DAB), digital video broadcasting (DVB) standards, and the wireless LAN standards, such as IEEE 802.11a, and HiperLAN2 [1]. The underlying idea is to divide the high rate data stream into parallel lower rate substreams, each of which modulates an orthogonal subcarrier that is transmitted in parallel to maintain the total desired data rate. With the appending of a cyclic prefix (CP), the lower rate substreams are essentially free of inter–symbol interference (ISI). Therefore, OFDM effectively converts the frequency selective fading channel into a set of parallel flat fading channels, thereby eliminating the need for complex time domain equalization.

For coherent detection in an OFDM system, it is necessary to estimate and track the channel. Indeed, a great number of algorithms have been proposed in the literature. In general, we can broadly divide these algorithms into two classes: i) blind and ii) pilot tone (known data) aided. A representative sample of various pilot tone aided techniques is given by [4], [9], [14], [21]. Compared to blind techniques, pilot tone aided algorithms are generally spectrally inefficient. Indeed, the inclusion of pilot tones wastes valuable bandwidth. For this reason, much research has focused on algorithms that blindly estimates the channel. As such, blind techniques detect the data without the use of any pilot tones. Some recent works are [16] and [20], which are blind algorithms for a perfectly synchronized OFDM system in an unknown frequency selective fading channel. Note that the assumption of a perfectly synchronized OFDM system is rather generous since omnipresent nuisances such as frequency offset and phase noise (PN) will inevitably lead to the loss of perfect synchronization. In practice, imperfect synchronization due to the aforementioned nonidealities may significantly degrade the performance of these algorithms, but more importantly, the achievable BER of the considered OFDM system.

A number of researchers have studied the performance of OFDM systems with imperfect synchronization. For instance, in [25], the authors studied the joint effects of frequency offset and PN, while in [24], [28], [29] the authors specifically concentrated on the effects of PN, all for the case of an additive white Gaussian noise (AWGN) channel.

In fact, OFDM systems suffer from increased sensitivity to random PN that is introduced by the local oscillator [25]. The effect of PN is two–fold. The first is a random phase rotation that is common to all subcarriers, which is appropriately referred to as the common phase error (CPE). The second is the introduction of intercarrier interference (ICI), which results from the loss of orthogonality between each subcarrier. In practice, both of these effects may significantly degrade the performance of the system. Thus, a number of PN compensation algorithms have been proposed in the literature.

In [23], [31], using pilot tones and assuming knowledge of the channel prior to PN compensation, it is proposed to counter rotate the received signal constellation via an estimate of the CPE term. The algorithms proposed in [11], [28] also exploit pilot tones for PN compensation, but with the additional restrictive assumption of an AWGN channel. In [10], both channel estimation and PN compensation were performed, but again, rely on the presence of pilot tones in each OFDM symbol. Reference [32] considers the effect of the CPE as well as the ICI, but again uses pilot tones and assumes knowledge of the frequency selective channel prior to PN compensation. Clearly, the aforementioned PN compensation schemes either rely on pilot tones, assume knowledge of the channel, or both when compensating for PN. As mentioned before, the use of pilot tones wastes valuable bandwidth. Thus, there is significant impetus to address the blind detection problem for an OFDM system in the presence of PN and an unknown multipath fading channel.

To this end, we propose a blind algorithm that need not rely on pilot tones, or even estimate the channel when attempting to recover the transmitted symbols of interest. The proposed algorithm is rooted in modern Bayesian theory and the main objective is to compute the maximum *a posteriori* (MAP) estimates of the transmitted symbols. The transmitted symbols are differentially encoded so as to remove the inherent phase ambiguity that is characteristic of any blind detector. However, the MAP estimates of interest are analytically intractable. Hence, we resort to the sequential Monte Carlo (SMC) method, otherwise known as particle filtering [7], for *online* detection of the transmitted data.

In recent years, the SMC method has enjoyed considerable success and has shown itself to be promising approach for solving a number of difficult problems in equalization. For instance, in [19], the authors successfully address the blind detection problem for a single carrier system in an unknown frequency selective channel with channel order that is fixed (time–invariant) and possibly unknown. In [3], [27], the authors tackle the blind detection problem for an unknown flat fading channel in the presence of additive non–Gaussian noise. In [34], we presented a pilot tone aided particle filter (PF) that jointly equalizes the multipath fading channel and compensates for the CPE in an OFDM system. In [33], the authors successfully develop a blind detector for a perfectly synchronized OFDM system in an unknown multipath fading channel. There, the channel is assumed to change independently between each OFDM symbol; thus, the developed detector is not entirely appropriate for the scenario of a time–correlated multipath fading channel. Also, the derivation of the former assumes the channel order to be known and time–invariant, and therefore, does not explicitly address the problem of channel order uncertainty.

The idea of the SMC method is to use a randomly weighted set of samples (particles) to recursively build in time a point–mass approximation of the true posterior probability density function (pdf). In practice, these methods provide a means for recursively approximating various optimal Bayesian estimators of interest, even for nonlinear, and possibly non–Gaussian dynamical systems. Indeed, as the number of particles becomes very large, the approximations approach the true optimal Bayesian estimators of interest.

As mentioned above, the PF offers several advantages. In fact, as shown in this paper, the adopted particle filtering framework also allows the handling of channel order uncertainty. In particular, we will also address the problem where the channel order is unknown and time–varying. Indeed, in mobile communications, the time–varying nature of the channel implies that the delay spread changes in time, consequently, the channel order also varies with time. In OFDM, the realistic scenario of an unknown channel order has received relatively little consideration. Although, the channel order can be upper bounded, a great number of detectors proposed in the literature rely on perfect knowledge of the channel order, which may be unrealistic in practice. In order to develop a practical detector, we ultimately solve the rather difficult blind detection problem for an OFDM system in the presence of PN and unknown multipath fading channel, even with channel order that is unknown and possibly varying over time.

The remainder of this paper is organized as follows. In Section II, we introduce the considered baseband OFDM system. In Section III, we introduce the dynamic state space model (DSSM) that is critical to the development of the proposed algorithm. Section IV reviews the underlying ideas of the standard SMC method. Section V provides a derivation of the proposed detector, the Rao–Blackwellized Sequential Monte Carlo (RB–SMC) detector, which is a combination of the standard SMC method and a variance reduction technique known as Rao–Blackwellization. Section VI introduces various modifications to the RB–SMC detector such that it can address the problem of channel order uncertainty. Section VII presents some experiments that validate the performance of the proposed algorithms. Finally, Section VIII concludes this paper.

Notation: We use $(\cdot)_{1:m}$ to indicate all the elements from time 1 to time m. The notations $\boldsymbol{I}_{m\times m}$ and $\mathbf{0}_{m\times n}$ denote an m by m identity and m by n all–zero matrix, respectively. $(\cdot)^*$ and $(\cdot)^H$ denote conjugation and Hermitian transpose, respectively. The notation $\mathrm{diag}(\boldsymbol{x})$ denotes the diagonal matrix with vector $\boldsymbol{x}$ on its diagonal. Finally, a N–dimensional complex Gaussian random vector $\boldsymbol{x}$ has a pdf of the form

$$\mathcal{N}_c(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\pi^N|\boldsymbol{\Sigma}|}e^{-(\boldsymbol{x}-\boldsymbol{\mu})^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are, respectively, the mean and covariance of $\boldsymbol{x}$.

II. OFDM SYSTEM MODEL

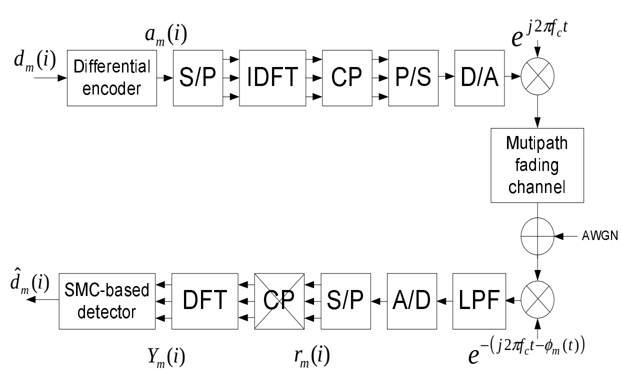


Fig. 1: Baseband OFDM System in the presence of PN and a multipath fading channel.

The considered OFDM M–ary differential phase–shift keying (MDPSK) system is implemented as shown in Fig. 1. The input to the differential encoder is a stream of *i.i.d* MPSK symbols, each of sample period T_s and taking on values from a finite alphabet set $\mathcal{A} = \{e^{j\frac{2\pi m}{M}}, m = 1, \ldots, M\}$. There are N subcarriers for each OFDM symbol. $d_m(i) \in \mathcal{A}$ denotes the MPSK symbol that will be associated with the i–th subcarrier of the m–th OFDM symbol. With differential encoding implemented in the *frequency domain* [15], *N-1* MPSK symbols $\{d_m(i)\}_{i=1}^{N-1}$ are encoded into N MDPSK symbols $\{a_m(i)\}_{i=0}^{N-1}$ such that[1]

$$a_m(0) = 1 \quad (1)$$

$$a_m(i) = a_m(i-1)d_m(i), \quad i = 1, \ldots, N-1 \quad (2)$$

where $a_m(i)$ denotes the MDPSK symbol at the i–th subcarrier of the m–th OFDM symbol.

Upon serial/parallel conversion, a N–point inverse DFT (IDFT) is applied to $\{a_m(i)\}_{i=0}^{N-1}$ which at the output produces

$$s_m(k) = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} a_m(i) e^{\frac{j2\pi ik}{N}}, \qquad k = 0, \ldots, N-1. \quad (3)$$

To remove ISI, a CP of length N_{cp} is appended to the front of each OFDM symbol, i.e., we precede $\{s_m(k)\}_{k=0}^{N-1}$ by $\{s_m(k)\}_{k=-N_{cp}}^{-1}$ where $s_m(-k) = s_m(N-k)$ for $k = 1, \ldots, N_{cp}$. At the output of the D/A converter, the m–th *cyclically extended* OFDM symbol $s_m(t)$ can be written as

$$s_m(t) = \sum_{k=-N_{cp}}^{N-1} s_m(k) g(t - kT_s) \quad (4)$$

where $g(t)$ is the impulse response of the transmitter D/A converter. We remark that the period of each cyclically extended OFDM symbol is $T = T_{CP} + T_{OFDM}$ where $T_{CP} = N_{cp}T_s$ and $T_{OFDM} = NT_s$ is the duration of the CP and the period of each OFDM symbol, respectively.

In this work, the time–varying frequency selective channel is assumed to be constant for the duration of one cyclically extended OFDM symbol. Thus, if we use $h_m(\tau)$ to denote the channel for the m-th time step, and assume that $h_m(\tau)$ is represented by a tapped delay model of length L [18], [26], $h_m(\tau)$ can be written as

$$h_m(\tau) = \sum_{l=0}^{L-1} h_{m,l} \delta(\tau - lT_s) \quad (5)$$

where $\delta(\cdot)$ denotes the Dirac delta function and $h_{m,l}$ denotes the l-th channel tap for the m-th OFDM symbol. Evidently, we also need to specify a statistical model for $h_{m,l}$. As discussed in Section III-A, we will adopt the Jakes' fading model [13] to describe $h_{m,l}$.

At the receiver, the *free–running* local oscillator introduces PN $\phi_m(t)$. As a result, the m–th received symbol $r_m(t)$ can be written as

$$r_m(t) = e^{j\phi_m(t)} \sum_{l=0}^{L-1} h_{m,l} \sum_{k=-N_{cp}}^{N-1} s_m(k) p(t - lT_s - kT_s) + n(t) \quad (6)$$

[1]In the sequel, m is a temporal index whereas i is a subcarrier index. Thus, for each m, i ranges from 0 to $N-1$.

where $n(t)$ is the AWGN, $p(t) = g(t) * f(t)$ and $f(t)$ is the impulse response of the receiver low pass filter (LPF). We assume perfect frequency and timing synchronization and that $p(t)$ satisfies the Nyquist criterion [26]. Therefore, after discarding of the CP with $N_{cp} \geq L-1$, the k–th sample of the m–th received symbol $r_m(k)$ can be written as

$$r_m(k) = \sum_{l=0}^{L-1} h_{m,l} s_m(k-l) e^{j\phi_m(k)} + n_m(k), \qquad k = 0, \ldots, N-1 \tag{7}$$

where $\phi_m(k)$ is the k–th sample of the PN for the m–th received symbol and $n_m(k)$ is a zero mean complex Gaussian random variable with variance σ^2.

With a DFT applied to $\{r_m(k)\}_{k=0}^{N-1}$, we have

$$Y_m(i) = a_m(i) H_m(i) \underbrace{I_m(0)}_{CPE} + \underbrace{ICI_m(i)}_{ICI} + N_m(i) \qquad i = 0, \ldots, N-1 \tag{8}$$

where $H_m(i) = \sum_{l=0}^{L-1} h_{m,l} e^{-j\frac{2\pi il}{N}}$, $I_m(i) = \frac{1}{N}\sum_{k=0}^{N-1} e^{j\phi_m(k)} e^{-j\frac{2\pi ik}{N}}$, $ICI_m(i) = \sum_{\substack{p=0 \\ p\neq i}}^{N-1} a_m(p) H_m(p) I_m(i-p)$ and $N_m(i)$ is the DFT of $n_m(k)$. In (8), $ICI_m(i)$ denotes the intercarrier interference (ICI), $I_m(0) = \frac{1}{N}\sum_{k=0}^{N-1} e^{j\phi_m(k)}$ denotes the the common phase error (CPE), while $H_m(i)$ denotes the frequency response at the i–th subcarrier of the m–th OFDM symbol. The CPE is a distortion common to all subcarriers, and interestingly, it can be approximated by [22], [23]

$$I_m(0) \approx \hat{I}_m(0) = e^{j\theta_m} \tag{9}$$

where

$$\theta_m = \frac{1}{N}\sum_{k=0}^{N-1} \phi_m(k) \tag{10}$$

is the average of the PN that was sampled during the period of the m–th OFDM symbol.

