Joint Channel Estimation and Data Detection in MIMO-OFDM Systems: A Sparse Bayesian Learning Approach

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

***Abstract*—The impulse response of wireless channels between the N_t transmit and N_r receive antennas of a MIMO-OFDM system are group approximately sparse (ga-sparse), i.e., the N_tN_r channels have a small number of significant paths relative to the channel delay spread and the time-lags of the significant paths between transmit and receive antenna pairs coincide. Often, wireless channels are also group *approximately cluster sparse* (gac-sparse), i.e., every ga-sparse channel consists of clusters, where a few clusters have all strong components while most clusters have all weak components. In this work, we cast the problem of estimating the ga-sparse and gac-sparse block-fading and time-varying channels in the Sparse Bayesian Learning (SBL) framework, and propose a bouquet of novel algorithms for pilot-based channel estimation and joint channel estimation and data detection in MIMO-OFDM systems. The proposed joint channel estimation and data detection schemes are capable of recovering ga-sparse and gac-sparse channels even when the measurement matrix is only partially known. Further, we employ a first order autoregressive modeling of the temporal variation of the wireless ga-sparse and gac-sparse channels and propose a recursive Kalman filtering and smoothing (KFS) technique for joint channel estimation, tracking and data detection. The KFS framework exploits the correlation structure in the time-varying channel. We also propose novel, parallel-implementation based, low complexity techniques for estimating gac-sparse channels. Monte Carlo simulations illustrate the efficacy of proposed techniques in terms of mean square error (MSE) and coded bit error rate (BER) performance. In particular, we demonstrate the performance benefits offered by algorithms that exploit the gac-sparse structure in the wireless channel.**

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

I. Introduction

Multiple Input Multiple Output (MIMO) combined with Orthogonal Frequency Division Multiplexing (OFDM) is the air-interface solution for next-generation broadband wireless systems and standards. Multiple antennas are employed at the transmitter and receiver of a MIMO-OFDM system in order to exploit the diversity and multiplexing advantages of a MIMO system, while OFDM provides resilience to the commonly encountered frequency-selective fading of a multipath wireless environment [2]. Most OFDM-based wireless standards such as DVB-T, IEEE 802.11a, IEEE 802.16e etc., employ pilot-based channel estimation techniques for accurately decoding the transmitted data bits. However, such methods necessitate the transmission of pilots symbols on a set of anchor subcarriers per transmit antenna, leading to severe overheads on the spectral efficiency. In this paper, we propose novel MIMO-OFDM channel estimation techniques using far fewer pilots compared to the conventional methods [3], [4], by exploiting the approximate sparsity of the wireless channel. We also extend the algorithms to exploit structure beyond sparsity, such as temporal correlation, and clustered multipath components.

In this work, we model the spatially uncorrelated N_tN_r MIMO-OFDM wireless channels as (a) group approximately-sparse (ga-sparse), and (b) group approximately cluster sparse (gac-sparse). Further, we formulate the channel estimation problem in block-fading and time-varying channels and investigate the problem of pilot-only channel estimation and joint channel estimation and data detection for ga-sparse and gac-sparse channels. Our focus is to design novel Sparse Bayesian Learning (SBL) type algorithms for joint ga-sparse and gac-sparse channel estimation and data detection in MIMO-OFDM systems.

A. Background and Literature Survey

In this subsection, we present the basic set-up of the coded MIMO-OFDM system considered in this work and formulate the problem of pilot-based channel estimation and joint channel estimation and data detection in a MIMO-OFDM system using the Multiple Measurement Vector (MMV) framework.

Figure 1 shows the block diagram of a typical MIMO-OFDM system with N subcarriers, N_t transmit antennas and N_r receive antennas. The transmissions take place through OFDM frames, where every frame consists of K OFDM symbols. For simplicity, we assume that the timing and frequency offsets between the transmitter and receiver are perfectly estimated and compensated for, prior to the start of communication. The residual offsets do result in a performance degradation, but from our numerical experiments, we have found that it does not change the relative performance of the different schemes considered in this work. A detailed study of

R. Prasad is with the Nanyang Technological University, Singapore. She was at the Dept. of Electrical Communication Engg. (ECE), Indian Institute of Science (IISc), Bangalore, India, during the course of this work. Email: ranjitha.p@gmail.com. C. R. Murthy is with the Dept. of ECE, IISc, Bangalore, India. Email: cmurthy@ece.iisc.ernet.in. B. D. Rao is with the Dept. of Electrical and Computer Engg., University of California, San Diego, USA. Email: brao@ucsd.edu.

This work was supported in part from the Tata Consultancy Services Research Scholar Program, the Dept. of Information Technology, and the Defense Research and Development Organization, Govt. of India, and NSF grant CCF-1115645.

This work has appeared in part in [1].

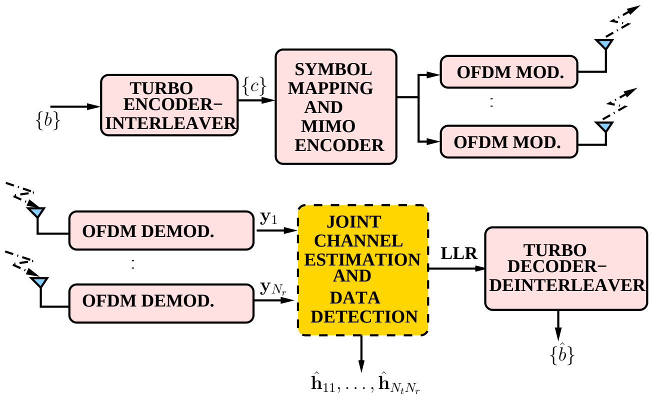

Figure 1. Turbo encoded/decoded transmitter and receiver chain of a MIMO-OFDM system. The dashed box (block shaded in yellow) highlights the proposed algorithms. Note that the quantities of interest are the channel estimates $\hat{\mathbf{h}}_{11}, \ldots, \hat{\mathbf{h}}_{N_t N_r}$ and output bits $\{\hat{b}\}$.

the effect of timing and frequency offsets, while interesting in its own right, is beyond the scope of this paper.

At the transmitter of the discrete-time MIMO-OFDM system, $\{b\}$ input bits are first encoded and interleaved into a new sequence of coded bits, $\{c\}$. The coded bits $\{c\}$ are mapped into an M-ary complex symbol sequence, which is further divided into N_t streams. In this work, we consider both the block-fading channel, where the channel coefficients remain fixed across the OFDM frame duration and vary in an i.i.d. fashion from frame to frame, and the slowly time-varying channel, where the channel coefficients can vary across the OFDM frame duration. At every transmit antenna, P_b pilots are inserted in an OFDM frame in the case of block-fading channels, and P_t pilots are inserted in every OFDM symbol of an OFDM frame in the case of time-varying channels. The pilot symbols along with coded data symbols $\{c\}$ are OFDM modulated and transmitted over the multipath fading channel of the k^{th} OFDM symbol, denoted by $\mathbf{h}_{n_t n_r,k} \in \mathbb{C}^{L\times 1}$. Here, n_t (n_r) denotes the transmit (receive) antenna index, and L is the length of the channel. After OFDM demodulation, the signal received at the n_r^{th} receive antenna of the k^{th} OFDM symbol is given by

$$\mathbf{y}_{n_r,k} = \sum_{n_t=1}^{N_t} \mathbf{X}_{n_t,k}\mathbf{F}\mathbf{h}_{n_t n_r,k} + \mathbf{v}_{n_r,k}, \quad n_r = 1, \ldots, N_r, \quad (1)$$

where the diagonal matrix $\mathbf{X}_{n_t,k} \in \mathbb{C}^{N\times N}$ consists of the pilot as well as data transmitted over the n_t^{th} transmit antenna and k^{th} OFDM symbol, and $\mathbf{F} \in \mathbb{C}^{N\times L}$ represents the matrix consisting of the first L columns of the $N \times N$ DFT matrix. Each component of $\mathbf{v}_{n_r,k} \in \mathbb{C}^{N\times 1}$ is an additive white circularly symmetric Gaussian noise with probability distribution $\mathcal{CN}(0, \sigma^2)$. In this work, we assume that the noise variance σ^2 is known at the receiver. When it is unknown, or when it is time-varying due to interference, one needs to dynamically estimate both the thermal noise as well as the interference power. It is possible to incorporate noise variance estimation within the Expectation Maximization (EM) framework of SBL [5]; however, we omit the details here due to lack of space.

In the complex baseband representation, the time domain channel impulse response between the n_t^{th} transmit antenna and the n_r^{th} receive antenna in the k^{th} symbol, denoted as $\tilde{h}_{n_t n_r,k}[t], t \in \mathbb{R}$, can be modeled as a stationary tapped delay line filter in the lag domain:

$$\tilde{h}_{n_t n_r,k}[t] = \sum_{l=1}^{\tilde{L}} \tilde{h}_{n_t n_r,k,l}\,\delta[t - \tau_l], \quad (2)$$

where $\delta[t]$ is the Dirac delta function, $\tilde{h}_{n_t n_r,k,l}$ and τ_l represent the attenuation and propagation delay on the path l corresponding to the n_t^{th} transmit and the n_r^{th} receive antenna, respectively, and $\tilde{L}$ is the number of resolvable paths [6]. Wireless channel models obtained using channel sounding experiments, on the other hand, exhibit *approximate* sparsity in the lag domain, for e.g., due to non-perfect low-pass filtering, e.g., raised cosine filtering [7]. Based on these practical considerations, we model the lag domain *filtered* channel impulse response as, $h_{n_t n_r,k}[t] = g_t[t] * \tilde{h}_{n_t n_r,k}[t] * g_r[t]$, where $g_t[t]$ and $g_r[t]$ represent the baseband transmit and receive filters employed at every transmit and receive antenna of the MIMO-OFDM system, and $*$ represents the convolution operation. Then, the corresponding discrete-time channel can be represented as, $h_{n_t n_r,k}(l) = h_{n_t n_r,k}[(l-1)T]$, where T is the baud interval. The overall channel is represented as $\mathbf{h}_{n_t n_r,k} = [h_{n_t n_r,k}(1), h_{n_t n_r,k}(2), \ldots, h_{n_t n_r,k}(L)]^T$. In addition, it is known that the sample-spaced representation of $\tilde{h}_{n_t n_r,k}[t]$ between different transmit and receive antenna pairs are group-sparse [8], [9], i.e., the locations of non-zero elements of the sparse vectors coincide. Since $g_t[t]$ and $g_r[t]$ are identical for every transmit and receive antenna, we deduce that the locations of the significant components in $\mathbf{h}_{n_t n_r,k}$ also coincide across the entire MIMO-OFDM system. In this work, we consider the following scenarios:

- $\mathbf{h}_{n_t n_r,k}$ is *group approximately-sparse* (ga-sparse), i.e., the $N_t N_r$ wireless channels consists of a few strong components and several weak components, and the time-lags of strong and weak components between transmit and receive antenna pairs coincide.
- $\mathbf{h}_{n_t n_r,k}$ is *group approximately cluster sparse* (gac-sparse), i.e., the $N_t N_r$ ga-sparse wireless channels consists of clusters such that the components of a given cluster are all strong or weak. In addition, there are a few clusters consisting of strong components.

To recover the ga-sparse and gac-sparse channels, we cast (1) in an MMV framework [10], [11]. Here, in the k^{th} OFDM symbol, the observations from the N_r receivers form the observation matrix, $\mathbf{Y}_k$, which is related to the vectors in the channel matrix, $\mathbf{H}_k$, through a common dictionary $\mathbf{\Phi}_k$, as follows:

$$\underbrace{[\mathbf{y}_{1,k}, \ldots, \mathbf{y}_{N_r,k}]}_{\mathbf{Y}_k \in \mathbb{C}^{N\times N_r}} = \underbrace{\mathbf{X}_k(\mathbf{I}_{N_t} \otimes \mathbf{F})}_{\mathbf{\Phi}_k \in \mathbb{C}^{N\times LN_t}} \underbrace{\begin{bmatrix} \mathbf{h}_{11,k} & \ldots & \mathbf{h}_{1N_r,k} \\ \vdots & \vdots & \\ \mathbf{h}_{N_t 1,k} & \ldots & \mathbf{h}_{N_t N_r,k} \end{bmatrix}}_{\mathbf{H}_k \in \mathbb{C}^{LN_t \times N_r}} + \underbrace{[\mathbf{v}_{1,k}, \ldots, \mathbf{v}_{N_r,k}]}_{\mathbf{V}_k \in \mathbb{C}^{N\times N_r}}, \quad (3)$$

where the overall transmit data matrix $\mathbf{X}_k \in \mathbb{C}^{N \times NN_t}$ is given by $\mathbf{X}_k \triangleq [\mathbf{X}_{1,k}, \mathbf{X}_{2,k}, \ldots, \mathbf{X}_{N_t,k}]$. At the P pilot subcarriers, the MIMO-OFDM system model can be written as

$$\mathbf{Y}_{p,k} = \mathbf{\Phi}_{p,k}\mathbf{H}_k + \mathbf{V}_{p,k}, \tag{4}$$

where $\mathbf{Y}_{p,k} \in \mathbb{C}^{P \times N_r}$, $\mathbf{\Phi}_{p,k} \in \mathbb{C}^{P \times LN_t}$ and $\mathbf{V}_{p,k} \in \mathbb{C}^{P \times N_r}$ are obtained by sampling $\mathbf{Y}_k$, $\mathbf{\Phi}_k$ and $\mathbf{V}_k$ at the pilot subcarriers, respectively.

