Joint Approximately Sparse Channel Estimation and Data Detection in OFDM Systems using Sparse Bayesian Learning

Ranjitha Prasad*, Chandra R. Murthy† *Senior Member, IEEE*, and Bhaskar D. Rao‡ *Fellow, IEEE*

*†Dept. of ECE, Indian Institute of Science, Bangalore, India, {ranjitha.p, cmurthy}@ece.iisc.ernet.in

‡Dept. of ECE, UC San Diego, La Jolla, CA, USA, brao@ucsd.edu

***Abstract*— It is well-known that the impulse response of a wide-band wireless channel is approximately sparse, in the sense that it has a small number of significant components relative to the channel delay spread. In this paper, we consider the estimation of the unknown channel coefficients and its support in OFDM systems using a Sparse Bayesian Learning (SBL) framework for exact inference. In a quasi-static, block-fading scenario, we employ the SBL algorithm for channel estimation, and propose a Joint SBL (J-SBL) and a low-complexity recursive J-SBL algorithm for joint channel estimation and data detection. In a time-varying scenario, we use a first order auto-regressive model for the wireless channel, and propose a novel, recursive, low-complexity Kalman filtering-based SBL (K-SBL) algorithm for channel estimation. We generalize the K-SBL algorithm to obtain the recursive Joint K-SBL algorithm that performs joint channel estimation and data detection. Our algorithms can efficiently recover a group of approximately sparse vectors even when the measurement matrix is partially unknown due to the presence of unknown data symbols. Moreover, the algorithms can fully exploit the correlation structure in the multiple measurements. Monte Carlo simulations illustrate the efficacy of the proposed techniques in terms of the mean square error and bit error rate performance.**

EDICS: MLR-BAYL, MLR-SLER, SPC-CEST, SPC-MULT, SPC-DETC

I. Introduction and System Model

Orthogonal Frequency Division Multiplexing (OFDM) is a well-known multi-carrier modulation technique used in several emerging communications standards, since it provides high spectral efficiency and resilience to multi-path distortion of the wireless channel [2]. Accurate decoding of the transmit data bits requires compensating for the channel distortion, which necessitates estimation of the wireless channel at the receiver. Typically, a set of anchor sub-carriers which carry known signals (pilots) are used to estimate the channel frequency response [3], [4].

In practice, wireless channels have a large delay spread with a few significant channel tap coefficients, and therefore, the channel is *approximately sparse* (a-sparse) in the lag domain. Several papers in literature have proposed sparse channel estimation techniques (see [5]–[7] and references therein). In the context of channel estimation for OFDM systems, spectrally efficient techniques (for which $P < L$, where P is the number of pilots and L is the length of the channel) that leverage this approximate sparsity using Compressed Sensing (CS) [8] have been proposed [7], [9]–[13]. In this work, we propose to formulate the problem of channel estimation in a Sparse Bayesian Learning (SBL) framework [14], [15]. In [1], we had proposed SBL-based channel recovery for OFDM systems; we expand on this previous work in this paper. Specifically, we design novel SBL algorithms for OFDM systems in the following scenarios: (i) The block-fading case, where the channel coefficients remain fixed across the OFDM frame duration and vary in an i.i.d. fashion from frame to frame; and (ii) the time-varying case, where the channel coefficients across successive OFDM symbols are temporally correlated but have a common support.

A. Problem Formulation and Contributions

In this subsection, we cast the channel estimation problem in the SBL framework and describe the contributions of this work. In an OFDM system with N subcarriers, the instantaneous received signal, denoted by $\mathbf{y} \in \mathbb{C}^{N\times 1}$, is mathematically represented as [2]

$$\mathbf{y} = \mathbf{XFh} + \mathbf{v}. \tag{1}$$

Here, $\mathbf{F} \in \mathbb{C}^{N\times L}$ $(N > L)$ contains the first L columns of the $N \times N$ Discrete Fourier Transform (DFT) matrix, $\mathbf{h} \in \mathbb{C}^{L}$ is the channel impulse response. The dictionary matrix is given by $\mathbf{\Phi} = \mathbf{XF}$, where the diagonal matrix $\mathbf{X} \in \mathbb{C}^{N\times N}$ contains the N transmitted symbols comprising both known pilot symbols and unknown M-PSK/M-QAM modulated data along the diagonal. Each component of $\mathbf{v} \in \mathbb{C}^{N\times 1}$ is a zero mean circularly symmetric additive white Gaussian noise with pdf denoted by $\mathcal{CN}(0, \sigma^2)$, where σ^2 is the noise variance. Typically, the communication between the transmitter and the receiver occurs in frames consisting of K consecutive OFDM symbols. Suppose that, in a given OFDM symbol, P of the N subcarrier locations are pilot subcarriers and the remaining $(N - P)$ subcarriers carry unknown data symbols. The system model pertaining to the *pilot* subcarriers can be written as

$$\mathbf{y}_p = \mathbf{X}_p\mathbf{F}_p\mathbf{h} + \mathbf{v}_p, \tag{2}$$

where $\mathbf{y}_p$ is a $P\times 1$ vector containing the entries of $\mathbf{y}$ sampled at pilot locations, $\mathbf{X}_p$ is a $P \times P$ diagonal matrix with the known pilot symbols along its diagonal, $\mathbf{F}_p$ is the $P \times L$ $(P < L)$ submatrix of $\mathbf{F}$ consisting of the rows corresponding to the pilot locations and $\mathbf{v}_p$ is a $P\times 1$ vector, again consisting of components of $\mathbf{v}$ sampled at pilot locations.

This work has appeared in part in [1].

In the complex baseband representation, the scalar channel impulse response $\tilde{h}[t], t \in \mathbb{R}$ can be modeled as a stationary tapped delay line filter in the lag-domain:

$$\tilde{h}[t] = \sum_{l=1}^{\tilde{L}} \tilde{h}_l \delta[t - \tau_l], \tag{3}$$

where $\delta[t]$ is the Dirac delta function, $\tilde{h}_l$ and τ_l represent the attenuation and propagation delay between the transmitter and the receiver path l, respectively, and $\tilde{L}$ is the number of paths [16]. It is known that the wireless channel models obtained using channel sounding experiments exhibit approximate sparsity in the lag-domain (for e.g., due to non-perfect low-pass filtering using raised cosine filtering), as the communication bandwidth and sampling frequency increase [13]. Hence, based on these practical considerations, we consider the lag-domain filtered channel impulse response, which can be represented as $h[t] = g_t[t] * \tilde{h}[t] * g_r[t]$, where $g_t[t]$ and $g_r[t]$ represent the baseband transmit and receive filters and $*$ represents the convolution operation [12]. Then, the discrete-time channel can be represented as, $h(l) = h[(l-1)T]$, where T is the baud interval. The overall channel is represented as $\mathbf{h} = (h(1), h(2), \ldots, h(L))^T$. Further, in an SBL framework, we model the channel as $\mathbf{h} \sim \mathcal{CN}(0, \mathbf{\Gamma})$, where $\mathbf{\Gamma} = \text{diag}(\gamma(1), \ldots, \gamma(L))$. Note that if $\gamma(l) \to 0$, then the corresponding $h(l) \to 0$ [15], [17].

In wireless channel modeling, the Multipath Intensity Profile (MIP) is defined as the averaged multipath power profile measured at a particular location on a measurement grid [18]. The traditional methods for channel estimation in OFDM systems assume knowledge of the MIP and use pilots for channel estimation and tracking [3], or employ iterative techniques based on the Expectation Maximization (EM) algorithms for joint channel estimation/tracking and data detection [19], [20]. CS techniques have been proposed for the estimation of the time-varying channel over all the symbols in a frame when the channel consists of a few significant nonzero entries but the path delays are unknown [21]–[23]. Further, approximate inference methods have been used to solve the problem of joint channel estimation and decoding in a BICM-OFDM system, where the time-varying sparse channel is modeled using a Bernoulli-Gaussian prior [12], [24]. In [25], the authors design variational message-passing algorithms based on hierarchical Bayesian prior models for pilot-assisted OFDM channel estimation.

In this paper, we propose SBL algorithms for exact inference[1] based channel estimation, channel tracking, and data detection. In addition to the monotonicity property of SBL by virtue of the EM framework, SBL offers guarantees such as convergence to the sparsest solution when the noise variance is zero, and converging to a sparse local minimum irrespective of the noise variance [15]. In contrast, approximate inference methods [29], although lower in computational complexity, do not offer such rigorous convergence guarantees. Given the prior distributions of the noise $\mathbf{v}_p$ and the channel $\mathbf{h}$ in (2), the a-sparse channel estimation problem is given by

$$(P1) \quad \hat{\mathbf{h}} = \underset{\mathbf{h}, \boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}}{\arg\min} \frac{\|\mathbf{y}_p - \mathbf{X}_p \mathbf{F}_p \mathbf{h}\|_2^2}{\sigma^2} + \log|\mathbf{\Gamma}| + \mathbf{h}^H \mathbf{\Gamma}^{-1} \mathbf{h}, \tag{4}$$

where[2] $\mathbf{\Gamma} \triangleq \text{diag}(\gamma(1), \ldots, \gamma(L))$ and $|\cdot|$ denotes the determinant of a matrix. In the objective function above, the first term originates from the data likelihood and the other terms are from the Gaussian prior (conditioned on $\boldsymbol{\gamma}$) assumed on the wireless channel. In this paper, we specifically address the problem of OFDM channel estimation.

Note that, the above problem addresses the estimation of the wireless channel using pilot subcarriers only. However, in the OFDM scenario, several subcarriers carry unknown data as well. In this work, we also consider the problem of joint channel estimation and data detection, which can be stated as

$$(P2) \quad \hat{\mathbf{h}}, \hat{\mathbf{X}} = \underset{\mathbf{h}, \boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}, \mathbf{X} \in \mathcal{S}}{\arg\min} \frac{\|\mathbf{y} - \mathbf{X}\mathbf{F}\mathbf{h}\|_2^2}{\sigma^2} + \log|\mathbf{\Gamma}| + \mathbf{h}^H \mathbf{\Gamma}^{-1} \mathbf{h}. \tag{5}$$

where $\mathcal{S} \subset \mathbb{C}$ denotes M-QAM/M-PSK constellation from which the symbol is transmitted.

Depending on the mobility of the receiver, the channel may remain essentially constant over the frame duration, or may be slowly time-varying. If the channel is constant, the a-sparse channel estimate can be obtained from the pilot subcarriers by solving $(P1)$. When the channel is time-varying, typically, the nonzero channel coefficients vary slowly and are temporally correlated, but the hyperparameters of the channel remain constant for several OFDM frames [30]. Consequently, the locations of the significant components coincide in successive channel instantiations, i.e., the channels are approximately *group-sparse* (a-group-sparse). In this work, we cast the channel estimation problem as a a-group-sparse channel estimation problem and devise exact Bayesian inference based solutions. Approximate inference techniques for estimating the time-varying sparse vector and support have been proposed in [31]. In the context of SBL, block-based methods such as Block SBL (BSBL) and Temporal SBL (TSBL) algorithms [32] have been proposed to estimate the time-varying correlated sparse vectors when the correlation among the group-sparse vectors is modeled using a general correlation structure. In contrast to the above-mentioned works, the autoregressive (AR) state space model has been employed to model the correlation among the group sparse vectors and approximate Kalman filtering techniques have been proposed [29]. Further, CS based Kalman filtering has been proposed in the context of sparse correlated vector estimation [33].

In this work, we adopt the Kalman Filter (KF) based exact inference, where the temporal variations of the channel are captured by an AR model. Moreover, it is known that the first order AR model accurately captures the local behavior of fading wireless channels [34]. The first order AR model for

[1] In the machine learning literature (e.g., [26]–[28]), “exact inference” is an attribute associated with algorithms that obtain the exact posterior distribution of the hidden/missing variable.