Although the PN, the multipath fading channel, possibly the channel order L are unknown, the main objective is to estimate $d_m(i)$. With this in mind, we aim to compute the MAP estimate of $d_m(i)$. More precisely,

$$\widehat{d}_m^{MAP}(i) = \arg\max_{d_m(i)\in A} \left\{ p(d_m(i)|\boldsymbol{Y}_{1:m}(i)) \right\} \tag{11}$$

where $\boldsymbol{Y}_{1:m}(i) = \{\boldsymbol{Y}_{1:m-1}, \boldsymbol{Y}_m(i)\}$, $\boldsymbol{Y}_m = \{Y_m(0), \ldots, Y_m(N-1)\}$ and $\boldsymbol{Y}_m(i) = \{Y_m(0), \ldots, Y_m(i)\}$. Equivalently, $p(d_m(i)|\boldsymbol{Y}_{1:m}(i))$ can be written as

$$\begin{aligned} & p(d_m(i)|\boldsymbol{Y}_{1:m}(i)) \\ &= \mathbb{E}_{p(a_m(i),a_m(i-1)|\boldsymbol{Y}_{1:m}(i))}\left[\mathcal{I}\left(a_m(i)a_m^*(i-1) - d_m(i)\right)\right] \\ &= \int \mathcal{I}\left(a_m(i)a_m^*(i-1) - d_m(i)\right) p(a_m(i), a_m(i-1)|\boldsymbol{Y}_{1:m}(i)) da_m(i) da_m(i-1) \end{aligned} \tag{12}$$

where

$$\mathcal{I}(x-a) = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{otherwise} \end{cases}.$$

However, $p(a_m(i), a_m(i-1)|\boldsymbol{Y}_{1:m}(i))$ is analytically intractable. Therefore, we propose to numerically approximate $p(a_m(i), a_m(i-1)|\boldsymbol{Y}_{1:m}(i))$ via particle filtering so that we can ultimately compute an estimate of $d_m^{MAP}(i)$.

III. State Space Model

The SMC–based detector requires a process equation and an observation equation, in other words, a DSSM. In this section, the DSSM of interest will be developed via the known statistics of the channel and the PN.

A. Channel Model

Under the wide–sense stationary and uncorrelated scattering model (WSSUS) model, the L channel taps $\{h_{m,l}\}_{l=0}^{L-1}$ are mutually uncorrelated, zero mean, wide–sense–stationary complex Gaussian processes. Because we are considering a scenario where the transmitter is fixed, the receiver is moving at a velocity v, and the transmitted signal is scattered by stationary objects surrounding the mobile, the time autocorrelation function (ACF) of the l–th channel tap $r_l(k)$ is given by Jakes' ACF [13]. That is,

$$r_l(k) = \mathbb{E}[h_{m,l}h_{m+k,l}^*] = r_l(0)J_0(2\pi f_d kT), \qquad l = 0, \ldots, L-1 \tag{13}$$

where $r_l(0)$ denotes the power of the l–th channel tap, $J_0(\cdot)$ is the zeroth–order Bessel function of the first kind, f_d denotes the maximum Doppler frequency, and T as mentioned before is the period of one cyclically extended OFDM symbol.

Unfortunately, Jakes' ACF is nonrational, and thus, it is not possible to exactly model $h_{m,l}$ via a discrete time process equation. Therefore, to circumvent this difficulty, we follow the lead of [12], and adopt a second order autoregressive AR(2) model to describe $h_{m,l}$. More precisely,

$$h_{m,l} = -\gamma_1 h_{m-1,l} - \gamma_2 h_{m-2,l} + v_{m,l}, \quad l = 0, \ldots, L-1 \tag{14}$$

where $v_{m,l}$ is a zero mean complex Gaussian random variable with variance σ_l^2.

In order to complete the model, we choose coefficients γ_1 and γ_2 such that the ACF of (14) closely matches Jakes' ACF. As shown in [30], a sensible choice is $\gamma_1 = -2r\cos(2\pi f_d T/\sqrt{2})$ and $\gamma_2 = r^2$ where $r \in [0.9, 0.999]$ is the pole radius of the AR(2) model. Finally, if we appeal to the Yule–Walker equations for (14), it can be shown that

$$\sigma_l^2 = r_l(0)\frac{(1-\gamma_2)\left((1+\gamma_2)^2 - \gamma_1^2\right)}{(1+\gamma_2)}, \quad l = 0, \ldots, L-1$$

where, because we have adopted a uniform power delay profile for $\{h_{m,l}\}_{l=0}^{L-1}$:

$$r_l(0) = \frac{1}{L} \tag{15}$$

for $l = 0, \ldots, L-1$.

B. Phase Noise Model

For a *free–running* local oscillator, a Wiener process is a reasonably accurate model for PN $\phi_m(t)$ [25], [29]. We will adopt such a model, and thus, the power spectrum of $c(t) = e^{j\phi_m(t)}$ in (6) turns out to be Lorenzian [5], i.e.,

$$S_c(f) = \frac{2}{\pi B}\frac{1}{[1+(\frac{2f}{B})^2]}$$

where B denotes the two–sided 3 dB bandwidth of $S_c(f)$. Based on this model, $\phi_m(i)$ is a discrete time Wiener process [32]. In (7), we have

$$\phi_m(i) = \phi_{m-1}(N-1) + \sum_{j=-N_{CP}}^{i} u[(m-1)(N+N_{CP})+j] \tag{16}$$

where $\phi_0(i) = 0$ for all $i = 0, \ldots, N-1$ and the $u(\cdot)'s$ denote zero mean Gaussian random variables with variance $\sigma_w^2 = 2\pi BT_s = 2\pi BT_{OFDM}/N$. For convenience, we will use BT_{OFDM} to express the effects of PN.

However, for reasons evident in Section III-D, we are mainly interested in obtaining a discrete time process equation for $\theta_m = \arg\{\hat{\mathrm{I}}_{\mathrm{m}}(0)\}$. And as first shown in [23], or by substituting (16) into (10):

$$\theta_m = \theta_{m-1} + \epsilon_m \tag{17}$$

where ϵ_m is a zero mean Gaussian random variable with variance $\sigma_{cpe}^2 = \left(\frac{2N^2+1}{3N} + N_{cp}\right)\sigma_w^2$.

C. Observation model

Using (9), the i–th post DFT sample of the m–th received symbol (8) can be approximated by[2]

$$\begin{aligned} Y_m(i) &\approx a_m(i)H_m(i)\hat{I}_m(0) + ICI_m(i) + N_m(i) \\ &= \boldsymbol{g}(a_m(i),\theta_m)^H\boldsymbol{x}_m + ICI_m(i) + N_m(i) \end{aligned} \tag{18}$$

where

$$\boldsymbol{g}(a_m(i),\theta_m) = a_m^*(i)\hat{I}_m^*(0)\boldsymbol{\rho}_i \quad \in \mathbb{C}^{2L\times 1} \tag{19}$$

$$\boldsymbol{x}_m = [\boldsymbol{h}_m^T, \boldsymbol{h}_{m-1}^T]^T \quad \in \mathbb{C}^{2L\times 1} \tag{20}$$

$$\boldsymbol{h}_m = [h_{m,0}, \ldots, h_{m,L-1}]^T \quad \in \mathbb{C}^{L\times 1} \tag{21}$$

and $\boldsymbol{\rho}_i = [1, e^{-j2\pi i/N}, \ldots, e^{-j2\pi i(L-1)/N}, \mathbf{0}_{1\times L}]^H$. $N_m(i)$ as defined before is the DFT of $n_m(k)$. Since the latter is Gaussian distributed, $N_m(i)$ is also Gaussian distributed with zero mean and variance σ^2. The ICI on the other hand is distributed according to some non-standard distribution $p(ICI_m(i))$, which does not readily admit a closed form expression. Thus, as a convenient means to characterize the ICI, we will in

[2]In the sequel, we replace the approximation with an equality.

the sequel, assume that $ICI_m(i)$ is complex Gaussian with zero mean and variance $\sigma^2_{ICI} = \pi BT_{OFDM}/3$ as computed in [32].

For convenience, we will consider the combined effect of $ICI_m(i)$ and $N_m(i)$. To this end, we define the *effective* observation noise $e_m(i) = ICI_m(i) + N_m(i)$, and its associated pdf: $p(e_m(i)) = \mathcal{N}_c(e_m(i); 0, R)$ where $R = \sigma^2_{ICI} + \sigma^2$.

D. Dynamic state space model

Equation (1) together with (2), (14), (17) and (18) lead to the considered DSSM. Specifically, for the i–th subcarrier of the m-th OFDM symbol, we have

$$
\begin{aligned}
a_m(i) &= \begin{cases} a_m(i-1)d_m(i), & i \geq 1 \\ a_m(0), & i = 0 \end{cases} && (22) \\
\boldsymbol{x}_m &= \boldsymbol{F}\boldsymbol{x}_{m-1} + \boldsymbol{C}\boldsymbol{v}_m && (23) \\
\theta_m &= \theta_{m-1} + \epsilon_m && (24) \\
Y_m(i) &= \boldsymbol{g}(a_m(i), \theta_m)^H \boldsymbol{x}_m + e_m(i) && (25)
\end{aligned}
$$

where

$$
\boldsymbol{F} = \begin{bmatrix} -\gamma_1 \boldsymbol{I}_{L\times L} & -\gamma_2 \boldsymbol{I}_{L\times L} \\ \boldsymbol{I}_{L\times L} & \boldsymbol{0}_{L\times L} \end{bmatrix} \in \mathbb{R}^{2L\times 2L}, \qquad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{I}_{L\times L} \\ \boldsymbol{0}_{L\times L} \end{bmatrix} \in \mathbb{R}^{2L\times L}
$$

and $\boldsymbol{v}_m = [v_{m,0}, \ldots, v_{m,L-1}]^T$ is a $L \times 1$ vector of white zero mean Gaussian noise with covariance matrix $\boldsymbol{Q} = \mathbb{E}[\boldsymbol{v}_m \boldsymbol{v}_m^H] = \mathrm{diag}([\sigma_0^2, \ldots, \sigma^2_{\mathrm{L-1}}]^{\mathrm{T}})$.

Lastly, let us define the required *a priori* distributions for $a_m(i)$, $\boldsymbol{x}_0$ and θ_0. Accordingly,

$$
p(a_m(i)) = \frac{1}{M}, \qquad p(\boldsymbol{x}_0) = \mathcal{N}_c(\boldsymbol{x}_0; \boldsymbol{\mu}_0, \boldsymbol{P}_0), \qquad p(\theta_0) = \mathcal{N}(\theta_0; 0, \sigma^2_{cpe}). \tag{26}
$$

where M as mentioned before is the size of the constellation; $\boldsymbol{\mu}_0$ and $\boldsymbol{P}_0$ are, respectively, the mean and covariance of $\boldsymbol{x}_0$.

IV. Sequential Monte Carlo Methods

In this section, we will provide a brief discussion of the basic SMC filter [7]. Central to the implementation of the SMC filter is the formulation of an appropriate DSSM. For this purpose, we assume a DSSM of the form

$$
\boldsymbol{\psi}_m = \boldsymbol{f}(\boldsymbol{\psi}_{m-1}) + \boldsymbol{\kappa}_m \tag{27}
$$

$$
\boldsymbol{z}_m = \boldsymbol{\beta}(\boldsymbol{\psi}_m) + \boldsymbol{n}_m \tag{28}
$$

where $\boldsymbol{f}(\cdot)$ is a possibly nonlinear process function, $\boldsymbol{\kappa}_m$ is a possibly non–Gaussian process noise, $\boldsymbol{\beta}(\cdot)$ is a possibly nonlinear measurement function and $\boldsymbol{n}_m$ is a possibly non–Gaussian measurement noise.

Generally, the objective is to recursively estimate the state $\boldsymbol{\psi}_m$ from its sequence of measurements $\boldsymbol{z}_{1:m}$. Usually, the Bayesian estimators of interest are the MAP or MMSE estimate of $\boldsymbol{\psi}_m$. Unfortunately, all Bayesian inferences rely on the availability $p(\boldsymbol{\psi}_{1:m}|\boldsymbol{z}_{1:m})$, which, in general, is analytically intractable. For this reason, alternative approaches must be considered.

One approach is the SMC filter, which in recent years, has emerged as a promising solution to the nonlinear non–Gaussian filtering problem. The SMC filter is based on the sequential importance sampling (SIS) method, which is nothing more than a constrained version of importance sampling (IS) [7]. The underlying idea is to use a randomly weighted set of samples or *particles* to recursively build in time, a point-mass approximation of the true posterior PDF.

According to the method of IS if: i) a set of particles $\{\boldsymbol{\psi}_{1:m}^{(n)}\}_{n=1}^{N_p}$ is drawn from an importance distribution $q(\boldsymbol{\psi}_{1:m}|\boldsymbol{z}_{1:m})$ where $p(\boldsymbol{\psi}_{1:m}|\boldsymbol{z}_{1:m}) > 0$ implies that $q(\boldsymbol{\psi}_{1:m}|\boldsymbol{z}_{1:m}) > 0$; and ii) each particle $\boldsymbol{\psi}_{1:m}^{(n)}$ is assigned a so–called importance weight $w_m^{(n)}$, the joint posterior pdf can be approximated by

$$\widehat{p}(\boldsymbol{\psi}_{1:m}|\boldsymbol{z}_{1:m}) = \frac{1}{W_m}\sum_{n=1}^{N_p} w_m^{(n)}\delta(\boldsymbol{\psi}_{1:m} - \boldsymbol{\psi}_{1:m}^{(n)}) \tag{29}$$

where

$$w_m^{(n)} = \frac{p(\boldsymbol{\psi}_{1:m}^{(n)}|\boldsymbol{z}_{1:m})}{q(\boldsymbol{\psi}_{1:m}^{(n)}|\boldsymbol{z}_{1:m})} \tag{30}$$

and $W_m = \sum_{n=1}^{N_p} w_m^{(n)}$. Of particular significance is that by adopting an importance distribution of the form

$$q(\boldsymbol{\psi}_{1:m}|\boldsymbol{z}_{1:m}) = q(\boldsymbol{\psi}_m|\boldsymbol{\psi}_{1:m-1}, \boldsymbol{z}_{1:m})q(\boldsymbol{\psi}_{1:m-1}|\boldsymbol{z}_{1:m-1}) \tag{31}$$

we can obtain the aforementioned SIS procedure. Indeed, it can be shown that

$$w_m^{(n)} \propto \frac{p(\boldsymbol{z}_m|\boldsymbol{\psi}_{1:m}^{(n)}, \boldsymbol{z}_{1:m-1})p(\boldsymbol{\psi}_m^{(n)}|\boldsymbol{\psi}_{1:m-1}^{(n)})}{q(\boldsymbol{\psi}_m^{(n)}|\boldsymbol{\psi}_{1:m-1}^{(n)}, \boldsymbol{z}_{1:m})} w_{m-1}^{(n)} \tag{32}$$

where $w_0^{(n)} = 1$ for all n, and $\boldsymbol{\psi}_m^{(n)} \sim q(\boldsymbol{\psi}_m^{(n)}|\boldsymbol{\psi}_{1:m-1}^{(n)}, \boldsymbol{z}_{1:m})$ for all time steps m.