Several channel estimation techniques for MIMO-OFDM systems have been proposed in literature. Conventional pilot-based interpolation techniques using on frequency domain Least Squares (LS) or Minimum Mean Square Error (MMSE) methods [3], [4] and lag domain LS and MMSE [4] do not provide reliable estimates when $P_b < L$, unless the prior knowledge of the the average multipath power profile measured at a particular location, also called as the Multipath Intensity Profile (MIP) of the channel [12], is known. In scenarios where the MIP is not known, blind methods [13] and techniques based on Compressed Sensing (CS) using group-sparse based formulation [8], [9] have been employed. Specifically, CS based simultaneous Orthogonal Matching Pursuit (OMP) [14], Modified OMP [15], Simultaneous Basis Pursuit Denoising and Simultaneous OMP [16] have been proposed for pilot-assisted ga-sparse channel estimation in MIMO-OFDM systems. Further, CS based Block OMP (BOMP) has been proposed for pilot-assisted gac-sparse MIMO-OFDM channel estimation [17]. In general, CS based methods recover an approximately sparse vector by recovering the s significant non-zero coefficients [18]: a large value of s guarantees recovery accuracy, but requires a correspondingly large number of measurements. Bayesian algorithms such as the Temporal SBL (TSBL) [19] have been proposed for recovery of temporally correlated group-sparse vectors, by modeling the correlation among the group-sparse vectors using a general correlation structure. However, due to the generality of the correlation structure assumed, the complexity of such algorithms quickly becomes prohibitive as the time-window over which estimation is performed increases, making these algorithms unsuitable for OFDM channel tracking.

If the MIP is known, incorporating the observations available at the data subcarriers into channel estimation by using joint data detection and channel estimation techniques can enhance the quality of channel estimates in MIMO-OFDM systems [11]. We showed that such joint approximately sparse channel estimation and data detection schemes enhance the quality of channel estimates in SISO-OFDM systems [20]. However, using SISO-OFDM estimators in parallel to obtain estimators in the MIMO-OFDM context does not lead to significant gains as they do not exploit the spatial ga-sparse and the gac-sparse nature of the channel. The novelty of this work is that our proposed Bayesian joint channel estimation and data detection techniques exploit the ga-sparse and gac-sparse structure in MIMO-OFDM channels, and in the case of both block-fading as well as time-varying channels. To the best of our knowledge, this is the first work in which such structure has been exploited for channel estimation and data detection in MIMO-OFDM systems.

B. Problem Formulation and Contributions

In this work, we address the problem of pilot-assisted and joint ga-sparse and gac-sparse channel estimation and data detection in MIMO-OFDM systems using the SBL framework. Among the known Bayesian sparse recovery techniques [21], [22], SBL exhibits the Expectation Maximization (EM) based monotonicity property, and offers guarantees such as convergence to the sparsest solution when the noise variance is zero, and convergence to a sparse local minimum, irrespective of the noise variance [23]. This motivates us to employ SBL [23], [24] based algorithms for recovery of the spatially uncorrelated ga-sparse and gac-sparse channels in MIMO-OFDM systems. In the SBL framework, we model the wireless channel as follows:

- In the case of ga-sparse channel, we model the channel as $\mathbf{h}_{n_tn_r,k} \sim \mathcal{CN}(0, \mathbf{\Gamma})$, where the hyperparameters $\mathbf{\Gamma} = \text{diag}(\gamma(1), \ldots, \gamma(L))$ are common for the N_tN_r channels for $0 \leq k \leq K$, i.e., the channels are spatially and temporally ga-sparse. Note that if $\gamma(l) \to 0$, then the corresponding $h_{n_tn_r,k}(l) \to 0$ for all the N_tN_rK channels [5], [23].
- In the case of gac-sparse channel, we assume that the L-length approximately sparse channel $\mathbf{h}_{n_tn_r,k}$ consists of B clusters, each of length M, as follows:

$$\mathbf{h}_{n_tn_r,k} = [\underbrace{h_{n_tn_r,k}(1), \ldots, h_{n_tn_r,k}(M)}_{\mathbf{h}_{n_tn_r,1k} \in \mathbb{C}^{1 \times M}}; \ldots; \underbrace{h_{n_tn_r,k}((M-1)B+1), \ldots, h_{n_tn_r,k}(MB)}_{\mathbf{h}_{n_tn_r,Bk} \in \mathbb{C}^{1 \times M}}], \tag{5}$$

  for $1 \leq k \leq K$. Here, the gac-sparse structure is imposed by modeling the b^{th} cluster of the channel as $\mathbf{h}_{n_tn_r,bk} \sim \mathcal{N}(0, \gamma(b)\mathbf{I}_M)$, where $\gamma(b)$ is an unknown hyperparameter such that when $\gamma(b) = 0$, the b^{th} block of $\mathbf{h}_{n_tn_r,k}$ is zero [25]. In addition, we note that different clusters of the gac-sparse channel are mutually uncorrelated, and hence, the overall covariance matrix of $\mathbf{h}_{n_tn_r,k}$ is a block-diagonal matrix with principal blocks given by $\gamma(b)\mathbf{I}_M$, $1 \leq b \leq B$.

Depending on the mobility of the receiver, the ga-sparse and the gac-sparse channels may remain constant over the frame duration (block-fading), or may be slowly time-varying. When the channel is time-varying, the nonzero channel coefficients vary slowly and are temporally correlated, but the locations of significant components of the channel remain constant for several OFDM frames [26]. In addition, it is known that the first order autoregressive (AR) model accurately captures the local behavior of fading wireless channels [27]. Hence, in this work, we employ a first order AR model for the time-varying channel in both the ga-sparse as well as the gac-sparse cases, and develop a Kalman Filter (KF) based framework for exact inference using the received pilot and data symbols. The first order AR model for the k^{th} channel tap is given by

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

where $\rho = J_0(2\pi f_d T_s)$ is the AR coefficient, $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, f_d is the Doppler

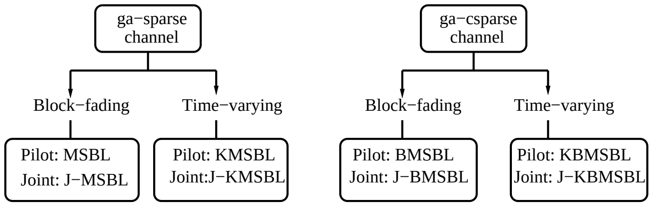

Figure 2. Figure depicting the algorithms proposed in this work.

frequency, and T_s is the OFDM symbol duration [28]. The driving noise $\mathbf{u}_{n_t n_r,k}$ is distributed as $\mathbf{u}_{n_t n_r,k} \sim \mathcal{CN}(0,(1-\rho^2)\gamma(i)\mathbf{I}_M)$.

As depicted in Fig. 2, the following algorithms are proposed for recovering ga-sparse channels in the MMV framework given in (3):

- In Sec. II, we adapt the multiple response SBL (MSBL) algorithm [10] and propose a novel Kalman MSBL (KMSBL) for pilot-based channel estimation in block-fading channels and time-varying channels, respectively.
- In Sec. III, we propose a Joint-MSBL (J-MSBL) algorithm and Joint-KMSBL (J-KMSBL) algorithm for joint channel estimation/tracking and data detection in block-fading and time-varying scenarios, respectively.

Further, in the context of recovering gac-sparse channels, we propose the following algorithms:

- In Sec. IV, we propose the Block MSBL (BMSBL) and Kalman BMSBL (KBMSBL) algorithm for pilot-based gac-sparse channel estimation for block-fading and time-varying channels, respectively.
- In Sec. V, we propose the and Joint-BMSBL (J-BMSBL) and Joint-KBMSBL (J-KBMSBL) algorithm for joint channel tracking and data detection in the block-fading and time-varying scenario, respectively.

The joint counterparts of the proposed algorithms, wherein a joint ML estimation of both the hyperparameters and the data is performed, leads to significant enhancement in the quality of channel estimates. In the M-step, this joint estimation problem separates as independent optimization problems, leading to a simple, computationally inexpensive maximization procedure, with no loss of optimality. This, in turn, leads to significant improvement in the coded Bit Error Rate (BER) performance compared to the pilot-based and conventional methods. In the context of gac-sparse channel estimation, we show that the proposed algorithms are more accurate than algorithms that do not exploit the cluster sparsity. The algorithms proposed to handle the time-varying channel conditions fully exploit the correlation structure of the channel, resulting in gains of $1-2$ dB in the coded BER performance, as illustrated using Monte Carlo simulations. Further, we propose novel, implementation-friendly structures which lead to a lower computational load for gac-sparse channels.

Notation: Boldface small letters denote vectors and boldface capital letters denote matrices. The symbols $(\cdot)^T$, $|\cdot|$ and $\mathrm{Tr}(\cdot)$ denote the transpose, determinant and the trace of a matrix, respectively. Also, $\mathrm{diag}(\mathbf{a})$ denotes a diagonal matrix with entries 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. $\mathcal{CN}(\cdot)$ denotes the complex Gaussian pdf. The expectation with respect to X is denoted as $\mathbb{E}_X(\cdot)$. The symbols $\mathbf{I}_L$ and $\mathbf{A}\otimes\mathbf{B}$ denotes an $L\times L$ identity matrix and Kronecker product of $\mathbf{A}$ and $\mathbf{B}$, respectively. The i^{th} entry of $\mathbf{a}$ and the $(i,j)^{\text{th}}$ entry of $\mathbf{A}$ 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 for Ga-sparse Channels

In this section, we propose algorithms for ga-sparse channel estimation and tracking, using the pilot-subcarriers $\mathbf{Y}_{p,k}$ in (4), in both block-fading and time-varying scenarios. First, we adapt the MSBL algorithm for block-fading channel estimation using P_b pilots placed in an equidistant manner over the time-frequency grid in a lattice structure, as prescribed by the standard [29], [30]. Subsequently, we propose the novel KMSBL algorithm for estimation and tracking of time-varying channels using P_t pilots placed in an equidistant lattice structure in every OFDM symbol.

A. The MSBL Algorithm

Here, we describe the MSBL algorithm for pilot-assisted channel estimation in MIMO-OFDM systems.

In the MSBL framework, multiple group-sparse vectors are recovered from multiple observation vectors [10] with a parameterized prior incorporated to obtain group-sparse solutions. The prior density is given by

$$p(\mathbf{H};\boldsymbol{\Gamma}) = \prod_{n_r=1}^{N_r} p(\mathbf{h}_{n_r};\boldsymbol{\Gamma}), \tag{7}$$

where $\mathbf{h}_{n_r}$ represents the n_r^{th} column of $\mathbf{H}$, given by $\mathbf{h}_{n_r} = [\mathbf{h}_{1n_r}^T,\ldots,\mathbf{h}_{N_t n_r}^T]^T$, with a prior pdf of $\mathbf{h}_{n_r} \sim \mathcal{CN}(0,\boldsymbol{\Gamma}_b)$, $\boldsymbol{\Gamma}_b = \mathbf{I}_{N_t}\otimes\boldsymbol{\Gamma}$ which control the variances of elements in $\mathbf{H}$. The hyperparameters in $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = [\gamma(1),\gamma(2),\ldots,\gamma(L)]^T$, can be estimated using the type-II ML procedure [24], i.e., by maximizing the marginalized pdf $p(\mathbf{y}_{p,n_r};\boldsymbol{\gamma})$[1] at a given receive antenna, as follows:

$$\gamma_{ML}(i) = \underset{\gamma(i)\in\mathbb{R}_+}{\arg\max}\, p(\mathbf{y}_{p,n_r};\boldsymbol{\gamma}), \quad 1\le i\le L, 1\le n_r\le N_r. \tag{8}$$

Since the above problem cannot be solved in closed form, iterative estimators such as the EM based[2] MSBL algorithm [10] are employed. In this approach, $\mathbf{H}$ is treated as the hidden variable, and the posterior distribution of $\mathbf{H}$ is obtained in the E-step and the ML estimate of $\boldsymbol{\gamma}$ is obtained in the M-step. The steps of the algorithm are given as

$$\mathrm{E}: Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}) = \mathbb{E}_{\mathbf{H}|\mathbf{Y}_p;\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{Y}_p,\mathbf{H};\boldsymbol{\gamma})] \tag{9}$$

$$\mathrm{M}: \gamma^{(r+1)}(i) = \underset{\gamma(i)\in\mathbb{R}_+}{\arg\max}\, Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}), \tag{10}$$

[1] Here, we describe the MSBL algorithm for $k=1$, and hence, we drop the subscript k in $\mathbf{y}_{p,n_r}$ and $\mathbf{Y}_p$.

