Sparse Doubly-Selective Channel Estimation Techniques for OSTBC MIMO-OFDM Systems: A Hierarchical Bayesian Kalman Filter Based Approach

Suraj Srivastava, *Student Member, IEEE,* Mahendrada Sarath Kumar, Amrita Mishra, *Member, IEEE,* Sanjana Chopra, Aditya K. Jagannatham, *Member, IEEE,* and Lajos Hanzo, *Fellow, IEEE*

Abstract—Hierarchical Bayesian Kalman filter (HBKF) based schemes are conceived for doubly-selective sparse channel estimation in orthogonal space-time block coded (OSTBC) multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) wireless systems. Initially, a pilot based multiple measurement vector (MMV) model is formulated for estimating the OSTBC MIMO-OFDM channel. This is followed by the development of a low-complexity, online pilot-based HBKF (P-HBKF) scheme for tracking the sparse time-varying frequency-selective channel. The salient advantages of the proposed P-HBKF technique are that it requires significantly lower number of pilot subcarriers, while also exploiting the inherent sparsity of the wireless channel. Subsequently, data detection is also incorporated in the proposed framework, leading to the development of a procedure for joint sparse doubly-selective channel estimation and symbol detection. Recursive Bayesian Cramér-Rao bounds and closed form expressions are also obtained for the asymptotic mean square error (MSE) based on the solution of the Riccati equation for the KF for benchmarking the performance. Simulation results are presented for validating the theoretical bounds and for comparing the performance of the proposed and existing techniques.

Index Terms—MIMO-OFDM, OSTBC, channel estimation, sparsity, sparse Bayesian learning, Riccati equation, hierarchical Bayesian Kalman filter, frequency-selective, Cramér-Rao bound

I. INTRODUCTION

Multiple-input multiple-output (MIMO) schemes combined with orthogonal frequency division multiplexing (OFDM), i.e., MIMO-OFDM, have found their way into various standardized wireless systems such as the Long Term Evolution (LTE/

S. Srivastava and A. K. Jagannatham are with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur, UP 208016, India (e-mail: ssrivast@iitk.ac.in, adityaj@iitk.ac.in). M. Sarath is with Qualcomm, Hyderabad, Telangana 500081, India (email: saramahe@qti.qualcomm.com). A. Mishra is with the Department of Electronics and Communication Engineering, Dr. SPM IIIT-Naya Raipur, Chhattisgarh, India (email: amrita@iiitnr.edu.in). S. Chopra is with Analog Devices, Bangalore, Karnataka 560093, India (email: sanjanachopra92@gmail.com). L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

LTEA), IEEE 802.11 ac, 802.11 ax, 5G NR and several others due to the twin benefits of spatial multiplexing gains from the MIMO schemes coupled with the robustness of OFDM to multipath fading [1], [2]. Furthermore, orthogonal space time block codes (OSTBCs) are also attractive for such systems due to their diversity gain and simplified receiver architecture [3], [4]. Therefore, OSTBC based MIMO-OFDM, which exploits the combined advantages of both MIMO-OFDM and OSTBCs, is eminently suitable for next generation wireless networks. The enhanced performance gain and spectral efficiency of such systems is critically dependent on the accuracy of the channel state information (CSI) available at the receiver [1]. However, channel estimation is quite challenging for high-Doppler wireless channels, since every time the Doppler frequency is doubled, we have to double the pilot overhead. Thus, it is imperative to design accurate channel estimation schemes for achieving the best performance in OSTBC-based MIMO-OFDM systems. Various existing contributions in the literature have therefore addressed this challenging problem, to varying degrees, using different approaches. A succinct review of these contributions is presented next.

A. Review of Existing Works

The straightforward pilot based channel estimation technique has been lavishly studied for employment in OFDM-based wireless standards, such as DVB-T [5]. However, the increased pilot overheads can lead to inefficiency in spectrum utilization, especially in doubly-selective channels, due to the rapid variation of the channel coupled with a wide bandwidth. Conventional techniques, such as frequency domain least squares (LS)/ minimum mean squared error (MMSE) [6] as well as time domain LS/ MMSE require the transmission of a large number of pilot symbols, especially in channels having high delay spreads. This is due to the fact that such orthodox techniques require an overdetermined system for channel estimation, and lead to an ill-posed problem when the number of pilot transmissions is lower than the maximum number of channel taps in the finite impulse response (FIR) filter modeling the frequency-selective channel. However, the conventional techniques do not exploit the sparsity of the multipath wireless channel arising from the presence of only a few significant non-zero channel taps in the channel response,

which has an otherwise large delay spread. Exploiting this sparsity can lead to a significant reduction in the number of pilot transmissions required. In this context, the compressed sensing (CS) techniques proposed in [7], [8] utilize a sparse signal recovery model for channel estimation. The authors of [9], [10], have demonstrated the applicability of techniques such as simultaneous orthogonal matching pursuit (SOMP) and modified OMP, respectively, for MIMO-OFDM systems. However, the schemes proposed in [10] are greedy in nature and their performance is sensitive both to the choice of the dictionary matrix as well as to the stopping criterion. The convex-relaxation based least absolute shrinkage and selection operator (Lasso) [11] and the focal underdetermined system solver (FOCUSS) [12], which are based on l_1 and $l_p, p < 1$ norm minimization, respectively, have also been successfully exploited for sparse signal estimation. These techniques were then extended to their respective versions for scenarios with multiple observation vectors [13] and [14], termed as the group Lasso (GRP-Lasso) and M-FOCUSS, respectively. However, while the performance of Lasso depends critically on the user-defined regularization parameter, that of FOCUSS suffers from convergence problems. An approximate message passing (AMP) approach has been developed in [15] with its application to sparse channel estimation shown in [16]. However, a key drawback of the AMP technique is that its performance deteriorates even for slightly ill-conditioned dictionary matrices [17], thereby limiting its applicability in many practical scenarios. To overcome these drawbacks, the sparse Bayesian learning (SBL) [18] technique and its multiple signal extension M-SBL [19] have been proposed for OFDM based systems in [20], [21] for simultaneous sparse signal recovery, wherein they were indeed seen to achieve improved results in comparison to the existing schemes, especially upon leveraging the expectation-maximization (EM) framework for sparse channel estimation. Subsequently, Mishra *et al.* [22] combined an SBL-based sparse channel estimation scheme for OSTBC MIMO-OFDM systems. However, the schemes proposed in all the above treatises consider a quasi-static scenario, wherein the MIMO-OFDM channel is assumed to be constant in each block with the different realizations being independent identically distributed (i.i.d.) across the blocks. Needless to say, such a model is impractical, since the wireless channel is time-selective in nature, which leads to significant temporal correlation between the channel realizations of successive blocks [23]. It is important to note that the conventional compressed sensing techniques discussed above are not suitable for such use cases due to their inability to exploit the temporal correlation. In this regard, recent contributions such as [21] have focused on the estimation of a time-selective MIMO-OFDM channel jointly over multiple transmission blocks. However, the schemes proposed therein exhibit an increased processing delay and also require storing the outputs corresponding to a large number of receive blocks for the subsequent data detection process. Furthermore, the algorithms in [21] are also computationally complex due to the additional processing required for the implementation of the Kalman filter smoother (KFS). Moreover, the waveform therein does not consider OSTBC, which is popular for

practical implementation due to its low complexity. In this context, the hierarchical Bayesian Kalman filter (HBKF) [24] based online schemes, which have been shown to have a lower processing delay and computational complexity, are further developed in this work for online sparse channel estimation in OSTBC-based MIMO-OFDM systems. The various contributions of this paper are summarized next.