More importantly is that (29) provides a means for recursively computing analytically intractable expectations of the form $\mathbb{E}_{p(\boldsymbol{\psi}_{1:m}|\boldsymbol{z}_{1:m})}[\Gamma(\boldsymbol{\psi}_m)]$, i.e.,

$$\widehat{\mathbb{E}}_{p(\boldsymbol{\psi}_m|\boldsymbol{z}_{1:m})}[\Gamma(\boldsymbol{\psi}_m)] = \frac{1}{W_m}\sum_{n=1}^{N_p} w_m^{(n)}\Gamma(\boldsymbol{\psi}_m) \tag{33}$$

where $\Gamma(\cdot)$ is an arbitrary function of $\boldsymbol{\psi}_m$. For this reason, SIS provides a viable alternative for *online* nonlinear possibly non-Gaussian optimal filtering. As an example, let us set $\boldsymbol{\psi}_m$ to $([a_m(0), \ldots, a_m(N-1)]^T, \boldsymbol{x}_m, \theta_m)$ and the measurement $\boldsymbol{z}_m$ to $\boldsymbol{Y}_m$ in (27) and (28), respectively, we can readily implement the aforementioned SIS procedure to detect $\{d_m(i)\}_{i=0}^{N-1}$.

In practice, however, the efficiency of the SIS procedure depends on the choice of the importance distribution as well as the presence of a resampling step. Indeed, the SIS procedure suffers from the degeneracy problem [7], i.e., after a few iterations, only a few particles possess significant weights, which ultimately gives rise to biased estimates with large variance. Therefore, as a means to mitigate this phenomenon and to thereby design a workable SMC filter, we introduce the resampling of particles. The underlying idea is to discard particles with weak importance weights and to multiply ones with sizable importance weights. There are a number of resampling strategies [8]. In this work, we will adopt the residual resampling strategy which as described in [17] is of complexity $O(N_p)$.

V. Rao-Blackwellized Sequential Monte Carlo detector with known channel order

For this particular DSSM, it is possible to design a better SMC filter that estimates $d_m(i)$ with smaller variance. The main idea is to exploit the linear sub-structure of the considered DSSM via a Kalman filter (KF), and thereby, marginalize out the nuisance process $\boldsymbol{x}_m$. Indeed, conditional on $(\boldsymbol{a}_{1:m}(i), \theta_{1:m})$, (23) and (25) form a linear Gaussian (LG) system in $\boldsymbol{x}_m$ for which the KF is the optimal estimator. Therefore, if we write the joint posterior pdf $p(\boldsymbol{x}_m, \boldsymbol{a}_{1:m}(i), \theta_{1:m}|\boldsymbol{Y}_{1:m}(i))$ as

$$p(\boldsymbol{x}_m, \boldsymbol{a}_{1:m}(i), \theta_{1:m}|\boldsymbol{Y}_{1:m}(i)) = \underbrace{p(\boldsymbol{x}_m|\boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i))}_{\text{Kalman filter}}\underbrace{p(\boldsymbol{a}_{1:m}(i), \theta_{1:m}|\boldsymbol{Y}_{1:m}(i))}_{\text{Particle filter}} \tag{34}$$

where $\boldsymbol{a}_{1:m}(i) = \{\boldsymbol{a}_{1:m-1}, \boldsymbol{a}_m(i)\}$, $\boldsymbol{a}_m = \{a_m(0), \ldots, a_m(N-1)\}$ and $\boldsymbol{a}_m(i) = \{a_m(0), \ldots, a_m(i)\}$, it is apparent that we can use the optimal KF to obtain the Gaussian pdf $p(\boldsymbol{x}_m|\boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i))$, and that we only require the PF to approximate the marginalized posterior pdf $p(\boldsymbol{a}_{1:m}(i), \theta_{1:m}|\boldsymbol{Y}_{1:m}(i))$.

In the literature, this approach has appeared in various forms, see e.g., [8], [3], and [12]. Intuitively, we may consider the proposed RB–SMC detector as an algorithm that uses particle filtering for the truly nonlinear state $(\boldsymbol{a}_{1:m}(i), \theta_{1:m})$, and *optimal* Kalman filtering for the conditionally linear state $\boldsymbol{x}_m$. Unlike the standard implementation of the PF, we do not employ a PF to approximate $p(\boldsymbol{x}_m, \boldsymbol{a}_{1:m}(i), \theta_{1:m}$ $|\boldsymbol{Y}_{1:m}(i))$, rather; we are using a PF to approximate a *lower dimensional* pdf $p(\boldsymbol{a}_{1:m}(i), \theta_{1:m}|\boldsymbol{Y}_{1:m}(i))$. Thus, it follows from intuition, that for a given number of particles the Rao-Blackwellized SMC detector will provide better estimates than the standard SMC detector. Formally, this observation has been proven in [8].

We now will provide a formal derivation of the proposed RB–SMC detector. To begin let us focus on the derivation of $p(\boldsymbol{x}_m|\boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i))$. As mentioned before, conditional on $(\boldsymbol{a}_{1:m}(i), \theta_{1:m})$, (23) and (25) forms a LG system in $\boldsymbol{x}_m$ for which the KF is the optimal estimator. Thus, it follows from Kalman filtering [2] that $p(\boldsymbol{x}_m|\boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i))$, $p(Y_m(i)|\boldsymbol{a}_{1:m}(i)\ , \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1))$ and $p(\boldsymbol{x}_m|\boldsymbol{a}_{1:m-1}, a_m(0), \theta_{1:m}, \boldsymbol{Y}_{1:m-1})$ are all Gaussian distributions that satisfy

$$p(\boldsymbol{x}_m|\boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i)) = \mathcal{N}_c(\boldsymbol{x}_m; \boldsymbol{x}_{m|m}(i), \boldsymbol{P}_{m|m}(i)) \tag{35}$$

$$p(Y_m(i)|\boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1)) = \mathcal{N}_c(Y_m(i); \overline{Y}_m(i), \Sigma_m(i)) \tag{36}$$

$$p(\boldsymbol{x}_m|\boldsymbol{a}_{1:m-1}, a_m(0), \theta_{1:m}, \boldsymbol{Y}_{1:m-1}) = \mathcal{N}_c(\boldsymbol{x}_m; \boldsymbol{x}_{m|m-1}, \boldsymbol{P}_{m|m-1}) \tag{37}$$

where

$$\boldsymbol{x}_{m|m}(i) = \boldsymbol{P}_{m|m}(i)\left[\boldsymbol{P}_{m|m-1}^{-1}\boldsymbol{x}_{m|m-1} + R^{-1}\sum_{k=0}^{i}\boldsymbol{g}(a_m(k),\theta_m)Y_m(k)\right] \tag{38}$$

$$\boldsymbol{P}_{m|m}(i) = \left[\boldsymbol{P}_{m|m-1}^{-1} + R^{-1}\sum_{k=0}^{i}\boldsymbol{g}(a_m(k),\theta_m)\boldsymbol{g}(a_m(k),\theta_m)^H\right]^{-1} \tag{39}$$

$$\overline{Y}_m(i) = \boldsymbol{g}(a_m(i),\theta_m)^H\boldsymbol{x}_{m|m}(i-1) \tag{40}$$

$$\Sigma_m(i) = \boldsymbol{g}(a_m(i),\theta_m)^H\boldsymbol{P}_{m|m}(i-1)\boldsymbol{g}(a_m(i),\theta_m) + R \tag{41}$$

$$\boldsymbol{x}_{m|m-1} = \boldsymbol{F}\boldsymbol{x}_{m-1|m-1}(N-1) \tag{42}$$

$$\boldsymbol{P}_{m|m-1} = \boldsymbol{F}\boldsymbol{P}_{m-1|m-1}(N-1)\boldsymbol{F}^H + \boldsymbol{C}\boldsymbol{Q}\boldsymbol{C}^H. \tag{43}$$

What remains is an estimate of the analytically intractable posterior pdf $p(\boldsymbol{a}_{1:m}(i),\theta_{1:m}|\boldsymbol{Y}_{1:m}(i))$. As mentioned before, this density is approximated by the PF. Consequently, $\hat{p}(\boldsymbol{a}_{1:m}(i),\theta_{1:m}|\boldsymbol{Y}_{1:m}(i))$ is given by

$$\hat{p}(\boldsymbol{a}_{1:m}(i),\theta_{1:m}|\boldsymbol{Y}_{1:m}(i)) = \frac{1}{W_m(i)}\sum_{n=1}^{N_p} w_m^{(n)}(i)\delta\big((\boldsymbol{a}_{1:m}(i),\theta_{1:m}) - (\boldsymbol{a}_{1:m}^{(n)}(i),\theta_{1:m}^{(n)})\big) \tag{44}$$

where the importance weight $w_m^{(n)}(i)$ satisfies

$$w_m^{(n)}(i) = \frac{p(\boldsymbol{a}_{1:m}^{(n)}(i),\theta_{1:m}^{(n)}|\boldsymbol{Y}_{1:m}(i))}{q(\boldsymbol{a}_{1:m}^{(n)}(i),\theta_{1:m}^{(n)}|\boldsymbol{Y}_{1:m}(i))} \tag{45}$$

and $W_m(i) = \sum_{n=1}^{N_p} w_m^{(n)}(i)$. Because our goal is to implement a SIS procedure, i.e., given a new set of received signals $\boldsymbol{Y}_m(i)$, we wish to obtain from $\{w_{m-1}^{(n)}(N-1)\}_{n=1}^{N_p}$ and $\{\boldsymbol{a}_{1:m-1}^{(n)},\theta_{1:m-1}^{(n)}\}_{n=1}^{N_p}$, a new set of weights $\{w_m^{(n)}(i)\}_{i=1}^{N_p}$ and particles $\{\boldsymbol{a}_{1:m}^{(n)}(i),\theta_{1:m}^{(n)}\}_{n=1}^{N_p}$, which together form an estimate of $p(\boldsymbol{a}_{1:m}(i),\theta_{1:m}|\boldsymbol{Y}_{1:m}(i))$, we choose an importance distribution of the form

$$\begin{aligned} &q(\boldsymbol{a}_{1:m}(i),\theta_{1:m}|\boldsymbol{Y}_{1:m}(i)) \\ &\quad= \prod_{k=1}^{i} q(a_m(k)|\boldsymbol{a}_{1:m}(k-1),\theta_{1:m},\boldsymbol{Y}_{1:m}(k)) \\ &\qquad\times q(a_m(0),\theta_m|\boldsymbol{a}_{1:m-1},\theta_{1:m-1},\boldsymbol{Y}_{1:m}(0))q(\boldsymbol{a}_{1:m-1},\theta_{1:m-1}|\boldsymbol{Y}_{1:m-1}). \end{aligned} \tag{46}$$

Before we make any additional progress, we must stress that the CPE $I_m(0)$ is constant throughout each OFDM symbol. It is for this reason that we only sample θ_m at $i=0$ and, in addition, also isolate $q(a_m(0),\theta_m|\boldsymbol{a}_{1:m-1},\theta_{1:m-1}\ ,\boldsymbol{Y}_{1:m}(0))$ from $\prod_{k=1}^{i} q(a_m(k)|\boldsymbol{a}_{1:m}(k-1),\theta_{1:m},\boldsymbol{Y}_{1:m}(k))$ in (46).

Now, by writing the marginalized posterior pdf as

$$\begin{aligned} &p(\boldsymbol{a}_{1:m}(i),\theta_{1:m}|\boldsymbol{Y}_{1:m}(i)) \\ &\propto \prod_{k=0}^{i} p(Y_m(k)|\boldsymbol{a}_{1:m}^{(n)}(k),\theta_{1:m}^{(n)},\boldsymbol{Y}_{1:m}(k-1)) \prod_{k=0}^{i} p(a_m(k))p(\theta_m|\theta_{m-1})p(\boldsymbol{a}_{1:m-1},\theta_{1:m-1}|\boldsymbol{Y}_{1:m-1}), \end{aligned} \tag{47}$$

we can substitute (46) and (47) into (45), and obtain the desired recursion for $w_m^{(n)}(i)$. That is,

$$w_m^{(n)}(i) \propto \begin{cases} \dfrac{p(Y_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))p(a_m^{(n)}(i))}{q(a_m^{(n)}(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i))} w_m^{(n)}(i-1), & i \geq 1 \quad \text{(48a)} \\ \\ \dfrac{p(Y_m(0)|\boldsymbol{a}_{1:m}^{(n)}(0), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m-1})p(a_m^{(n)}(0))p(\theta_m^{(n)}|\theta_{m-1}^{(n)})}{q(a_m^{(n)}(0), \theta_m^{(n)}|\boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m}(0))} w_{m-1}^{(n)}(N-1), & i = 0 \quad \text{(48b)} \end{cases}$$

where $w_0^{(n)}(N-1) = 1$ for all n, $a_m^{(n)}(i) \sim q(a_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i))$ and $(a_m^{(n)}(0), \theta_m^{(n)}) \sim q(a_m(0), \theta_m|\boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m}(0))$ for all time steps m. Note that in calculating $w_m^{(n)}(i)$, the prior $p(\theta_m|\theta_{m-1}^{(n)})$ can be obtained from (24); the predicted likelihood $p(Y_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i)\ , \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))$ can be computed by a KF conditioned on $(\boldsymbol{a}_{1:m}^{(n)}(i), \theta_{1:m}^{(n)})$ and the set of received signals $\boldsymbol{Y}_{1:m}(i-1)$ (see (36)). In short, a KF is associated with $(\boldsymbol{a}_{1:m}^{(n)}(i), \theta_{1:m}^{(n)})$ for $n = 1, \ldots, N_p$.