[2] Note that all the algorithms proposed in the paper use EM-based updates, and hence, they have a convergence guarantee to a local optima, with the likelihood increasing in each iteration [31].

for $1 \le i \le L$, and the E and M steps are iterated until convergence. The E-step requires the posterior distribution $p(\mathbf{H}|\mathbf{Y}_p;\boldsymbol{\gamma}^{(r)})$, which can be obtained from the likelihood at the n_r^{th} receiver, as follows:

$$p(\mathbf{y}_{p,n_r}|\mathbf{h}_{n_r}) = \frac{1}{(\pi\sigma^2)^{N_r}} \exp\left(-\frac{\|\mathbf{y}_{p,n_r} - \mathbf{\Phi}_p\mathbf{h}_{n_r}\|_2^2}{\sigma^2}\right). \tag{11}$$

Combining the likelihood and the prior distribution, the posterior distribution of $\mathbf{h}_{n_r}$ is given by $p\left(\mathbf{h}_{n_r}|\mathbf{y}_{p,n_r};\boldsymbol{\gamma}^{(r)}\right) \sim \mathcal{CN}(\boldsymbol{\mu}_{n_r},\mathbf{\Sigma})$, with mean and covariance given by

$$\boldsymbol{\mu}_{n_r} = \sigma^{-2}\mathbf{\Sigma}\mathbf{\Phi}_p^H\mathbf{y}_{p,n_r},\ \mathbf{\Sigma} = \left(\frac{\mathbf{\Phi}_p^H\mathbf{\Phi}_p}{\sigma^2} + \mathbf{\Gamma}_b^{(r)^{-1}}\right)^{-1} \tag{12}$$

Here, $\mathbf{\Gamma}_b^{(r)}$ is the hyperparameter value in the r^{th} iteration and $\mathbf{\Sigma}$ is common to all receive antennas, and hence, independent of the subscript n_r.

The M-step, given by (10), can be simplified to obtain the update equation for $\boldsymbol{\gamma}$ as

$$\gamma^{(r+1)}(i) = \frac{1}{N_tN_r}\sum_{n_r=1}^{N_r}\sum_{n_t=0}^{N_t-1}\left(|\boldsymbol{\mu}_{n_r}(i+n_tL)|^2 + \mathbf{\Sigma}(i+n_tL, i+n_tL)\right). \tag{13}$$

Note that, in the above equation, the ga-sparse nature of the channel results in the update of $\boldsymbol{\gamma}$ which is *averaged* over the N_tN_r channels of the MIMO-OFDM system. For a SISO-OFDM system, $N_t = N_r = 1$, and the above expression simplifies to the one obtained in [20].

The MSBL algorithm consists of executing the E and the M steps iteratively, until the algorithm reaches convergence, i.e., the difference $\|\boldsymbol{\gamma}^{(r)} - \boldsymbol{\gamma}^{(r-1)}\|_2^2 \le \epsilon$, where ϵ is a user defined parameter. The E-step involves computing the posterior mean and variance of the ga-sparse MIMO-OFDM channel as given in (12), incurring a computational complexity given by $\mathcal{O}(P_b^2L)$ [10], while M-step computes the hyperparameter update as given in (13), incurring a computational complexity of $\mathcal{O}(N_tN_rL)$. In practice, it is found that an initial estimate for $\mathbf{\Gamma}$ given by

$$\mathbf{\Gamma}^{(0)} = \mathbf{I}_{L\times L}, \tag{14}$$

is sufficient for the MSBL algorithm.

In the case of multiple OFDM symbols in a block-fading channel, the channel remains constant for the K OFDM symbols. The system model in (4) can be used for channel estimation, such that the number of observations corresponding to pilot subcarriers is P_b.

The MSBL algorithm, in the current form, is not capable of exploiting the correlation that exists in time-varying channels across OFDM symbols. In the following subsection, we extend MSBL algorithm to obtain the recursive KMSBL algorithm which exploits the temporal correlation across OFDM symbols, resulting in a significant performance improvement when the channel is time-varying.

B. The KMSBL Algorithm

In this subsection, we describe the KMSBL algorithm which tracks the N_tN_r ga-sparse MIMO-OFDM channels by exploiting both the group-sparsity and the temporal channel correlation using a Kalman filter and smoother (KFS) based recursive framework.

In the time-varying scenario, the state space equations for $k = 1, 2, \ldots, K-1$ are as follows:

$$\mathbf{Y}_{p,k} = \mathbf{\Phi}_{p,k}\mathbf{H}_k + \mathbf{V}_{p,k}, \tag{15}$$

$$\mathbf{H}_{k+1} = \rho\mathbf{H}_k + \mathbf{U}_{k+1}, \tag{16}$$

where $\mathbf{\Phi}_{p,k} = [\mathbf{\Phi}_{p,1,k}, \ldots, \mathbf{\Phi}_{p,N_t,k}]$, and $\mathbf{\Phi}_{p,n_t,k} \in \mathbb{C}^{P_t\times L}$ is given by $\mathbf{\Phi}_{p,n_t,k} \triangleq \mathbf{X}_{p,n_t,k}\mathbf{F}_{p,n_t}$. Here, $\mathbf{X}_{p,n_t,k} \in \mathbb{C}^{P_t\times P_t}$ is a diagonal matrix consisting of pilots symbols transmitted from the n_t^{th} antenna in the k^{th} OFDM symbol, and $\mathbf{F}_{p,n_t} \in \mathbb{C}^{P_t\times L}$ is a truncated DFT matrix consisting of the first L columns and the P_t rows corresponding to the pilot subcarriers of the n_t^{th} transmit antenna. Further, $\mathbf{H}_k$ consists of the N_r channels corresponding to the k^{th} OFDM symbol, i.e., $\mathbf{H}_k = [\mathbf{h}_{1,k}, \ldots, \mathbf{h}_{N_r,k}]$ where $\mathbf{h}_{n_r,k} = [\mathbf{h}_{1n_r,k}^T, \ldots, \mathbf{h}_{N_tn_r,k}^T]^T$. In the above equation, we define $\mathbf{H}_0 \triangleq \mathbf{0}_{N_tL\times N_r}$, the $N_tL \times N_r$ matrix of zeros. Note that columns of the matrix $\mathbf{U}_{k+1}$ consists of the driving noise vectors, $\mathbf{u}_{n_r,k+1}$ which consists of independent components $\mathbf{u}_{n_r,k+1}(i) \sim \mathcal{CN}(0,(1-\rho^2)\gamma(i))$. The initial condition for the a-sparse channel is given by $\mathbf{h}_1 \sim \mathcal{CN}(0,\mathbf{\Gamma})$.

The EM update equations corresponding to the KMSBL algorithm are as follows:

$$\begin{aligned} &\text{E} : Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right) = \\ &\qquad \mathbb{E}_{\mathbf{H}_1,\ldots,\mathbf{H}_K|\mathbf{Y}_p;\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{Y}_p, \mathbf{H}_1, \ldots, \mathbf{H}_K; \boldsymbol{\gamma})] \\ &\text{M} : \boldsymbol{\gamma}^{(r+1)} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}}{\arg\max}\, Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right). \end{aligned} \tag{17}$$

In the above expression, $\mathbf{Y}_p \triangleq [\mathbf{Y}_{p,1}, \ldots, \mathbf{Y}_{p,K}]$ represents the overall observation matrix.

To compute the E-step given above, we require the posterior distribution of the unknown ga-sparse channel $\mathbf{H}_k$. For this, we employ the Kalman based recursive update equations. The KFS update equations for K OFDM symbols are as follows [20], [32], [33]:

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

Prediction: $\hat{\mathbf{H}}_{k|k-1} = \rho\hat{\mathbf{H}}_{k-1|k-1}$ (18)

$$\mathbf{P}_{k|k-1} = \rho^2\mathbf{P}_{k-1|k-1} + (1-\rho^2)\mathbf{\Gamma}_b \tag{19}$$

Filtering:

$$\mathbf{G}_k = \mathbf{P}_{k|k-1}\mathbf{\Phi}_{p,k}^H\left(\sigma^2\mathbf{I}_{P_t} + \mathbf{\Phi}_{p,k}\mathbf{P}_{k|k-1}\mathbf{\Phi}_{p,k}^H\right)^{-1} \tag{20}$$

$$\hat{\mathbf{H}}_{k|k} = \hat{\mathbf{H}}_{k|k-1} + \mathbf{G}_k(\mathbf{y}_{p,k} - \mathbf{\Phi}_{p,k}\hat{\mathbf{H}}_{k|k-1}) \tag{21}$$

$$\mathbf{P}_{k|k} = (\mathbf{I}_{N_tL} - \mathbf{G}_k\mathbf{\Phi}_{p,k})\mathbf{P}_{k|k-1} \tag{22}$$

end (23)

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

Smoothing:

$$\hat{\mathbf{H}}_{j-1|K} = \hat{\mathbf{H}}_{j-1|j-1} + \mathbf{J}_{j-1}(\hat{\mathbf{H}}_{j|K} - \hat{\mathbf{H}}_{j|j-1}) \tag{24}$$

$$\mathbf{P}_{j-1|K} = \mathbf{P}_{j-1|j-1} + \mathbf{J}_{j-1}(\mathbf{P}_{j|K} - \mathbf{P}_{j|j-1})\mathbf{J}_{j-1}^H \tag{25}$$

end (26)

where $\mathbf{J}_{j-1} \triangleq \rho\mathbf{P}_{j-1|j-1}\mathbf{P}_{j|j-1}^{-1}$ and $\mathbf{G}_k$ is the Kalman gain matrix. In the above, the symbols $\hat{\mathbf{H}}_{k|k-1}$, $\mathbf{P}_{k|k-1}$, etc. have their usual meanings as in the KF literature [33]. For example, $\hat{\mathbf{H}}_{k|k-1}$ is the channel estimate at the k^{th} OFDM symbol given the observations $\mathbf{Y}_{p,k-1}$; $\mathbf{P}_{k|k-1}$ is the covariance of the k^{th} channel estimate given $\mathbf{Y}_{p,k-1}$, etc. The above KFS equations are initialized by setting $\hat{\mathbf{H}}_{0|0} = \mathbf{0}$ and $\mathbf{P}_{0|0} = \mathbf{\Gamma}_b$, where $\mathbf{\Gamma}_b = \mathbf{\Gamma} \otimes \mathbf{I}_{N_t}$.

In order to simplify (17), we use the joint pdf of the observations $\mathbf{Y}_p$ and K channel instantiations, $\mathbf{H}_1, \ldots, \mathbf{H}_K$, given by

$$p(\mathbf{Y}_p, \mathbf{H}_1, \ldots, \mathbf{H}_K; \boldsymbol{\gamma}) = \prod_{k=1}^{K} p(\mathbf{Y}_p | \mathbf{H}_1, \ldots, \mathbf{H}_K) \, p(\mathbf{H}_k | \mathbf{H}_{k-1}; \boldsymbol{\gamma}). \quad (27)$$

Since $\mathbf{H}_k$ consists of columns $\mathbf{h}_{n_r,k}$ for $1 \leq n_r \leq N_r$, the M-step results in the following optimization problem:

$$\boldsymbol{\gamma}^{(r+1)} = \arg\max_{\boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}} \mathbb{E}_{\mathbf{H}_1,\ldots,\mathbf{H}_K | \mathbf{Y}_p; \boldsymbol{\gamma}^{(r)}} [K N_r \log |\mathbf{\Gamma}_b| + \frac{1}{(1-\rho^2)} \sum_{n_r=1}^{N_r} \sum_{k=2}^{K} \left[(\mathbf{h}_{n_r,k} - \rho \mathbf{h}_{n_r,k-1})^H \mathbf{\Gamma}_b^{-1} (\mathbf{h}_{n_r,k} - \rho \mathbf{h}_{n_r,k-1}) + \mathbf{h}_{n_r,1}^H \mathbf{\Gamma}_b^{-1} \mathbf{h}_{n_r,1} \right]] . \quad (28)$$

We see that the M-step requires the computation of $\hat{\mathbf{H}}_{j|K} \triangleq \mathbb{E}_{\mathbf{H}_1,\ldots,\mathbf{H}_K | \mathbf{Y}_p; \boldsymbol{\gamma}^{(r)}} [\mathbf{H}_j]$, and the covariance $\mathbb{E}_{\mathbf{H}_1,\ldots,\mathbf{H}_K | \mathbf{Y}_p; \boldsymbol{\gamma}^{(r)}} [\mathbf{H}_j \mathbf{H}_j^H] \triangleq \mathbf{P}_{j|K} + \hat{\mathbf{H}}_{j|K} \hat{\mathbf{H}}_{j|K}^H$ for $j = 1, \ldots, K$, which is obtained from (18)-(26). The M-step also requires the computation of $\mathbb{E}_{\mathbf{H}_1,\ldots,\mathbf{H}_K | \mathbf{Y}_p; \boldsymbol{\gamma}^{(r)}} [\mathbf{H}_j \mathbf{H}_{j-1}^H] \triangleq \mathbf{P}_{j,j-1|K} + \hat{\mathbf{H}}_{j|K} \hat{\mathbf{H}}_{j-1|K}^H$ for $j = K, K-1, \ldots, 2$, which we obtain from [33] as follows:

$$\mathbf{P}_{j-1,j-2|K} = \mathbf{P}_{j-1|j-1} \mathbf{J}_{j-2}^H + \mathbf{J}_{j-1}^H (\mathbf{P}_{j,j-1|K} - \rho \mathbf{P}_{j-1|j-1}) \mathbf{J}_{j-2}. \quad (29)$$