[2] Due to the one-to-one correspondence between the vector $\boldsymbol{\gamma}$ and the diagonal matrix $\mathbf{\Gamma}$, we use them interchangeably.

the k^{th} channel tap is given by

$$\mathbf{h}_k = \rho \mathbf{h}_{k-1} + \mathbf{u}_k, \tag{6}$$

where the Jakes' Doppler spectrum leads to $\rho = J_0(2\pi f_d T_s) \in \mathbb{R}$ where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, f_d is the Doppler frequency, and T_s is the OFDM symbol duration [35]. The driving noise $\mathbf{u}_k$ consists of independent components $\mathbf{u}_k(i) \sim \mathcal{CN}(0, (1-\rho^2)\gamma(i))$. The initial condition for the a-sparse channel is given by $\mathbf{h}_1 \sim \mathcal{CN}(0, \boldsymbol{\Gamma})$.

When the hyperparameters are known, a KF approach has been used for channel tracking using the pilot symbols [36]. The EM based KF has also been proposed for joint channel tracking and data detection in OFDM systems [35], [37]. However, these algorithms are not applicable in scenarios where the hyperparameters are unknown and need to be estimated along with the channel tap coefficients and the data symbols. In contrast, we use the exact inference techniques employed for linear dynamical systems [26], [38] to exploit the known correlation structure of the channel. We note that by using an AR state space model, it is possible to significantly reduce the computational complexity compared to the block-based a-sparse estimation techniques such as the ARSBL [39].

Since the unknown channels have a common hyperparameter vector, the joint pdf of the K received OFDM signals and the a-group-sparse temporally correlated channels is given by

$$p(\mathbf{Y}_{p,K}, \mathbf{h}_1, \ldots, \mathbf{h}_K; \boldsymbol{\gamma}) = \prod_{m=1}^{K} p(\mathbf{y}_{p,m}|\mathbf{h}_m) p(\mathbf{h}_m|\mathbf{h}_{m-1}; \boldsymbol{\gamma}), \tag{7}$$

where $\mathbf{Y}_{p,K} = [\mathbf{y}_{p,1}, \ldots, \mathbf{y}_{p,K}]$, and, by convention, we use $p(\mathbf{h}_1|\mathbf{h}_0; \boldsymbol{\gamma}) \triangleq p(\mathbf{h}_1; \boldsymbol{\gamma})$ where $\mathbf{h}_1 \sim \mathcal{CN}(0, \boldsymbol{\Gamma})$. To obtain the optimization problem, we consider $-\log p(\mathbf{Y}_{p,K}, \mathbf{h}_1, \ldots, \mathbf{h}_K; \boldsymbol{\gamma})$ and neglect the terms that are constant w.r.t. $\mathbf{h}$ and $\boldsymbol{\gamma}$, to obtain

$$\begin{aligned} f(\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma}) &= \sum_{m=1}^{K} \frac{\|\mathbf{y}_{p,m} - \mathbf{X}_{p,m}\mathbf{F}_{p,m}\mathbf{h}_m\|_2^2}{\sigma^2} \\ +K \log|\boldsymbol{\Gamma}| &+ \sum_{m=2}^{K} \frac{(\mathbf{h}_m - \rho\mathbf{h}_{m-1})^H \boldsymbol{\Gamma}^{-1} (\mathbf{h}_m - \rho\mathbf{h}_{m-1})}{(1-\rho^2)} \\ &+ \mathbf{h}_1^H \boldsymbol{\Gamma}^{-1} \mathbf{h}_1. \end{aligned} \tag{8}$$

From the equation above, the pilot-based channel estimation problem for K OFDM symbols can be written as

$$(P3) \quad \hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K = \underset{\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}}{\arg\min} f(\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma}). \tag{9}$$

Problem $(P3)$ addresses the estimation of *time-varying* wireless channels using only the pilot subcarriers. However, as mentioned earlier, several subcarriers in each of the OFDM symbols carry unknown data. Hence, we can also consider the problem of joint time-varying channel estimation and data detection, by modifying (9) as follows:

$$\begin{aligned} (P4) \quad & \hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K, \hat{\mathbf{X}}_1, \ldots, \hat{\mathbf{X}}_K = \\ & \underset{\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}, \mathbf{X}_1, \ldots, \mathbf{X}_K \in \mathcal{S}}{\arg\min} g(\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma}, \mathbf{X}_1, \ldots, \mathbf{X}_K) \end{aligned} \tag{10}$$

where

$$\begin{aligned} g(\mathbf{h}_1, \ldots, \mathbf{h}_K, \boldsymbol{\gamma}, \mathbf{X}_1, \ldots, \mathbf{X}_K) &= \sum_{m=1}^{K} \frac{\|\mathbf{y}_m - \mathbf{X}_m\mathbf{F}\mathbf{h}_m\|_2^2}{\sigma^2} \\ + \sum_{m=2}^{K} \frac{(\mathbf{h}_m - \rho\mathbf{h}_{m-1})^H \boldsymbol{\Gamma}^{-1} (\mathbf{h}_m - \rho\mathbf{h}_{m-1})}{(1-\rho^2)} &+ K\log|\boldsymbol{\Gamma}| \\ + \mathbf{h}_1^H \boldsymbol{\Gamma}^{-1} \mathbf{h}_1. \end{aligned} \tag{11}$$

Contributions: In this work, we propose a practical and principled approach for joint a-group-sparse channel estimation and data detection in SISO-OFDM systems, that evaluates well in Monte-Carlo simulations. First, we show that the problem in $(P1)$ can be solved using the SBL framework of [15]. We next generalize the SBL framework to obtain the J-SBL algorithm as a solution to $(P2)$. A key feature of the J-SBL algorithm is that the observations from both the data and the pilot subcarriers are incorporated to jointly estimate the a-sparse channel as well as the unknown data. We also propose a low complexity, recursive J-SBL (RJ-SBL) algorithm to solve $(P2)$. We show that the joint estimation procedure leads to a significant improvement in the Mean Square Error (MSE) of the channel estimate at SNRs of practical interest. Further, we propose a novel, low-complexity K-SBL algorithm as a recursive solution to $(P3)$. We enhance the K-SBL algorithm to obtain the JK-SBL algorithm, which is a recursive solution to $(P4)$. The results are summarized in the Table I.

Although our work focuses on a-sparse channel estimation for OFDM systems using the SBL framework, the algorithms we develop are important in their own right due to several reasons. This is the first paper in the literature that proposes recursive techniques for exact inference in sparse signal recovery. We show that the joint problems of hyperparameter estimation and data detection separate out in the M-step. This leads to a simple maximization procedure in the M-step, with no loss of optimality. The joint algorithms involve estimation of the unknown data symbols, which necessitates the development of techniques that are capable of handling partially unknown dictionary matrices.[3] Finally, the recursive versions of the algorithms have the advantage of computational simplicity compared to other exact inference methods, while retaining the performance advantages of SBL estimators.

The rest of this paper is organized as follows. In Sec. II, we propose algorithms for a-sparse channel estimation using pilots. In Sec. III, the joint channel estimation and data detection algorithms are proposed and the implementation issues are discussed. The efficacy of the proposed techniques is demonstrated through simulation results in Sec. IV. We offer some concluding remarks in Sec. V.

Notation: Boldface small letters denote vectors and boldface capital letters denote matrices. The symbols $(\cdot)^T$ and $|\cdot|$ denote the transpose and determinant of a matrix, respectively. Also, diag$(\mathbf{a})$ denotes a diagonal matrix with entries on the diagonal given by $\mathbf{a}$. The pdf of the random variable X is represented as $p(x)$ and the random variables and deterministic parameters in the pdf are separated using a semicolon. The

[3]That is, the algorithms are capable of handling the fact that, due to the $N-P$ unknown data symbols in $\mathbf{X}$, the measurement matrix $\boldsymbol{\Phi} = \mathbf{X}\mathbf{F}$ is partially unknown.

TABLE I
THE MAIN CONTRIBUTIONS OF THIS PAPER.

Sl. no.	Novel algorithms proposed and Section number	Goal	Applicability
1	SBL in Sec. II	Joint channel, hyperparameter estimation	Block-fading channels ($P1$)
2	J-SBL in Sec. III and Recursive J-SBL	Joint channel, hyperparameter estimation and data detection	Block-fading channels ($P2$)
3	K-SBL in Sec. II	Recursive joint channel and hyperparameter estimation	Time-varying and Block-fading channels ($P3$)
4	JK-SBL in Sec. III	Recursive joint channel, hyperparameter estimation and data detection	Time-varying and Block-fading channels ($P4$)

expectation with respect to a random variable X is denoted as $\mathbb{E}_X(\cdot)$. The $L \times L$ identity matrix is represented as $\mathbf{I}_L$ and $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. The i^{th} entry of a vector $\mathbf{a}$ and the $(i,j)^{\text{th}}$ entry of a matrix $\mathbf{A}$ are represented as $a(i)$ and $A(i,j)$, respectively. Throughout the paper, p as a subscript refers to pilots and (r) in the superscript refers to the iteration number.

II. CHANNEL ESTIMATION AND TRACKING USING PILOT SUBCARRIERS

In this section, we propose SBL algorithms for a-group-sparse channel estimation in OFDM systems using pilot symbols, for both block-fading and time-varying channels. First, we discuss the SBL algorithm to solve $(P1)$, i.e., the problem of a-sparse channel estimation using P_b pilots in the entire OFDM frame when the channel is block-fading. Subsequently, we consider the time-varying channel using P_t pilots in every symbol, and propose a novel, recursive approach for a-group-sparse channel estimation, i.e., a solution to $(P3)$.

A. The SBL Algorithm: Block-Fading Case

Here, we propose the SBL algorithm for channel estimation using pilot subcarriers in a single OFDM symbol; this forms the basis for the algorithms developed in the sequel. The observation model is given by (2). SBL uses a parametrized prior to obtain sparse solutions, given by

$$p(\mathbf{h};\boldsymbol{\gamma}) = \prod_{i=1}^{L} (\pi\gamma(i))^{-1} \exp\left(-\frac{|h(i)|^2}{\gamma(i)}\right). \qquad (12)$$

Typically, the hyperparameters $\boldsymbol{\gamma}$ can be estimated using the type-II ML procedure [14], i.e., by maximizing the marginalized pdf $p(\mathbf{y}_p;\boldsymbol{\gamma})$ as

$$\hat{\boldsymbol{\gamma}}_{ML} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}}{\arg\max}\, p(\mathbf{y}_p;\boldsymbol{\gamma}). \qquad (13)$$

Since the above problem cannot be solved in closed form, iterative estimators such as the EM based SBL algorithm [15] have to be employed. The sparse channel $\mathbf{h}$ is considered as the hidden variable and the ML estimate of $\boldsymbol{\gamma}$ is obtained in the M-step. The steps of the algorithm can be given as

$$\text{E-step}: Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right) = \mathbb{E}_{\mathbf{h}|\mathbf{y}_p;\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y}_p, \mathbf{h}; \boldsymbol{\gamma})] \qquad (14)$$

$$\text{M-step}: \boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}}{\arg\max}\, Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right). \qquad (15)$$

The E-step above requires the posterior density of the sparse vector with the hyperparameter $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(r)}$, which can be expressed as

$$p\left(\mathbf{h}|\mathbf{y}_p;\boldsymbol{\gamma}^{(r)}\right) = \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad (16)$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}^{(r)} - \boldsymbol{\Gamma}^{(r)}\boldsymbol{\Phi}_p^H\left(\sigma^2\mathbf{I}_{P_b} + \boldsymbol{\Phi}_p\boldsymbol{\Gamma}^{(r)}\boldsymbol{\Phi}_p^H\right)^{-1}\boldsymbol{\Phi}_p\boldsymbol{\Gamma}^{(r)}$, and $\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}_p^H\mathbf{y}_p$, where $\boldsymbol{\Phi}_p = \mathbf{X}_p\mathbf{F}_p$. Notice that the EM algorithm given by the steps in (14), (15) also solves $(P1)$, where we obtain a MAP estimate of the a-sparse channel, i.e., $\hat{\mathbf{h}} = \boldsymbol{\mu}$ with $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma}^{(r)})$. The M-step in (15) can be simplified, to obtain

$$\gamma^{(r+1)}(i) = \underset{\gamma(i)\in\mathbb{R}_+}{\arg\max}\, \mathbb{E}_{\mathbf{h}|\mathbf{y}_p;\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{h};\boldsymbol{\gamma})] \qquad (17)$$

$$= \mathbb{E}_{\mathbf{h}|\mathbf{y}_p;\boldsymbol{\gamma}^{(r)}}\left[|h(i)|^2\right] = \Sigma(i,i) + |\mu(i)|^2. \qquad (18)$$

In (17), the term $\mathbb{E}_{\mathbf{h}|\mathbf{y}_p;\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y}_p|\mathbf{h};\boldsymbol{\gamma})]$ has been dropped, as it is not a function of $\gamma(i)$. Note that, since all the algorithms proposed in this paper use the EM updates, they have monotonicity property, i.e., the likelihood is guaranteed to increase at each iteration [40], [41].[4]

In the case of multiple OFDM symbols in a block-fading channel, since the channel remains constant for the K OFDM symbols, the system model in (2) is modified as

$$\mathbf{y}_{p,m} = \mathbf{X}_{p,m}\mathbf{F}_{p,m}\mathbf{h} + \mathbf{v}_{p,m}, \quad m = 1, \ldots, K. \qquad (19)$$

The equation above has a one-to-one correspondence with (2), since $\mathbf{y}_{p,m}$ denotes the observations corresponding to pilot subcarriers in the m^{th} OFDM symbol and $\boldsymbol{\Phi}_{p,m} = \mathbf{X}_{p,m}\mathbf{F}_{p,m}$ denotes the matrix consisting of measurements corresponding to pilot subcarriers in the m^{th} OFDM symbol. The steps of the SBL algorithm for block-fading channel estimation are given in the Appendix A.