However, in view of (48a) and (48b), it is apparent that we must recursively compute $p(Y_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i)\ , \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))$. Looking at (36), we note that both $\boldsymbol{x}_{m|m}(i)$ and $\boldsymbol{P}_{m|m}(i)$ must be recursively computed as well, i.e., upon the receipt of the new observation $Y_m(i)$, we wish to obtain from $\boldsymbol{x}_{m|m}(i-1)$ and $\boldsymbol{P}_{m|m}(i-1)$, a new estimate of the channel $\boldsymbol{x}_{m|m}(i)$ and covariance $\boldsymbol{P}_{m|m}(i)$, which together enable the computation of $p(Y_m(i+1)|\boldsymbol{a}_{1:m}^{(n)}(i+1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i))$. With this in mind, we equivalently write (38) and (39) as

$$\boldsymbol{x}_{m|m}(i) = \boldsymbol{x}_{m|m}(i-1) + \boldsymbol{P}_{m|m}(i)\boldsymbol{g}(a_m(i), \theta_m)R^{-1}\left[Y_m(i) - \boldsymbol{g}(a_m(i), \theta_m)^H \boldsymbol{x}_{m|m}(i-1)\right] \quad (49)$$

$$\boldsymbol{P}_{m|m}(i) = \left[\boldsymbol{P}_{m|m}^{-1}(i-1) + R^{-1}\boldsymbol{\rho}_i\boldsymbol{\rho}_i^H\right]^{-1} \quad (50)$$

where $i = 0, \ldots, N-1$, $\boldsymbol{x}_{m|m}(-1) = \boldsymbol{x}_{m|m-1}$ and $\boldsymbol{P}_{m|m}(-1) = \boldsymbol{P}_{m|m-1}$. We note that $\boldsymbol{P}_{m|m}(i)$ is independent of $a_m(i)$ and θ_m for any $0 \leq i \leq N-1$. If the aforementioned bank of KF's are initialized with identical covariances, it follows that $\boldsymbol{P}_{m|m}^{(n)}(i) = \boldsymbol{P}_{m|m}(i)$ for $n = 1, \ldots, N_p$, and therefore, (50) need only be computed once for each iteration of the algorithm. Of course, this leads to a substantial reduction in the computational complexity of the algorithm.

Now, we will focus on the selection of the importance distribution. In practice, there are many choices with the only constraint being that the support of the former includes the support of the posterior density. Often, we can adopt the prior as the importance distribution, that is,

$$q(\theta_m|\boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m}(0)) = p(\theta_m|\theta_{m-1}^{(n)}) \quad (51)$$

$$q(a_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i)) = p(a_m(i)) \quad (52)$$

for sampling θ_m and $a_m(i)$, respectively. However, the prior proposes samples without knowledge of the set of received signals $\boldsymbol{Y}_{1:m}(i)$, and as a consequence, may lead to an inefficient algorithm, i.e., we may require more particles to adequately approximate $p(\boldsymbol{a}_{1:m}(i), \theta_{1:m}|\boldsymbol{Y}_{1:m}(i))$, and thus, an estimate of $d_m^{MAP}(i)$. Alternatively, we may adopt the so-called optimal importance distribution (OID) for sampling

$(a_m(i), \theta_m)$:

$$q_{OPT}(\theta_m | \boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m}(0)) = p(\theta_m | \boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m}(0)) \tag{53}$$

$$q_{OPT}(a_m(i) | \boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i)) = p(a_m(i) | \boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i)) \tag{54}$$

where

$$\begin{aligned} &p(\theta_m | \boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m}(0)) \\ &\quad = \frac{p(Y_m(0) | \boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \theta_m, \boldsymbol{Y}_{1:m-1}) p(\theta_m | \theta_{m-1}^{(n)})}{\int p(Y_m(0) | \boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \theta_m, \boldsymbol{Y}_{1:m-1}) p(\theta_m | \theta_{m-1}^{(n)}) d\theta_m} \end{aligned} \tag{55}$$

$$\begin{aligned} &p(a_m(i) | \boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i)) \\ &\quad = \frac{p(Y_m(i) | \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) p(a_m(i))}{\sum_{a_m(i) \in \mathcal{A}} p(Y_m(i) | \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) p(a_m(i))}. \end{aligned} \tag{56}$$

Unlike the prior distributions, (53) and (54) with $i = 0$ minimizes the variance of $w_m(0)$ conditional upon $(\boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)})$ and $\boldsymbol{Y}_{1:m}(0)$. As for $i > 0$, the OID (54) minimizes the variance of $w_m(i)$ conditional upon $(\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)})$ and $\boldsymbol{Y}_{1:m}(i)$. Consequently, the pair of OID's improve the efficiency of the algorithm. However, as can be seen from (36), the predicted likelihood $p(\boldsymbol{Y}_m(0) | \boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \theta_m, \boldsymbol{Y}_{1:m-1})$ depends on θ_m via a nonlinear observation function $\boldsymbol{G}(\cdot, \theta_m)$ such that it precludes the combining of (55) into a single, "easy–to–sample" Gaussian distribution. Therefore, to circumvent this difficulty, we resort to the prior (51) for sampling θ_m, and for reasons mentioned above, choose the OID (54) as the importance distribution for $a_m(i)$. With the chosen importance distributions, (48a) and (48b) simplify to

$$w_m^{(n)}(i) \propto \begin{cases} \displaystyle\sum_{a_m(i) \in \mathcal{A}} p(Y_m(i) | \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) p(a_m(i)) w_m^{(n)}(i-1), & i \geq 1 \quad \text{(57a)} \\[2ex] \displaystyle\sum_{a_m(0) \in \mathcal{A}} p(Y_m(0) | \boldsymbol{a}_{1:m-1}^{(n)}, a_m(0), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m-1}) p(a_m(0)) w_{m-1}^{(n)}(N-1), & i = 0 \quad \text{(57b)} \end{cases}$$

where $w_0^{(n)}(N-1) = 1$ for all n. In order to design an efficient but tractable algorithm, we have effectively adopted a compromise between the "full" prior and the "full" OID for sampling $(a_m(i), \theta_m)$.

Given (44), $\widehat{p}(a_m(i), a_m(i-1) | \boldsymbol{Y}_{1:m}(i))$ can now be obtained by

$$\begin{aligned} &\widehat{p}(a_m(i), a_m(i-1) | \boldsymbol{Y}_{1:m}(i)) \\ &\quad = \int \cdots \int \widehat{p}(\boldsymbol{a}_{1:m}(i), \theta_{1:m} | \boldsymbol{Y}_{1:m}(i)) d\boldsymbol{a}_{1:m-1} d\boldsymbol{a}_m(i-2) d\theta_{1:m} \\ &\quad = \frac{1}{W_m(i)} \sum_{n=1}^{N_p} w_m^{(n)}(i) \delta(a_m(i) - a_m^{(n)}(i))(a_m(i-1) - a_m^{(n)}(i-1)). \end{aligned} \tag{58}$$

Substituting (58) into (12), we finally obtain the desired estimate of $d_m^{MAP}(i)$. That is,

$$\widehat{d}_m^{MAP}(i) = \arg \max_{d_m(i) \in A} \left\{ \frac{1}{W_m(i)} \sum_{n=1}^{N_p} w_m^{(n)}(i) \mathcal{I}\left(a_m^{(n)}(i) a_m^{(n)*}(i-1) - d_m(i) \right) \right\}. \tag{59}$$

A. Complexity of the RB–SMC detector

At each iteration, the computational complexity and memory requirements of the RB–SMC detector, are approximately equal to that of $|\mathcal{A}|N_p$ one-step KF measurement prediction formulas, each calculating the mean and covariance of $Y_m(i)$ conditional upon $(\boldsymbol{a}_{1:m}^{(n)}(i), \theta_{1:m}^{(n)})$ and the set of received signals $\boldsymbol{Y}_{1:m}(i-1)$ (see (56), (57a) and (57b)). Based on this observation, it follows that both the computational complexity and storage requirements of this algorithm are $O(|\mathcal{A}|N_p)$. However, if the bank of KF's are initialized with identical covariances, we need not run N_p full KF's, as (50) need only be calculated once per iteration of the algorithm. Lastly, like other particle filtering algorithms, the RB–SMC detector is well suited for implementation via massively parallel hardware technology.

VI. Rao-Blackwellized Sequential Monte Carlo detector with unknown possibly time-varying channel order

In an OFDM–MDPSK system, the scenario of an unknown and possibly time-varying channel order has received relatively little consideration. Indeed, in most works, the typical assumption is a fixed and known channel order L. But, in practice, the time-varying nature of the channel implies that the delay spread changes in time, consequently, the channel order also varies with time.

The RB–SMC detector described in Section V assumes knowledge of the channel order. In this section, we relax this assumption and extend the applicability of the RB–SMC detector to the scenario where the channel order is unknown and time-varying. For convenience, we will refer to the resulting algorithm as the extended RB–SMC (E–RB–SMC) detector.

To begin, recall that, within a Bayesian framework, we must specify appropriate prior information for all unknowns. Hence, in addition to the predefined priors for the symbols, phase noise, and channel, we must define the following prior probability mass function (pmf) for the initial channel order l_0, that is,

$$p(l_0) = \frac{1}{L_{MAX}}, \qquad l_0 \in \mathcal{L} = \{1, ..., L_{MAX}\} \tag{60}$$

where L_{MAX} is the maximum channel order for the considered multipath fading channel. Now, to account for the evolution of the channel order, we will introduce a discrete–time, first–order Markov chain. Denoting l_m as the channel order for the m–th OFDM symbol, l_m will have time–invariant transition probabilities $p_{u,v}$ such that

$$p_{u,v} = p(l_m = v | l_{m-1} = u), \qquad u, v \in \mathcal{L} \tag{61}$$

where $p(l_m = v | l_{m-1} = u)$ is the probability that the channel switches to length v at time m given that it was length u at time $m-1$. Of course, $\sum_{v=1}^{L_{MAX}} p_{u,v} = 1$ for each u in $\mathcal{L}$.

Similar to the multipath channel $\boldsymbol{x}_m$, the channel order l_m is treated as a nuisance process, and thus, we proceed to marginalize the latter. Under this approach, it is sufficient to only draw samples for

$(a_m(i), \theta_m)$. Therefore, we proceed as in Section V, and adopt the proposed importance distributions, i.e., (51) and (54), for sampling θ_m and $a_m(i)$, respectively:

$$q(\theta_m|\boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m}(0)) = p(\theta_m|\theta_{m-1}^{(n)}) \quad (62)$$

$$q_{OPT}(a_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i)) = p(a_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i)) \quad (63)$$

where

$$p(a_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i)) \propto p(Y_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))p(a_m(i)). \quad (64)$$

Now, to account for channel order uncertainty, let us write $p(Y_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1)\ , a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))$ as

$$\begin{aligned} &p(Y_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) \\ &= \sum_{l_m \in \mathcal{L}} p(Y_m(i)|l_m, \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))p(l_m|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) \end{aligned} \quad (65)$$

where $p(Y_m(i)|l_m, \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))$ is the order–conditioned predicted likelihood, and $p(l_m\ |\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))$ is the *a posterior* pmf of the channel order, which itself can be further expanded as

$$\begin{aligned} &p(l_m|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) \\ &\quad = \frac{p(\boldsymbol{Y}_{1:m}(i-1), l_m, \boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)})}{p(\boldsymbol{Y}_{1:m}(i-1), \boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)})} \\ &\quad = \frac{\prod_{k=0}^{i-1} p(Y_m(k)|l_m, \boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(k-1))p(l_m|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m-1})}{p(\boldsymbol{Y}_m(i-1)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m-1})} \\ &\quad \propto \prod_{k=0}^{i-1} p(Y_m(k)|l_m, \boldsymbol{a}_{1:m}^{(n)}(k), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(k-1))p(l_m|\boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m-1}) \end{aligned} \quad (66)$$

where

$$p(l_m = v|\boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m-1}) = \sum_{u \in \mathcal{L}} p_{u,v} p(l_{m-1} = u|\boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m-1}) \quad (67)$$

is the *predicted* posterior pmf of l_m.

We note, from (65) and (66), that the evaluation of $p(Y_m(i)|l_m, \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}\ , \boldsymbol{Y}_{1:m}(i-1))$ is of critical importance, which through Bayes' theorem, is given as

$$\begin{aligned} &p(Y_m(i)|l_m, \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) \\ &\quad \propto \sum_{l_1} \cdots \sum_{l_{m-1}} \prod_{k=1}^{i} p(Y_m(k)|l_{1:m-1}, l_m, \boldsymbol{a}_{1:m}^{(n)}(k-1), a_m(k), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(k-1)) \\ &\qquad \times p(Y_m(0)|l_{1:m-1}, l_m, \boldsymbol{a}_{1:m}^{(n)}(0), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m-1})p(l_{1:m-1}|l_m, \boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m-1}) \end{aligned} \quad (68)$$

where

$$p(l_{1:m-1}|l_m, \boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m-1}) \propto p(l_m|l_{m-1})p(l_{1:m-1}|\boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m-1}). \tag{69}$$

Evidently, we require an exponentially increasing number of KF's to compute $p(Y_m(i)|l_m\ , \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))$, each calculating $p(Y_m(k)|l_{1:m-1}, l_m, \boldsymbol{a}_{1:m}^{(n)}(k-1), a_m(k)\ , \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(k-1))$ for a particular sequence of $l_{1:m}$ in (68). Clearly, this is not feasible in practice. Fortunately, as detailed in a later section of this paper, it is possible to recursively approximate (68), thereby, allowing the derivation of a suboptimal but more practical SMC-based filter. Assuming that we have an approximation of (68), i.e., $\widehat{p}(Y_m(i)|l_m, \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))$, and thus an approximation of $p(l_m|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))$ through (66), we can substitute (65) into (64), and obtain an approximation of the desired importance distribution for $a_m(i)$, i.e.,

$$\begin{aligned}
&\widehat{p}(a_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i)) \\
&\quad\propto \sum_{l_m \in \mathcal{L}} \widehat{p}(Y_m(i)|l_m, \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) \\
&\qquad\qquad \times \widehat{p}(l_m|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))p(a_m(i)).
\end{aligned} \tag{70}$$

Equation (70) is an unnormalized pmf in $a_m(i) \in \mathcal{A}$. In practice, $\widehat{p}(a_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i))$ can be sampled through the discrete version of the inverse transform method [6].