B. Contributions of the Work

- Novel hierarchical Bayesian Kalman filter (HBKF) based sparse channel estimation schemes are developed in this work for doubly-selective OSTBC MIMO-OFDM systems, along with the supporting theoretical foundation. These schemes are motivated by the MMSE optimality of the KF for time-selective channel estimation [25] coupled with the remarkable performance of SBL [18] for sparse signal estimation.
- Moreover, since none of the existing contributions consider an online approach for exploiting the temporal correlation, this paper fills this gap in the literature. First, the channel estimation model is developed, followed by the pilot-based HBKF (P-HBKF) technique conceived for doubly-selective MIMO-OFDM channel estimation.
- The recursive Bayesian Cramér-Rao lower bound (BCRB) is also derived for analyzing the efficacy of the proposed doubly-selective channel estimation scheme.
- Another novel contribution of the paper is to develop an analytical framework for characterizing the asymptotic MSE performance of the proposed scheme based on the solution of the discrete-time Riccati equation for the KF.
- Subsequently, the D-HBKF is derived for joint channel estimation and data detection, which successfully incorporates the OSTBC based maximum-likelihood (ML) symbol detector in the HBKF framework for improved performance. This is one of the key advantages of the proposed HBKF framework, since none of the conventional compressed sensing techniques are suitable for such integration.
- Our simulation results demonstrate the benefits of the proposed schemes over several performance benchmarks.

C. Organization of the paper

This paper is organized as follows. Section II presents the OSTBC MIMO-OFDM system model followed by the doubly-selective sparse channel model. Section III develops the proposed P-HBKF technique for doubly-selective sparse channel estimation. The recursive BCRB and asymptotic MSE expressions characterizing the performance are derived in Section IV and Section V, respectively. The joint doubly-selective channel estimation and data detection framework is followed by the proposed D-HBKF technique developed in Section VI. Finally, our simulation results and concluding remarks are offered in Sections VII and VIII, respectively.

Notation: The notation $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary parts of complex value. The notations $\|\mathbf{h}\|_2$ and $\|\mathbf{H}\|_F$ denote l_2-norm and Frobenius-norm of vector $\mathbf{h}$ and matrix $\mathbf{H}$, respectively. $\text{vec}(\mathbf{H})$ denotes the vector equivalent
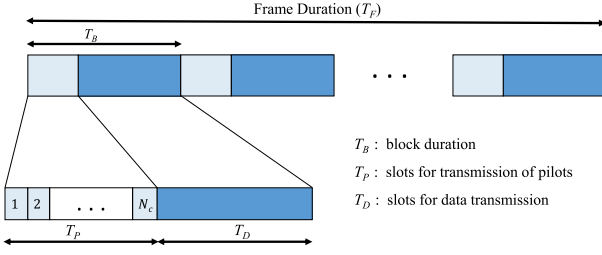
Fig. 1. Frame structure for doubly-selective channel estimation

of the matrix $\mathbf{H}$ formed by stacking its columns as a single column vector. $\mathrm{diag}(\mathbf{h})$ denotes the diagonal matrix with the elements of vector $\mathbf{h}$ on its principal diagonal. The notations $\mathbf{I}_N$, $\mathbf{0}_{u \times v}$ and $\otimes$ are used to represent an $N \times N$ identity matrix, $u \times v$ zero matrix and matrix Kronecker product.

II. SYSTEM AND CHANNEL MODEL

Consider an OSTBC-based MIMO-OFDM system [22], having N_T transmit antennas (TAs), N_R receive antennas (RAs) and N subcarriers. The transmission frame structure is shown in Fig. 1. Each frame in this system is comprised of R consecutive transmission blocks (TBs). Let N_c denote the number of slots occupied by a single OSTBC codeword. The OSTBC codeword matrix $\mathbf{X}_n(m) \in \mathbb{C}^{N_T \times N_c}$ corresponding to the mth subcarrier in the nth TB can be represented as [26]

$$\mathbf{X}_n(m) = \alpha \sum_{n_s=1}^{N_s} \mathbf{P}_{n_s} \Re\{s_{n_s,n}(m)\} + j\mathbf{Q}_{n_s}\Im\{s_{n_s,n}(m)\},$$
(1)

where $s_{n_s,n}(m), \forall\ 1 \leq n_s \leq N_s$, denotes the symbol that is drawn from a suitable constellation, $\mathbf{P}_{n_s}$ and $\mathbf{Q}_{n_s}$ represent fixed matrices of size $N_T \times N_c$ that characterize the OSTBC and N_s denotes the number of symbols per OSTBC codeword. The symbols are normalized to have unit energy, i.e., $E_s = \mathbb{E}\{|s_{n_s,n}(m)|^2\} = 1$. Only the initial N_c time slots corresponding to a single codeword on a chosen set of pilot subcarriers of each TB are reserved for pilot transmission. The remaining time slots on all the subcarriers are used for data transmission in each TB. In order to constrain the transmit power to unity, the normalization factor α is set to $\alpha = \sqrt{\frac{N_c}{N_s N_T}}$, which ensures $\mathbb{E}\left\{\mathrm{Tr}\left(\mathbf{X}_n(m)\mathbf{X}_n^H(m)\right)\right\} = 1$.

Let the vector $\mathbf{h}_{t,r,n}$ of length L denote the channel taps of the FIR filter corresponding to the frequency-selective channel between the tth TA and rth RA in the nth TB. Let the complex scalar quantity $X_{t,n_c,n}(m)$ denote the (t, n_c)th element of the OSTBC codeword matrix $\mathbf{X}_n(m)$. For mathematical convenience, these can be stacked across all the N subcarriers to obtain the vector $\mathbf{x}_{t,n_c,n} \in \mathbb{C}^{N \times 1}$ that is defined as

$$\mathbf{x}_{t,n_c,n} = [X_{t,n_c,n}(1), X_{t,n_c,n}(2), \ldots, X_{t,n_c,n}(N)]^T. \quad (2)$$

After cyclic prefix (CP) removal at the receiver followed by the fast Fourier transform (FFT) operation, the signal $\mathbf{y}_{r,n_c,n} \in \mathbb{C}^{N \times 1}$ received by the rth RA in the nth TB can be expressed as

$$\mathbf{y}_{r,n_c,n} = [\mathrm{diag}(\mathbf{x}_{1,n_c,n}), \ldots, \mathrm{diag}(\mathbf{x}_{N_T,n_c,n})] \times$$
$$(\mathbf{I}_{N_T} \otimes \mathbf{F})\, \mathbf{h}_{r,n} + \mathbf{w}_{r,n_c,n}, \quad (3)$$

where the concatenated channel vector $\mathbf{h}_{r,n} \in \mathbb{C}^{LN_T \times 1}$ across all the TAs is defined as

$$\mathbf{h}_{r,n} = \left[\mathbf{h}_{1,r,n}^T, \mathbf{h}_{2,r,n}^T, \ldots, \mathbf{h}_{N_T,r,n}^T\right]^T. \quad (4)$$