We note that the SBL algorithm proposed in this section is not equipped to use the correlations between the channel across successive OFDM symbols in a time-varying channel. A straightforward approach to exploit the correlation is to use a block-based method, where the estimates of all the K channel vectors are obtained jointly using the observations for the K OFDM symbols [32], [39]. However, this joint processing of all K OFDM symbols is computationally expensive, as it requires inverting matrices of the size $KP_t \times KP_t$. In the next subsection, we propose a recursive approach that is not only

[4] We have found, empirically, that the straightforward initialization such as $\boldsymbol{\Gamma}^{(0)} = \mathbf{I}_L$ leads to accurate solutions.

low-complexity compared to the block-based techniques, but also exploits the temporal channel correlation across symbols, resulting in an enhanced channel tracking performance.

B. The K-SBL Algorithm: Time-Varying Case

In this subsection, we derive algorithms for tracking the *slowly time-varying* channel using an SBL framework to learn the hyperparameters along with the channel coefficients, i.e., we solve $(P3)$. We derive recursive techniques based on the Kalman Filter and Smoother (KFS), with an AR model for the temporal evolution of the channel. Interestingly, the framework developed in this section can also be used to accommodate detection of the unknown data (i.e., a solution to $(P4)$), as we show in the next section.

In the time-varying case, the measurement equation given by the OFDM system model, and the state equation given by the first order AR channel model, for K consecutive symbols, are as follows:

$$\mathbf{y}_{p,m} = \mathbf{\Phi}_{p,m}\mathbf{h}_m + \mathbf{v}_{p,m}, \tag{20}$$

$$\mathbf{h}_{m+1} = \rho\mathbf{h}_m + \mathbf{u}_{m+1}, \quad m = 1, 2, \ldots, K, \tag{21}$$

where $\mathbf{\Phi}_{p,m} = \mathbf{X}_{p,m}\mathbf{F}_{p,m}$. Typically, in a KF approach to $(P3)$, the goal is to recursively estimate the channel state and its covariance matrix using forward and backward recursions, given the observations $\mathbf{y}_{p,1}, \ldots, \mathbf{y}_{p,K}$ sampled at the P_t pilot subcarriers. In the forward recursion, for each OFDM symbol, the KF operates on the received symbol to obtain the estimates of the a-sparse channel as a weighted average of the previous estimate and the current received symbol. These weights are given by the the Kalman gain matrix, and are updated for each OFDM symbol. In the backward recursion, the Kalman *smoother* ensures that the observations until the K^{th} OFDM symbol are included in the estimation of the a-sparse channel corresponding to the m^{th} symbol for $1 \leq m < K$. Hence, it improves the accuracy of the estimates of the previous channel states in every recursion.

For the moment, if we assume that $\mathbf{\Gamma}$ is known, and if we denote the posterior mean and the covariance matrix of channel in the m^{th} OFDM symbol by $\hat{\mathbf{h}}_{m|m}$ and $\mathbf{P}_{m|m}$, respectively, for $1 \leq m \leq K$, then the KFS update equations are as follows [38], [42]:

for $m = 1, \ldots, K$ **do**

Prediction: $\hat{\mathbf{h}}_{m|m-1} = \rho\hat{\mathbf{h}}_{m-1|m-1}$ (22)

$$\mathbf{P}_{m|m-1} = \rho^2\mathbf{P}_{m-1|m-1} + (1-\rho^2)\mathbf{\Gamma} \tag{23}$$

Filtering:

$$\mathbf{G}_m = \mathbf{P}_{m|m-1}\mathbf{\Phi}_{p,m}^H\left(\sigma^2\mathbf{I}_{P_t} + \mathbf{\Phi}_{p,m}\mathbf{P}_{m|m-1}\mathbf{\Phi}_{p,m}^H\right)^{-1} \tag{24}$$

$$\hat{\mathbf{h}}_{m|m} = \hat{\mathbf{h}}_{m|m-1} + \mathbf{G}_m(\mathbf{y}_{p,m} - \mathbf{\Phi}_{p,m}\hat{\mathbf{h}}_{m|m-1}) \tag{25}$$

$$\mathbf{P}_{m|m} = (\mathbf{I}_L - \mathbf{G}_m\mathbf{\Phi}_{p,m})\mathbf{P}_{m|m-1} \tag{26}$$

end

for $j = K, K-1, \ldots, 2$ **do**

Smoothing: $\hat{\mathbf{h}}_{j-1|K} = \hat{\mathbf{h}}_{j-1|j-1} + \mathbf{J}_{j-1}(\hat{\mathbf{h}}_{j|K} - \hat{\mathbf{h}}_{j|j-1})$ (27)

$$\mathbf{P}_{j-1|K} = \mathbf{P}_{j-1|j-1} + \mathbf{J}_{j-1}(\mathbf{P}_{j|K} - \mathbf{P}_{j|j-1})\mathbf{J}_{j-1}^H, \tag{28}$$

end

where $\mathbf{J}_{j-1} \triangleq \rho\mathbf{P}_{j-1|j-1}\mathbf{P}_{j|j-1}^{-1}$ and $\mathbf{G}_m$ is the Kalman gain. In the above, the symbols $\hat{\mathbf{h}}_{m|m-1}$, $\mathbf{P}_{m|m-1}$, etc. have their usual meanings as in the KF literature [38]. For example, $\hat{\mathbf{h}}_{m|m-1}$ is the channel estimate at the m^{th} OFDM symbol given the observations $\mathbf{Y}_{p,m-1} = [\mathbf{y}_{p,1}, \ldots, \mathbf{y}_{p,m-1}]$ and $\mathbf{P}_{m|m-1}$ is the covariance of the m^{th} channel estimate given $\mathbf{Y}_{p,m-1}$. The above KFS equations are initialized by setting $\hat{\mathbf{h}}_{0|0} = \mathbf{0}$ and $\mathbf{P}_{0|0} = \mathbf{\Gamma}$. They track the channel in the forward direction using the prediction and the filtering equations in (22)-(26) and *smooth* the obtained channel estimates using the backward recursions in (27)-(28). However, in the a-sparse channel tracking problem, $\mathbf{\Gamma}$ is unknown. Hence, we propose the K-SBL algorithm, which simultaneously estimates the channel coefficients and also learns the unknown $\mathbf{\Gamma}$.

Recall that the a-group-sparse channel has a common hyperparameter set. The joint pdf of the received signals and the a-group-sparse channel for K OFDM symbols is given by (7), which leads to the optimization problem as given by $(P3)$. We propose the K-SBL algorithm using the EM updates, as follows:

$$\text{E-step}: Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right) = \mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{Y}_{p,K}, \mathbf{h}_1, \ldots, \mathbf{h}_K; \boldsymbol{\gamma})] \tag{29}$$

$$\text{M-step}: \boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}}{\arg\max}\, Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right). \tag{30}$$

To compute the E-step given above, we require the posterior distribution of the unknown a-sparse channel, which is obtained using the recursive update equations given by (22)-(28). In order to obtain an ML estimate of $\boldsymbol{\gamma}$, K-SBL incorporates an M-step, which, in turn, utilizes the mean and covariance of the posterior distribution from the E-step. From (7), the M-step results in the following optimization problem:

$$\boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}}{\arg\max}\, \mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}\Big[c - K\log|\mathbf{\Gamma}| - \sum_{j=2}^{K}\frac{(\mathbf{h}_j - \rho\mathbf{h}_{j-1})^H\mathbf{\Gamma}^{-1}(\mathbf{h}_j - \rho\mathbf{h}_{j-1})}{(1-\rho^2)} - \mathbf{h}_1^H\mathbf{\Gamma}^{-1}\mathbf{h}_1\Big], \tag{31}$$

where c is a constant independent of $\boldsymbol{\gamma}$. As mentioned earlier, we see that the M-step requires the computation of $\hat{\mathbf{h}}_{j|K} \triangleq \mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}[\mathbf{h}_j]$, and covariance $\mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}[\mathbf{h}_j\mathbf{h}_j^H] \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j|K}^H$ for $j = 1, \ldots, K$, which is obtained from (22)-(28). The M-step also requires the computation of $\mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\boldsymbol{\gamma}^{(r)}}[\mathbf{h}_j\mathbf{h}_{j-1}^H] \triangleq \mathbf{P}_{j,j-1|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j-1|K}^H$ for $j = K, K-1, \ldots, 2$, which we obtain from [38] as follows:

$$\mathbf{P}_{j-1,j-2|K} = \mathbf{P}_{j-1|j-1}\mathbf{J}_{j-2}^H + \mathbf{J}_{j-1}^H(\mathbf{P}_{j,j-1|K} - \rho\mathbf{P}_{j-1|j-1})\mathbf{J}_{j-2}. \tag{32}$$

The above recursion is initialized using $\mathbf{P}_{K,K-1|K} = \rho(\mathbf{I}_L - \mathbf{G}_K\mathbf{\Phi}_{p,K})\mathbf{P}_{K-1|K-1}$. Using the above expressions, (31) simplifies as

$$\gamma^{(r+1)} = \underset{\gamma \in \mathbb{R}_+^{L\times 1}}{\arg\max} \left\{ c' - K \log |\Gamma| - \text{Trace}(\Gamma^{-1}\mathbf{M}_{1|K}) - \frac{1}{(1-\rho^2)} \sum_{j=2}^{K} \text{Trace}(\Gamma^{-1}\mathbf{M}_{j|K}) \right\}, \quad (33)$$

where c' is a constant independent of γ, $\mathbf{M}_{j|K} \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j|K}^H + \rho^2(\mathbf{P}_{j-1|K} + \hat{\mathbf{h}}_{j-1|K}\hat{\mathbf{h}}_{j-1|K}^H) - 2\rho\text{Re}(\mathbf{P}_{j,j-1|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j-1|K}^H)$ and $\mathbf{M}_{1|K} \triangleq \mathbf{P}_{1|K} + \hat{\mathbf{h}}_{1|K}\hat{\mathbf{h}}_{1|K}^H$. Differentiating (33) w.r.t. $\gamma(i)$ and setting the resulting equation to zero gives the update for the i^{th} hyperparameter as follows:

$$\gamma^{(r+1)}(i) = \frac{1}{K}\left(\sum_{j=2}^{K} \frac{M_{j|K}(i,i)}{(1-\rho^2)} + M_{1|K}(i,i) \right), \quad (34)$$

for $i = 1, \ldots, L$. Thus the K-SBL algorithm learns γ in the M-step and provides low-complexity and recursive estimates of the a-sparse channel in the E-step. This completes the EM based solution to $(P3)$. The algorithmic representation of the K-SBL algorithm is given in Appendix B.