Since (62) and (70) are the chosen importance distributions, the associated importance weights $w_m^{(n)}(i)$ are also in the form specified by (57a) and (57b). Unlike the case of known channel order, $p(Y_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))$ is approximated by

$$\begin{aligned}
&\widehat{p}(Y_m(i)|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) \\
&\quad= \sum_{a_m(i) \in \mathcal{A}} \sum_{l_m \in \mathcal{L}} \widehat{p}(Y_m(i)|l_m, \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) \\
&\qquad\qquad \times \widehat{p}(l_m|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))p(a_m(i)),
\end{aligned} \tag{71}$$

consequently,

$$w_m^{(n)}(i) \propto \begin{cases}
\displaystyle\sum_{a_m(i) \in \mathcal{A}} \sum_{l_m \in \mathcal{L}} \widehat{p}(Y_m(i)|l_m, \boldsymbol{a}_{1:m}^{(n)}(i-1), a_m(i), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1)) & \\
\qquad \times \widehat{p}(l_m|\boldsymbol{a}_{1:m}^{(n)}(i-1), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))p(a_m(i))w_m^{(n)}(i-1), & i \geq 1 \quad \text{(72a)} \\
\displaystyle\sum_{a_m(0) \in \mathcal{A}} \sum_{l_m \in \mathcal{L}} \widehat{p}(Y_m(0)|l_m, \boldsymbol{a}_{1:m-1}^{(n)}, a_m(0), \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m-1}) & \\
\qquad \times \widehat{p}(l_m|\boldsymbol{a}_{1:m-1}^{(n)}, \theta_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m-1})p(a_m(0))w_{m-1}^{(n)}(N-1), & i = 0 \quad \text{(72b)}
\end{cases}$$

where $w_0^{(n)}(N-1) = 1$ for all n.

The estimate of $d_m^{MAP}(i)$ is given by (59), but with $(72a)$ and $(72b)$ as the importance weights. If necessary, we can also obtain an estimate of l_m^{MAP} as follows:

$$\begin{aligned}\widehat{l}_m^{MAP} &= \arg\max_{l_m \in \mathcal{L}} \{\hat{p}(l_m|\boldsymbol{Y}_{1:m})\} \\ &= \arg\max_{l_m \in \mathcal{L}} \left\{ \int \widehat{p}(l_m|\boldsymbol{a}_{1:m}, \theta_{1:m}, \boldsymbol{Y}_{1:m})\widehat{p}(\boldsymbol{a}_{1:m}, \theta_{1:m}|\boldsymbol{Y}_{1:m})d\boldsymbol{a}_{1:m}d\theta_{1:m} \right\} \\ &= \arg\max_{l_m \in \mathcal{L}} \left\{ \frac{1}{W_m(N-1)} \sum_{n=1}^{N_p} w_m^{(n)}(N-1)\widehat{p}(l_m|\boldsymbol{a}_{1:m}^{(n)}, \theta_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}) \right\}. \end{aligned} \tag{73}$$

So far, we have assumed the availability of $\widehat{p}(Y_m(i)|l_m, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1))$. Now, we will discuss a method to recursively approximate $p(Y_m(i)|l_m, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1))$, which in view of (25) can be expanded as

$$\begin{aligned}&p(Y_m(i)|l_m = v, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1)) \\ &\quad = \int p(Y_m(i)|\boldsymbol{x}_m, l_m = v, a_m(i), \theta_m)p(\boldsymbol{x}_m|l_m = v, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1))d\boldsymbol{x}_m \end{aligned} \tag{74}$$

where $v \in \mathcal{L}$, $p(Y_m(i)|\boldsymbol{x}_m, l_m = v, a_m(i), \theta_m) = \mathcal{N}_c(Y_m(i); \boldsymbol{g}(a_m(i), \theta_m)^H \boldsymbol{x}_m, R)$ is the likelihood and $p(\boldsymbol{x}_m\ |l_m = v, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1))$ is the order-conditioned posterior pdf of the channel $\boldsymbol{x}_m$. In the likelihood, $\boldsymbol{g}(a_m(i), \theta_m)^H$ is of dimension $1 \times 2v$ while $\boldsymbol{x}_m = [\boldsymbol{h}_m^T, \boldsymbol{h}_{m-1}^T]^T$ is of dimension $2v \times 1$ (the size of the vectors depend on l_m). $p(\boldsymbol{x}_m|l_m = v, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1))$, through Bayes' theorem is given by

$$\begin{aligned}&p(\boldsymbol{x}_m|l_m = v, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1)) \\ &\quad \propto \prod_{k=0}^{i-1} p(Y_m(k)|l_m = v, \boldsymbol{x}_m, a_m(k), \theta_m)p(\boldsymbol{x}_m|l_m = v, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) \end{aligned} \tag{75}$$

where $p(\boldsymbol{x}_m|l_m = v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})$ is the order–conditioned *predicted* posterior pdf of the channel $\boldsymbol{x}_m$.

Notice that, if $p(\boldsymbol{x}_m|l_m = v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})$ is Gaussian, we can use a $2v$ dimensional KF measurement update recursion to obtain the mean and covariance of (75). And given the previous result, we can then use a $2v$ dimensional KF measurement prediction recursion to compute the sufficient statistics of (74). Unfortunately, $p(\boldsymbol{x}_m|l_m = v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})$ in (75) is not Gaussian, but rather, is given by

$$\begin{aligned}&p(\boldsymbol{x}_m|l_m = v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) \\ &\quad = \int p(\boldsymbol{x}_m|l_m = v, \boldsymbol{x}_{m-1})p(\boldsymbol{x}_{m-1}|l_m = v, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})d\boldsymbol{x}_{m-1} \end{aligned} \tag{76}$$

where $p(\boldsymbol{x}_m|l_m = v, \boldsymbol{x}_{m-1}) = \mathcal{N}_c(\boldsymbol{x}_m; \boldsymbol{F}\boldsymbol{x}_{m-1}, \boldsymbol{CQC}^H)$ is the state transition prior[3] and $p(\boldsymbol{x}_{m-1}|l_m =$

[3]In view of (23), recognize that $\boldsymbol{F}$ and $\boldsymbol{CQC}^H$ in (76) are both of dimension $2v \times 2v$.

$v, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})$ is a non–standard distribution of the form

$$\begin{aligned} & p(\boldsymbol{x}_{m-1}|l_m = v, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) \\ & \quad = \sum_{l_{m-1}\in\mathcal{L}} p(\boldsymbol{x}_{m-1}|l_{m-1}, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) p(l_{m-1}|l_m = v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}). \end{aligned} \tag{77}$$

Equation (77) is non–standard in the sense that

$$p(l_{m-1}|l_m = v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) \propto p(l_m = v|l_{m-1}) p(l_{m-1}|\boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}), \tag{78}$$

while, $p(\boldsymbol{x}_{m-1}|l_{m-1}, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})$ is a Gaussian mixture model such that the number of mixands exponentially increases with m. In fact, we need a growing symphony of KF's to calculate the latter, each corresponding to one mixand of $p(\boldsymbol{x}_{m-1}|l_{m-1}, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})$. Thus, to limit computational complexity and thereby use a $2v$ dimensional KF to efficiently calculate (74), we propose to approximate (77) with a single Gaussian distribution.

In preparation, (77) is first simplified as follows:

$$\begin{aligned} & p(\boldsymbol{x}_{m-1}|l_m = v, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) \\ & \quad = \sum_{l_{m-1}\in\mathcal{L}} p(\boldsymbol{x}_{m-1}|l_{m-1}, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) p(l_{m-1}|l_m = v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) \\ & \quad \approx \sum_{l_{m-1}\in\mathcal{L}} p(\boldsymbol{x}_{m-1}|l_{m-1}, \{\boldsymbol{x}^{(r)}_{m-1|m-1}(N-1), \boldsymbol{P}^{(r)}_{m-1|m-1}(N-1)\}_{r=1}^{L_{max}}) \\ & \qquad \times p(l_{m-1}|l_m = v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) \\ & \quad = \sum_{l_{m-1}\in\mathcal{L}} p(\boldsymbol{x}_{m-1}|l_{m-1}, \{\boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1), \boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1)\}) \\ & \qquad \times p(l_{m-1}|l_m = v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}). \end{aligned} \tag{79}$$

The second line of (79) reflects the approximation that the past through $m-1$, i.e., $\{\boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}\}$, is summarized by L_{max} order-conditioned channel estimates and covariances, $\{\boldsymbol{x}^{(r)}_{m-1|m-1}(N-1), \boldsymbol{P}^{(r)}_{m-1|m-1}(N-1)\}_{r=1}^{L_{max}}$. *Accordingly, the latter are calculated by L_{max} KF's, each making use of $\{\boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}\}$, but of dimension $2r$ for $m = 1, \ldots, N_{OFDM}$ where N_{OFDM} is the number of OFDM symbols.* Therefore, with the assumption that (79) is a mixture of Gaussian pdfs, i.e., $p(\boldsymbol{x}_{m-1}|l_{m-1}, \{\boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1), \boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1)\}) = \mathcal{N}_c(\boldsymbol{x}_{m-1}; \boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1), \boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1))$ for $l_{m-1} = 1, \ldots, L_{MAX}$, and then approximating (79) with a moment–matched single Gaussian distribution, it follows that

$$\widehat{p}(\boldsymbol{x}_{m-1}|l_m = v, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) = \mathcal{N}_c(\boldsymbol{x}_{m-1}; \widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1), \widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)) \tag{80}$$

where $\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$ and $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$ are the mean and the covariance of (79), respectively.

At first glance, the mean of the mixture $\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$ is easily given by the weighted sum of the means of the component densities, i.e., $\{\boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1)\}_{l_{m-1}=1}^{L_{MAX}}$, whereas, the covariance of the mixture $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$ is given by a term known as the *spread of means* plus the weighted sums of the covariances, i.e., $\{\boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1)\}_{l_{m-1}=1}^{L_{MAX}}$ [2].[4] Actually, $\boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1)$ and $\boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1)$ in (79) are of dimension $2l_{m-1}\times 1$ and $2l_{m-1}\times 2l_{m-1}$, respectively, i.e., the size of the matrices depend on the channel order $l_{m-1}\in\mathcal{L}$, and thus, it is not possible to straight forwardly compute $\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$ and $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$ as required in the Gaussian approximation shown in (80).

Fortunately, if we define a new mean $\tilde{\boldsymbol{x}}^{(l_{m-1})}_{m-1|m-1}(N-1)$ and covariance $\tilde{\boldsymbol{P}}^{(l_{m-1})}_{m-1|m-1}(N-1)$ that is constant in dimension for $l_{m-1}=1,\dots,L_{MAX}$, i.e., $\tilde{\boldsymbol{x}}^{(l_{m-1})}_{m-1|m-1}(N-1)\in\mathbb{C}^{2L_{MAX}\times 1}$ and $\tilde{\boldsymbol{P}}^{(l_{m-1})}_{m-1|m-1}(N-1)\in\mathbb{C}^{2L_{MAX}\times 2L_{MAX}}$ for all $l_{m-1}\in\mathcal{L}$, it is possible to calculate $\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$ and $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$ as required in (80). Indeed, when calculating $\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$, all the component channel means $\{\boldsymbol{x}^{(r)}_{m-1|m-1}(N-1)\}_{r=1}^{L_{max}}$ must have conforming dimensions; similarly for the calculation of $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$ and its associated set of component covariances $\{\boldsymbol{P}^{(r)}_{m-1|m-1}(N-1)\}_{r=1}^{L_{max}}$. So in view of the structure of $\boldsymbol{x}_m=[\boldsymbol{h}_m,\boldsymbol{h}_{m-1}]^T$, let us define with the aid of MATLAB notation:

$$\tilde{\boldsymbol{x}}^{(l_{m-1})}_{m-1|m-1}(N-1)=\begin{cases}\begin{bmatrix}\boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1)[1:l_{m-1},1]\\ \mathbf{0}_{L_{MAX}-l_{m-1}\times 1}\\ \boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1)[l_{m-1}+1:2l_{m-1},1]\\ \mathbf{0}_{L_{MAX}-l_{m-1}\times 1}\end{bmatrix} & \text{if } l_{m-1}<L_{MAX} \quad \text{(81a)}\\[2ex] \boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1) & \text{if } l_{m-1}=L_{MAX} \quad \text{(81b)}\end{cases}$$

where $\boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1)[1:l_{m-1},1]$ and $\boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1)[l_{m-1}+1:2l_{m-1},1]$ are, respectively, the first and last l_{m-1} elements of $\boldsymbol{x}^{(l_{m-1})}_{m-1|m-1}(N-1)$; a new covariance matrix

$$\tilde{\boldsymbol{P}}^{(l_{m-1})}_{m-1|m-1}(N-1)=\begin{cases}\begin{bmatrix}\boldsymbol{A}_{11} & \mathbf{0}_{L_{MAX}\times L_{MAX}}\\ \mathbf{0}_{L_{MAX}\times L_{MAX}} & \boldsymbol{A}_{22}\end{bmatrix} & \text{if } l_{m-1}<L_{MAX} \quad \text{(82a)}\\[2ex] \boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1) & \text{if } l_{m-1}=L_{MAX} \quad \text{(82b)}\end{cases}$$

such that

$$\begin{aligned}\boldsymbol{A}_{11} &= \begin{bmatrix}\boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1)[1:l_{m-1},1:l_{m-1}] & \mathbf{0}_{l_{m-1}\times L_{MAX}-l_{m-1}}\\ \mathbf{0}_{L_{MAX}-l_{m-1}\times l_{m-1}} & \boldsymbol{I}_{L_{MAX}-l_{m-1}\times L_{MAX}-l_{m-1}}\end{bmatrix}\\ \boldsymbol{A}_{22} &= \begin{bmatrix}\boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1)[l_{m-1}+1:2l_{m-1},l_{m-1}+1:2l_{m-1}] & \mathbf{0}_{l_{m-1}\times L_{MAX}-l_{m-1}}\\ \mathbf{0}_{L_{MAX}-l_{m-1}\times l_{m-1}} & \boldsymbol{I}_{L_{MAX}-l_{m-1}\times L_{MAX}-l_{m-1}}\end{bmatrix}\end{aligned}$$

where $\boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1)[1:l_{m-1},1:l_{m-1}]$ and $\boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1)[l_{m-1}+1:2l_{m-1},l_{m-1}+1:2l_{m-1}]$ are, respectively, the upper left $l_{m-1}\times l_{m-1}$ and lower right $l_{m-1}\times l_{m-1}$ sub-matrices of $\boldsymbol{P}^{(l_{m-1})}_{m-1|m-1}(N-1)$.