The above recursion is initialized using $\mathbf{P}_{K,K-1|k} = \rho(\mathbf{I}_{N_t L} - \mathbf{G}_K \mathbf{\Phi}_{p,K}) \mathbf{P}_{K-1|K-1}$. Using the above expressions, the optimization problem in (28) can be written as

$$\boldsymbol{\gamma}^{(r+1)} = \arg\min_{\boldsymbol{\gamma} \in \mathbb{R}_+^{L \times 1}} \left\{ K N_t N_r \log |\mathbf{\Gamma}| + \sum_{n_t=1}^{N_t} \mathrm{Tr}(\mathbf{\Gamma}^{-1} \mathbf{M}_{n_t,1|K}) + \tfrac{1}{(1-\rho^2)} \sum_{j=2}^{K} \sum_{n_t=1}^{N_t} \mathrm{Tr}(\mathbf{\Gamma}^{-1} \mathbf{M}_{n_t,j|K}) \right\}, \quad (30)$$

where $\mathbf{M}_{n_t,1|K} \in \mathbb{C}^{L \times L}$ is the submatrix consisting of rows and columns $(n_t - 1)L$ through $n_t L$ from the matrix $\mathbf{M}_{1|K} \triangleq N_r \mathbf{P}_{j|K} + \hat{\mathbf{H}}_{j|K} \hat{\mathbf{H}}_{j|K}^H + \rho^2 (N_r \mathbf{P}_{j-1|K} + \hat{\mathbf{H}}_{j-1|K} \hat{\mathbf{H}}_{j-1|K}^H) - 2\rho \mathrm{Re}(N_r \mathbf{P}_{j,j-1|K} + \hat{\mathbf{H}}_{j|K} \hat{\mathbf{H}}_{j-1|K}^H)$. Similarly, $\mathbf{M}_{n_t,j|K} \in \mathbb{C}^{L \times L}$ is the submatrix of $\mathbf{M}_{j|K} \triangleq N_r \mathbf{P}_{1|K} + \hat{\mathbf{H}}_{1|K} \hat{\mathbf{H}}_{1|K}^H$, consisting of rows and columns $(n_t - 1)L$ through $n_t L$. Since the individual channel components of $\mathbf{h}_{n_r,k}$ given by $\mathbf{h}_{n_t n_r,k}$ for $1 \leq n_t \leq N_t$ are governed by $\boldsymbol{\gamma}$, we note that the update of $\boldsymbol{\gamma}$ is averaged over the N_t components via the summation over n_t. Differentiating (30) w.r.t. $\gamma(i)$ and setting the resulting expression to zero and solving for $\boldsymbol{\gamma}$ gives the update for the i^{th} hyperparameter as follows:

$$\gamma^{(r+1)}(i) = \left[\tfrac{1}{K N_t N_r} \left(\sum_{j=2}^{K} \sum_{n_t=1}^{N_t} \frac{M_{n_t,j|K}(i,i)}{(1-\rho^2)} + M_{n_t,1|K}(i,i) \right) \right]^+, \quad (31)$$

for $i = 1, \ldots, L$. Thus the KMSBL algorithm learns $\boldsymbol{\gamma}$ in the M-step and provides low-complexity and recursive estimates of the ga-sparse channel in the E-step.

Remarks: When $\rho = 1$, $\mathbf{H}_1 = \ldots = \mathbf{H}_K$ and hence, the channel is constant across the OFDM frame, i.e., the channel is block-fading. The results from Sec. II-A demonstrate that P_b pilots per OFDM symbol are sufficient for recovering $\mathbf{H}$ from $\mathbf{Y}_p$. Substituting $\rho = 1$ in (18)-(26), the KFS update equations collapse to the following three equations:

$$\mathbf{G}_k = \mathbf{P}_{k-1|k-1} \mathbf{\Phi}_{p,k}^H (\sigma^2 \mathbf{I}_{P_t} + \mathbf{\Phi}_{p,k} \mathbf{P}_{k-1|k-1} \mathbf{\Phi}_{p,k}^H)^{-1} \quad (32)$$

$$\hat{\mathbf{H}}_{k|k} = \hat{\mathbf{H}}_{k-1|k-1} + \mathbf{G}_k (\mathbf{Y}_{p,k} - \mathbf{\Phi}_{p,k} \hat{\mathbf{H}}_{k-1|k-1}) \quad (33)$$

$$\mathbf{P}_{k|k} = (\mathbf{I}_{N_t L} - \mathbf{G}_k \mathbf{\Phi}_{p,k}) \mathbf{P}_{k-1|k-1}. \quad (34)$$

Further, when $\rho = 1$, the M-step of (28) simplifies to the M-step of MSBL given in (10).

The KMSBL algorithm proposed in this section is a generalized version of the KSBL algorithm proposed in [20] for pilot-based SISO-OFDM channel estimation, i.e., setting $N_t = N_r = 1$ in the KMSBL algorithm leads to the KSBL algorithm. However, in contrast to the KSBL algorithm, the KMSBL algorithm incorporates the spatial sparsity that exists in the MIMO-OFDM framework, and tracks N_r correlated channel vectors governed by a common $\boldsymbol{\gamma}$.

In order to estimate the wireless channel when the data is observed up to the K^{th} OFDM symbol, (18)-(22) are applied recursively until we reach the K^{th} OFDM symbol 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 $j = 0, \ldots, K$ in the forward recursion. Next, we apply the backward recursion using the Kalman smoother given by (24)-(26), 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.

Using a flop-count analysis [34], the computations of the KMSBL algorithm is 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 per receive antenna. We see that if the number of OFDM symbols to be tracked are such that $KP_t > L$, the complexity of the block-based ARSBL algorithm [19] is larger than the KMSBL algorithm. In other words, the KMSBL algorithm is a good choice among the exact inference techniques when the number of OFDM symbols to be tracked is large [20].

The algorithms proposed in this section do not utilize the information available from the data subcarriers in estimating the channel. In the following section, we propose joint channel estimation and data detection schemes for ga-sparse channels.

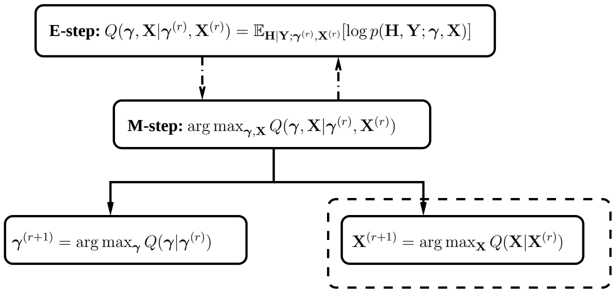

Figure 3. The J-MSBL algorithm: E-step computes the expectation over the posterior density of $\mathbf{H}$. The joint maximization in the M-step simplifies into two independent maximizations over $\boldsymbol{\gamma}$ and $\mathbf{X}$. The dashed box indicates the novelty in the J-MSBL approach.

III. Joint Channel Estimation/Tracking and Data Detection Using Pilot and Data Subcarriers for Ga-sparse Channels

In this section, we present the novel J-MSBL and J-KMSBL algorithm that generalizes the pilot-based MSBL and KMSBL algorithms for joint ga-sparse channel estimation and data detection in MIMO-OFDM systems. Further, using the recursive J-KMSBL algorithm, we show that a low-complexity recursive variant of J-MSBL can be derived using the KFS update equations given in (18)-(26).

A. The J-MSBL Algorithm

In this subsection, we derive the J-MSBL algorithm for joint estimation of the ga-sparse channels and the transmit data in a MIMO-OFDM system. To derive this algorithm, we modify the MSBL framework such that the unknown variables are not only the hyperparameters but also the data transmitted in K OFDM symbols.

We consider $\mathbf{H}$ in (3) as the hidden variable, and, in contrast to the MSBL setup, we consider $\left(\boldsymbol{\gamma}, \mathbf{X} \triangleq [\mathbf{X}_{11}, \ldots, \mathbf{X}_{n_t k}, \ldots, \mathbf{X}_{N_t K}]\right)$ as parameters to be estimated. Here, $\mathbf{X}_{n_t k}$ consists of the data corresponding to the n_t^{th} antenna in the k^{th} OFDM symbol. The E and the M-steps of the J-MSBL algorithm can be given as

$$\begin{aligned}
&\text{E} : Q(\boldsymbol{\gamma}, \mathbf{X}|\boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}) = \mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)}}[\log p(\mathbf{Y}, \mathbf{H}; \boldsymbol{\gamma}, \mathbf{X})] \\
&\text{M} : \left(\boldsymbol{\gamma}^{(r+1)}, \mathbf{X}^{(r+1)}\right) = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1}, \mathbf{X}: x_i \in \mathcal{S}}{\arg\max} Q(\boldsymbol{\gamma}, \mathbf{X}|\boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}),
\end{aligned} \tag{35}$$

where x_i is an element in $\mathbf{X}$, and $\mathcal{S}$ is the constellation from which the symbols are transmitted. The E-step of J-MSBL consists of computing the posterior distribution at every receive antenna, and is given as $p(\mathbf{h}_{n_r}|\mathbf{y}_{n_r}; \boldsymbol{\gamma}^{(r)}, \mathbf{X}^{(r)}) \sim \mathcal{CN}(\boldsymbol{\mu}_{n_r}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu}_{n_r} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}_b^H \mathbf{y}_{n_r k}, \;\; \boldsymbol{\Sigma} = \left(\sigma^{-2}\boldsymbol{\Phi}_b^H\boldsymbol{\Phi}_b + \boldsymbol{\Gamma}^{(r)^{-1}}\right)^{-1}, \tag{36}$$

for K OFDM symbols in a frame. In the above equation, $\boldsymbol{\Phi}_b = [\boldsymbol{\Phi}_1^T, \ldots, \boldsymbol{\Phi}_K^T]^T$, and for $1 \leq k \leq K$, $\mathbf{F}_b = \mathbf{1}_{N_t} \otimes \mathbf{F}$, $\boldsymbol{\Phi}_k = \mathbf{F}_b\text{blkdiag}(\mathbf{X}_{1k}^{(r)}, \ldots, \mathbf{X}_{N_t k}^{(r)})$ and $\mathbf{y}_{n_r,k} = [\mathbf{y}_{1,k}^T, \ldots, \mathbf{y}_{N_r,k}^T]^T$.

At the outset, solving the optimization problem in the M-step in (35) might appear to be an uphill task, as it involves joint optimization over $\mathbf{X}$ and $\boldsymbol{\gamma}$. However, we see that, in (35), the objective function w.r.t. $\boldsymbol{\gamma}$ and $\mathbf{X}$ can be decoupled as the sum of two independent terms, $Q(\mathbf{X}|\mathbf{X}^{(r)}) \triangleq \mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)},\mathbf{X}^{(r)}}[\log p(\mathbf{Y}|\mathbf{H}; \mathbf{X})]$ and $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}) \triangleq \mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)},\mathbf{X}^{(r)}}[\log p(\mathbf{H}; \boldsymbol{\gamma})]$. This is schematically illustrated in Fig. 3.[3] Further, we see that $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)})$ of the MSBL algorithm and the J-MSBL algorithm are identical, and hence, upon optimizing $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)})$ with respect to $\gamma(i)$, we obtain the expression for $\gamma^{(r+1)}(i)$ as in the MSBL algorithm, given by (10). Further, the objective function to obtain $\mathbf{X}$, i.e., $Q(\mathbf{X}|\mathbf{X}^{(r)})$, can be derived as follows:

$$\begin{aligned}
Q(\mathbf{X}|\mathbf{X}^{(r)}) &= \mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)},\mathbf{X}^{(r)}}\left[\log \prod_{n_r=1}^{N_r} p(\mathbf{y}_{n_r,k}|\mathbf{h}_{n_r}; \mathbf{X})\right] \\
&= -\mathbb{E}_{\mathbf{H}|\mathbf{Y};\boldsymbol{\gamma}^{(r)},\mathbf{X}^{(r)}}\left[\sum_{n_r=1}^{N_r} \|\mathbf{y}_{n_r,k} - \boldsymbol{\Phi}_b\mathbf{h}_{n_r}\|_2^2\right].
\end{aligned} \tag{37}$$

and hence, the optimization problem for $\mathbf{X}$ is given by

$$\begin{aligned}
&X_{11}^{(r+1)}(i,i), \ldots, X_{N_t K}^{(r+1)}(i,i) = \underset{x_1,\ldots,x_{N_t}\in\mathcal{S}}{\arg\min} C(i,i) \\
&+ \sum_{n_r=1}^{N_r} |y_{n_r,k}(i) - \sum_{n_t=1}^{N_t} x_{n_t,k}\mathbf{F}_b(i,:)\boldsymbol{\mu}_{n_r}|^2,
\end{aligned} \tag{38}$$

where $i \in \mathcal{D}$, $\mathcal{D}$ is an index set consisting of the data subcarrier locations, $\mathbf{C} = \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^H$, $\mathbf{F}_b(i,:)$ is the i^{th} row of the $\mathbf{F}_b$ matrix, $\boldsymbol{\mu}_{n_r}$ and $\boldsymbol{\Sigma}$ are given in (36). The computational complexity of this algorithm is dominated by the inverse operation in (36), and is $\mathcal{O}(K^2N^2LN_t)$.