Remarks: When $\rho = 1$, the AR model simplifies to $\mathbf{h} = \mathbf{h}_1 = \ldots = \mathbf{h}_K$, and hence, it reduces to the block-fading channel scenario. The recursive updates in the E-step are given by the KFS equations (22)-(28), and the M-step is given by

$$Q(\gamma|\gamma^{(r+1)}) = \mathbb{E}_{\mathbf{h}|\mathbf{Y}_{p,K};\gamma^{(r)}}[c' - (\mathbf{h}^H\Gamma^{-1}\mathbf{h} + \log|\Gamma|)], \quad (35)$$

which results in the same M-step as that of the SBL algorithm in the block-fading case. Hence, this algorithm provides a low-complexity recursive solution to the SBL problem in the block-fading scenario, which we discuss in detail in Sec. III-B. At the other extreme, when $\rho = 0$, the AR model simplifies to $\mathbf{h}_m = \mathbf{u}_m$ for $m = 1, \ldots, K$, i.e., the channels for OFDM symbols are mutually independent of each other. In this case, the prediction equations of the KFS equations simplify as $\hat{\mathbf{h}}_{m|m-1} = \mathbf{0}$ and $\mathbf{P}_{m|m-1} = \Gamma$, and the expressions for $\hat{\mathbf{h}}_{m|m}$ and $\mathbf{P}_{m|m}$ simplify to the mean and covariance matrix, as obtained in the SBL algorithm for a single OFDM symbol. The smoothing equations simplify to $\hat{\mathbf{h}}_{m-1|m} = \hat{\mathbf{h}}_{m-1|m-1}$ and $\mathbf{P}_{m-1|m} = \mathbf{P}_{m-1|m-1}$, i.e., the smoothed mean and covariance at the $(m-1)^{\text{th}}$ symbol depend only on observations of the $(m-1)^{\text{th}}$ OFDM symbol, as expected.

Although the algorithms proposed in this section are easy to implement and computationally simple due to their recursive nature, they do not utilize all the information available from the observation vectors $\mathbf{y}_1, \ldots, \mathbf{y}_K$. Only the pilot subcarriers are used for channel estimation. Hence, in the next section, we extend the SBL framework developed in this section to detect the unknown data. We show how these decisions can be coalesced into the EM iterations, leading to joint channel estimation and data detection.

III. Joint Channel Estimation and Data Detection Using Pilot and Data Subcarriers

In this section, we start by deriving the J-SBL and the RJ-SBL algorithm for joint estimation of the unknown a-sparse channel and transmit data in a block-fading OFDM system. Subsequently, we consider the time-varying channel, and generalize the K-SBL to obtain the JK-SBL for jointly estimating the unknown data and tracking the a-group-sparse channel. Our proposed algorithms solve the problems $(P2)$ and $(P4)$ using an SBL framework.

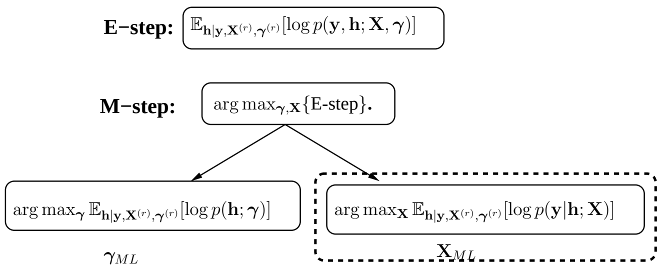

Fig. 1. The J-SBL algorithm: the E-step computes the expectation over the posterior density of $\mathbf{h}$. The joint maximization in the M-step simplifies into two independent maximizations over $\gamma(i)$ and $\mathbf{X}$. The step inside the dashed box indicates the new ingredient in the J-SBL algorithm.

A. The J-SBL Algorithm: Block-Fading Case

To derive the algorithm for an OFDM frame consisting of K OFDM symbols, we consider $\mathbf{h}$ as a hidden variable and $[\gamma, \mathbf{X}_1, \ldots, \mathbf{X}_K]$ as the parameters to be estimated. The E and the M-steps of the J-SBL algorithm can be given as

$$\text{E-step}: \; Q(\mathbf{X},\gamma|\mathbf{X}^{(r)},\gamma^{(r)}) = \mathbb{E}_{\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\gamma^{(r)}}[\log p(\mathbf{y},\mathbf{h};\mathbf{X},\gamma)] \quad (36)$$

$$\text{M-step}: \; \left(\mathbf{X}^{(r+1)},\gamma^{(r+1)}\right) = \underset{\mathbf{X},\gamma\in\mathbb{R}_+^{L\times 1}}{\arg\max}\, Q\left(\mathbf{X},\gamma|\mathbf{X}^{(r)},\gamma^{(r)}\right), \quad (37)$$

where $\mathbf{X} \in \mathbb{C}^{NK\times NK}$ is a block diagonal matrix consisting of the matrices $\mathbf{X}_1, \ldots, \mathbf{X}_K$ whose diagonal entries consist of symbols from the transmit constellation, and $\mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_K^T]^T$. The posterior density computed in the E-step is $p\left(\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\gamma^{(r)}\right) = \mathcal{CN}(\mu,\Sigma)$, where

$$\begin{aligned} \mu &= \sigma^{-2}\Sigma\mathbf{F}_b^H{\mathbf{X}^{(r)}}^H\mathbf{y} \\ \Sigma &= \left(\sigma^{-2}\mathbf{F}_b^H{\mathbf{X}^{(r)}}^H\mathbf{X}^{(r)}\mathbf{F}_b + {\Gamma^{(r)}}^{-1}\right)^{-1}, \end{aligned} \quad (38)$$

where $\mathbf{F}_b \in \mathbb{C}^{NK\times L}$ with $\mathbf{F}_b = \mathbf{1}_K \otimes \mathbf{F}$, where $\mathbf{1}_K$ is a vector of all ones. Notice that (38) and (16) are different since the former uses the known pilot symbols, $\mathbf{X}_p \in \mathbb{C}^{P\times P}$, whereas the latter uses the pilot symbols along with the estimated transmit data, together given by $\mathbf{X}^{(r)}$ in the r^{th} iteration. The proposed algorithm is pictorially depicted in Fig. 1.

The objective function in the M-step given in (37) can be written as

$$Q\left(\mathbf{X},\gamma|\mathbf{X}^{(r)},\gamma^{(r)}\right) = c'' - \mathbb{E}_{\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\gamma^{(r)}}\left[\frac{\|\mathbf{y}-\mathbf{X}\mathbf{F}_b\mathbf{h}\|_2^2}{\sigma^2} + \log|\Gamma| + \mathbf{h}^H\Gamma^{-1}\mathbf{h}\right] \quad (39)$$

where c'' is a constant independent of γ and $\mathbf{X}$. The objective function given above is the sum of two independent functions, $\mathbb{E}_{\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\gamma^{(r)}}[\log p(\mathbf{y}|\mathbf{h};\mathbf{X})]$ and $\mathbb{E}_{\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\gamma^{(r)}}[\log p(\mathbf{h};\gamma)]$. The key aspect of the M-step below is that the function

$\mathbb{E}_{\mathbf{h}|\mathbf{y};\mathbf{X}^{(r)},\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{y}|\mathbf{h};\mathbf{X})]$ is maximized over $\mathbf{X}$, which incorporates the information discarded in the M-step of the SBL algorithm presented in Sec. II-A. Now, since the first term does not depend on $\boldsymbol{\gamma}$, we optimize the second function with respect to $\gamma(i)$ to obtain $\gamma^{(r+1)}(i)$ as in the SBL algorithm, given by (18). On the other hand, the first function can be optimized by solving the following problem:

$$X^{(r+1)}(i,i) = \underset{x_i \in \mathcal{S}}{\arg\min} \left\{ C_b(i,i)|x_i|^2 + |y(i) - x_i \mathbf{F}_b(i,:)\boldsymbol{\mu}|^2 \right\} \tag{40}$$

where $i \in \mathcal{D}$, $\mathcal{D}$ is an index set consisting of the data subcarrier locations, $\mathbf{C}_b = \mathbf{F}_b \boldsymbol{\Sigma} \mathbf{F}_b^H$, $\mathbf{F}_b(i,:)$ is the i^{th} row of the $\mathbf{F}_b$ matrix, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given in (38) and $\mathcal{S}$ is the constellation from which the symbol is transmitted. Due to the above maximization, $Q(\mathbf{X}, \boldsymbol{\gamma}|\mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)})$ increases monotonically for $1 \leq r \leq r_{\text{max}}$, where r_{max} is the maximum number of iterations. That is,

$$\begin{aligned} &Q(\mathbf{X}^{(r+1)}, \boldsymbol{\gamma}^{(r+1)}|\mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}) \geq \\ &Q(\mathbf{X}^{(r)}, \boldsymbol{\gamma}^{(r)}|\mathbf{X}^{(r-1)}, \boldsymbol{\gamma}^{(r-1)}), \quad \text{for} \quad 1 \leq r \leq r_{\text{max}}. \end{aligned} \tag{41}$$

Note that the above function $Q(.)$ monotonically (in $\mathbf{X}$ and $\boldsymbol{\gamma}$) approaches the likelihood function, which in turn is bounded. This guarantees the convergence of the proposed J-SBL algorithm. Further, by the same reasoning, the convergence guarantee holds good for the JK-SBL algorithm which will be presented in the sequel.

The J-SBL requires initial estimates of the unknown parameters $\boldsymbol{\gamma}$ and $\mathbf{X}$. The initial estimate of $\boldsymbol{\Gamma}$ is taken to be the identity matrix, as in the previous section. The initialization of the $(KN - P_b)$ non-pilot data in turn requires an initial channel estimate. Channel estimates using methods like LS and MMSE cannot be used here, as they require knowledge of the support and the hyperparameters, respectively. Hence, the initialization of $\mathbf{X}$ can be obtained from the channel estimate obtained from a few iterations of the SBL algorithm from the $P_b = P$ pilots (denoted as $\hat{\mathbf{h}}_{SBL}$). The ML data detection problem is given by

$$X^{(0)}(i,i) = \underset{x_i \in \mathcal{S}}{\arg\min} |y(i) - x_i \mathbf{F}_b(i,:)\hat{\mathbf{h}}_{SBL}|^2, \quad i \in \mathcal{D}. \tag{42}$$

The algorithmic representation of the J-SBL algorithm is provided in Appendix C. J-SBL algorithm is a block-based algorithm, and hence, the complexity of the algorithm is dominated by the E-step, which incurs a complexity of $\mathcal{O}(N^2 L K^3)$ [43]. In the next subsection, we derive a low-complexity, recursive version of the J-SBL algorithm, using the K-SBL algorithm with $\rho = 1$.

B. Recursive J-SBL Algorithm: Block-fading Case

In this subsection, we derive the recursive joint SBL algorithm which is mathematically equivalent to the J-SBL algorithm proposed in Sec. III-A, using the framework of the K-SBL algorithm with $\rho = 1$, i.e., for the block-fading channel. Hence, we solve the problem $(P3)$, using a low-complexity RJ-SBL algorithm.