The matrix $\mathbf{F} \in \mathbb{C}^{N \times L}$ is the truncated discrete Fourier transform (DFT) matrix obtained by taking the first L columns of the $N \times N$ DFT matrix that has $\omega^{(i-1)(k-1)}$, $\omega = \exp(-j2\pi/N)$, as its (i, k)th entry. The vector $\mathbf{w}_{r,n_c,n} \in \mathbb{C}^{N \times 1}$ in (3) denotes the additive noise, which is assumed to be comprised of symmetric zero-mean i.i.d. complex Gaussian noise samples of variance σ^2. Let the vectors $\mathbf{y}_{r,n} \in \mathbb{C}^{NN_c \times 1}$ and $\mathbf{w}_{r,n} \in \mathbb{C}^{NN_c \times 1}$ represent the concatenated receive and noise vectors corresponding to the rth RA across all the N_c time instants in each codeword, respectively, which are defined as

$$\mathbf{y}_{r,n} = \left[\mathbf{y}_{r,1,n}^T, \mathbf{y}_{r,2,n}^T, \ldots, \mathbf{y}_{r,N_c,n}^T\right]^T,$$
$$\mathbf{w}_{r,n} = \left[\mathbf{w}_{r,1,n}^T, \mathbf{w}_{r,2,n}^T, \ldots, \mathbf{w}_{r,N_c,n}^T\right]^T. \quad (5)$$

Finally, let $\mathbf{Y}_n \in \mathbb{C}^{NN_c \times N_R}$, $\mathbf{H}_n \in \mathbb{C}^{LN_t \times N_R}$ and $\mathbf{W}_n \in \mathbb{C}^{NN_c \times N_R}$ denote the overall concatenated received signal, channel and noise matrices for the nth TB that can be defined as

$$\mathbf{Y}_n = [\mathbf{y}_{1,n}, \ldots, \mathbf{y}_{N_R,n}],$$
$$\mathbf{H}_n = [\mathbf{h}_{1,n}, \ldots, \mathbf{h}_{N_R,n}],$$
$$\mathbf{W}_n = [\mathbf{w}_{1,n}, \ldots, \mathbf{w}_{N_R,n}]. \quad (6)$$

The OSTBC-based MIMO-OFDM system model corresponding to the nth TB can now be succinctly represented as

$$\mathbf{Y}_n = \mathbf{\Psi}_n \mathbf{H}_n + \mathbf{W}_n, \quad (7)$$

where the MIMO-OFDM codeword matrix $\widetilde{\mathbf{X}}_n \in \mathbb{C}^{NN_c \times NN_T}$ and the dictionary matrix $\mathbf{\Psi}_n \in \mathbb{C}^{NN_c \times LN_t}$ of the nth TB are given by:

$$\widetilde{\mathbf{X}}_n = \begin{bmatrix} \mathrm{diag}(\mathbf{x}_{1,1,n}) & \cdots & \mathrm{diag}(\mathbf{x}_{N_T,1,n}) \\ \vdots & \ddots & \vdots \\ \mathrm{diag}(\mathbf{x}_{1,N_c,n}) & \cdots & \mathrm{diag}(\mathbf{x}_{N_T,N_c,n}) \end{bmatrix},$$
$$\mathbf{\Psi}_n = \widetilde{\mathbf{X}}_n (\mathbf{I}_{N_t} \otimes \mathbf{F}). \quad (8)$$

The above model can be recast in the convenient equivalent vectorized format of

$$\mathbf{y}_n = \mathrm{vec}(\mathbf{Y}_n) = \mathbf{\Phi}_n \mathbf{h}_n + \mathbf{w}_n, \quad (9)$$

where $\mathbf{\Phi}_n = (\mathbf{I}_{N_R} \otimes \mathbf{\Psi}_n) \in \mathbb{C}^{NN_c N_R \times LN_T N_R}$, $\mathbf{h}_n = \mathrm{vec}(\mathbf{H}_n) \in \mathbb{C}^{LN_T N_R \times 1}$ denote the effective dictionary matrix and the stacked channel vector, respectively, for the nth TB. The quantity $\mathbf{w}_n = \mathrm{vec}(\mathbf{W}_n) \in \mathbb{C}^{NN_c N_R \times 1}$ denotes the equivalent noise vector. In such scenarios, the time-evolution of the sparse channel vector $\mathbf{h}_n$ corresponding to the nth TB can be readily represented by using the first order autoregressive (AR1) model [27]

$$\mathbf{h}_n = \rho \mathbf{h}_{n-1} + \sqrt{1-\rho^2}\mathbf{u}_n, \quad (10)$$

where $\mathbf{h}_{n-1}$ denotes the sparse channel vector of the $(n-1)$st block and $\mathbf{u}_n \in \mathbb{C}^{LN_R N_T \times 1}$ represents the state innovation noise process that has a sparsity profile identical to that of

the channel vector $\mathbf{h}_n$. The latter assumption is justified by the fact that the locations of the dominant components of the channel vector $\mathbf{h}_n$ do not change for several OFDM frames [23], [28], [29], since the power delay profile is constant, and has naturally been used in several contributions such as [20], [21], [30]. Furthermore, upon considering the TB duration T_B to be well within the coherence time interval, we can assume the channel $\mathbf{h}_n$ to be constant over each TB. The quantity ρ denotes the temporal correlation coefficient corresponding to the channel vector $\mathbf{h}_n$ and can be evaluated from Jake's model, as described later in Section VII. The next section develops a pilot-based scheme for sparse doubly-selective channel estimation in OSTBC-based MIMO-OFDM systems.