As stated in the previous section, the initial estimate of $\boldsymbol{\Gamma}$ is taken to be an identity matrix. The initialization of the $(KNN_t - P_bN_t)$ non-pilot data in turn requires an initial channel estimate. Channel estimates using methods like LS and MMSE cannot be used, as they require knowledge of the MIP. Hence, the initialization of $\mathbf{X}$ is set to be the channel estimate obtained from a few iterations of the MSBL algorithm from the P_b pilots (denoted as $\hat{\mathbf{h}}_{MSBL}$). The ML data detection problem for obtaining the initial data estimates is given by

$$\begin{aligned}
&X_1^{(0)}(i,i), \ldots, X_{N_t}^{(0)}(i,i) = \\
&\underset{x_1,\ldots,x_{N_t}\in\mathcal{S}}{\arg\min} |y_{n_r,k}(i) - \sum_{n_t=1}^{N_t} x_{n_t}\mathbf{F}_b(i,:)\hat{\mathbf{h}}_{MSBL}|^2, \quad i \in \mathcal{D}.
\end{aligned} \tag{39}$$

In order to obtain the solution for both (38) and (39), we need to find the vector $[x_1, \ldots, x_{N_t}]$ that jointly minimizes (38). Although we can solve this problem with moderate complexity for MIMO-OFDM systems with N_t up to 4 [29], the complexity of this problem is high for larger N_t. In such scenarios, one can use sphere decoding [11].

[3] Notice that (10) and (35) are different, since the former uses the measurement matrix containing only the known pilot symbols, $\boldsymbol{\Phi}_p$, whereas the latter uses measurement matrices which consist of pilot symbols along with the estimated data, together given by $\boldsymbol{\Phi}^{(r)}$.

In the following section, we discuss the pilot-based and joint channel estimation and data detection for time-varying ga-sparse MIMO-OFDM channels.

B. The J-KMSBL Algorithm

In this section, we generalize the KMSBL algorithm of Sec. II-B to obtain the J-KMSBL algorithm, which utilizes the observations available at all the N subcarriers of the K OFDM symbols, and performs data detection at the $(N-P_t)$ data subcarriers of each OFDM symbol. Generalizing the J-MSBL to the J-KMSBL algorithm involves incorporating an E-step that exploits the correlation in the time-varying channels such that the algorithm is recursive in nature, and the smoothed channel estimates obtained for the K OFDM symbols are used to jointly estimate the ga-sparse channel and the unknown data of the K OFDM symbols.

Our starting point, again, is the state space model given by (16). The EM update equations in this context are given by

$$\begin{aligned}
&\text{E}: Q(\boldsymbol{\gamma},\mathbf{X}|\boldsymbol{\gamma}^{(r)},\mathbf{X}^{(r)}) = \mathbb{E}_{\mathbf{H}_1,\ldots,\mathbf{H}_K|\mathbf{Y};\boldsymbol{\gamma}^{(r)}} \\
&\qquad [\log p(\mathbf{Y},\mathbf{H}_1,\ldots,\mathbf{H}_K;\boldsymbol{\gamma},\mathbf{X})] \\
&\text{M}: \left(\boldsymbol{\gamma}^{(r+1)},\mathbf{X}^{(r+1)}\right) = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^{L\times 1},\mathbf{X}:x_i\in\mathcal{S}}{\arg\max}\ Q(\boldsymbol{\gamma},\mathbf{X}|\boldsymbol{\gamma}^{(r)},\mathbf{X}^{(r)}),
\end{aligned} \tag{40}$$

where $\mathbf{X}$ comprises the data transmitted on the K OFDM symbols, as defined in the previous subsection. Since the J-KMSBL algorithm uses the observations available at all the N subcarriers of each OFDM symbol, recursive updates of the posterior mean and covariance are given by (18)-(26), with $\mathbf{Y}_p$ and $\boldsymbol{\Phi}_p$ replaced by $\mathbf{Y}$ and $\boldsymbol{\Phi}$, respectively. Further, since $\boldsymbol{\Gamma}$ and data at the non-pilot subcarriers are unknown, the SBL framework leads to the objective function for K OFDM symbols in the M-step given by

$$\begin{aligned}
&Q\left(\mathbf{X},\boldsymbol{\gamma}|\mathbf{X}^{(r)},\boldsymbol{\gamma}^{(r)}\right) = c - KN_r\log|\boldsymbol{\Gamma}_b| - \mathbb{E}_{\mathbf{H}_1,\ldots,\mathbf{H}_K|\mathbf{Y};\mathbf{X},\boldsymbol{\gamma}^{(r)}} \\
&\Bigg[\sum_{j=1}^{K}\sum_{n_r=1}^{N_r}\sigma^{-2}\|\mathbf{y}_{n_r,j}-\sum_{n_t=1}^{N_t}\mathbf{X}_{n_t,j}\mathbf{F}\mathbf{h}_{n_tn_r,j}\|^2 \\
&-\sum_{j=2}^{K}\sum_{n_r=1}^{N_r}\frac{(\mathbf{h}_{n_r,j}-\rho\mathbf{h}_{n_r,j-1})^H\boldsymbol{\Gamma}_b^{-1}(\mathbf{h}_{n_r,j}-\rho\mathbf{h}_{n_r,j-1})}{(1-\rho^2)} \\
&-\mathbf{h}_{n_r,1}^H\boldsymbol{\Gamma}_b^{-1}\mathbf{h}_{n_r,1}\Bigg],
\end{aligned} \tag{41}$$

where c is a constant independent of $\boldsymbol{\gamma}$ and $\mathbf{X}$. The expression above is a sum of terms which are independent functions of $\boldsymbol{\gamma}$ and $\mathbf{X}_k \triangleq [\mathbf{X}_{1,k},\ldots,\mathbf{X}_{N_t,k}]$ for $1\le k\le K$, denoted as $Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right)$ and $Q\left(\mathbf{X}_k|\mathbf{X}_k^{(r)}\right), 1\le k\le K$, respectively. Further, we see that $Q\left(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(r)}\right)$ is the same as (30). Hence, the learning rule for $\boldsymbol{\gamma}$ follows from the M-step of the KMSBL algorithm, and is given by (31). The expression for $Q\left(\mathbf{X}_k|\mathbf{X}_k^{(r)}\right)$ is given by

$$\begin{aligned}
&Q\left(\mathbf{X}_k|\mathbf{X}_k^{(r)}\right) = \mathbb{E}_{\mathbf{H}_k|\mathbf{Y};\mathbf{X}^{(r)},\boldsymbol{\gamma}^{(r)}} \\
&\left[c-\sum_{n_r=1}^{N_r}\sigma^{-2}\|\mathbf{y}_{n_r,k}-\sum_{n_t=1}^{N_t}\mathbf{X}_{n_t,k}\mathbf{F}\mathbf{h}_{n_tn_r,k}\|^2\right].
\end{aligned} \tag{42}$$

The M-step requires $\hat{\mathbf{H}}_{k|K} \triangleq \mathbb{E}_{\mathbf{H}_k|\mathbf{Y};\mathbf{X}^{(r)},\boldsymbol{\gamma}^{(r)}}[\mathbf{H}_k]$ and $\mathbf{P}_{k|K} \triangleq \mathbb{E}_{\mathbf{H}_k|\mathbf{Y};\mathbf{X}^{(r)},\boldsymbol{\gamma}^{(r)}}[\mathbf{H}_k\mathbf{H}_k^H]$, which are given by the KFS equations of the E-step. The maximization of $Q\left(\mathbf{X}_k|\mathbf{X}_k^{(r)}\right)$ in (42) leads to the following optimization problem for $\mathbf{X}_k$:

$$\begin{aligned}
&X_{1,k}^{(r+1)}(i,i),\ldots,X_{N_t,k}^{(r+1)}(i,i) = \underset{x_1,\ldots,x_{N_t}\in\mathcal{S}}{\arg\min}\ C(i,i) \\
&+\sum_{n_r=1}^{N_r}|y_{n_r}(i)-\sum_{n_t=1}^{N_t}x_{n_t}\mathbf{F}(i,:)\hat{\mathbf{h}}_{n_r,k|K}|^2,
\end{aligned} \tag{43}$$

where $i\in\mathcal{D}$, $\mathcal{D}$ is an index set consisting of the data subcarrier locations, $\mathbf{C}=\boldsymbol{\Phi}\mathbf{P}_{k|K}\boldsymbol{\Phi}^H$, $\mathbf{F}(i,:)$ is the i^{th} row of $\mathbf{F}$ and $\hat{\mathbf{h}}_{n_r,k|K}$ is the n_r^{th} column of $\hat{\mathbf{H}}_{k|K}$. Note that, in contrast to the expression for $\mathbf{C}$ in (38), the above expression is a function of $\mathbf{P}_{k|k}$ since the covariance is computed recursively.

Data detection in the M-step results in the measurement matrix $\boldsymbol{\Phi}_k^{(r)}$ in the r^{th} iteration and k^{th} OFDM symbol. Hence, the iterations of the J-KMSBL are comprised of KFS update equations that incorporate $\boldsymbol{\Phi}_k^{(r)}$ instead of the pilot-only $\boldsymbol{\Phi}_{p,k}$ used in the KSBL algorithm. Further, the data detection in the M-step necessitates the initialization of transmit data, $\mathbf{X}_k^{(0)}$ for $0\le k\le K$. We use the channel estimate obtained from a few iterations of the KMSBL algorithm from the P_t pilots (denoted as $\hat{\mathbf{h}}_{KMSBL}$) to obtain the initial estimate $\mathbf{X}_k^{(0)}$ for $0\le k\le K$ and $i\in\mathcal{D}$ as

$$\begin{aligned}
&X_{1,k}^{(0)}(i,i),\ldots,X_{N_t,k}^{(0)}(i,i) = \\
&\underset{x_1,\ldots,x_{N_t}\in\mathcal{S}}{\arg\min}\ |y_{n_r}(i)-\sum_{n_t=1}^{N_t}x_{n_t}\mathbf{F}(i,:)\hat{\mathbf{h}}_{KMSBL}|^2.
\end{aligned} \tag{44}$$

As mentioned in Sec. II-B, when $\rho=1$, the channel is block-fading in nature. Employing P_b pilots in an OFDM frame, we can emulate the block-fading scenario described in Sec. III-A, and hence implement the J-MSBL algorithm recursively using KFS equations given by (34). Further, the M-step of the J-KMSBL algorithm is given by (31) and (44).

Until now, we focussed on recovering the block-fading and time-varying ga-sparse channels using pilot-only and joint techniques. In the sequel, we design pilot-only and joint channel estimation and data detection algorithms for group approximately *cluster-sparse* block-fading and time-varying channels.

IV. Channel Estimation and Tracking Using Pilot Subcarriers for Gac-sparse Channels

In this section, we model the channel as gac-sparse, i.e., the entries of the approximately sparse channel are constrained to lie in a few clusters. Each cluster of the gac-sparse channel $\mathbf{h}_{n_tn_r}$ with B blocks of length M each, consists of all strong or all weak components and the strong component clusters are few in number. The parametric prior modeling in SBL can be extended to the gac-sparse channels by assigning a hyperparameter $\gamma_c(i)$ to the i^{th} cluster, $1\le i\le B$, instead of the i^{th} component, as given in Sec. II-A. That is, the B length hyperparameter vector $\boldsymbol{\gamma}_c$ is associated with the pdf of $\mathbf{h}_{n_tn_r}$, such that every M length cluster of the channel is distributed as $\mathcal{CN}(0,\gamma_c(i))$.

First, we propose the Block MSBL (BMSBL) for pilot-based gac-sparse block-fading channel estimation in a MIMO-OFDM framework. We implement the BMSBL algorithm using the parallel cluster MSBL (PCMSBL) approach [35], which is same as the BMSBL in performance but has the advantage of lower computational complexity as it allows for the parallel implementation of the algorithm. Thereafter, we propose the Kalman-BMSBL (KBMSBL) algorithm for pilot-based gac-sparse time-varying channel estimation, and propose to implement the algorithm using the low-complexity Nested MSBL (NMSBL) approach [35].

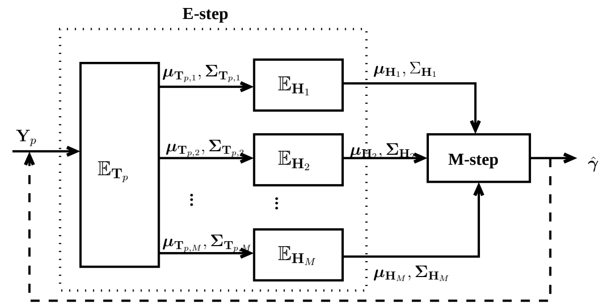

Figure 4. Block Diagram of the PCMSBL algorithm depicting M parallel branches.

A. The BMSBL Algorithm

In this subsection, we propose the Block MSBL (BMSBL) algorithm for pilot-based sparse channel estimation in block-fading channels. We propose to recover the gac-sparse channel, $\mathbf{h}_{n_t n_r}$, by generalizing the BSBL algorithm [25] to the multiple measurement scenario, i.e., we recover the N_r ga-sparse channels, $\mathbf{h}_{n_r,m}$ from N_r observation vectors, $\mathbf{y}_{n_r,m}$. Note that setting $N_r = N_t = 1$ leads to the SISO-OFDM problem, making the proposed algorithm backward compatible for the SISO-OFDM gac-sparse channel estimation.