Consider the state space model in (20) and (21) in the block-fading case, where the channel remains constant for K OFDM symbols, with $\mathbf{h} = \mathbf{h}_1 = \cdots = \mathbf{h}_K$. The prediction equations of the KFS update equations in (22)-(26) simplify as $\hat{\mathbf{h}}_{m|m-1} = \hat{\mathbf{h}}_{m-1|m-1} \triangleq \hat{\mathbf{h}}_{m-1}$ and $\mathbf{P}_{m|m-1} = \mathbf{P}_{m-1|m-1} \triangleq \mathbf{P}_{m-1}$, for $m = 1, \ldots, K$. Moreover, for $\rho = 1$, the smoothing equations in (27)-(28) simplify as $\mathbf{h}_{j-1|K} = \mathbf{h}_{j|K}$ and $\mathbf{P}_{j-1|K} = \mathbf{P}_{j|K}$ for $j = K, \ldots, 1$. Hence, the filtering equations of the KFS updates suffice to describe the recursions, as follows. For $m = 1, \ldots, K$, the E-step of the J-SBL algorithm can be replaced by

$$\mathbf{G}_m = \mathbf{P}_{m-1} \boldsymbol{\Phi}_m^H \left(\sigma^2 \mathbf{I}_N + \boldsymbol{\Phi}_m \mathbf{P}_{m-1} \boldsymbol{\Phi}_m^H \right)^{-1} \tag{43}$$

$$\hat{\mathbf{h}}_m = \hat{\mathbf{h}}_{m-1} + \mathbf{G}_m (\mathbf{y}_m - \boldsymbol{\Phi}_m \hat{\mathbf{h}}_{m-1}) \tag{44}$$

$$\mathbf{P}_m = (\mathbf{I}_L - \mathbf{G}_m \boldsymbol{\Phi}_m) \mathbf{P}_{m-1}, \tag{45}$$

where $\boldsymbol{\Phi}_k$ denotes the measurement matrix of the k^{th} OFDM symbol given by $\boldsymbol{\Phi}_k = \mathbf{X}_k \mathbf{F}$. However, since $\boldsymbol{\Gamma}$ is unknown and $\mathbf{X}_k$ is known only at pilot locations, the SBL framework is incorporated to learn the unknown $\boldsymbol{\Gamma}$ and unknown data in $\mathbf{X}_k$. Hence, the update equations given above form the E-step, while the M-step is the same as that of the J-SBL algorithm, given by (37). The update for $\boldsymbol{\gamma}$ is given by,

$$\gamma^{(r+1)}(i) = P_K(i,i) + \left| \hat{h}_K(i) \right|^2, \tag{46}$$

where $\hat{\mathbf{h}}_K$ and $\mathbf{P}_K$ are given by (44) and (45), respectively. The unknown data can be detected by solving the following optimization problem:

$$\begin{aligned} X^{(r+1)}(i,i) = \underset{x_i \in \mathcal{S}}{\arg\min} &\left\{ |x_i|^2 C(i,i) \right. \\ &\left. + |y(i) - x_i \mathbf{F}_b(i,:)\hat{\mathbf{h}}_K|^2 \right\} \end{aligned} \tag{47}$$

where $i \in \mathcal{D}$, $\mathbf{C} = \mathbf{F}_b \mathbf{P}_K \mathbf{F}_b^H$. The initialization of $\boldsymbol{\gamma}$ and $\mathbf{X}^{(0)}$ is the same as the J-SBL algorithm of Sec. III-A. The steps of the RJ-SBL algorithm are listed in Appendix D. The complexity of the RJ-SBL algorithm is dominated by the computation of $\mathbf{G}_k$, and is given by $\mathcal{O}(NL^2K)$. Hence, for large K, the RJ-SBL algorithm is computationally significantly cheaper than the J-SBL algorithm.

The E-step of the RJ-SBL is a recursive implementation of the E-step of the J-SBL algorithm, and the M-steps of the algorithms are the same. Hence, the algorithms are mathematically equivalent if the same initializations are employed. This is illustrated via simulations in Sec. IV (see Fig. 4).

C. The JK-SBL Algorithm: Time-Varying Case

In this section, we generalize the K-SBL algorithm of Sec. II-B to obtain the JK-SBL algorithm, which utilizes the observations available at all the N subcarriers and performs data detection at the $(N - P_t)$ data subcarriers of the OFDM symbol. The algorithm is recursive in nature, and the channel estimates for K OFDM symbols are used to jointly estimate the a-sparse channel and the unknown data of the $m^{th}, 1 \leq m \leq K$ OFDM symbol. In essence, we solve the problem given by $(P4)$.

Our starting point, again, is the state space model given by (20) and (21). Using the observations $\mathbf{Y}_K = [\mathbf{y}_1, \ldots, \mathbf{y}_K]$, the recursive updates of the mean and the covariance of the posterior distribution are given by (22)-(28), with $\mathbf{y}_{p,m}$ and $\mathbf{\Phi}_{p,m}$ replaced by $\mathbf{y}_m$ and $\mathbf{\Phi}_m$, respectively. Thus, the JK-SBL algorithm uses the observations available at all the N subcarriers. Further, since $\mathbf{\Gamma}$ and data at the non-pilot subcarriers are unknown, the SBL framework leads to the objective function for the M-step given by

$$\begin{aligned}Q\left(\mathbf{X}_1,\ldots,\mathbf{X}_K,\boldsymbol{\gamma}|\mathbf{X}_1^{(r)},\ldots,\mathbf{X}_K^{(r)},\boldsymbol{\gamma}^{(r)}\right) = c'''\\ -\mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_K;\mathbf{X}_1^{(r)},\ldots,\mathbf{X}_K^{(r)},\boldsymbol{\gamma}^{(r)}}\Bigg[\sum_{j=1}^{K}\frac{\|\mathbf{y}_j-\mathbf{X}_j\mathbf{F}\mathbf{h}_j\|^2}{\sigma^2}\\ +K\log|\mathbf{\Gamma}|+\sum_{j=2}^{K}\frac{(\mathbf{h}_j-\rho\mathbf{h}_{j-1})^H\mathbf{\Gamma}^{-1}(\mathbf{h}_j-\rho\mathbf{h}_{j-1})}{(1-\rho^2)}\\ +\mathbf{h}_1^H\mathbf{\Gamma}^{-1}\mathbf{h}_1\Bigg],\end{aligned} \quad (48)$$

where c''' is a constant independent of $\boldsymbol{\gamma}$ and $\mathbf{X}_1,\ldots,\mathbf{X}_K$. The expression above is a sum of terms which are independent functions of $\boldsymbol{\gamma}$ and $\mathbf{X}_K$, denoted as $Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right)$ and $Q\left(\mathbf{X}_1,\ldots,\mathbf{X}_K|\mathbf{X}_1^{(r)},\ldots,\mathbf{X}_K^{(r)}\right)$, respectively. Further, we see that $Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right)$ is the same as the expression in (33). Hence, the learning rule for $\boldsymbol{\gamma}$ follows from the M-step of the K-SBL algorithm, and is given by (34). The expression for $Q\left(\mathbf{X}_1,\ldots,\mathbf{X}_K|\mathbf{X}_1^{(r)},\ldots,\mathbf{X}_K^{(r)}\right)$ is given by

$$\begin{aligned}Q\left(\mathbf{X}_1,\ldots,\mathbf{X}_K|\mathbf{X}_1^{(r)},\ldots,\mathbf{X}_K^{(r)}\right) = c\\ -\mathbb{E}_{\mathbf{h}_1,\ldots\mathbf{h}_K|\mathbf{Y}_K;\mathbf{X}_1^{(r)},\ldots,\mathbf{X}_K^{(r)},\boldsymbol{\gamma}^{(r)}}\left[\sum_{m=1}^{K}\frac{\|\mathbf{y}_m-\mathbf{X}_m\mathbf{F}\mathbf{h}_m\|^2}{\sigma^2}\right].\end{aligned} \quad (49)$$

As mentioned earlier, the M-step requires the computation of $\hat{\mathbf{h}}_{j|K} \triangleq \mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\mathbf{X}_1^{(r)},\ldots,\mathbf{X}_K^{(r)},\boldsymbol{\gamma}^{(r)}}[\mathbf{h}_j]$, and covariance $\mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_K|\mathbf{Y}_{p,K};\mathbf{X}_1^{(r)},\ldots,\mathbf{X}_K^{(r)},\boldsymbol{\gamma}^{(r)}}[\mathbf{h}_j\mathbf{h}_j^H] \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j|K}^H$ for $j = 1,\ldots,K$, which are given by the KFS equations of the E-step. The maximization of $Q\left(\mathbf{X}_1,\ldots,\mathbf{X}_K|\mathbf{X}_1^{(r)},\ldots,\mathbf{X}_K^{(r)}\right)$ in (49) leads to the following optimization problem for $\mathbf{X}_m$:

$$\begin{aligned}X_m^{(r+1)}(i,i) = \arg\min_{x_i\in\mathcal{S}}\Big\{|x_i|^2C_m(i,i)\\ +|y_m(i)-x_i\mathbf{F}(i,:)\hat{\mathbf{h}}_{m|K}|^2\Big\},\ 0\le m\le K,\ i\in\mathcal{D}\end{aligned} \quad (50)$$

where $\mathbf{C}_m = \mathbf{F}\mathbf{P}_{m|K}\mathbf{F}^H$ and $\mathbf{F}(i,:)$ represents the i^{th} row of the matrix $\mathbf{F}$. The steps of the JK-SBL algorithm are listed in Appendix E. The iterations of the JK-SBL proceed similar to the K-SBL algorithm, except for the additional M-step to estimate the unknown data. Also, the measurement matrix is given by $\mathbf{\Phi}_m^{(r)}$ in the r^{th} iteration of the m^{th} OFDM symbol, instead of the $\mathbf{\Phi}_{p,m}$ used in the K-SBL algorithm, which consisted of pilot subcarriers only. We provide a pictorial representation of the overall JK-SBL algorithm in Fig. 2. We use the channel estimate obtained from a few iterations of the K-SBL algorithm using P_t pilots (denoted as $\hat{\mathbf{h}}_{KSBL}$) to obtain the initial estimate $\mathbf{X}_m^{(0)}$ for $0\le m\le K$ as

$$X_m^{(0)}(i,i) = \arg\min_{x_i\in\mathcal{S}}|y_m(i)-x_i\mathbf{F}(i,:)\hat{\mathbf{h}}_{KSBL}|^2, i\in\mathcal{D}. \quad (51)$$

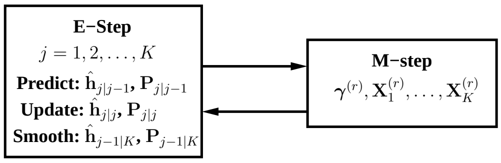

Fig. 2. Block diagram depicting the JK-SBL algorithm. The a-sparse channel is estimated and tracked in the E-step, while the M-step learns the unknown hyperparameters $\boldsymbol{\gamma}$ and detects the unknown transmit data $\mathbf{X}_1,\ldots,\mathbf{X}_K$.

Thus far, we proposed algorithms for joint a-sparse channel estimation and data detection in block-fading and time-varying channels in OFDM systems. We now discuss some implementation aspects of the proposed algorithms.

D. Discussion

In this subsection, we discuss the implementation of the proposed exact inference algorithms, and contrast their complexity with the block-based Autoregressive-SBL (ARSBL) algorithm [39] and the approximate inference algorithm [29].

Consider the estimation of the wireless channels when the data is observed up to the K^{th} OFDM symbol. First, in the forward recursion, (22)-(26) are applied recursively until we reach the K^{th} OFDM symbol. Hence, in the forward recursion, we store the values of $\hat{\mathbf{h}}_{j|j}$, $\hat{\mathbf{h}}_{j|j-1}$, $\mathbf{P}_{j|j}$ and $\mathbf{P}_{j|j-1}$ for $j = 1,2,\ldots,K$. Next, we apply the backward recursion using the Kalman smoother given by (27)-(28), i.e., KFS is applied to the whole sequence of observations before updating $\boldsymbol{\gamma}$. The Kalman smoother helps to utilize all the information available in both the past and future symbols, and hence improves the channel estimates. For the K-SBL and JK-SBL algorithms, the smoothed mean and covariance are required for the computation of the M-step.

The K-SBL and JK-SBL algorithms are iterative in nature, and the filtering and smoothing equations are executed in the E-step of every iteration using the hyperparameters obtained in the M-step of the previous iteration and the unknown data for K symbols. Hence, the E-step performs exact inference, by obtaining the exact posterior distribution of the a-sparse channel, given the estimate of the hyperparameters. Exact inference ensures that the likelihood function increases at each EM iteration. However, the price paid for the exact inference methods is their higher complexity, as has been well-demonstrated by the simulation results in [29].