III. PILOT-BASED DOUBLY-SELECTIVE CHANNEL ESTIMATION

Let the number of pilot subcarriers be denoted by N_P and let $\mathcal{P} \subseteq \{1, 2, \ldots, N\}$ represent the subset of pilot subcarriers. Furthermore, let $\mathbf{y}_{\mathcal{P},r,n_c,n} \in \mathbb{C}^{N_P \times 1}$ denote the received pilot output vector corresponding to the rth RA that is formed by extracting the elements of the vector $\mathbf{y}_{r,n_c,n}$ corresponding to the pilot locations $\mathcal{P}$, which can be determined as

$$\mathbf{y}_{\mathcal{P},r,n_c,n} = [\mathrm{diag}\,(\mathbf{x}_{\mathcal{P},1,n_c}), \ldots, \mathrm{diag}\,(\mathbf{x}_{\mathcal{P},N_T,n_c})] \times \\ (\mathbf{I}_{N_T} \otimes \mathbf{F}_{\mathcal{P}})\, \mathbf{h}_{r,n} + \mathbf{w}_{\mathcal{P},r,n_c,n}. \quad (11)$$

The vectors $\mathbf{x}_{\mathcal{P},t,n_c} \in \mathbb{C}^{N_P \times 1}$, $\mathbf{w}_{\mathcal{P},r,n_c,n} \in \mathbb{C}^{N_P \times 1}$ similarly denote the subvectors corresponding to the pilot locations extracted from $\mathbf{x}_{t,n_c,n}$, $\mathbf{w}_{r,n_c,n}$, respectively, and $\mathbf{F}_{\mathcal{P}} \in \mathbb{C}^{N_P \times L}$ is a submatrix formed by extracting the corresponding rows of the matrix $\mathbf{F}$. Similar to (5), the concatenated pilot output vector at the rth RA denoted by $\mathbf{y}_{\mathcal{P},r,n} \in \mathbb{C}^{N_P N_c \times 1}$ and the corresponding noise vector $\mathbf{w}_{\mathcal{P},r,n} \in \mathbb{C}^{N_P N_c \times 1}$ are given as

$$\mathbf{y}_{\mathcal{P},r,n} = \left[\mathbf{y}_{\mathcal{P},r,1,n}^T, \mathbf{y}_{\mathcal{P},r,2,n}^T, \ldots, \mathbf{y}_{\mathcal{P},r,N_c,n}^T\right]^T,$$
$$\mathbf{w}_{\mathcal{P},r,n} = \left[\mathbf{w}_{\mathcal{P},r,1,n}^T, \mathbf{w}_{\mathcal{P},r,2,n}^T, \ldots, \mathbf{w}_{\mathcal{P},r,N_c,n}^T\right]^T.$$

The equivalent input-output model can be obtained corresponding to the pilot locations similar to (54), using the various quantities defined above, as

$$\mathbf{Y}_{\mathcal{P},n} = \mathbf{\Psi}_{\mathcal{P}}\mathbf{H}_n + \mathbf{W}_{\mathcal{P},n}, \quad (12)$$

where the pilot output matrix $\mathbf{Y}_{\mathcal{P},n} \in \mathbb{C}^{N_P N_c \times N_R}$, noise matrix $\mathbf{W}_{\mathcal{P},n} \in \mathbb{C}^{N_P N_c \times N_r}$ and the matrix $\mathbf{\Psi}_{\mathcal{P}} \in \mathbb{C}^{N_P N_c \times L N_t}$ are given as

$$\mathbf{Y}_{\mathcal{P},n} = [\mathbf{y}_{\mathcal{P},1,n}, \ldots, \mathbf{y}_{\mathcal{P},N_R,n}],$$
$$\mathbf{W}_{\mathcal{P},n} = [\mathbf{w}_{\mathcal{P},1,n}, \ldots, \mathbf{w}_{\mathcal{P},N_R,n}],$$
$$\mathbf{\Psi}_{\mathcal{P}} = \widetilde{\mathbf{X}}_{\mathcal{P}}\,(\mathbf{I}_{N_t} \otimes \mathbf{F}_{\mathcal{P}}). \quad (13)$$

The MIMO-OFDM codeword matrix $\widetilde{\mathbf{X}}_{\mathcal{P}} \in \mathbb{C}^{N_P N_c \times N_P N_T}$ corresponding to the pilot subcarriers is constructed as

$$\widetilde{\mathbf{X}}_{\mathcal{P}} = \begin{bmatrix} \mathrm{diag}\,(\mathbf{x}_{\mathcal{P},1,1}) & \ldots & \mathrm{diag}\,(\mathbf{x}_{\mathcal{P},N_T,1}) \\ \vdots & \ddots & \vdots \\ \mathrm{diag}\,(\mathbf{x}_{\mathcal{P},1,N_c}) & \ldots & \mathrm{diag}\,(\mathbf{x}_{\mathcal{P},N_T,N_c}) \end{bmatrix}. \quad (14)$$

The compact vectorized input-output model of the pilot symbol locations is obtained following similar lines to (9) as

$$\mathbf{y}_{\mathcal{P},n} = \mathbf{\Phi}_{\mathcal{P}}\mathbf{h}_n + \mathbf{w}_{\mathcal{P},n}, \quad (15)$$

where $\mathbf{\Phi}_{\mathcal{P}} = (\mathbf{I}_{N_R} \otimes \mathbf{\Psi}_{\mathcal{P}}) \in \mathbb{C}^{N_P N_c N_R \times L N_R N_T}$, $\mathbf{y}_{\mathcal{P},n} = \mathrm{vec}(\mathbf{Y}_{\mathcal{P},n}) \in \mathbb{C}^{N_P N_c N_R \times 1}$ is the stacked received pilot vector and $\mathbf{w}_{\mathcal{P},n} = \mathrm{vec}(\mathbf{W}_{\mathcal{P},n}) \in \mathbb{C}^{N_P N_c N_R \times 1}$ denotes the stacked noise vector.

The conventional KF requires an overdetermined system, which has $N_P \geq \frac{L N_T}{N_c}$, for the estimation of the time-selective channel vector $\mathbf{h}_n$, which in turn leads to higher overheads and reduced spectral efficiency. Interestingly, the orthodox KF does not exploit the sparsity of the channel vector $\mathbf{h}_n$ arising due to the sparse multipath delay profile of a typical wireless channel. This implies that the number of non-zero channel impulse response (CIR) taps L_s is in practice much smaller than the total number L of CIR taps. Needless to say, exploiting this sparsity can lead to performance benefits associated with a reduced pilot overhead. This motivates the development of an online P-HBKF scheme for sparse doubly-selective MIMO-OFDM channel estimation as described next.

A. P-HBKF for online doubly-selective channel estimation

Using the standard KF notation, let $\widehat{\mathbf{h}}_{n|n-1}$ and $\mathbf{M}_{n|n-1}$ denote the prediction and the associated error covariance matrix, respectively, of the unknown sparse channel vector $\mathbf{h}_n$ in the nth filtering block. Similarly, let $\widehat{\mathbf{h}}_{n|n}$ and $\mathbf{M}_{n|n}$ represent the updated channel estimate and the respective minimum MSE matrix, in the nth block. These can be obtained using the standard Kalman filter for the nth block, as detailed in Appendix A. The P-HBKF procedure of estimating the doubly-selective sparse channel during the nth TB is described next. Let $\widehat{\mathbf{y}}_{\mathcal{P},n|n-1} = \mathbb{E}\{\mathbf{y}_{\mathcal{P},n} \mid \mathbf{y}_{\mathcal{P},0}, \cdots, \mathbf{y}_{\mathcal{P},n-1}\} \in \mathbb{C}^{N_P N_c N_R \times 1}$ denote the predicted measurement vector corresponding to the pilot subcarrier locations in the nth TB. The measurement error vector $\mathbf{y}_{\mathcal{P},e,n} \in \mathbb{C}^{N_P N_c N_R \times 1}$ can be defined and subsequently simplified as shown below

$$\mathbf{y}_{\mathcal{P},e,n} \triangleq \mathbf{y}_{\mathcal{P},n} - \widehat{\mathbf{y}}_{\mathcal{P},n|n-1}$$
$$= \rho\mathbf{\Phi}_{\mathcal{P}}\left(\mathbf{h}_{n-1} - \widehat{\mathbf{h}}_{n-1|n-1}\right) + \sqrt{1-\rho^2}\mathbf{\Phi}_{\mathcal{P}}\mathbf{u}_n + \mathbf{w}_{\mathcal{P},n} \quad (16)$$
$$\approx \sqrt{1-\rho^2}\mathbf{\Phi}_{\mathcal{P}}\mathbf{u}_n + \mathbf{w}_{\mathcal{P},n}. \quad (17)$$