The EM algorithm for obtaining the ML estimate of the unknown parameter $\boldsymbol{\gamma}_c$, in the BMSBL framework is as follows:

$$\begin{aligned} &\text{E} : Q\left(\boldsymbol{\gamma}_c|\boldsymbol{\gamma}_c^{(r)}\right) = \mathbb{E}_{\mathbf{H}|\mathbf{Y}_p;\boldsymbol{\gamma}_c^{(r)}}\left[\log p(\mathbf{Y}_p, \mathbf{H}; \boldsymbol{\gamma}_c)\right] \\ &\text{M} : \boldsymbol{\gamma}_c^{(r+1)} = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B\times 1}}{\arg\max}\, Q\left(\boldsymbol{\gamma}_c|\boldsymbol{\gamma}_c^{(r)}\right). \end{aligned} \tag{45}$$

The posterior distribution in the E-step above can be derived as $p(\mathbf{h}_{n_r}|\mathbf{y}_{p,n_r}; \boldsymbol{\gamma}_c^{(r)}) \sim \mathcal{CN}(\boldsymbol{\mu}_{c,n_r}, \boldsymbol{\Sigma}_c)$, where

$$\begin{aligned} &\boldsymbol{\mu}_{c,n_r} = \boldsymbol{\Sigma}_c \boldsymbol{\Phi}_p^H \mathbf{y}_{p,n_r}, \\ &\qquad \boldsymbol{\Sigma}_c = \sigma^{-2}\left(\frac{\boldsymbol{\Phi}_p^H \boldsymbol{\Phi}_p}{\sigma^2} + (\boldsymbol{\Gamma}_c \otimes \mathbf{I}_M)^{-1}\right)^{-1}. \end{aligned} \tag{46}$$

Observe that the MSBL Sec. II-A and the BMSBL algorithms differ in the prior distribution of $\mathbf{H}$. The logarithm of the pdf of the gac-sparse channel $\mathbf{H}$ is given by

$$\begin{aligned} &\log p(\mathbf{H}; \boldsymbol{\gamma}_c) = c' - \\ &N_t N_r \log|(\boldsymbol{\Gamma}_c \otimes \mathbf{I}_M)| - \sum_{n_t=1}^{N_t}\sum_{n_r=1}^{N_r} \mathbf{h}_{n_t n_r}^H (\boldsymbol{\Gamma}_c \otimes \mathbf{I}_M)^{-1} \mathbf{h}_{n_t n_r}, \end{aligned} \tag{47}$$

where c' is a constant independent of $\boldsymbol{\gamma}_c$. Maximizing $Q\left(\boldsymbol{\gamma}_c|\boldsymbol{\gamma}_c^{(r)}\right)$ in (45) w.r.t. $\boldsymbol{\gamma}_c$, we obtain the following

$$\begin{aligned} &\boldsymbol{\gamma}_c^{(r+1)}(i) = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+}{\arg\min}\, M N_t N_r \log|\boldsymbol{\Gamma}_c| \\ &+ \mathbb{E}_{\mathbf{H}|\mathbf{Y}_p;\boldsymbol{\gamma}_c^{(r)}} \sum_{n_r=1}^{N_r}\left[\sum_{n_t=1}^{N_t} (\boldsymbol{\Gamma}_c \otimes \mathbf{I}_M)^{-1} \mathrm{Tr}[\mathbf{h}_{n_t n_r}\mathbf{h}_{n_t n_r}^H]\right]. \end{aligned} \tag{48}$$

Simplifying the above, we obtain

$$\begin{aligned} &\boldsymbol{\gamma}_c^{(r+1)}(i) = \frac{1}{M N_t N_r}\sum_{m=1}^{M}\sum_{n_r=1}^{N_r}\sum_{n_t=1}^{N_t} \boldsymbol{\Sigma}_{c,n_t n_r}(m,m) \\ &+ |\boldsymbol{\mu}_{c,n_t n_r}(m)|^2. \end{aligned} \tag{49}$$

Note that, in contrast to (13), we obtain the averaging over the size of the cluster, since $\boldsymbol{\gamma}_c^{(r+1)}(i)$ is common to the entries of the cluster. Further, since the vectors are gac-sparse over N_t transmit and N_r receive antenna, we obtain the update, $\boldsymbol{\gamma}_c^{(r+1)}(i)$ which is averaged over $N_t N_r$ channels of the MIMO-OFDM system.

Implementation of BMSBL: Here, we discuss the implementation of the BMSBL algorithm. We employ the PCSBL approach [35], which significantly reduces the complexity of the proposed BMSBL algorithm.

The complexity of the BMSBL algorithm is dominated by the computation of the posterior covariance matrix $\boldsymbol{\Sigma}_c$, which incurs a computational load of $\mathcal{O}(N^2 M B)$. In [35], we proposed an approach for estimating cluster-sparse signals and showed that by using a Parallel Cluster SBL approach, the block-based algorithm [25] is amenable to parallel implementation.

We employ the PCSBL approach to handle multiple measurements, as depicted in Fig. 4, where the gac-sparse channel is recovered by solving M parallel problems. The M-step is simply the average of the hyperparameter updates obtained from the M parallel problems per receive antenna. The multiple measurement PCSBL incurs a maximum computational load of $\mathcal{O}(P_b^3)$, i.e., the complexity does not scale with $L = MB$.

The BMSBL algorithm is designed for block-fading channels, and hence, is not capable of exploiting the correlation seen in time-varying channels. Hence, in the following subsection, we design a recursive KBMSBL algorithm for time-varying gac-sparse channel estimation in order to exploit the temporal correlation.

B. The KBMSBL Algorithm

In this subsection, we derive an algorithm for tracking the *slowly time-varying* gac-sparse MIMO-OFDM channel using the SBL framework. As in Sec. II-B, we employ an AR model for the temporal evolution of the gac-sparse channel and derive recursive KFS based techniques. In addition, we propose a nested SBL approach [35] which facilitates the implementation of the proposed algorithm using M parallel Kalman filters/smoothers.

We formulate the gac-sparse channel estimation problem by modeling the channel corresponding to the k^{th} OFDM symbol as $\mathbf{h}_{n_r,k} \sim \mathcal{CN}(0, (\mathbf{\Gamma}_c \otimes \mathbf{I}_M))$. We model the temporal variation of the gac-sparse channel using the first order AR model as given in (16), i.e., the temporal evolution of every cluster in the gac-sparse channel follows a first order AR model given by

$$\mathbf{h}_{n_t n_r, mk} = \rho \mathbf{h}_{n_t n_r, mk-1} + \mathbf{u}_{n_t n_r, mk}, \tag{50}$$

for $1 \leq m \leq M$, where $\mathbf{u}_{n_t n_r, mk}$ is temporally and spatially white, i.e., $\mathbf{u}_{n_t n_r, mk} \sim \mathcal{CN}(0, (1-\rho^2)\gamma(m)\mathbf{I}_M)$. The EM algorithm for pilot-based gac-sparse channel estimation is given by

$$\begin{aligned} &\text{E} : Q\left(\boldsymbol{\gamma}_c | \boldsymbol{\gamma}_c^{(r)}\right) \\ &= \mathbb{E}_{\mathbf{H}_1,\ldots,\mathbf{H}_K|\mathbf{Y}_p;\boldsymbol{\gamma}_c^{(r)}}[\log p(\mathbf{Y}_p, \mathbf{H}_1, \ldots, \mathbf{H}_K; \boldsymbol{\gamma}_c)] \\ &\text{M} : \boldsymbol{\gamma}_c^{(r+1)} = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B\times 1}}{\arg\max}\, Q\left(\boldsymbol{\gamma}_c | \boldsymbol{\gamma}_c^{(r)}\right). \end{aligned} \tag{51}$$

Now, the posterior distribution of $\mathbf{H}_1, \ldots, \mathbf{H}_K$ can be efficiently evaluated using the Kalman Filter and Smoother (KFS) equations given in (18) - (26), by replacing $\mathbf{\Gamma}_b$ by $\mathbf{\Gamma}_{cb} \triangleq (\mathbf{I}_{N_t} \otimes (\mathbf{\Gamma}_c \otimes \mathbf{I}_M))$.

The logarithm of the conditional prior distribution is given by

$$\begin{aligned} &\log p(\mathbf{H}_k | \mathbf{H}_{k-1}; \mathbf{\Gamma}_c) = K N_r \log |\mathbf{\Gamma}_{cb}| \\ &- \frac{\sum_{n_r=1}^{N_r} \sum_{k=2}^{K} (\mathbf{h}_{n_r,k} - \rho \mathbf{h}_{n_r,k-1})^H \mathbf{\Gamma}_{cb}^{-1} (\mathbf{h}_{n_r,k} - \rho \mathbf{h}_{n_r,k-1})}{(1-\rho^2)} \\ &- \sum_{n_r=1}^{N_r} \mathbf{h}_{n_r,1}^H \mathbf{\Gamma}_{cb}^{-1} \mathbf{h}_{n_r,1}, \end{aligned} \tag{52}$$

Note that, in the above expression, $\mathbf{\Gamma}_{cb}^{-1}$ imposes the gac-sparse structure on the channel for K OFDM symbols. The M-step of KBMSBL can be simplified as follows:

$$\begin{aligned} &\boldsymbol{\gamma}_c^{(r+1)} = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B\times 1}}{\arg\min}\, \mathbb{E}_{\mathbf{H}_1,\ldots,\mathbf{H}_K|\mathbf{Y}_p;\mathbf{\Gamma}_c^{(r)}}[K N_r \log |\mathbf{\Gamma}_{cb}| \\ &+ \sum_{k=2}^{K} \sum_{n_r=1}^{N_r} \frac{(\mathbf{h}_{n_r,k} - \rho \mathbf{h}_{n_r,k-1})^H \mathbf{\Gamma}_{cb}^{-1} (\mathbf{h}_{n_r,k} - \rho \mathbf{h}_{n_r,k-1})}{(1-\rho^2)} \\ &+ \sum_{n_r=1}^{N_r} \mathbf{h}_{n_r,1}^H \mathbf{\Gamma}_{cb}^{-1} \mathbf{h}_{n_r,1}]. \end{aligned} \tag{53}$$

Using the prior distribution given in (52), and invoking the fact that $\mathbf{\Gamma}_{cb} = (\mathbf{I}_{N_t} \otimes (\mathbf{\Gamma}_c \otimes \mathbf{I}_M))$, we can simplify (53) as

$$\begin{aligned} &\boldsymbol{\gamma}_c^{(r+1)} = \underset{\boldsymbol{\gamma}_c \in \mathbb{R}_+^{B\times 1}}{\arg\min}\, K M N_r N_t \log |\mathbf{\Gamma}_c| + \sum_{n_t=1}^{N_t} \text{Tr}(\mathbf{\Gamma}_{cb}^{-1} \mathbf{M}_{cn_t,1|K}) \\ &+ \tfrac{1}{(1-\rho^2)} \sum_{k=2}^{K} \sum_{n_t=1}^{N_t} \text{Tr}(\mathbf{\Gamma}_{cb}^{-1} \mathbf{M}_{cn_t,k|K}), \end{aligned} \tag{54}$$

where $\mathbf{M}_{cn_t,j|K}$ consists of rows and columns $(n_t-1)L$ through $n_t L$ from the matrix $\mathbf{M}_{c,j|K} \triangleq N_r \mathbf{P}_{j|K} + \hat{\mathbf{H}}_{j|K}\hat{\mathbf{H}}_{j|K}^H + \rho^2(N_r \mathbf{P}_{j-1|K} + \hat{\mathbf{H}}_{j-1|K}\hat{\mathbf{H}}_{j-1|K}^H) - 2\rho\text{Re}(N_r \mathbf{P}_{j,j-1|K} + \hat{\mathbf{H}}_{j|K}\hat{\mathbf{H}}_{j-1|K}^H)$. Likewise, $\mathbf{M}_{cn_t,1|k}$ consists of rows and columns $(n_t-1)L$ through $n_t L$ from the matrix $\mathbf{M}_{c,1|k} \triangleq N_r \mathbf{P}_{1|k} + \hat{\mathbf{H}}_{1|k}\hat{\mathbf{H}}_{1|k}^H$. In the above expressions, $\mathbf{P}_{j|k}$ and $\hat{\mathbf{H}}_{j|k}$ are a function of $\boldsymbol{\gamma}_c$, unlike (30), where the expressions are a function of $\boldsymbol{\gamma}$. Differentiating (54) w.r.t. $\gamma_c(i)$ and setting the resulting equation to zero gives the update for the i^{th} hyperparameter as follows:

$$\begin{aligned} \gamma_c^{(r+1)}(i) = \tfrac{1}{MKN_tN_r} &\left(\sum_{n_t=1}^{N_t} \sum_{k=2}^{K} \sum_{m=1}^{M} \tfrac{\mathbf{M}_{m,cn_t,j|K}}{(1-\rho^2)} \right. \\ &\left. + \sum_{n_t=1}^{N_t} \sum_{m=1}^{M} \mathbf{M}_{m,cn_t,1|K} \right), \end{aligned} \tag{55}$$

where $\mathbf{M}_{m,cn_t,j|K}$ consists of rows and columns $(B-1)M$ through BM from the matrix $\mathbf{M}_{cn_t,j|K}$, and $\mathbf{M}_{m,cn_t,1|K}$ consists of rows and columns $(B-1)M$ through BM from the matrix $\mathbf{M}_{m,cn_t,1|K}$. Thus, the KBMSBL algorithm learns $\boldsymbol{\gamma}_c$ in the M-step and provides low-complexity and recursive estimates of the time-varying gac-sparse channel in the E-step using the KFS framework.