[4]If $\boldsymbol{x}\sim p(\boldsymbol{x})$ where $p(\boldsymbol{x})=\sum_{j=1}^{M}p_j\mathcal{N}_c(\boldsymbol{x};\boldsymbol{\mu}^j,\boldsymbol{P}^j)$, it can be shown that the mean and covariance of the mixture are $\mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}]=\sum_{j=1}^{M}p_j\boldsymbol{\mu}^j$, and $cov_{p(\boldsymbol{x})}[\boldsymbol{x}]=\sum_{j=1}^{M}p_j\boldsymbol{P}^j+\underbrace{\sum_{j=1}^{M}p_j(\mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}]-\boldsymbol{\mu}^j)(\mathbb{E}_{p(\boldsymbol{x})}[\boldsymbol{x}]-\boldsymbol{\mu}^j)^H}_{\text{spread of means}}$, respectively.

In total, we have defined L_{MAX} new posterior means and covariances, i.e., $\{\tilde{\boldsymbol{x}}^{(l_{m-1})}_{m-1|m-1}(N-1)\}^{L_{MAX}}_{l_{m-1}=1}$ and $\{\tilde{\boldsymbol{P}}^{(l_{m-1})}_{m-1|m-1}(N-1)\}^{L_{MAX}}_{l_{m-1}=1}$, respectively. More importantly, by using (81a) to (82b), we can now calculate $\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1) \in \mathbb{C}^{2L_{MAX}\times 1}$ and $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1) \in \mathbb{C}^{2L_{MAX}\times 2L_{MAX}}$ according to (see Footnote 4)

$$\begin{aligned}
&\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)\\
&\quad= \sum_{l_{m-1}\in\mathcal{L}} p(l_{m-1}|l_m=v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})\tilde{\boldsymbol{x}}^{(l_{m-1})}_{m-1|m-1}(N-1) \qquad (83)\\
&\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)\\
&\quad= \sum_{l_{m-1}\in\mathcal{L}} p(l_{m-1}|l_m=v, a_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})\Big[\tilde{\boldsymbol{P}}^{(l_{m-1})}_{m-1|m-1}(N-1)\\
&\qquad+\Big(\tilde{\boldsymbol{x}}^{(l_{m-1})}_{m-1|m-1}(N-1)-\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)\Big)\Big(\tilde{\boldsymbol{x}}^{(l_{m-1})}_{m-1|m-1}(N-1)-\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)\Big)^H\Big] \qquad (84)
\end{aligned}$$

and thereby compute the Gaussian approximation described in (80).

With a Gaussian expression for (80), the analytical solutions to (76), (75), and (74) also turn out to be Gaussian. Provided that certain assumptions hold, the latter can be efficiently computed by the KF. Actually, for $l_m = v$ in (76), the system matrix $\boldsymbol{F}$ as mentioned before is of dimension $2v \times 2v$, while $\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$ and $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$ are of dimension $2L_{MAX}\times 1$ and $2L_{MAX}\times 2L_{MAX}$. Thus, it is not possible to straightforwardly evaluate (76) with a $2v$ dimensional standard KF prediction recursion. Indeed, both $\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$ and $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$ must conform to the dimension of $\boldsymbol{F}$. To accommodate this constraint, let us also define, using MATLAB notation:

$$\dot{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1) = \begin{bmatrix} \widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)[1:v,1] \\ \widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)[v+1:2v,1] \end{bmatrix} \in \mathbb{C}^{2v\times 1} \qquad (85)$$

and

$$\begin{aligned}
&\dot{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)\\
&\quad= \begin{bmatrix} \widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)[1:v,1:v] & \boldsymbol{0}_{v\times v} \\ \boldsymbol{0}_{v\times v} & \widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)[v+1:2v,v+1:2v] \end{bmatrix} \in \mathbb{C}^{2v\times 2v} \qquad (86)
\end{aligned}$$

where $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)[1:v,1:v]$ and $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)[v+1:2v,v+1:2v]$ are, respectively, the upper left $v\times v$ and lower right $v\times v$ sub-matrices of $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$.

Having constructed (85) and (86), we can now readily compute (76) for $v = 1,\ldots,L_{MAX}$, i.e.,

$$\widehat{p}(\boldsymbol{x}_m|l_m=v, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) = \mathcal{N}_c(\boldsymbol{x}_m; \boldsymbol{x}^{(v)}_{m|m-1}, \boldsymbol{P}^{(v)}_{m|m-1}) \qquad (87)$$

where $\boldsymbol{x}^{(v)}_{m|m-1} = \boldsymbol{F}\dot{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$ and $\boldsymbol{P}^{(v)}_{m|m-1} = \boldsymbol{F}\dot{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)\boldsymbol{F}^H + \boldsymbol{C}\boldsymbol{Q}\boldsymbol{C}^H$. Given (87), we can now use the standard KF measurement update recursion to calculate the product specified by (75):

$$\widehat{p}(\boldsymbol{x}_m|l_m=v, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i)) = \mathcal{N}_c(\boldsymbol{x}_m; \boldsymbol{x}^{(v)}_{m|m}(i), \boldsymbol{P}^{(v)}_{m|m}(i)) \qquad (88)$$

where $\boldsymbol{x}^{(v)}_{m|m}(i) = \boldsymbol{P}^{(v)}_{m|m}(i)\left[[\boldsymbol{P}^{(v)}_{m|m-1}]^{-1}\boldsymbol{x}^{(v)}_{m|m-1} + R^{-1}\sum_{k=0}^{i}\boldsymbol{g}(a_m(k),\theta_m)Y_m(k)\right]$ and $\boldsymbol{P}^{(v)}_{m|m}(i) = \left[[\boldsymbol{P}^{(v)}_{m|m-1}]^{-1} + R^{-1}\sum_{k=0}^{i}\boldsymbol{g}(a_m(k),\theta_m)\boldsymbol{g}(a_m(k),\theta_m)^H\right]^{-1}$. More importantly, by using (88), we can finally straightforwardly calculate the order–conditioned predicted likelihood (74) as required in the E–RB–SMC detector, i.e.,

$$\widehat{p}(Y_m(i)|l_m = v, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1)) = \mathcal{N}_c(Y_m(i); \overline{Y}^{(v)}_m(i), \Sigma^{(v)}_m(i)) \tag{89}$$

where $\overline{Y}^{(v)}_m(i) = \boldsymbol{g}(a_m(i),\theta_m)^H\boldsymbol{x}^{(v)}_{m|m}(i-1)$ and $\Sigma^{(v)}_m(i) = \boldsymbol{g}(a_m(i),\theta_m)^H\boldsymbol{P}^{(v)}_{m|m}(i-1)\boldsymbol{g}(a_m(i),\theta_m)+R$.

In summary, the KF conditioned on $\{\boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}\}$, and matched to possible value of $l_m = v \in \mathcal{L}$, proceeds at time m as follows:

1) Using $\{\boldsymbol{x}^{(r)}_{m-1|m-1}(N-1), \boldsymbol{P}^{(r)}_{m-1|m-1}(N-1)\}_{r=1}^{L_{max}}$, compute the Gaussian approximation of $p(\boldsymbol{x}_{m-1} \,|l_m = v, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})$ according to (80) through (84).
2) Assuming that we have computed $\widehat{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$ and $\widehat{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$ from step 1, use (85) to form $\dot{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$ and (86) to form $\dot{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$.
3) Using $\dot{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1)$ and $\dot{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)$ obtained from step 2, compute $\widehat{p}(\boldsymbol{x}_m|l_m = v, \boldsymbol{a}_{1:m-1}\,, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})$, $\widehat{p}(\boldsymbol{x}_m|l_m = v, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i))$, and $\widehat{p}(Y_m(i)|l_m = v, \boldsymbol{a}_{1:m}(i)\,, \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1))$ according to

$$\widehat{p}(\boldsymbol{x}_m|l_m = v, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1}) = \mathcal{N}_c(\boldsymbol{x}_m; \boldsymbol{x}^{(v)}_{m|m-1}, \boldsymbol{P}^{(v)}_{m|m-1}) \tag{90}$$

$$\widehat{p}(\boldsymbol{x}_m|l_m = v, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i)) = \mathcal{N}_c(\boldsymbol{x}_m; \boldsymbol{x}^{(v)}_{m|m}(i), \boldsymbol{P}^{(v)}_{m|m}(i)) \tag{91}$$

$$\widehat{p}(Y_m(i)|l_m = v, \boldsymbol{a}_{1:m}(i), \theta_{1:m}, \boldsymbol{Y}_{1:m}(i-1)) = \mathcal{N}_c(Y_m(i); \overline{Y}^{(v)}_m(i), \Sigma^{(v)}_m(i)) \tag{92}$$

where

$$\boldsymbol{x}^{(v)}_{m|m-1} = \boldsymbol{F}\dot{\boldsymbol{x}}^{(v)}_{m-1|m-1}(N-1) \tag{93}$$

$$\boldsymbol{P}^{(v)}_{m|m-1} = \boldsymbol{F}\dot{\boldsymbol{P}}^{(v)}_{m-1|m-1}(N-1)\boldsymbol{F}^H + \boldsymbol{C}\boldsymbol{Q}\boldsymbol{C}^H \tag{94}$$

$$\boldsymbol{x}^{(v)}_{m|m}(i) = \boldsymbol{x}^{(v)}_{m|m}(i-1) + \boldsymbol{P}^{(v)}_{m|m}(i)\boldsymbol{g}(a_m(i),\theta_m)R^{-1}\left(Y_m(i) - \overline{Y}^{(v)}_m(i)\right) \tag{95}$$

$$\boldsymbol{P}^{(v)}_{m|m}(i) = \left[[\boldsymbol{P}^{(v)}_{m|m}(i-1)]^{-1} + R^{-1}\boldsymbol{\rho}_i\boldsymbol{\rho}_i^H\right]^{-1} \tag{96}$$

$$\overline{Y}^{(v)}_m(i) = \boldsymbol{g}(a_m(i),\theta_m)^H\boldsymbol{x}^{(v)}_{m|m}(i-1) \tag{97}$$

$$\Sigma^{(v)}_m(i) = \boldsymbol{g}(a_m(i),\theta_m)^H\boldsymbol{P}^{(v)}_{m|m}(i-1)\boldsymbol{g}(a_m(i),\theta_m) + R \tag{98}$$

and $\boldsymbol{x}^{(v)}_{m|m}(-1) = \boldsymbol{x}^{(v)}_{m|m-1}$ and $\boldsymbol{P}^{(v)}_{m|m}(-1) = \boldsymbol{P}^{(v)}_{m|m-1}$.

1) Complexity of the E–RB–SMC detector: At each iteration, the computational complexity and memory requirements of this algorithm are approximately equal to that of $|\mathcal{A}|L_{MAX}N_p$ KF measurement prediction formulas, each calculating the mean and covariance of $Y_m(i)$ conditional upon $(\boldsymbol{a}^{(n)}_{1:m}(i), \theta^{(n)}_{1:m}, \boldsymbol{Y}_{1:m}(i-1))$ and a realization of l_m (see (70), (72a) and (72b)). Thus, at each iteration of the algorithm, both the computational complexity and storage requirements of this algorithm are of $O(|\mathcal{A}|L_{MAX}N_p)$. Although the proposed algorithm may be computationally expensive, the E–RB–SMC detector like the RB–SMC detector, is well suited for implementation via massively parallel hardware technology.

VII. Simulations

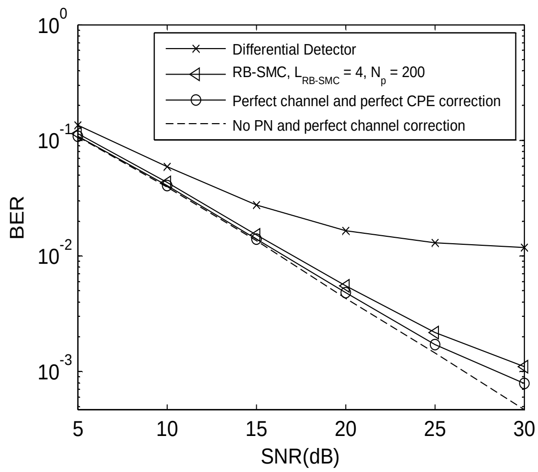
The considered OFDM–MDPSK system employs $N = 64$ subcarriers, a CP of length $N_{cp} = 5$ and total channel bandwidth of $B_w = 1\ MHz$. The adopted modulation scheme is differential BPSK, i.e., $M = 2$. Of course, there are no theoretical obstacles preventing the consideration of more complex high order *differentially encoded* modulation schemes, i.e., $M > 2$. For the simulations, we set the maximum channel order L_{MAX} to 4, the time-Doppler fading rate $f_d T$ to 0.045, and generated each tap of the corresponding multipath fading channel with the improved sum–of–sinusoids simulation model introduced in [35]. Regarding the simulation of the PN, we generated each sample of $\phi_m(i)$ in accordance with the discrete Wiener PN model presented in Section III-B (see (16) for specific details).