The simplification from (16) to (17) above is based on the assumption that the estimation error, i.e., $\mathbf{h}_{n-1} - \widehat{\mathbf{h}}_{n-1|n-1}$, progressively decreases as n increases. The HBKF framework assigns a parameterized Gaussian prior to the innovation noise $\mathbf{u}_n$ in contrast to the standard Kalman filter that assumes it to be fixed. Let $\mathbf{u}_{t,r,n} \in \mathbb{C}^{L \times 1}$ denote the innovation noise vector corresponding to the tth TA and rth RA, obtained by extracting the subvector of $\mathbf{u}_n$ corresponding to the elements from $[(r-1)N_T + t - 1]L$ to $[(r-1)N_T + t]L$. The parameterized prior for the same can be defined in the HBKF framework as

$$p(\mathbf{u}_{t,r,n}; \mathbf{\Gamma}_n) = \prod_{l=1}^{L}(\pi\gamma_{l,n})^{-1}\exp\left(-\frac{|\mathbf{u}_{t,r,n}(l)|^2}{\gamma_{l,n}}\right), \quad (18)$$

where the hyperparameter $\gamma_{l,n}$, $1 \leq l \leq L$, denotes the prior variance corresponding to the lth component of $\mathbf{u}_{t,r,n}$ and $\boldsymbol{\Gamma}_n \triangleq \mathrm{diag}(\boldsymbol{\gamma}_n) \in \mathbb{R}^{L \times L}$ is the diagonal matrix of hyperparameters corresponding to the nth TB, where $\boldsymbol{\gamma}_n = [\gamma_{1,n}, \gamma_{2,n}, \cdots, \gamma_{L,n}]^T \in \mathbb{R}^{L \times 1}$. Furthermore, it can be noted that the covariance matrix $\boldsymbol{\Gamma}_n$ is considered to be unknown and it is estimated together with the sparse CIR $\mathbf{h}_n$. The parameterized prior corresponding to the concatenated innovation noise vector $\mathbf{u}_n$ is formulated as

$$p(\mathbf{u}_n; \boldsymbol{\Gamma}_n) = \prod_{r=1}^{N_R} \prod_{t=1}^{N_T} p(\mathbf{u}_{t,r,n}; \boldsymbol{\Gamma}_n). \tag{19}$$

Using (17) and (18), the *a posteriori* probability density function of the innovation noise vector $\mathbf{u}_n$ can be evaluated as $p(\mathbf{u}_n|\mathbf{y}_{\mathcal{P},e,n}, \boldsymbol{\Gamma}_n) \sim \mathcal{CN}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$, where the quantities $\boldsymbol{\mu}_n \in \mathbb{C}^{LN_RN_T \times 1}$ and $\boldsymbol{\Sigma}_n \in \mathbb{C}^{LN_RN_T \times LN_RN_T}$ are determined as [25]

$$\boldsymbol{\mu}_n = \sigma^{-2} \boldsymbol{\Sigma}_n \sqrt{1 - \rho^2} \boldsymbol{\Phi}_{\mathcal{P}}^H \mathbf{y}_{\mathcal{P},e,n},$$
$$\boldsymbol{\Sigma}_n = \left(\sigma^{-2} \left(1 - \rho^2\right) \boldsymbol{\Phi}_{\mathcal{P}}^H \boldsymbol{\Phi}_{\mathcal{P}} + (\mathbf{I}_{N_RN_T} \otimes \boldsymbol{\Gamma}_n)^{-1}\right)^{-1}, \tag{20}$$

which can be seen to depend on the hyperparameter matrix $\boldsymbol{\Gamma}_n$. The HBKF approach chooses the hyperparameter matrix $\widehat{\boldsymbol{\Gamma}}_n$ and in turn the prior $p\left(\mathbf{u}_n; \widehat{\boldsymbol{\Gamma}}_n\right)$ that maximizes the Bayesian evidence $p(\mathbf{y}_{\mathcal{P},e,n}; \boldsymbol{\Gamma}_n)$, given as $\log(\mathbf{y}_{\mathcal{P},e,n}; \boldsymbol{\Gamma}_n) = \kappa - \log[\det(\boldsymbol{\Sigma}_{\mathcal{P},e,n})] - \mathbf{y}_{\mathcal{P},e,n}^H \boldsymbol{\Sigma}_{\mathcal{P},e,n} \mathbf{y}_{\mathcal{P},e,n}$, where $\kappa = -N_P N_c N_R \log(\pi)$ and $\boldsymbol{\Sigma}_{\mathcal{P},e,n} = \mathbf{I}_{N_P N_c N_R} + (1 - \rho^2) \boldsymbol{\Phi}_{\mathcal{P}} (\mathbf{I}_{N_R N_T} \otimes \boldsymbol{\Gamma}_n) \boldsymbol{\Phi}_{\mathcal{P}}^H$. As it can be readily seen, the optimization objective above is non-concave. Hence, direct maximization of the above cost-function for the estimation of the hyperparameters $\gamma_{l,n}$ becomes intractable. This impediment can be overcome using the mathematical framework of iterative expectation-maximization (EM) for achieving the above likelihood maximization iteratively, which is eminently suitable for practical implementation. As demonstrated in [18], upon convergence of the EM procedure, the hyperparameter estimates $\gamma_{l,n}$ corresponding to the zero locations of the innovation vector $\mathbf{u}_{t,r,n}(l)$ are driven to zero, thereby leading to a sparse estimate of the innovation vector. Since the innovation noise vector $\mathbf{u}_n$ and the channel vector $\mathbf{h}_n$ share a common sparsity profile, this in turn leads to a sparse estimate of the channel vector $\mathbf{h}_n$. The key steps of EM procedure are described next. Let $\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}$ denote the estimate of the hyperparameter matrix $\boldsymbol{\Gamma}_n$ in the $(i-1)$st EM iteration. The expectation (E-step) in the ith EM iteration evaluates the average log-likelihood $\mathcal{L}\left(\boldsymbol{\Gamma}_n|\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}\right)$ of the complete data $\{\mathbf{y}_{\mathcal{P},e,n}, \mathbf{u}_n\}$, which is formulated as:

$$\mathcal{L}\left(\boldsymbol{\Gamma}_n|\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}\right) = \mathbb{E}_{\mathbf{u}_n|\mathbf{y}_{\mathcal{P},e,n};\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}} \{\log p(\mathbf{y}_{\mathcal{P},e,n}, \mathbf{u}_n; \boldsymbol{\Gamma}_n)\}$$
$$= \mathbb{E}_{\mathbf{u}_n|\mathbf{y}_{\mathcal{P},e,n};\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}} \{\log p(\mathbf{y}_{\mathcal{P},e,n}|\mathbf{u}_n; \boldsymbol{\Gamma}_n)\}$$
$$+ \mathbb{E}_{\mathbf{u}_n|\mathbf{y}_{\mathcal{P},e,n};\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}} \{\log p(\mathbf{u}_n; \boldsymbol{\Gamma}_n)\}.$$