Using a flop-count analysis [44], for K ($K > 1$) OFDM symbols, the computations of the K-SBL and JK-SBL algorithms are dominated by the computation of the $\mathbf{J}_{K-1}$ term in the smoothing step, which has a complexity of $\mathcal{O}(KL^3)$ per iteration. In a block-based method such as the ARSBL, the computation of the covariance matrix $\mathbf{\Sigma}$ incurs a complexity of $\mathcal{O}(K^3P_t^2L)$ per iteration. Hence, we see that if the number of OFDM symbols to be tracked are such that $KP_t > L$, the complexity of the ARSBL algorithm is larger than the K-SBL algorithm. In other words, the K-SBL algorithm is a good choice among the exact inference techniques when the number of OFDM symbols to be tracked is large.

The proposed recursive algorithms are very flexible. For example, a pruning step, where small channel coefficients or hyperparameters are set to zero, can be incorporated between iterations. This leads to a reduced support set, which in turn results in faster convergence and lower complexity [1]. However, pruning may eliminate some of the basis vectors of the measurement matrix before achieving convergence and result in support recovery errors.

The improved channel estimation accuracy achieved by using the SBL techniques can lead to performance enhancements in different ways. As will be demonstrated in the next section, the BER performance can be improved, in both uncoded and coded systems. An additional approach could be to reduce, or optimize, the pilot density, with the aim of maximizing the outage capacity [7], [13].

IV. Simulation Results

In this section, we demonstrate the performance of the proposed channel estimation algorithms through Monte Carlo simulations. We consider the parameters in the 3GPP/LTE broadband standard [45], [46]. We use a 3MHz OFDM system with 256 subcarriers, with a sampling frequency of $f_s = 3.84$MHz, resulting in an OFDM symbol duration of $\sim 83.3\mu s$ with Cyclic Prefix (CP) of $16.67\mu s$ (64 subcarriers). The length of the a-sparse channel (L) is taken to be equal to the length of the CP. Each OFDM frame consists of $K = 7$ OFDM symbols, which is also known as an OFDM slot. The data is transmitted using a rate $1/2$ Turbo code with QPSK modulation. For the Turbo code generation, we use the publicly available software [47], which uses a maximum of 10 Turbo iterations.

A sample instantiation of the a-sparse channel used in the simulations and the filtered MIP are depicted in Fig. 3. The figure captures the leakage effect due to finite bandwidth sampling and practical filtering. To generate the plot, we have used the Pedestrian B channel model [48] with Rayleigh fading. We have also used raised cosine filtering at the receiver and transmitter with a roll-off factor of 0.5 [46]. At the sampling frequencies considered, the number of significant channel taps are far fewer than the weak channel taps in the filtered impulse response, as seen in Fig. 3. In the following subsections, we present the simulation results for the block fading and time varying scenarios.

A. Block-fading channel

In this subsection, we consider a block-fading channel and use $P_b = 44$ pilot subcarriers, uniformly placed in each OFDM symbol. Each OFDM frame consists of $K = 7$ OFDM symbols. We implement the SBL and the J-SBL algorithm and plot the MSE performance of both the algorithms in Fig. 4, using a convergence criteria of $\epsilon = 10^{-9}$ and $r_{max} = 200$ for both the algorithms. We compare the MSE performance of the proposed algorithms with the CS based channel estimation technique [10], and the MIP-aware methods: pilot-only MIP-aware estimation [3] and the MIP-aware joint data and channel estimation algorithm, which we refer to as the EM-OFDM algorithm [19]. From Fig. 4, we observe that the SBL

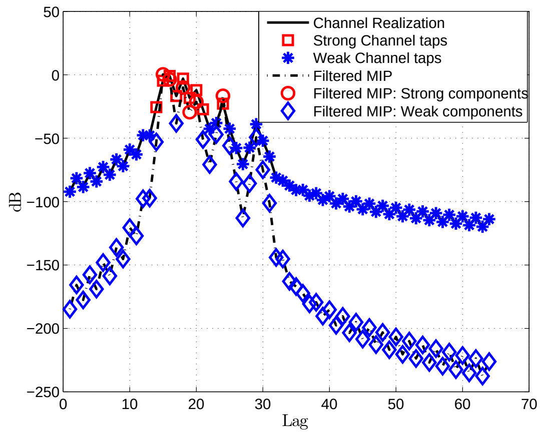

Fig. 3. One sample channel realization of the a-sparse channel, along with the filtered MIP, i.e., the MIP when raised cosine filters are employed at the transmitter and receiver. The plot also shows the strong (> -30 dB) and weak (< -30 dB) channel taps and filtered-MIP components, to illustrate that the channel can indeed be modeled as being approximately sparse.

algorithms perform better than the MIP-unaware, non-iterative schemes such as the Frequency Domain Interpolation (FDI) technique. Among the iterative methods, the J-SBL algorithm performs an order of magnitude better than the SBL algorithm, especially at higher values of SNR, while being within 3 dB from the MIP-aware EM-OFDM algorithm. The J-SBL jointly detects the $(KN - P_b)$ data symbols along with the estimating channel, resulting in a significantly lower overall MSE. As mentioned earlier, the RJ-SBL is mathematically equivalent to, and computationally simpler than, the J-SBL algorithm. Hence, they have the same performance.

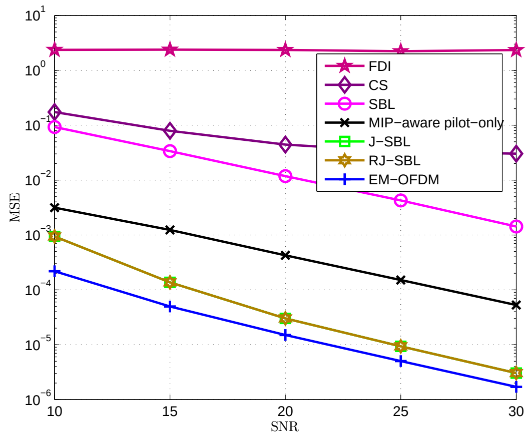

Fig. 4. MSE performance of SBL, J-SBL/RJ-SBL algorithms compared to FDI [4], CS [10], MIP-aware pilot-only [3] and EM [19] schemes in a block-fading channel, with $P_b = 44$ pilot subcarriers, as a function of SNR in dB.

The coded and the uncoded BER performance of the EM, J-SBL and a genie receiver, i.e., a receiver with perfect knowledge of the channel (labeled as `Genie`), is shown in Fig. 5. We also compare the performance with SBL and MIP-aware pilot-only channel estimation followed by data detection. The BER performance of the RJ-SBL is superior that of the SBL and CS algorithms in both coded and uncoded cases. The MIP-aware pilot-only estimation method has a better BER performance compared to RJ-SBL for SNRs < 15 dB, in both coded and the uncoded cases. Also, the MIP-aware EM-OFDM algorithm outperforms the proposed RJ-SBL algorithm by 3 dB. This is because, in the block-fading case, J-SBL algorithm suffers due to error propagation from the large number of data symbols that are simultaneously detected.

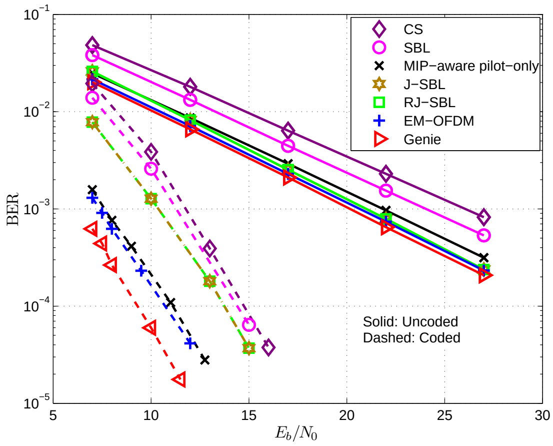

Fig. 5. BER performance of the proposed algorithms in a block-fading channel, with $P_b = 44$ pilot subcarriers, as a function of E_b/N_0.

B. *Slowly time-varying channel*

In this section, we consider a slowly time-varying channel, simulated according to a Jakes' model [49] with a normalized fade rate of $f_d T_s = 0.001$ and $P_t = 44$ pilot subcarriers in every OFDM symbol. The MSE performance of the K-SBL and the JK-SBL algorithms are plotted against SNR in Fig. 6 and compared with the per-symbol MIP-unaware FDI [4], and the per-symbol J-SBL and the SBL algorithm. Figure 6 also shows the performance of the optimal MIP-aware Kalman tracking algorithm [35] which considers all the subcarriers as carrying pilot symbols. The SBL and the J-SBL algorithms are not designed to exploit the temporal correlation in the channel, and hence, they perform 7-8 dB poorer than their recursive counterparts, the K-SBL and the JK-SBL algorithms. At higher SNR, we observe that the performance of the JK-SBL algorithm is only 2 dB worse than the MIP-aware Kalman tracking algorithm with all subcarriers being pilot subcarriers.

In Fig. 7, we depict the BER performance of the proposed algorithms. We see that, in the coded case, while the JK-SBL performs about 2 dB better than the J-SBL algorithm, it is only a fraction of a dB away from performance of the genie receiver which has perfect channel knowledge. The JK-SBL outperforms pilots-only based channel estimation using the K-SBL and the SBL algorithms by a large (4-5 dB) margin. Further, it outperforms the MIP-aware EM-OFDM algorithm, since the latter is unaware of the channel correlation, and performs channel estimation on a per-OFDM symbol basis; while the JK-SBL algorithm exploits its knowledge of the channel correlation to improve the channel estimates.

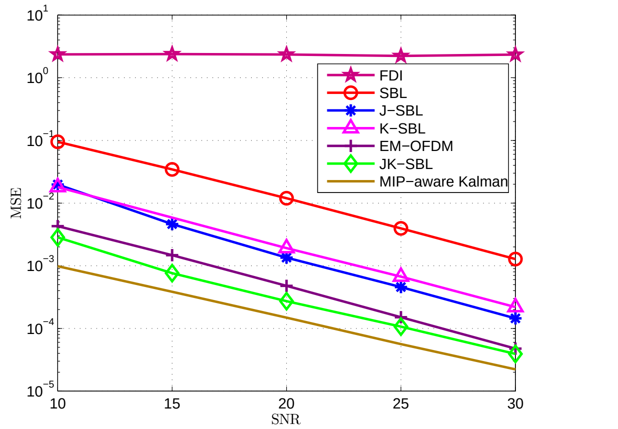

Fig. 6. MSE performance of different schemes in a time-varying channel, compared to the optimal Kalman tracker [35] with $f_d T_s = 0.001$ and $P_t = 44$, as a function of SNR in dB.

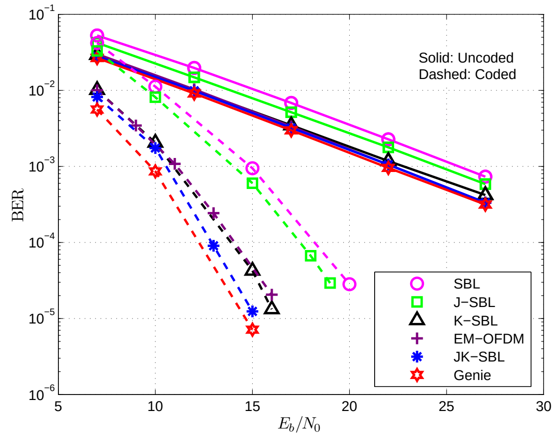

Fig. 7. BER performance of different schemes in a time-varying channel with $f_d T_s = 0.001$ and $P_t = 44$, as a function of E_b/N_0.