Implementation of KBMSBL: Here, we discuss the implementation details of the KBMSBL algorithm and propose a low complexity solution based on the nested EM algorithm.

The complexity of the KBMSBL algorithm is dominated by the $\mathbf{J}_{k-1}$ term, whose computational complexity is given by $\mathcal{O}(KL^3)$. In [35], we proposed a nested SBL approach for estimating cluster-sparse signals and showed that the nested SBL approach has low complexity.

In the nested SBL approach [35], we restructure the problem by introducing auxiliary variables $\mathbf{t}_{n_t n_r, k} \in \mathbb{C}^{N\times 1}$, such that

$$\mathbf{t}_{n_t n_r, k} = \mathbf{\Phi}_{n_t,k} \mathbf{h}_{n_t n_r, kc} + \mathbf{z}_{n_t n_r, kc}. \tag{56}$$

The structuring of the vectors $\mathbf{h}_{n_t n_r, kc}$ is crucial for the nested SBL algorithm since it directly affects the computational complexity. Here, we construct a vector $\mathbf{h}_{n_t n_r, kc}$, such that it consists of sub-vectors governed by a common hyperparameter vector $\boldsymbol{\gamma}_c$, i.e.,

$$\begin{aligned} \mathbf{h}_{n_t n_r, kc} = [&h_{n_t n_r, 1k}(1), h_{n_t n_r, 2k}(1), \ldots, h_{n_t n_r, Bk}(1), \\ &\ldots, h_{n_t n_r, 1k}(M), h_{n_t n_r, 2k}(M), \ldots, h_{n_t n_r, Bk}(M)]. \end{aligned} \tag{57}$$

Accordingly, $\mathbf{\Phi}_{n_t,k}$ consists of the columns of $\mathbf{\Phi}_k$ corresponding to entries of $\mathbf{h}_{n_t n_r, kc}$. Although $\mathbf{z}_{n_t n_r, kc}$ cannot explicitly obtained, we note that its covariance can be written as $\mathbf{z}_{n_t n_r, kc} \sim \mathcal{CN}(0, \beta_m \sigma^2 \mathbf{I}_N)$ where, $0 \leq \beta_m \leq 1$ and $\sum_{m=1}^{M} \beta_m = 1$. [35]. Further, we using $\mathbf{t}_{n_r,k} = \left[\mathbf{t}_{1n_r,k}^T, \ldots, \mathbf{t}_{N_t n_r,k}^T\right]$, we construct the matrix $\mathbf{T}_k \in \mathbb{C}^{N_t ML \times N_r}$ by stacking $\mathbf{t}_{1,k}, \ldots, \mathbf{t}_{N_r,k}$ as its columns. The auxiliary variable matrix $\mathbf{T}_k$ decomposes the problem of tracking gac-sparse channels into a problem of tracking M length ga-sparse channel component vectors.

The NSBL technique is implemented using two EM loops, one nested within the other, as depicted in Fig. 5. The outer EM loop consists of updating the posterior distribution of $\mathbf{T}_k$

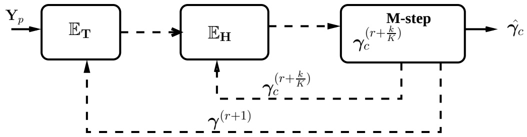

Figure 5. Block Diagram of the NSBL approach.

for $1 \leq k \leq K$, and the inner EM loop consists of updating the posterior distribution of the gac-sparse channel using the KFS framework across the K OFDM symbols.

After the posterior distribution of $\mathbf{T}_k$ is obtained in the outer EM loop, the inner EM loop in the NSBL algorithm is amenable to parallel implementation as M parallel Kalman filter and smoother chains. Each Kalman filter and smoother chain incurs a computational load of $\mathcal{O}(KB^3)$, since the parallel chains track a vector in a lower dimension (B). The computational complexity of the outer loop of the nested SBL approach dominates the overall complexity of the algorithm, and hence, the complexity of NSBL is given by $\mathcal{O}(K^3M^2P_t^2L)$. Note that, in contrast to KBMSBL which incurs a computational complexity of $\mathcal{O}(KL^3)$, the complexity of the NSBL approach scales linearly in L. Hence, the NSBL approach leads to efficient implementation of the proposed KBMSBL algorithm for large L.

In the following subsection, we generalize the proposed BMSBL and KBMSBL approaches for performing joint channel estimation and data detection in time-varying gac-sparse MIMO-OFDM channels.

V. Joint Channel Estimation/Tracking and Data Detection Using Pilot and Data Subcarriers for Gac-sparse Channels

In this section, we derive the novel J-BMSBL and J-KBMSBL algorithm that generalize the pilot-based BMSBL and KBMSBL algorithms for joint channel estimation and data detection in MIMO-OFDM systems.

To derive these algorithms, we modify the BMSBL and KBMSBL framework such that the unknown variables are not only the hyperparameters but also the unknown transmit data symbols in the entire OFDM frame. We consider $\mathbf{H}$ as the hidden variable, and, in contrast to BMSBL and KBMSBL, we consider $[\boldsymbol{\gamma}_c, \mathbf{X}]$ where $\mathbf{X} \triangleq [\mathbf{X}_1, \ldots, \mathbf{X}_{N_t}]$ as the parameters to be estimated. An important observation here is that the BMSBL/KBMSBL algorithm differs from the ga-sparse based MSBL/KMSBL algorithm due to the channel modeling, which in turn affects the posterior mean and variance of the channel. These posterior statistics affect the updates of $\boldsymbol{\gamma}_c$ as shown in (48)/(53). The updates of the transmit data $[\mathbf{X}_1, \ldots, \mathbf{X}_{N_t}]$ in the case of J-BMSBL, and $[\boldsymbol{\gamma}_c, \mathbf{X}_1, \ldots, \mathbf{X}_k]$ in the case of J-KBMSBL, can be obtained from the posterior estimates of the gac-sparse channel, from the E-step. Hence, the update equation for the transmit data remains the same as (38)/(43).

VI. Simulation Results

In this section, we demonstrate the performance of the proposed channel estimation algorithms using Monte Carlo simulations. We consider the parameters in the 3GPP/LTE broadband standard [29], [30]. We use a 3MHz 2×2 MIMO-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 $16.67\mu s$. The length of component channel vectors of the ga-sparse channel (L) is taken to be equal to the length of the CP. Each frame of the MIMO-OFDM system consists of $K = 7$ OFDM symbols. 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 [36], which uses a maximum of 10 Turbo iterations. We use a convergence criteria of $\epsilon = 10^{-9}$ and $r_{max} = 200$ for all the algorithms. We note that these values are not meant to be completely compliant with the LTE standard. We have chosen the parameter settings so as to facilitate visual comparison between the different schemes, illustrate the underlying performance tradeoffs, while at the same time being close to realistic settings used in the standard.

We use the Pedestrian B channel model [37] with Rayleigh fading. Further, we consider raised cosine filtering in every receive and transmit antenna chain with a roll-off factor of 0.5 [30]. This leads to the channel vectors being ga-sparse (see [20] for an illustration).

In the following subsections, we present the simulation results for the performance of the proposed algorithms in estimating block-fading and time-varying ga-sparse and gac-sparse wireless channels.

A. Block-fading Ga-sparse and Gac-sparse Channels

In this subsection, we consider the pilot-only channel estimation and joint channel estimation and data detection in block-fading ga-sparse and gac-sparse channels. Each OFDM frame consists of $K = 7$ OFDM symbols, with $P_b = 44$ pilots placed in an equidistant lattice structure in an OFDM frame of each transmitter. We implement the MSBL and the J-MSBL algorithm for ga-sparse and BMSBL and J-BMSBL algorithms (with block sizes of 4 and 6) in the case of gac-sparse block-fading channels, and plot the MSE and the coded BER performance of the algorithms in Fig. 6 and Fig. 7, respectively. We compare the performance of the proposed algorithms with the CS based Simultaneous OMP (SOMP) [38] using 50 pilots, MIP-aware methods: pilot-only MIP-aware estimation [4] and the MIP-aware joint data and channel estimation algorithm, which we refer to as the EM-OFDM algorithm [11].

From the top half of Fig. 6, we observe that the MSBL algorithms performs at least 1 dB better than the CS based SOMP technique. Since the proposed MSBL technique exploits spatial joint sparsity, MSBL performs 5 dB better than the per-symbol SBL algorithm proposed in [20]. We also observe that since BMSBL exploits the cluster-sparse structure, it performs $2 - 2.5$ dB better than the MSBL technique. The bottom half of Fig. 6 depicts the MSE performance of joint data detection techniques that detect the $(KN - P_b)$ data symbols along with estimating the channel, resulting in a significantly lower overall MSE compared to pilot-only schemes. We see that among the joint SBL based iterative methods, the J-MSBL

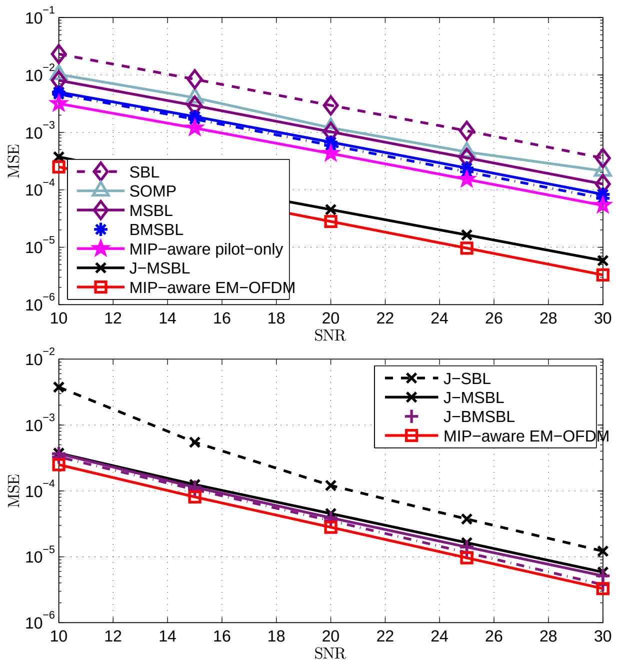

Figure 6. MSE performance in block-fading channels as a function of SNR in dB: Top: per-symbol SBL vs. MSBL vs. BMSBL. Bottom: J-SBL vs. J-MSBL vs. J-BMSBL. $P_b = 44$, gac-sparse: Solid curves - block size = 4, Dashed-dot curves - block size = 6.

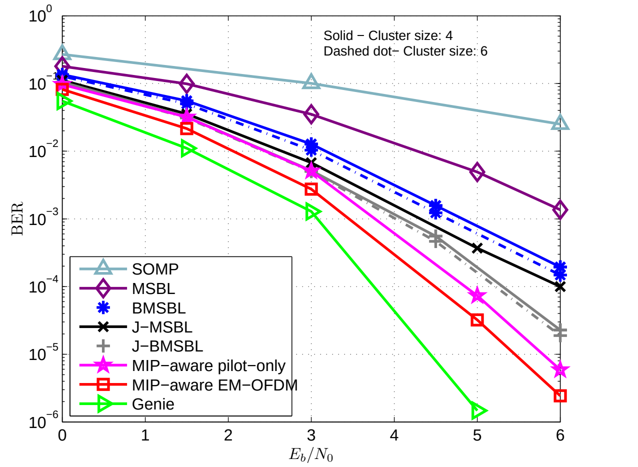

Figure 7. Coded BER performance of the proposed algorithms in a block-fading channel, with $P_b = 44$ pilot subcarriers, as a function of E_b/N_0.

algorithm performs an order of magnitude better than the MSBL algorithm, especially at higher values of SNR. Further, we see that J-BMSBL has a superior performance compared to J-MSBL and the per-symbol J-SBL [20]. Note that J-BMSBL is less than a dB from the MIP-aware EM-OFDM algorithm.

The coded BER performance of the proposed schemes are compared to the EM-OFDM, and a genie receiver, i.e., a receiver with perfect knowledge of the channel (labeled as `Genie`), in Fig. 7. We also compare the performance with MSBL, BMSBL and MIP-aware pilot-only channel estimation followed by data detection. First, we observe that the MSBL algorithm performs 2 dB better than the SOMP scheme, while being more than a dB worse than the BMSBL scheme. Further, the J-BMSBL technique, performs 1 dB better than the BMSBL scheme and 0.5 dB better than the J-MSBL scheme, and only 0.5 dB worse than the MIP-aware pilot-only technique. Since the MIP-aware pilot-only technique estimates the channel from an overdetermined system of equations, it outperforms the MIP-unaware pilot-only techniques. Moreover, for the SNRs between $0 - 10$ dB, the joint channel estimation and data detection techniques are prone to errors in the detected transmit data. Hence, they are outperformed by MIP-unaware pilot-only techniques.