Subsequently, the maximization (M-step), which maximizes $\mathcal{L}\left(\boldsymbol{\Gamma}_n|\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}\right)$ with respect to the hyperparameter vector $\boldsymbol{\gamma}_n$

can be formulated as

$$\widehat{\boldsymbol{\gamma}}_n^{(i)} = \arg \max_{\boldsymbol{\gamma}_n} \mathbb{E}_{\mathbf{u}_n|\mathbf{y}_{\mathcal{P},e,n};\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}} \{\log p(\mathbf{y}_{\mathcal{P},e,n}|\mathbf{u}_n; \boldsymbol{\Gamma}_n)\}$$
$$+ \mathbb{E}_{\mathbf{u}_n|\mathbf{y}_{\mathcal{P},e,n};\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}} \{\log p(\mathbf{u}_n; \boldsymbol{\Gamma}_n)\}. \tag{21}$$

The quantity within the first $\mathbb{E}\{\cdot\}$ above, simplifies as

$$-N_P N_c N_R \log\left(\pi\sigma^2\right) - \sigma^{-2} \left(\mathbf{y}_{\mathcal{P},e,n} - \sqrt{1 - \rho^2} \boldsymbol{\Phi}_{\mathcal{P}} \mathbf{u}_n\right)^H$$
$$\left(\mathbf{y}_{\mathcal{P},e,n} - \sqrt{1 - \rho^2} \boldsymbol{\Phi}_{\mathcal{P}} \mathbf{u}_n\right),$$

which is observed to be independent of the hyperparameter vector $\boldsymbol{\gamma}_n$ and can thus be ignored for the maximization. Therefore, the equivalent optimization problem used for the estimation of the hyperparameter vector $\boldsymbol{\gamma}_n$ is given as

$$\widehat{\boldsymbol{\gamma}}_n^{(i)} = \arg \max_{\boldsymbol{\gamma}_n} \mathbb{E}_{\mathbf{u}_n|\mathbf{y}_{\mathcal{P},e,n};\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}} \{\log p(\mathbf{u}_n; \boldsymbol{\Gamma}_n)\}$$
$$= \arg \max_{\boldsymbol{\gamma}_n} \sum_{r=1}^{N_R} \sum_{t=1}^{N_T} \sum_{l=1}^{L} c_1 - \log \gamma_{l,n} - \frac{\boldsymbol{\Sigma}_n^{(i)}(d,d) + \left|\boldsymbol{\mu}_n^{(i)}(d)\right|^2}{\gamma_{l,n}}, \tag{22}$$

where the constant c_1 can be derived as $c_1 = -\log \pi$ and the index d is defined as $d = [(r-1)N_T + t - 1]L + l$. The quantities $\boldsymbol{\mu}_n^{(i)}$, $\boldsymbol{\Sigma}_n^{(i)}$ denote the mean and covariance matrix corresponding to the *a posteriori* probability density function (PDF) $p(\mathbf{u}_n|\mathbf{y}_{\mathcal{P},e,n};\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}) = \mathcal{CN}\left(\boldsymbol{\mu}_n^{(i)}, \boldsymbol{\Sigma}_n^{(i)}\right)$, which can be obtained by substituting $\widehat{\boldsymbol{\Gamma}}_n^{(i-1)}$ in lieu of $\boldsymbol{\Gamma}_n$ in (20). It can now be observed that the maximization problem formulated above for the estimation of the hyperparameter vector $\boldsymbol{\gamma}_n$ is decoupled with respect to each individual $\gamma_{l,n}$. This can therefore be readily solved for obtaining the estimates $\widehat{\gamma}_{l,n}^{(i)}$ in the ith iteration of the EM algorithm as

$$\widehat{\gamma}_{l,n}^{(i)} = \frac{1}{N_R N_T} \sum_{r=1}^{N_R} \sum_{t=1}^{N_T} \boldsymbol{\Sigma}_n^{(i)}(d,d) + \left|\boldsymbol{\mu}_n^{(i)}(d)\right|^2. \tag{23}$$

The estimate of the hyperparameter matrix at the ith EM iteration in the nth TB follows as $\widehat{\boldsymbol{\Gamma}}_n^{(i)} = \mathrm{diag}\left(\widehat{\boldsymbol{\gamma}}_n^{(i)}\right)$. Upon convergence, the resultant hyperparameter matrix estimate is denoted by $\widehat{\boldsymbol{\Gamma}}_n^c$. Note that $\widehat{\boldsymbol{\Gamma}}_n^c$ also yields an estimate of the sparsity-profile of the channel vector $\mathbf{h}_n$. This is subsequently employed in (74) by setting the covariance of the innovation $\mathbf{u}_n$ equal to the converged estimate $\widehat{\boldsymbol{\Gamma}}_n^c$ of the hyperparameter matrix. Finally, the channel estimate corresponding to the nth TB is obtained using Equation (77) in the KF as $\widehat{\mathbf{h}}_{P,n} = \widehat{\mathbf{h}}_{n|n}$. A schematic representation of the P-HBKF approach is given in Fig. 2. In this figure, the EM block estimates the hyperparameter matrix $\boldsymbol{\Gamma}_n$ aiding the update process of the sparse channel estimate $\widehat{\mathbf{h}}_{n|n}$. Furthermore, the online nature of the proposed algorithm helps to track variations in the sparsity-profile of $\mathbf{h}_n$. A compact representation of the step-by-step procedure for the proposed online P-HBKF technique to estimate the doubly-selective sparse channel is presented in Algorithm 1. The initialization procedure for the P-HBKF scheme is described in the following subsection.