In Fig. 8, we study the MSE performance of K-SBL and the JK-SBL algorithm across the OFDM frame as a function of the OFDM symbol index for SNRs of 10 and 30 dB. It is observed that after an initial reduction in the MSE, the MSE tends to remain more or less unchanged throughout the frame, especially at an SNR of 30 dB, indicating that the algorithms learn the hyperparameters within the first few OFDM symbols. Hence, this study shows that at a given SNR, it is possible to restrict the number of OFDM symbols over which the proposed algorithms need to learn the hyperparameters. After the hyperparameters are estimated, channel tracking can be accomplished using the conventional MIP-aware Kalman tracking algorithm. This can lead to additional reduction in the computational complexity of the algorithms.

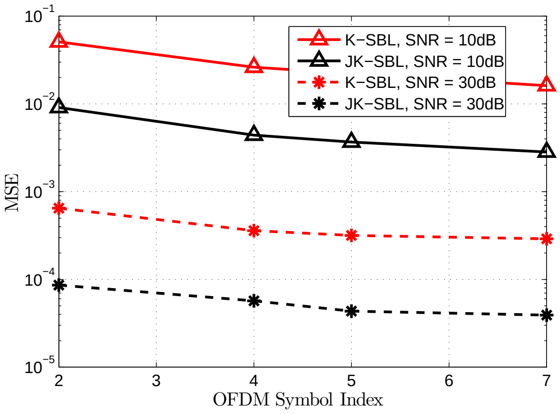

Fig. 8. MSE performance of the K-SBL and the JK-SBL algorithms, as a function of the OFDM symbol index with $f_d T_s = 0.001$ and $P_t = 44$.

Acknowledgments

The authors gratefully acknowledge the detailed feedback from the reviewers, which has significantly improved the presentation.

V. Conclusions

In this paper, we considered the joint approximately sparse channel estimation and data detection for block-fading and time-varying channels in SISO-OFDM systems, from the perspective of SBL. To estimate the a-sparse block-fading channel, we proposed the SBL algorithm and generalized it to obtain the J-SBL algorithm for joint a-sparse channel estimation and data detection. Furthermore, we obtained a mathematically equivalent low-complexity RJ-SBL algorithm. For the time-varying channels, we used a first order AR model to capture the temporal correlation of the a-sparse channel and proposed a novel K-SBL algorithm, using which we tracked the a-sparse channel. We generalized the K-SBL algorithm to obtain the JK-SBL algorithm for joint channel estimation and data detection. We discussed the implementation issues of the recursive algorithms and showed that the proposed algorithms entail a significantly lower computational complexity compared to the previously known SBL techniques. Simulation results showed that the proposed recursive techniques exploit the temporal correlation of the channel, leading to an enhanced channel estimation and data detection capability compared to the per-symbol SBL and J-SBL algorithms, and also learn the hyperparameters within a few OFDM symbols.

Appendix

A. Algorithmic Representation: SBL

Input: $\mathbf{y}_p$, $\mathbf{\Phi}_p$, r_{max} and ϵ.
Initialize $\boldsymbol{\gamma}^{(0)} = \mathbf{I}_L$, Set difference $= 1$, $r = 0$
while (difference $> \epsilon$ and $r < r_{max}$)
E-step: $\boldsymbol{\mu} = \sigma^{-2}\mathbf{\Sigma}\mathbf{\Phi}_p^H\mathbf{y}_p$
$\mathbf{\Sigma} = \mathbf{\Gamma}^{(r)} - \mathbf{\Gamma}^{(r)}\mathbf{\Phi}_p^H\left(\sigma^2\mathbf{I}_{P_b} + \mathbf{\Phi}_p\mathbf{\Gamma}^{(r)}\mathbf{\Phi}_p^H\right)^{-1}\mathbf{\Phi}_p\mathbf{\Gamma}^{(r)}$
M-step: $\gamma^{(r+1)}(i) = \Sigma(i,i) + |\mu(i)|^2$ for $i = 1, 2, \ldots, L$
difference $\triangleq \|\boldsymbol{\gamma}^{(r+1)} - \boldsymbol{\gamma}^{(r)}\|_2^2$, $r \leftarrow r+1$
end
Output: $\boldsymbol{\mu}$, $\boldsymbol{\gamma}^{(r)}$

B. Algorithmic Representation: K-SBL

Input: $\mathbf{y}_{p,1}, \ldots, \mathbf{y}_{p,K}$; $\mathbf{\Phi}_{p,1}, \ldots, \mathbf{\Phi}_{p,K}$; r_{max} and ϵ
Set difference $= 1$, $r = 0$, $\boldsymbol{\gamma}^{(0)} \triangleq \mathbf{I}_L$
while (difference $> \epsilon$ and $r < r_{max}$)
E-step: Set $\mathbf{P}_{0|0} = \mathbf{\Gamma}^{(r)}$; $\hat{\mathbf{h}}_{0|0} = \mathbf{0}$
for $m = 1, \ldots, K$ **do**
Prediction: $\hat{\mathbf{h}}_{m|m-1} = \rho\hat{\mathbf{h}}_{m-1|m-1}$
$\mathbf{P}_{m|m-1} = \rho^2\mathbf{P}_{m-1|m-1} + (1-\rho^2)\mathbf{\Gamma}^{(r)}$
Filtering: $\mathbf{G}_m = \mathbf{P}_{m|m-1}\mathbf{\Phi}_{p,m}^H\left(\sigma^2\mathbf{I}_{P_t} + \mathbf{\Phi}_{p,m}\mathbf{P}_{m|m-1}\mathbf{\Phi}_{p,m}^H\right)^{-1}$
$\hat{\mathbf{h}}_{m|m} = \hat{\mathbf{h}}_{m|m-1} + \mathbf{G}_m(\mathbf{y}_{p,m} - \mathbf{\Phi}_{p,m}\hat{\mathbf{h}}_{m|m-1})$
$\mathbf{P}_{m|m} = (\mathbf{I}_L - \mathbf{G}_m\mathbf{\Phi}_{p,m})\mathbf{P}_{m|m-1}$
end for
Smoothing: Set $\mathbf{P}_{K,K-1|K} = \rho(\mathbf{I}_L - \mathbf{G}_K\mathbf{\Phi}_{p,K})\mathbf{P}_{K-1|K-1}$
for $j = K, K-1, \ldots, 2$ **do**
$\mathbf{J}_{j-1} = \rho\mathbf{P}_{j-1|j-1}\mathbf{P}_{j|j-1}^{-1}$ and $\mathbf{J}_{j-2} = \rho\mathbf{P}_{j-2|j-2}\mathbf{P}_{j-1|j-2}^{-1}$
$\hat{\mathbf{h}}_{j-1|K} = \hat{\mathbf{h}}_{j-1|j-1} + \mathbf{J}_{j-1}(\hat{\mathbf{h}}_{j|K} - \hat{\mathbf{h}}_{j|j-1})$
$\mathbf{P}_{j-1|K} = \mathbf{P}_{j-1|j-1} + \mathbf{J}_{j-1}(\mathbf{P}_{j|K} - \mathbf{P}_{j|j-1})\mathbf{J}_{j-1}^H$
$\mathbf{P}_{j-1,j-2|K} = \mathbf{P}_{j-1|j-1}\mathbf{J}_{j-2}^H + \mathbf{J}_{j-1}^H(\mathbf{P}_{j,j-1|K}$
$-\rho\mathbf{P}_{j-1|j-1})\mathbf{J}_{j-2}^H$
$\mathbf{M}_{j|K} \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j|K}^H + \rho^2(\mathbf{P}_{j-1|K} + \hat{\mathbf{h}}_{j-1|K}\hat{\mathbf{h}}_{j-1|K}^H) -$
$2\rho\text{Re}(\mathbf{P}_{j,j-1|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j-1|K}^H)$
end for
M-step: $\mathbf{M}_{1|K} \triangleq \mathbf{P}_{1|K} + \hat{\mathbf{h}}_{1|K}\hat{\mathbf{h}}_{1|K}^H$
$\gamma^{(r+1)}(i) = \frac{1}{K}\left(\sum_{j=2}^{K}\frac{M_{j|K}(i,i)}{(1-\rho^2)} + M_{1|K}(i,i)\right)$ for $i = 1, \ldots, L$.
Compute the difference $\triangleq \|\boldsymbol{\gamma}^{(r+1)} - \boldsymbol{\gamma}^{(r)}\|_2^2$, $r \leftarrow r+1$
end while
Output: $\hat{\mathbf{h}}_{m|K}, 1 \leq m \leq K$; $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^{(r)}$

C. Algorithmic Representation: J-SBL

Input: $\mathbf{y}$, $\mathbf{F}_b$, r_{max} and ϵ
Initialize $\boldsymbol{\gamma}^{(0)} = \mathbf{I}_L$, $\hat{\mathbf{X}}_1^{(0)}, \ldots, \hat{\mathbf{X}}_K^{(0)}$ using (42), set difference $= 1$, $r = 0$
while (difference $> \epsilon$ and $r < r_{max}$)
E-step: $\mathbf{\Phi}^{(r)} = \mathbf{F}_b\mathbf{X}^{(r)\,H}$; $\boldsymbol{\mu} = \sigma^{-2}\mathbf{\Sigma}\mathbf{\Phi}^{(r)\,H}\mathbf{y}$
$\mathbf{\Sigma} = \mathbf{\Gamma}^{(r)} - \mathbf{\Gamma}^{(r)}\mathbf{\Phi}^{(r)\,H}\left(\sigma^2\mathbf{I}_{NK} + \mathbf{\Phi}^{(r)}\mathbf{\Gamma}^{(r)}\mathbf{\Phi}^{(r)\,H}\right)^{-1}\mathbf{\Phi}^{(r)}\mathbf{\Gamma}^{(r)}$
M-step: $\gamma^{(r+1)}(i) = \Sigma(i,i) + |\mu(i)|^2$ for $i = 1, 2, \ldots, L$
$\mathbf{C}_b = \mathbf{F}_b\mathbf{\Sigma}\mathbf{F}_b^H$; For $i \in \mathcal{D}$,
$X^{(r+1)}(i,i) = \arg\min_{x_i \in \mathcal{S}}\left\{|x_i|^2 C_b(i,i) + |y(i) - x_i\mathbf{F}_b(i,:)\boldsymbol{\mu}|^2\right\}$
difference $\triangleq \|\boldsymbol{\gamma}^{(r+1)} - \boldsymbol{\gamma}^{(r)}\|_2^2$, $r \leftarrow r+1$
end
Output: $\boldsymbol{\mu}$, $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^{(r)}$, $\hat{\mathbf{X}} = \mathbf{X}^{(r)}$

D. Algorithmic Representation: RJ-SBL

Input: $\mathbf{y}$, $\mathbf{F}$, $\mathbf{F}_b$, r_{max}, and ϵ
Initialize $\boldsymbol{\gamma}^{(0)} = \mathbf{I}_L$, $\mathbf{X}^{(0)}$ using (42), set difference $= 1$, $r = 0$
while (difference $> \epsilon$ and $r < r_{max}$)
$\mathbf{\Phi}_j^{(r)} = \mathbf{X}_j^{(r)}\mathbf{F}$, for $j = 1, \ldots, K$
E-step: Set $\mathbf{P}_0 = \mathbf{\Gamma}^{(r)}$; $\hat{\mathbf{h}}_0 = \mathbf{0}$
for $m = 1, \ldots, K$ **do**
$\mathbf{G}_m = \mathbf{P}_{m-1}\mathbf{\Phi}_m^{(r)\,H}\left(\sigma^2\mathbf{I}_N + \mathbf{\Phi}_m^{(r)}\mathbf{P}_{m-1}\mathbf{\Phi}_m^{(r)\,H}\right)^{-1}$
$\hat{\mathbf{h}}_m = \hat{\mathbf{h}}_{m-1} + \mathbf{G}_m(\mathbf{y}_m - \mathbf{\Phi}_m^{(r)}\hat{\mathbf{h}}_{m-1})$
$\mathbf{P}_m = \left(\mathbf{I}_L - \mathbf{G}_m\mathbf{\Phi}_m^{(r)}\right)\mathbf{P}_{m-1}$
end for
M-step: $\gamma^{(r+1)}(i) = P_K(i,i) + |h_K(i)|^2$ for $i = 1, 2, \ldots, L$
$\mathbf{C} = \mathbf{F}_b\mathbf{P}_K\mathbf{F}_b^H$; For $i \in \mathcal{D}$,
$X^{(r+1)}(i,i) = \arg\min_{x_i \in \mathcal{S}}\left\{|x_i|^2 C(i,i) + |y(i) - x_i\mathbf{F}_b(i,:)\hat{\mathbf{h}}_K|^2\right\}$
Compute the difference $\triangleq \|\boldsymbol{\gamma}^{(r+1)} - \boldsymbol{\gamma}^{(r)}\|_2^2$, $r \leftarrow r+1$
end while
Output: $\hat{\mathbf{h}}_{K|K}$, $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^{(r)}$