The various receiver schemes under consideration are as follows:

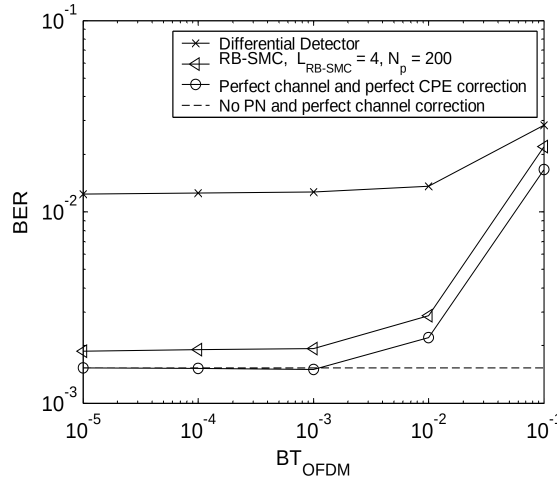
1) *Differential detector*: For differential detection in the frequency domain [15], no attempt is made to estimate the CPE or the corresponding frequency response at the particular subcarrier in question. Thus, the detected BPSK symbol at the i–th subcarrier of the m–th OFDM symbol is given by $\widehat{d}_m(i) = \text{sign}(\text{Re}\{Y_m(i)Y_m^*(i-1)\})$, or in view of (8):

$$\begin{aligned}\widehat{d}_m(i) = \text{sign}(\text{Re}\{&d_m(i)H_m(i)H_m^*(i-1)|I_m(0)|^2 + a_m(i)H_m(i)I_m(0)ICI_m^*(i-1)\\ &+ a_m(i)H_m(i)I_m(0)N_m^*(i-1) + ICI_m(i)Y_m^*(i-1) + N_m(i)Y_m^*(i-1)\}). \end{aligned} \tag{99}$$

2) *No PN and perfect channel correction bound*: For this particular receiver, the system under consideration is a perfectly synchronized OFDM system, i.e., $BT_{OFDM} = 0$. Also, the receiver is assumed to have perfect knowledge of the considered multipath fading channel. As such, the detected BPSK symbol is given by $\hat{d}_m(i) = \hat{a}_m(i)\hat{a}_m^*(i-1)$ where $\hat{a}_m(i) = \text{sign}(\text{Re}\{H_m^*(i)Y_m(i)\})$.
3) *Perfect channel and perfect CPE correction bound*: Here, we assume the receiver to have perfect knowledge of both the CPE and the frequency response. Thus, the detected BPSK symbol is given by $\hat{d}_m(i) = \hat{a}_m(i)\hat{a}_m^*(i-1)$ where $\hat{a}_m(i) = \text{sign}(\text{Re}\{I_m^*(0)H_m^*(i)Y_m(i)\})$.
4) *RB–SMC detector*: The associated algorithm is summarized by Table 1. For all the ensuing experiments, unless specified otherwise, the number of particles is set to $N_p = 200$, and the assumed channel order is set to the maximum channel order, i.e., $L_{RB-SMC} = L_{MAX}$.
5) *E–RB–SMC detector*: The E–RB–SMC dector is summarized by Table 2. Unless specified otherwise, the number of particles is set to $N_p = 200$. The maximum channel order, as mentioned before, is set to $L_{MAX} = 4$, i.e., $\mathcal{L} = \{1, ..., 4\}$.
6) *MM–RB–SMC detector*: The multiple model RB–SMC (MM–RB–SMC) detector is an alternative to the E–RB–SMC detector. The main idea in handling channel order uncertainty is to apply the SMC method to approximate the more complex posterior distribution $p(\boldsymbol{a}_{1:m}(i), \theta_{1:m}, l_{1:m}|\boldsymbol{Y}_{1:m}(i))$. In



(a) BER versus SNR.



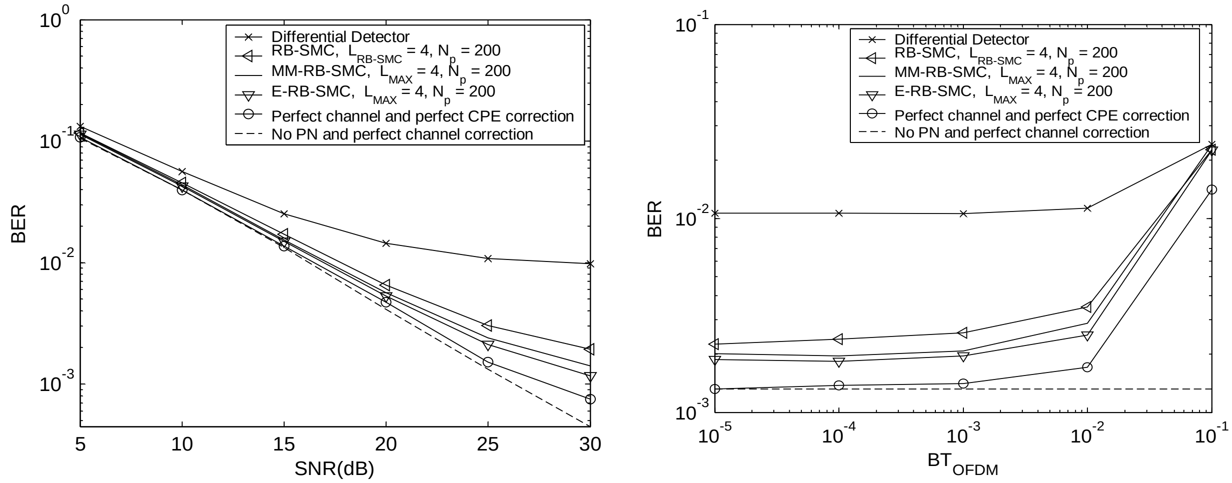
(b) BER versus BT_{OFDM} at a SNR of 25 dB.

Fig. 2: Obtained results are averaged over 6000 OFDM symbols. The true channel order was set to $L = 4$. For the RB–SMC detector, we set $N_p = 200$ and the assumed channel order to $L_{RB-SMC} = 4$.

the context of target tracking, a similar approach has been adopted in [8], where a time-varying integer index is used to denote the DSSM at each time instance of interest. Although, the MM–RB–SMC detector is not detailed in this paper, it can be derived in a manner that is similar to the derivation of the RB–SMC detector. Conceptually, the key difference between the RB–SMC detector and the MM–RB–SMC detector is the latter also samples l_m. Thus, the MM–RB–SMC detector estimates the channel order by sampling. This is unlike the E–RB–SMC detector which aims to marginalize both $\boldsymbol{x}_m$ and l_m. In the following experiments, unless specified otherwise, the number of particles is set to $N_p = 200$. The maximum channel order is set to $L_{MAX} = 4$, that is, $\mathcal{L} = \{1, ..., 4\}$.

In Fig. 2(a), the BER is averaged over 6000 OFDM symbols, the PN is set to a level corresponding to $BT_{OFDM} = 0.005$, and the considered multipath fading channel is characterized by $L_{MAX} = 4$ resolvable paths, i.e., $L = L_{MAX}$. From the figure, it is clear that the BER of the differential detector does not decrease below a certain irreducible error floor. Rightfully so, as there is a random multiplicative distortion $H_m(i)H_m^*(i-1)|I_m(0)|^2$ on symbol $d_m(i)$ from subcarrier $i-1$ to i (see (99)). Compared to the BER of the proposed RB–SMC detector, it can be seen that the RB–SMC detector significantly outperforms the scheme based on simple differential detection. More significant is that the attained BER of the RB–SMC detector is very close to that which is achieved by the receiver with perfect knowledge of the CPE and the considered multipath fading channel, i.e., the perfect channel and perfect CPE correction bound.

Figure 2(b) shows the attained BER for the differential detector, and the RB–SMC detector for different PN levels at a SNR of 25 dB. Also shown are the BER curves for the perfect channel correction bounds



(a) BER versus SNR.

(b) BER versus BT_{OFDM} at a SNR of 25 dB.

Fig. 3: Obtained results are averaged over 6000 OFDM symbols. All SMC-based detectors were implemented with $N_p = 200$ particles. The assumed channel order for the RB–SMC detector is $L_{RB-SMC} = 4$, and the maximum channel order for both the E–RB–SMC detector and the MM–RB–SMC detector is $L_{MAX} = 4$.

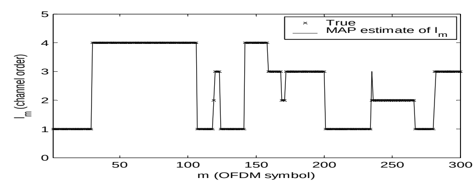
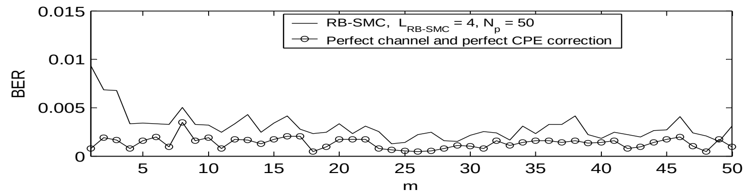


Fig. 4: True and estimated trajectories of l_m at a SNR of 15 dB. For the E–RB–SMC detector, we set $N_p = 200$ and the maximum channel order to $L_{MAX} = 4$.

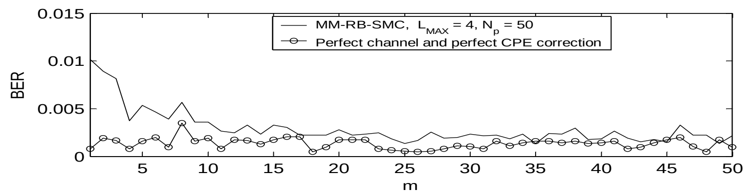
with and without the effects of PN. In all the considered scenarios, the differential detector performs poorly when compared to the BER of the RB–SMC detector. Indeed, even in the presence of severe PN, the latter achieves a BER that is close to the perfect channel and perfect CPE correction bound.

Now, we will examine the performance of the E–RB–SMC detector. To simulate the evolution of the unknown channel order l_m, we set the transition probabilities to $p_{u,u} = 0.96$ and $p_{u,v} = 0.0133$ for $u \neq v$. Figure 3(a) shows the resulting BER curves for 6000 OFDM symbols. As before, the PN is set to a level that corresponds to $BT_{OFDM} = 0.005$. Again, the RB–SMC detector assuming $L_{RB-SMC} = L_{MAX}$ significantly outperforms the detector based on differential detection. However, the RB–SMC detector invokes the assumption of a fixed channel order $L_{RB-SMC} = L_{MAX}$, and as a consequence, typically overestimates the true channel order l_m, which inevitably limits the performance of the former. Indeed, if we explicitly account for channel order uncertainty via the MM–RB–SMC detector or the E–RB–SMC detector, we obtain, respectively, about 4 or 5.5 dB gain in performance over the RB–SMC detector at high SNR. For SNR's in excess of 25 dB, the E–RB–SMC detector outperforms the MM–RB–SMC detector by about 1.5 dB.

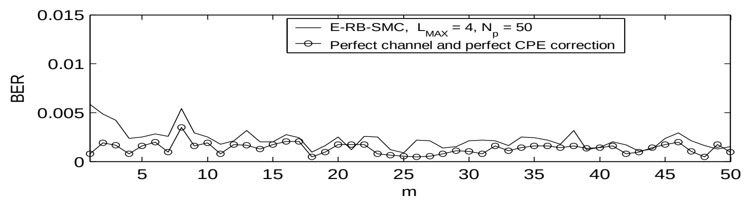
Figure 3(b) shows the attained BER for the differential detector, the RB–SMC detector, MM–RB–SMC



(a)



(b)



(c)

Fig. 5: BER for 50 OFDM symbols averaged over 400 Monte Carlo simulations at a SNR of 25 dB. All SMC-based detectors were implemented with $N_p = 50$ particles. The assumed channel order for the RB–SMC detector is $L_{RB-SMC} = 4$, and the maximum channel order for both the E–RB–SMC detector and the MM–RB–SMC detector is $L_{MAX} = 4$. As before, $p_{u,u} = 0.96$ and $p_{u,v} = 0.0133$ for $u \neq v$.

detector, and the E–RB–SMC detector for different PN levels at a SNR of 25 dB. As before, also shown are the BER curves for the perfect channel correction bounds with and without the effects of PN. Evidently, the differential detector performs poorly when compared to the performances of the SMC–based detectors. For the considered PN levels, both the MM–RB–SMC and the E–RB–SMC detector outperforms the RB–SMC detector. On the other hand, the E–RB–SMC detector also outperforms the MM–RB–SMC detector. Thus, the E–RB–SMC detector successfully addresses the blind detection problem for an OFDM system in the presence of PN, multipath fading, and channel order uncertainty, albeit, at a higher computational complexity. With regard to the estimate of the channel order, a typical and estimated trajectory of l_m at an SNR of 15 dB for 300 OFDM symbols is shown in Figure 4. Evidently, the proposed E–RB–SMC detector faithfully tracks the true trajectory of l_m.

Now, we will address an important characteristic of these algorithms, namely, the speed of convergence of the considered algorithms. For this purpose, we present figures that show the evolution of the BER as a function of time, i.e., the BER at the m-th time step corresponds to the BER of the m-th OFDM symbol. In gathering meaningful BER curves, we conducted 400 Monte Carlo simulations, each of a length that corresponds to 50 OFDM symbols. The results are summarized in Figure 5. Rather surprising is that within 8 symbols, all the BER's of the considered SMC–based detector have reached a steady state value. Of particular interest is that the proposed E–RB–SMC detector exhibits the smallest acquisition time among the considered algorithms. Indeed, within only 4 OFDM symbols, the proposed E–RB–SMC detector achieves a BER that is close to the perfect channel and perfect CPE correction bound. Of course, we can further reduce the acquisition time by increasing the number of particles, but this is at the cost of higher computational complexity. In general, if $N_p > 200$, we observed no significant improvement in performance, while for $N_p < 30$, the performance of the SMC–based detectors degraded considerably.

Finally, we can improve the performance of the proposed SMC–based detectors if we are able to compensate for the ICI, particularly at very high SNR's. However, this topic is beyond the scope of the paper and will be investigated in future work.

VIII. Conclusion

In this paper, we address the blind detection problem for an OFDM system in the presence of PN and an unknown multipath fading channel with uncertainty in the channel order. In our approach, the popular Jakes' model is approximated by an AR(2) model, and the sampled PN process is modelled as a discrete time Wiener process. With such a model, we derived a blind MAP detector for an OFDM system in the presence of PN and an unknown multipath fading channel. The proposed detector is based on the SMC method, and a technique known as Rao–Blackwellization for variance reduction in the estimates. However, the channel order is assumed to be known and time–invariant, which in practice, may be a generous assumption for a number of scenarios. Thus, with the aim of relaxing this assumption, we model channel order uncertainty via a first order Markov process, and subsequently, introduce appropriate extensions to the proposed SMC–based detector. The performance of the novel SMC–based detectors are validated through computer simulations. The resulting BER curves are close to that which is achieved by the receiver with perfect knowledge of the CPE and the considered multipath fading channel. In comparison to the RB–SMC and the MM–RB–SMC detector, the proposed E–RB–SMC detector is also shown to have the smallest acquisition time.