Algorithm 1 P-HBKF algorithm for sparse doubly-selective channel estimation

Input: Observation vector $\mathbf{y}_{\mathcal{P},n}$, Dictionary matrix $\mathbf{\Phi}_{\mathcal{P}}$, Correlation coefficient ρ, Noise variance σ^2, Stopping Parameters ϵ and $i_{\max}$

Initialization: $\widehat{\mathbf{\Gamma}}_{-1}^c = \mathbf{I}_L, \widehat{\mathbf{h}}_{-1|-1}^{(0)} = \mathbf{0}_{LN_R N_T \times 1}$

for $n = 0, 1, 2 \cdots$

    Set counter $i = 0$ and $\widehat{\mathbf{\Gamma}}_n^{(-1)} = \mathbf{0}$ and $\widehat{\mathbf{\Gamma}}_n^{(0)} = \widehat{\mathbf{\Gamma}}_{n-1}^c$

    while $\left(\left\| \widehat{\boldsymbol{\gamma}}_n^{(i)} - \widehat{\boldsymbol{\gamma}}_n^{(i-1)} \right\|_2 > \epsilon \;\; \&\& \;\; i < i_{\max} \right)$ **do**

    $i \leftarrow i + 1$

    Compute the measurement error $\mathbf{y}_{\mathcal{P},e,n}$ using (76) in Appendix A

    E-step:

$$\mathbf{\Sigma}_n^{(i)} = \left(\sigma^{-2} \left(1 - \rho^2 \right) \mathbf{\Phi}_{\mathcal{P}}^H \mathbf{\Phi}_{\mathcal{P}} + \left(\mathbf{I} \otimes \widehat{\mathbf{\Gamma}}_n^{(i-1)} \right)^{-1} \right)^{-1}$$

$$\boldsymbol{\mu}_n^{(i)} = \sigma^{-2} \mathbf{\Sigma}_n^{(k)} \sqrt{1 - \rho^2} \mathbf{\Phi}_{\mathcal{P}}^H \mathbf{y}_{\mathcal{P},e,n}$$

    M-step: Update hyperparameters $\widehat{\gamma}_{l,n}^{(i)}$ using (23), set $\widehat{\mathbf{\Gamma}}_n^c = \widehat{\mathbf{\Gamma}}_n^{(i)} = \mathrm{diag}\left(\widehat{\gamma}_{1,n}^{(i)}, \widehat{\gamma}_{2,n}^{(i)}, \cdots, \widehat{\gamma}_{L,n}^{(i)} \right)$

    end while

    Evaluate the filtered estimate $\widehat{\mathbf{h}}_{n|n}$ using (77) in Appendix A
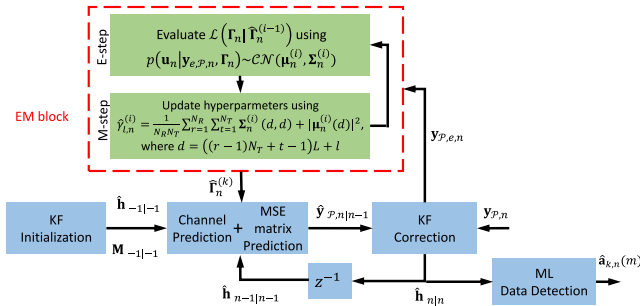
end for

Fig. 2. Block diagram representation of the P-HBKF scheme for doubly-selective sparse channel estimation in OSTBC MIMO-OFDM systems

B. P-HBKF Initialization

In the conventional HBKF framework [24], the estimate of the channel, the MSE and the hyperparameter matrices are initialized as

$$\widehat{\mathbf{h}}_{-1|-1} = \mathbf{0}_{LN_R N_T \times 1}, \mathbf{M}_{-1|-1} = \mathbf{I}_{LN_R N_T}, \widehat{\mathbf{\Gamma}}_n^{(0)} = \mathbf{I}_L, \forall n.$$

However, even though this straightforward initialization procedure leads to fairly accurate estimates, a large number of blocks n is required for MSE convergence, as also seen from the simulation results of Section VII. In order to overcome this shortcoming, a novel initialization method [31] is described below for the P-HBKF.

The MSE matrix $\mathbf{M}_{-1|-1}$ can be initialized with the estimate of the hyperparameter matrix $\widehat{\mathbf{\Gamma}}_0^c$, i.e., the converged estimate of the hyperparameter matrix for the 0th block. Furthermore, for the hyperparameter update procedure of Algorithm 1, the hyperparameter matrix $\widehat{\mathbf{\Gamma}}_n^{(0)}$ of the nth block is initialized to the converged hyperparameter matrix obtained

from the previous block, i.e. we have $\widehat{\mathbf{\Gamma}}_n^{(0)} = \widehat{\mathbf{\Gamma}}_{n-1}^c$. This initialization scheme is seen to achieve a significantly lower MSE and requires much fewer blocks as well as EM iterations for convergence to the steady state MSE in comparison to the conventional HBKF initialization, as shown later in Section VII. The BCRB for the proposed P-HBKF technique is presented next.

IV. BAYESIAN CRAMÉR-RAO LOWER BOUNDS

Let $\mathbf{J}_n \in \mathbb{C}^{LN_T N_R \times LN_T N_R}$ denote the Bayesian Fisher information matrix (BFIM) for the doubly-selective channel vector $\mathbf{h}_n$ to be estimated in the nth TB. Using the results of [32], this can be recursively determined as

$$\mathbf{J}_n = \mathbf{G}_n^{22} - \mathbf{G}_n^{21} \left(\mathbf{J}_{n-1} + \mathbf{G}_n^{11} \right)^{-1} \mathbf{G}_n^{12}, \quad (24)$$

where the matrices $\mathbf{G}_n^{11}, \mathbf{G}_n^{12}, \mathbf{G}_n^{21}$ and $\mathbf{G}_n^{22}$, all of size $LN_T N_R \times LN_T N_R$, are derived as

$$\mathbf{G}_n^{11} = -\mathbb{E}\left\{ \frac{\partial^2 \mathcal{L}(\mathbf{h}_n \mid \mathbf{h}_{n-1})}{\partial \mathbf{h}_{n-1} \partial \mathbf{h}_{n-1}^H} \right\}, \quad (25)$$

$$\mathbf{G}_n^{12} = -\mathbb{E}\left\{ \frac{\partial^2 \mathcal{L}(\mathbf{h}_n \mid \mathbf{h}_{n-1})}{\partial \mathbf{h}_{n-1} \partial \mathbf{h}_n^H} \right\} = \left(\mathbf{G}_n^{21} \right)^H, \quad (26)$$

$$\mathbf{G}_n^{22} = -\mathbb{E}\left\{ \frac{\partial^2 \mathcal{L}(\mathbf{h}_n \mid \mathbf{h}_{n-1})}{\partial \mathbf{h}_n \partial \mathbf{h}_n^H} \right\} - \mathbb{E}\left\{ \frac{\partial^2 \mathcal{L}(\mathbf{y}_{\mathcal{P},n} \mid \mathbf{h}_n)}{\partial \mathbf{h}_n \partial \mathbf{h}_n^H} \right\}. \quad (27)$$