B. *Time-varying Ga-sparse and Gac-sparse Channels*

In this section, we consider a slowly time-varying channel, simulated according to a Jakes' model [39] 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 proposed algorithms as a function of SNR is depicted in Fig. 8. In the top half of the plot, we demonstrate that the pilot-only KMSBL algorithm performs $5 - 7$ dB better than the MSBL and per-symbol J-SBL algorithms, since the KMSBL algorithm exploits the temporal correlation and joint sparsity in time-varying channels. Further, we demonstrate that the KBMSBL technique, which exploits the approximate cluster-sparsity, performs $1.5 - 2.5$ dB better than the KMSBL algorithm, while being $5 - 6$ dB away from the optimal MIP-aware Kalman tracking algorithm [28]. However, the J-KBMSBL algorithm performs 5 dB better than its pilot-only counterpart, i.e., the KBMSBL algorithm, while being less than a dB away from the optimal MIP-aware Kalman tracking algorithm. The optimal MIP-aware algorithm performs joint channel estimation and data detection, i.e., uses an MIP-aware EM algorithm, which implements the channel estimation in the E-step using a Kalman tracker and detects the transmit data in the M-step.

In the bottom half of Fig. 8, we demonstrate the performance of joint channel estimation and data detection schemes in time-varying channels. First, we observe that the per-symbol J-SBL algorithm that is not designed to exploit the temporal correlation performs 5-6 dB poorer than the recursive KMSBL and JKMSBL algorithms. At higher SNR, we observe that the performance of the JKMSBL algorithm is only 2 dB worse than the MIP-aware Kalman tracking algorithm. In contrast to pilot-only schemes, J-KMSBL and J-KBMSBL have the same performance while being 1 dB away from the MIP-aware Kalman tracking algorithm, especially at higher values of SNR, i.e., the advantage of modeling the channel as being cluster-sparse diminishes, as we see higher number of pilots due to accurate detection of transmit data at high SNRs.

In Fig. 9, we depict the coded BER performance of the proposed algorithms. We see that, while the proposed algorithms perform better than the SOMP algorithm by a margin larger than 2.5 dB, the JKBMSBL is only a fraction of a dB away from performance of the MIP-aware Kalman and the genie receiver which has perfect channel knowledge. The J-KSBL outperforms the pilot-only based KMSBL by a margin

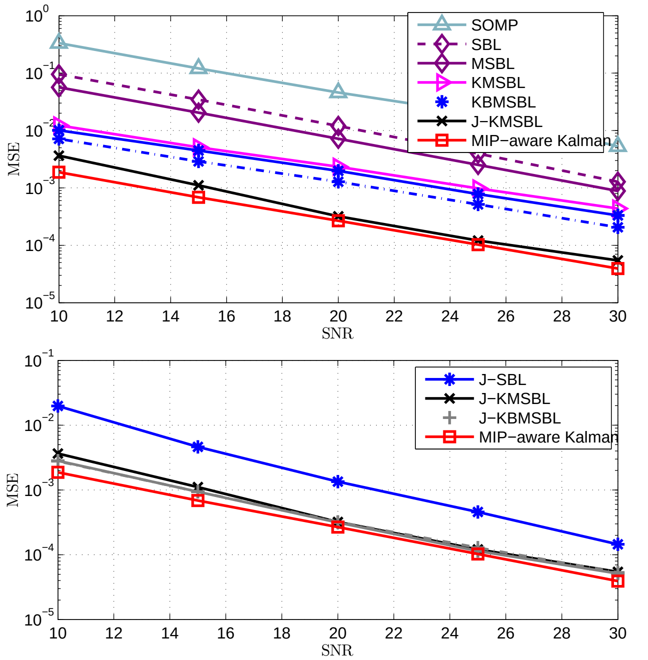

Figure 8. MSE performance in time-varying channels as a function of SNR in dB: Top: per-symbol SBL vs. MSBL vs. KMSBL vs. KBMSBL, Bottom: J-SBL vs. J-MSBL vs. J-KMSBL vs. J-KBMSBL as compared to the optimal Kalman tracker [28]. $f_dT_s = 0.001$ and $P_t = 44$. Cluster-sparse: Solid curves - block size = 4, Dashed-dot curves - block size = 6.

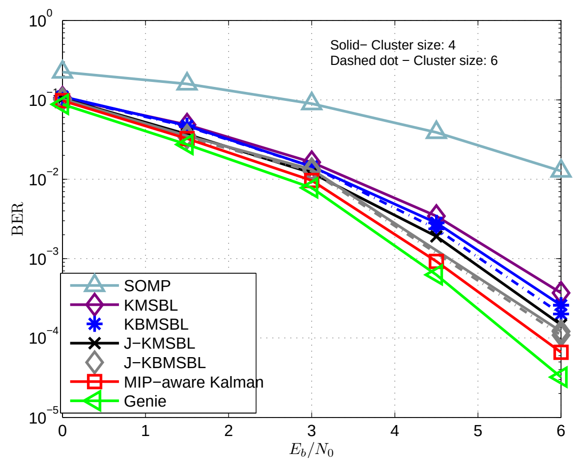

Figure 9. Coded BER performance of different schemes in a time-varying channel with $f_dT_s = 0.001$ and $P_t = 44$, as a function of E_b/N_0.

of 0.5 dB. Further, the gac-sparse KBMSBL and J-KBMSBL algorithms perform better than their ga-sparse counterparts, i.e., the KMSBL and J-KMSBL algorithms, by a margin of 0.5 dB.

In both block-fading and time-varying channel conditions, we note that the performance of algorithms proposed for gac-sparse channels was better than their ga-sparse counterparts, ascertaining that the cluster-sparse channel modeling is indeed useful for estimating wireless OFDM channels. Intuitively, modeling the channel using $\gamma \in \mathbb{C}^{L\times 1}$ leads to overfitting, which is overcome by modeling the channel using $\gamma_c \in \mathbb{C}^{B\times 1}$ where $B < L$. We also observe that the performance is better for a block-size of 6 compared to the block-size of 4 . This is because, in this example, a block-size of 6 is more accurate, and avoids overfitting, as compared to a block-size of 4.

VII. Conclusions

In this paper, we considered the pilot-only channel estimation and joint ga-sparse and gac-sparse channel estimation and data detection for block-fading and time-varying channels in MIMO-OFDM systems, using the SBL framework. To estimate the ga-sparse and gac-sparse block-fading channels, we adapted the existing MSBL and BMSBL algorithms and generalized it to obtain the J-MSBL and J-BMSBL algorithms, respectively, for joint ga-sparse and gac-sparse channel estimation and data detection. We used a first order AR model to capture the temporal correlation of the ga-sparse and gac-sparse channels and proposed the pilot-only KMSBL and KBMSL algorithms, respectively. We generalized these algorithms to obtain the J-KMSBL and J-KBMSBL algorithms, respectively, for joint channel estimation and data detection. We discussed the computational aspects of the proposed algorithms and showed that the proposed recursive algorithms entail a significantly lower computational complexity compared to the previously known SBL based techniques. Further, we also discussed efficient implementation structures for gac-sparse channels in block-fading and time-varying scenarios. Simulation results showed that (i) joint algorithms outperformed their pilot-only counterparts, (ii) recursive techniques outperformed the per-symbol algorithms, and (iii) algorithms proposed in the context of gac-sparse channels outperformed their ga-sparse counterparts.

References

[1] R. Prasad and C. R. Murthy, “Joint channel estimation and data detection in MIMO-OFDM systems using sparse Bayesian learning,” in *Proc. Nat. Conf. Commun. (NCC)*, Feb 2014, 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] 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.

[4] 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.

[5] M. E. Tipping, “The relevance vector machine,” in *Advances in NIPS*, vol. 12, 2000.

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

[7] 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.

[8] A. Gilbert and J. Tropp, “Applications of sparse approximation in communications,” in *Proc. IEEE Int. Symp. Inf. Theory*. IEEE, 2005, pp. 1000–1004.

[9] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, “Estimation of sparse MIMO channels with common support,” *Arxiv preprint arXiv:1107.1339*, 2011.

[10] 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.

[11] M. Wang, D. Quevedo, and G. Goodwin, “Joint data detection and channel estimation for MIMO-OFDM systems via EM algorithm and sphere decoding,” in *Proc. Globecom*. IEEE, Dec. 2006, pp. 1 –5.

[12] 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. Tech.*, vol. 54, no. 4, pp. 1235–1244, 2005.

[13] F. Wan, W.-P. Zhu, and M. Swamy, "Semiblind sparse channel estimation for mimo-ofdm systems," *IEEE Trans. Veh. Tech.*, vol. 60, no. 6, pp. 2569–2582, 2011.

[14] Y. Peng, X. Yang, X. Zhang, W. Wang, and B. Wu, "Compressed mimo-ofdm channel estimation," in *Proc. of ICCT*. IEEE, 2010, pp. 1291–1294.

[15] M. Khojastepour, K. Gomadam, and X. Wang, "Pilot-assisted channel estimation for mimo ofdm systems using theory of sparse signal recovery," in *Proc. ICASSP*. IEEE, 2009, pp. 2693–2696.

[16] D. Eiwen, G. Taubock, F. Hlawatsch, and H. Feichtinger, "Group sparsity methods for compressive channel estimation in doubly dispersive multicarrier systems," in *Proc. SPAWC*. IEEE, 2010, pp. 1–5.

[17] G. Gui, N. Zheng, N. Wang, A. Mehbodniya, and F. Adachi, "Compressive estimation of cluster-sparse channels." *Progress In Electromagnetics Research C*, vol. 24, 2011.

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

[19] 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.

[20] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse Bayesian learning," *IEEE Trans. Signal Process.*, 2014.

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

[22] S. Babacan, R. Molina, and A. Katsaggelos, "Bayesian Compressive Sensing using Laplace Priors," *IEEE Trans. Image Process.*, vol. 19, pp. 53–64, 2010.

[23] 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.

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

[25] Z. Zhang and B. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, 2013.

[26] 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.

[27] 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.

[28] 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.

[29] J. Zyren and W. McCoy, "Overview of the 3GPP long term evolution physical layer," *Freescale Semiconductor, Inc., white paper*, 2007.

[30] "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.

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

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

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

[34] R. Hunger, "Floating point operations in matrix-vector calculus," *Munich University of Technology, TUM-LNS-TR-05-05*, 2005.

[35] R. Prasad, C. Murthy, and B. D. Rao, "Nested sparse Bayesian learning for block-sparse signals with intra-block correlation," in *Proc. ICASSP*, 2014, pp. 1–6.

[36] 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.

[37] Recommended ITU-R M.1225, "Guidelines for evaluation of radio transmission technologies for IMT-2000," 1997.

[38] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[39] 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.

Ranjitha Prasad received the B.E. degree in Electronics and Communication Engineering from National Institute of Engineering, Mysore, India, in 2004, and the M.S. degree in Electrical Engineering, Indian Institute of Technology Madras, Chennai, India, in 2009. From July 2004-2006, she worked as a senior design engineer at Tata Elxsi, Bangalore, India. She obtained the Ph.D. from the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India, in 2015. She is currently pursuing a post-doctoral fellowship at the Nanyang Technological University, Singapore. Her research interests include signal processing for communications, adaptive filter theory, sparse Bayesian learning and compressive Sensing.

Chandra R. Murthy (S'03–M'06 – SM'11) received the B. Tech. degree in Electrical Engineering from the Indian Institute of Technology, Madras in 1998, the M. S. and Ph. D. degrees in Electrical and Computer Engineering from Purdue University and the University of California, San Diego, in 2000 and 2006, respectively. From 2000 to 2002, he worked as an engineer for Qualcomm Inc., where he worked on WCDMA baseband transceiver design and 802.11b baseband receivers. From Aug. 2006 to Aug. 2007, he worked as a staff engineer at Beceem Communications Inc. on advanced receiver architectures for the 802.16e Mobile WiMAX standard. In Sept. 2007, he joined the Department of Electrical Communication Engineering at the Indian Institute of Science, where he is currently working as an Associate Professor.

His research interests are in the areas of Cognitive Radio, Energy Harvesting Wireless Sensors, MIMO systems with channel-state feedback, and sparse signal recovery techniques applied to wireless communications. His paper won the best paper award in the Communications Track in the National Conference on Communications 2014. He is currently serving as the Chair of the IEEE Signal Processing Society, Bangalore Chapter, and as an associate editor for the IEEE Signal Processing Letters. He is an elected member of the IEEE SPCOM Technical Committee for the years 2014-16.

Bhaskar D. Rao (S'80–M'83–SM'91–F'00) received the B. Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in 1979 and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, in 1981 and 1983, respectively. Since 1983, he has been with the University of California at San Diego, La Jolla, where he is currently a Professor with the Electrical and Computer Engineering Department. His interests are in the areas of digital signal processing, estimation theory, and optimization theory, with applications to digital communications, speech signal processing, and human-computer interactions.

He is the holder of the Ericsson endowed chair in Wireless Access Networks and was the Director of the Center for Wireless Communications (2008-2011). He has made numerous contributions to the problem of sparse signal recovery and co-organized the first special session on the topic of sparsity at ICASSP 1998. His research group has received several paper awards. A paper he coauthored with B. Song and R. Cruz received the 2008 Stephen O. Rice Prize Paper Award in the Field of Communications Systems and a paper he coauthored with D. P. Wipf received the 2012 SP society best paper award. He was elected to the fellow grade in 2000 for his contributions in high resolution spectral estimation. Prof. Rao has been a member of the Statistical Signal and Array Processing technical committee, the Signal Processing Theory and Methods technical committee, and the Signal Processing for Communications technical committee of the IEEE Signal Processing Society. He is currently a member of the Machine Learning for Signal Processing technical committee.