E. Algorithmic Representation: JK-SBL

Input: $\mathbf{y}_1, \ldots, \mathbf{y}_K$, $\mathbf{F}$, r_{max} and ϵ.
Initialize $\mathbf{X}_1^{(0)}, \ldots, \mathbf{X}_K^{(0)}$ using (51), $\mathbf{\Phi}_m^{(0)} = \mathbf{X}_m^{(0)}\mathbf{F}$
Set difference $= 1$, $r = 0$, $\boldsymbol{\gamma}^{(0)} \triangleq \mathbf{I}_L$
while (difference $> \epsilon$ and $r < r_{max}$)
E-step: Set $\mathbf{P}_{0|0} = \mathbf{\Gamma}^{(r)}$, $\hat{\mathbf{h}}_{0|0} = \mathbf{0}$
for $m = 1, \ldots, K$ **do**
Prediction: $\hat{\mathbf{h}}_{m|m-1} = \rho\hat{\mathbf{h}}_{m-1|m-1}$
$\mathbf{P}_{m|m-1} = \rho^2\mathbf{P}_{m-1|m-1} + (1-\rho^2)\mathbf{\Gamma}^{(r)}$
Filtering: $\mathbf{G}_m = \mathbf{P}_{m|m-1}\mathbf{\Phi}_m^{(r)\,H}\left(\sigma^2\mathbf{I}_N + \mathbf{\Phi}_m^{(r)}\mathbf{P}_{m|m-1}\mathbf{\Phi}_m^{(r)\,H}\right)^{-1}$
$\hat{\mathbf{h}}_{m|m} = \hat{\mathbf{h}}_{m|m-1} + \mathbf{G}_m(\mathbf{y}_{p,m} - \mathbf{\Phi}_m^{(r)}\hat{\mathbf{h}}_{m|m-1})$
$\mathbf{P}_{m|m} = (\mathbf{I}_L - \mathbf{G}_m\mathbf{\Phi}_m^{(r)})\mathbf{P}_{m|m-1}$
end for
Smoothing: Set $\mathbf{P}_{K,K-1|K} = \rho(\mathbf{I}_L - \mathbf{G}_K\mathbf{\Phi}_K^{(r)})\mathbf{P}_{K-1|K-1}$
for $j = K, K-1, \ldots, 2$ **do**
$\mathbf{J}_{j-1} = \rho\mathbf{P}_{j-1|j-1}\mathbf{P}_{j|j-1}^{-1}$ and $\mathbf{J}_{j-2} = \rho\mathbf{P}_{j-2|j-2}\mathbf{P}_{j-1|j-2}^{-1}$
$\hat{\mathbf{h}}_{j-1|K} = \hat{\mathbf{h}}_{j-1|j-1} + \mathbf{J}_{j-1}(\hat{\mathbf{h}}_{j|K} - \hat{\mathbf{h}}_{j|j-1})$
$\mathbf{P}_{j-1|K} = \mathbf{P}_{j-1|j-1} + \mathbf{J}_{j-1}(\mathbf{P}_{j|K} - \mathbf{P}_{j|j-1})\mathbf{J}_{j-1}^H$
$\mathbf{P}_{j-1,j-2|K} = \mathbf{P}_{j-1|j-1}\mathbf{J}_{j-2}^H + \mathbf{J}_{j-1}^H(\mathbf{P}_{j,j-1|K} - \rho\mathbf{P}_{j-1|j-1})\mathbf{J}_{j-2}^H$
$\mathbf{M}_{j|K} \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j|K}^H + \rho^2(\mathbf{P}_{j-1|K} + \hat{\mathbf{h}}_{j-1|K}\hat{\mathbf{h}}_{j-1|K}^H)$
$-2\rho\text{Re}(\mathbf{P}_{j,j-1|K} + \hat{\mathbf{h}}_{j|K}\hat{\mathbf{h}}_{j-1|K}^H)$
end for
M-step: $\mathbf{M}_{1|K} \triangleq \mathbf{P}_{1|K} + \hat{\mathbf{h}}_{1|K}\hat{\mathbf{h}}_{1|K}^H$
$\gamma^{(r+1)}(i) = \frac{1}{K}\left(\sum_{j=2}^K \frac{M_{j|K}(i,i)}{(1-\rho^2)} + M_{1|K}(i,i)\right)$ for $i = 1, \ldots, L$.
$X_m^{(r+1)}(i) = \arg\min_{x_i \in \mathcal{S}} \left\{|x_i|^2 C_m(i,i)\right.$
$\left. + |y_m(i) - x_i\mathbf{F}(i,:)\hat{\mathbf{h}}_{m|K}|^2\right\}, i \in \mathcal{D}$, $\mathbf{\Phi}_m^{(r+1)} = \mathbf{X}_m^{(r+1)}\mathbf{F}$, $1 \leq m \leq k$
end for
Compute the difference $\triangleq \|\boldsymbol{\gamma}^{(r+1)} - \boldsymbol{\gamma}^{(r)}\|_2^2$, $r \leftarrow r+1$
end while
Output: $\hat{\mathbf{h}}_{m|K}, 1 \leq m \leq K$, $\hat{\boldsymbol{\gamma}}_K = \boldsymbol{\gamma}^{(r)}$, $\hat{\mathbf{X}}_m = \mathbf{X}_m^{(r)}, 1 \leq m \leq K$

REFERENCES

[1] R. Prasad and C. Murthy, "Bayesian learning for joint sparse OFDM channel estimation and data detection," in *Proc. Globecom*, 2010, pp. 1–6.

[2] L. Hanzo, M. Munster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-user communications, WLANs and Broadcasting*. John Wiley, 2003.

[3] J. Van de Beek, O. Edfors, M. Sandell, S. Wilson, and P. Borjesson, "On channel estimation in OFDM systems," in *Proc. VTC*, vol. 2, 1995, pp. 815–819.

[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] S.-J. Hwang and P. Schniter, "Efficient multicarrier communication for highly spread underwater acoustic channels," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 9, pp. 1674–1683, 2008.

[6] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[7] P. Schniter, "Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels," *Physical Commun.*, vol. 5, no. 2, pp. 91–101, 2012.

[8] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[9] G. Taubock and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Proc. ICASSP*, 2008, pp. 2885 –2888.

[10] M. Sharp and A. Scaglione, "Application of sparse signal recovery to pilot-assisted channel estimation," in *Proc. ICASSP*. IEEE, 2008, pp. 3469–3472.

[11] C. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Comm. Magazine*, vol. 48, no. 11, pp. 164–174, 2010.

[12] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, 2011.

[13] A. P. Kannu and P. Schniter, "On communication over unknown sparse frequency-selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6619–6632, 2011.

[14] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–214, 2001.

[15] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, 2004.

[16] D. Tse, and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[17] M. E. Tipping, "The relevance vector machine," in *Advances in Neural Information Processing Systems*, vol. 12, 2000.

[18] S. S. Ghassemzadeh, L. J. Greenstein, T. Sveinsson, A. Kavcic, and V. Tarokh, "UWB delay profile models for residential and commercial indoor environments," *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1235–1244, 2005.

[19] X. Ma, H. Kobayashi, and S. Schwartz, "An EM-based estimation of OFDM signals," in *Proc. WCNC*. IEEE, 2002, pp. 228–232.

[20] T.Y. Al Naffouri, A. Bahai, and A. Paulraj, "An EM-based OFDM receiver for time-variant channels," in *Proc. Globecom*, vol. 1, 2002.

[21] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255–271, 2010.

[22] C. Wu and D. Lin, "A group matching pursuit algorithm for sparse channel estimation for OFDM transmission," in *Proc. ICASSP*, vol. 4, 2006.

[23] S. Barembruch, E. Moulines, and A. Scaglione, "A sparse EM algorithm for blind and semi-blind identification of doubly selective OFDM channels," in *Proc. SPAWC*, 2010, pp. 1–5.

[24] D. Meng, "Approximate message passing for multi-carrier transmission over doubly selective channels," Ph.D. dissertation, Ohio State University, 2012.

[25] N. L. Pedersen, C. N. Manchón, D. Shutin, and B. H. Fleury, "Application of Bayesian hierarchical prior modeling to sparse channel estimation," in *Proc. ICC*, 2012, pp. 3487–3492.

[26] K. Murphy, "Dynamic Bayesian networks: representation, inference and learning," Ph.D. dissertation, University of California, 2002.

[27] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, University of London, 2003.

[28] B. J. Frey and N. Jojic, "A comparison of algorithms for inference and learning in probabilistic graphical models," *IEEE Trans. on Patt. Analysis and Machine Intelli.*, vol. 27, no. 9, pp. 1392–1416, 2005.

[29] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE J. Sel. Topics Signal Process.*, vol. 61, no. 2, pp. 340–354, 2013.

[30] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay-subspace," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 315–325, 2004.

[31] J. Ziniel, L. C. Potter, and P. Schniter, "Tracking and smoothing of time-varying sparse signals via approximate belief propagation," in *Proc. Asilomar Conf. on Signals, Syst., and Comput.*, 2010, pp. 808–812.

[32] Z. Zhang and B. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.

[33] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. ICIP*, 2008, pp. 893–896.

[34] T. Feng, T. Field, and S. Haykin, "Stochastic differential equation theory applied to wireless channels," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1478–1483, 2007.

[35] W. Chen and R. Zhang, "Kalman-filter channel estimator for OFDM systems in time and frequency-selective fading environment," in *Proc. ICASSP*, vol. 4, 2004, p. 377.

[36] Z. Yuanjin, "A novel channel estimation and tracking method for wireless OFDM systems based on pilots and Kalman filtering," *IEEE Trans. Consum. Electron.*, vol. 49, no. 2, pp. 275–283, 2003.

[37] T. Y. Al-Naffouri, "An EM-based forward-backward Kalman filter for the estimation of time-variant channels in OFDM," *IEEE Trans. Signal Process.*, vol. 55, no. 7, p. 3924, 2007.

[38] Z. Ghahramani and G. E. Hinton, "*Parameter estimation for linear dynamical systems*," Tech. Rep., 1996.

[39] Z. Zhang and B. Rao, "Sparse signal recovery in the presence of correlated multiple measurement vectors," in *Proc. ICASSP*, 2010, pp. 3986–3989.

[40] C. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, 1983.

[41] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*. Wiley New York, 1997, vol. 274.

[42] B. Anderson and J. Moore, *Optimal filtering*. Courier Dover Publications, 2005.

[43] D. Wipf and B. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, 2007.

[44] R. Hunger, "Floating point operations in matrix-vector calculus," *Munich University of Technology, TUM-LNS-TR-05-05*, 2005.
[45] J. Zyren and W. McCoy, "Overview of the 3GPP long term evolution physical layer," *Freescale Semiconductor, Inc., white paper*, 2007.
[46] "Universal Mobile Telecommunications System (UMTS), Selection procedures for the choice of radio transmission technologies of the UMTS," ETSI technical report, UMTS 21.01 version 3.0.1, Nov., 1997.
[47] C. Studer, C. Benkeser, S. Belfanti, and Q. Huang, "Design and implementation of a parallel turbo-decoder asic for 3gpp-lte," *IEEE J. Solid State Circuits.*, vol. 46, no. 1, pp. 8–17, 2011.
[48] ITU-R Rec. M.1225, "Guidelines for evaluation of radio transmission technologies (RTTs) for IMT-2000," 1997.
[49] Y. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, 2003.