References

[1] A. R. S. Bahai, B. R. Saltzberg, and M. Ergen, *Multi-Carrier Digital Communications: Theory and Applications of OFDM*, 2nd ed. New York, NY: Springer Science+Business Media Inc., 2004.

[2] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Algorithms and Software for Information Extraction*. New York: Wiley, 2001.

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

[4] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, 2002.

[5] E. Costa and S. Pupolin, "M–QAM–OFDM system performance in the presence of a nonlinear amplifier and phase noise," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 462–472, Mar. 2002.

[6] L. Devroye, *Non-Uniform Random Variate Generation*. New York: Springer-Verlag, 1986.

[7] A. Doucet, S. J. Godsill, and C. Andrieu, “On sequential Monte Carlo sampling methods for Bayesian filtering,” *Statistics and Computing*, vol. 10, pp. 197–208, Dec. 2000.

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

[9] O. Edfors, M. Sandell, J. van de Beek, S. K. Wilson, and P. O. Borjesson, “OFDM channel estimation by singular value decomposition,” *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, 1998.

[10] M. S. El-Tanany, Y. Wu, and L. Hazy, “Analytical modeling and simulation of phase noise interference in OFDM–based digital television terrestrial broadcasting systems,” *IEEE Trans. Broadcast.*, vol. 47, no. 1, pp. 20–31, Mar. 2001.

[11] M. R. Gholami, S. Nader-Esfahani, and A. A. Eftekhar, “A new method of phase noise compensation in OFDM,” in *Proc. IEEE Int. Conf Commun*, May 2003, pp. 3443–3446.

[12] Y. Huang and P. M. Djurić, “A blind particle filtering detector of signals transmitted over flat fading channels,” *IEEE Trans. Signal Processing*, vol. 52, no. 7, pp. 1891–1900, 2004.

[13] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

[14] Y. Li, L. J. Cimini, and N. R. Sollengerger, “Robust channel estimation for OFDM sysems with rapid dispersive fading channels,” *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, 1998.

[15] D. Lin, P. Chiang, and H. Li, “Performance of noncoherent maximum–likelihood sequence reception for differential OFDM systems with diversity reception,” *IEEE Trans. Broadcast.*, accepted for future publication.

[16] R. Lin and A. P. Petropulu, “Linear precoding assisted blind channel estimation in OFDM systems,” *IEEE Trans. Veh. Technol.*, vol. 54, no. 3, pp. 983–995, May 2005.

[17] J. S. Liu and R. Chen, “Sequential Monte Carlo methods for dynamic systems,” *J. Am. Stat. Assoc.*, vol. 93, pp. 1032–1044, 1998.

[18] B. Lu, X. Wang, and Y. Li, “Iterative receivers for space–time block–coded OFDM systems in dispersive fading channels,” *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 213–225, 2002.

[19] J. Miguez and P. M. Djurić, “Blind equalization of frequency–selective channels by sequential importance sampling,” *IEEE Trans. Signal Processing*, vol. 52, no. 10, pp. 2738–2748, Oct. 2004.

[20] B. Muquet, M. Courville, and P. Duhamel, “Subspace–based blind and semi–blind channel estimation for OFDM systems,” *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1699–1712, 2002.

[21] R. Negi and J. Cioffi, “Pilot tone selection for channel estimation in a mobile OFDM system,” *IEEE Trans. Consumer Electron.*, vol. 44, no. 3, pp. 1122–1128, Aug. 1998.

[22] K. Nikitopoulos and A. Polydoros, “Phase–impairment effects and compensation algorithms for OFDM systems,” *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 698–707, 2005.

[23] D. Petrovic, W. Rave, and G. Fettweis, “Phase Noise suppression in OFDM using a Kalman filter,” in *Proc. IEEE Wireless Personal Multimedia Communications*, Yokosuka, Japan, Oct. 2003.

[24] L. Piazzo and P. Mandarini, “Analysis of phase noise effects in OFDM modems,” *IEEE Trans. Commun.*, vol. 50, no. 10, pp. 1696–1705, Oct. 2002.

[25] T. Pollet, M. V. Bladel, and M. Moeneclaey, “BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise,” *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 191–193, Feb./March/April 1995.

[26] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[27] E. Punskaya, C. Andrieu, A. Doucet, and W. J. Fitzgerald, “Particle filtering for demodulation in fading channels with non–Gaussian additive noise,” *IEEE Trans. Commun.*, vol. 49, no. 4, pp. 579–582, 2001.

[28] P. Robertson and S. Kaiser, “Analysis of the effects of phase–noise in orthogonal frequency division multiplex (OFDM) systems,” in *Proc. IEEE Int. Conf Commun*, Seattle, WA, June 1995, pp. 1652–1657.

[29] L. Tomba, “On the effect of Wiener phase noise in OFDM systems,” *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 580–583, May 1998.

[30] P. H. Y. Wu and A. Duel-Hallen, “Multiuser detectors with disjoint Kalman channel estimators for synchronous CDMA mobile radio channels,” *IEEE Trans. Commun.*, vol. 48, no. 5, pp. 752–756, May 2000.

[31] S. Wu and Y. Bar-Ness, “A phase noise suppression algorithm for OFDM–based WLANs,” *IEEE Commun. Lett.*, vol. 6, no. 12, pp. 535–537, Dec. 2002.

[32] ——, “OFDM systems in the presence of phase noise: consequences and solutions,” *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1988–1996, Nov. 2004.

[33] Z. Yang and X. Wang, “A sequential Monte Carlo blind receiver for OFDM systems in frequency–selective fading channels,” *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 271–280, Feb. 2002.

[34] D. Yee, J. P. Reilly, and T. Kirubarajan, “Channel equalization and phase noise suppression for OFDM systems in a time-varying frequency selective channel using particle filtering,” in *Proc. IEEE ICASSP*, 2005, pp. 777–780.

[35] Y. R. Zheng and C. Xiao, “Improved models for the generation of multiple uncorrelated rayleigh fading waveforms,” *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 256–258, 2002.

Table 1: Summary of the RB–SMC detector

Initialization: Draw $\theta_0^{(n)} \sim p(\theta_0)$. We set $\boldsymbol{x}_{0|0}^{(n)}(N-1) = \boldsymbol{\mu}_0 = \mathbf{0}_{2L_{RB-SMC}\times 1}$ and $\boldsymbol{P}_{0|0}^{(n)}(N-1) = \boldsymbol{P}_0 = \boldsymbol{I}_{2L_{RB-SMC}\times 2L_{RB-SMC}}$ for $n = 1, \ldots, N_p$.

For $m = 1$ to N_{OFDM} (total number of OFDM symbols)

For $i = 0$ to $N-1$ (N is the total number of subcarriers)

If $i = 0$, compute $\boldsymbol{P}_{m|m-1}$ and $\boldsymbol{P}_{m|m}(i)$ using (43) and (50), respectively, otherwise just compute $\boldsymbol{P}_{m|m}(i)$ using (50).

For $n = 1$ to N_p (total number of particles)

If $i = 0$

Set $\tilde{\boldsymbol{a}}_{1:m-1}^{(n)} = \boldsymbol{a}_{1:m-1}^{(n)}$ and $\tilde{\theta}_{1:m-1}^{(n)} = \theta_{1:m-1}^{(n)}$.

$\angle$ **CPE proposal**: Draw $\tilde{\theta}_m^{(n)} \sim p(\theta_m|\tilde{\theta}_{m-1}^{(n)})$.

KF time prediction: Compute $\tilde{\boldsymbol{x}}_{m|m-1}^{(n)}$ using (42). Set $\tilde{\boldsymbol{P}}_{m|m-1}^{(n)} = \boldsymbol{P}_{m|m-1}$.

Symbol proposal: Draw $\tilde{a}_m^{(n)}(0) \sim p(a_m(0)|\tilde{\boldsymbol{a}}_{1:m-1}^{(n)}, \tilde{\theta}_{1:m}^{(n)}, \boldsymbol{Y}_{1:m-1}, Y_m(0))$.

else $1 \leq i \leq N-1$

Set $\tilde{\boldsymbol{a}}_{1:m}^{(n)}(i-1) = \boldsymbol{a}_{1:m}^{(n)}(i-1)$ and $\tilde{\theta}_{1:m}^{(n)} = \theta_{1:m}^{(n)}$.

Symbol proposal: Draw $\tilde{a}_m^{(n)}(i) \sim p(a_m(i)|\tilde{\boldsymbol{a}}_{1:m}^{(n)}(i-1), \tilde{\theta}_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i))$.

end

KF measurement update: Set $\tilde{\boldsymbol{P}}_{m|m}^{(n)}(i) = \boldsymbol{P}_{m|m}(i)$ then compute $\tilde{\boldsymbol{x}}_{m|m}^{(n)}(i)$ using (49).

end

Calculate importance weights: Use (57a) or (57b) to compute $w_m^{(n)}(i)$ up to a normalizing constant.

Resampling: Multiply $\{\tilde{\boldsymbol{a}}_{1:m}^{(n)}(i)\}_{n=1}^{N_p}$ and $\{\tilde{\theta}_{1:m}^{(n)}\}_{n=1}^{N_p}$, with respect to $\{w_m^{(n)}(i)/W_m(i)\}_{n=1}^{N_p}$ to obtain $\{\boldsymbol{a}_{1:m}^{(n)}(i))\}_{n=1}^{N_p}$ and $\{\theta_{1:m}^{(n)}\}_{n=1}^{N_p}$, respectively. Set $w_m^{(n)}(i) = 1/N_p$ for all n.

Estimates: Compute $\hat{d}_m^{MAP}(i)$ using (59).

end

end

Table 2: Summary of the E–RB–SMC detector

Initialization: Draw $\theta_0^{(n)} \sim p(\theta_0)$, set $\boldsymbol{x}_{0|0}^{(n,l_0)}(N-1) = \boldsymbol{\mu}_0 = \boldsymbol{0}_{2l_0\times 1}$, and $\boldsymbol{P}_{0|0}^{(n,l_0)}(N-1) = \boldsymbol{P}_0 = \boldsymbol{I}_{2l_0\times 2l_0}$ for $n = 1,\ldots,N_p$ and $l_0 = 1,\ldots,L_{MAX}$.

For $m = 1$ to N_{OFDM} (total number of OFDM symbols)

For $i = 0$ to $N-1$ (N is the total number of subcarriers)

For $n = 1$ to N_p (total number of particles)

If $i = 0$

Set $\tilde{\boldsymbol{a}}_{1:m-1}^{(n)} = \boldsymbol{a}_{1:m-1}^{(n)}$ and $\tilde{\theta}_{1:m-1}^{(n)} = \theta_{1:m-1}^{(n)}$.

$\angle$ **CPE proposal**: Draw $\tilde{\theta}_m^{(n)} \sim p(\theta_m|\tilde{\theta}_{m-1}^{(n)})$.

For $l_m = 1$ to L_{MAX} (maximum channel order)

Predicted posterior pmf of l_m: Compute $\widehat{p}(l_m|\tilde{\boldsymbol{a}}_{1:m-1}^{(n)}, \tilde{\theta}_{1:m-1}^{(n)}, \boldsymbol{Y}_{1:m-1})$.

Compute Gaussian approximation of (79): Compute $\hat{p}(\boldsymbol{x}_{m-1}|l_m = v, \boldsymbol{a}_{1:m-1}, \theta_{1:m-1}, \boldsymbol{Y}_{1:m-1})$ according to (80) through (84).

KF time prediction: Compute $\tilde{\boldsymbol{x}}_{m|m-1}^{(n,l_m)}$ and $\tilde{\boldsymbol{P}}_{m|m-1}^{(n,l_m)}$ using (93) and (94), respectively.

end

Symbol proposal: Draw $\tilde{a}_m^{(n)}(0) \sim \widehat{p}(a_m(0)|\tilde{\boldsymbol{a}}_{1:m-1}^{(n)}, \tilde{\theta}_{1:m}^{(n)}, \boldsymbol{Y}_{1:m-1}, Y_m(0))$.

else $1 \leq i \leq N-1$

Set $\tilde{\boldsymbol{a}}_{1:m}^{(n)}(i-1) = \boldsymbol{a}_{1:m}^{(n)}(i-1)$ and $\tilde{\theta}_{1:m}^{(n)} = \theta_{1:m}^{(n)}$.

For $l_m = 1$ to L_{MAX} (maximum channel order)

Posterior pmf of l_m: Compute $\widehat{p}(l_m|\tilde{\boldsymbol{a}}_{1:m}^{(n)}(i-1), \tilde{\theta}_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i-1))$.

end

Symbol proposal: Draw $\tilde{a}_m^{(n)}(i) \sim \widehat{p}(a_m(i)|\tilde{\boldsymbol{a}}_{1:m}^{(n)}(i-1), \tilde{\theta}_{1:m}^{(n)}, \boldsymbol{Y}_{1:m}(i))$.

end

For $l_m = 1$ to L_{MAX} (maximum channel order)

KF measurement update: Compute $\tilde{\boldsymbol{x}}_{m|m}^{(n,l_m)}(i)$ and $\tilde{\boldsymbol{P}}_{m|m}^{(n,l_m)}(i)$ using (95) and (96).

end

end

Calculate importance weights: Use (72a) or (72b) to compute $w_m^{(n)}(i)$ up to a normalizing constant.

Resampling: Multiply $\{\tilde{\boldsymbol{a}}_{1:m}^{(n)}(i)\}_{n=1}^{N_p}$ and $\{\tilde{\theta}_{1:m}^{(n)}\}_{n=1}^{N_p}$, with respect to $\{w_m^{(n)}(i)/W_m(i)\}_{n=1}^{N_p}$ to obtain $\{\boldsymbol{a}_{1:m}^{(n)}(i))\}_{n=1}^{N_p}$ and $\{\theta_{1:m}^{(n)}\}_{n=1}^{N_p}$, respectively. Set $w_m^{(n)}(i) = 1/N_p$ for all n.

Estimates: Compute $\hat{d}_m^{MAP}(i)$ using (59).

end

end