Let $\mathbf{R}_u = \mathbb{E}\left\{ \mathbf{u}_n \mathbf{u}_n^H \right\} \in \mathbb{C}^{LN_T N_R \times LN_T N_R}$ denote the co-variance matrix of the innovation noise. Using the state and measurement model Equations in (10) and (15), respectively, the quantities $\mathcal{L}(\mathbf{h}_n \mid \mathbf{h}_{n-1})$ and $\mathcal{L}(\mathbf{y}_{\mathcal{P},n} \mid \mathbf{h}_n)$ can be evaluated as

$$\mathcal{L}(\mathbf{h}_n \mid \mathbf{h}_{n-1}) = \kappa_1 - \frac{(\mathbf{h}_n - \rho \mathbf{h}_{n-1})^H \mathbf{R}_u^{-1} (\mathbf{h}_n - \rho \mathbf{h}_{n-1})}{1 - \rho^2}, \quad (28)$$

$$\mathcal{L}(\mathbf{y}_{\mathcal{P},n} \mid \mathbf{h}_n) = \kappa_2 - \frac{(\mathbf{y}_{\mathcal{P},n} - \mathbf{\Phi}_{\mathcal{P}} \mathbf{h}_n)^H (\mathbf{y}_{\mathcal{P},n} - \mathbf{\Phi}_{\mathcal{P}} \mathbf{h}_n)}{\sigma^2}, \quad (29)$$

where the constants $\kappa_1 = -LN_R N_T \log \pi \left(1 - \rho^2 \right) - \log \det(\mathbf{R}_u)$ and $\kappa_2 = -N_P N_c N_R \log \left(\pi \sigma^2 \right)$. Substituting the quantities $\mathcal{L}(\mathbf{h}_n \mid \mathbf{h}_{n-1})$ and $\mathcal{L}(\mathbf{y}_{\mathcal{P},n} \mid \mathbf{h}_n)$ from (28), (29) into (25)-(27) determines the matrices $\mathbf{G}_n^{11}, \mathbf{G}_n^{12}, \mathbf{G}_n^{21}$ and $\mathbf{G}_n^{22}$. These are in turn substituted into (24) followed by simplification to obtain the matrix $\mathbf{J}_n$ corresponding to the P-HBKF estimate as

$$\mathbf{J}_n = \left(\rho^2 \mathbf{J}_{n-1} + \left(1 - \rho^2 \right) \mathbf{R}_u \right)^{-1} + \frac{1}{\sigma^2} \mathbf{\Phi}_{\mathcal{P}}^H \mathbf{\Phi}_{\mathcal{P}}. \quad (30)$$

The BCRB for the MSE of the P-HBKF based channel estimate is finally expressed as

$$\mathrm{MSE}\left(\widehat{\mathbf{h}}_{\mathrm{P},n} \right) \triangleq \mathbb{E}\left\{ \left\| \widehat{\mathbf{h}}_{\mathrm{P},n} - \mathbf{h}_n \right\|_2^2 \right\} \geq \mathrm{Tr}\{\mathbf{J}_n^{-1}\}. \quad (31)$$

$$\bar{\mathbf{M}}_{n+1|n} = \rho^2 \bar{\mathbf{M}}_{n|n-1} - \rho^2 \bar{\mathbf{M}}_{n|n-1} \boldsymbol{\Phi}_n^T \left(\sigma^2 \mathbf{I} + \boldsymbol{\Phi}_n \bar{\mathbf{M}}_{n|n-1} \boldsymbol{\Phi}_n^T \right)^{-1} \boldsymbol{\Phi}_n \bar{\mathbf{M}}_{n|n-1} + \left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1}$$

$$= \rho^2 \left(\bar{\mathbf{M}}_{n|n-1} - \bar{\mathbf{M}}_{n|n-1} \boldsymbol{\Phi}_n^T \left(\sigma^2 \mathbf{I} + \boldsymbol{\Phi}_n \bar{\mathbf{M}}_{n|n-1} \boldsymbol{\Phi}_n^T \right)^{-1} \boldsymbol{\Phi}_n \bar{\mathbf{M}}_{n|n-1} \right) + \left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1} \qquad (32)$$

$$\bar{\mathbf{M}}_{n+1|n+1} = \left(\mathbf{I} - \bar{\mathbf{M}}_{n+1|n} \boldsymbol{\Phi}_{n+1}^T \left(\sigma^2 \mathbf{I} + \boldsymbol{\Phi}_{n+1} \bar{\mathbf{M}}_{n|n-1} \boldsymbol{\Phi}_{n+1}^T \right)^{-1} \boldsymbol{\Phi}_{n+1} \right) \bar{\mathbf{M}}_{n+1|n}$$

$$= \bar{\mathbf{M}}_{n+1|n} - \bar{\mathbf{M}}_{n+1|n} \boldsymbol{\Phi}_{n+1}^T \left(\sigma^2 \mathbf{I} + \boldsymbol{\Phi}_{n+1} \bar{\mathbf{M}}_{n|n-1} \boldsymbol{\Phi}_{n+1}^T \right)^{-1} \boldsymbol{\Phi}_{n+1} \bar{\mathbf{M}}_{n+1|n} \qquad (33)$$

$$\bar{\mathbf{M}}_{n+1|n}^{-1} = \left(\left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1} \right)^{-1} - \left(\left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1} \right)^{-1} \rho^2 \left[\bar{\mathbf{M}}_{n|n}^{-1} + \rho^2 \left(\left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1} \right)^{-1} \right]^{-1} \times$$

$$\left(\left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1} \right)^{-1} \qquad (34)$$

V. ASYMPTOTIC MSE

This section derives the closed form expressions of the asymptotic MSE for the P-HBKF scheme, as the block length obeys $n \to \infty$. Hence, the rate at which the MSE of the proposed technique converges to the respective asymptotic MSE value characterizes the efficiency of the proposed estimation technique. Let $\bar{\mathbf{h}}_n \in \mathbb{R}^{2LN_T N_R \times 1}$ denote the real-valued equivalent of the CIR vector $\mathbf{h}_n$ that is obtained by stacking the real and imaginary components as $\bar{\mathbf{h}}_n = \left[\left(\Re(\mathbf{h}_n) \right)^T \left(\Im(\mathbf{h}_n) \right)^T \right]^T$. Also, let the real equivalent dictionary matrix $\boldsymbol{\Phi}_n$ corresponding to the P-HBKF scheme is given as

$$\boldsymbol{\Phi}_n = \begin{bmatrix} \Re(\boldsymbol{\Phi}_{\mathcal{P}}) & -\Im(\boldsymbol{\Phi}_{\mathcal{P}}) \\ \Im(\boldsymbol{\Phi}_{\mathcal{P}}) & \Re(\boldsymbol{\Phi}_{\mathcal{P}}) \end{bmatrix}. \qquad (35)$$

Using the KF notation similar to Appendix A, the prediction error covariance matrix $\bar{\mathbf{M}}_{n|n-1}$, the Kalman gain matrix $\bar{\mathbf{K}}_n$ and the estimation error covariance matrix $\bar{\mathbf{M}}_{n|n}$ are defined as

$$\bar{\mathbf{M}}_{n|n-1} = \rho^2 \bar{\mathbf{M}}_{n-1|n-1} + \left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_n \qquad (36)$$

$$\bar{\mathbf{K}}_n = \bar{\mathbf{M}}_{n|n-1} \boldsymbol{\Phi}_n^H \left(\sigma^2 \mathbf{I} + \boldsymbol{\Phi}_n \bar{\mathbf{M}}_{n|n-1} \boldsymbol{\Phi}_n^H \right)^{-1} \qquad (37)$$

$$\bar{\mathbf{M}}_{n|n} = \left(\mathbf{I} - \bar{\mathbf{K}}_n \boldsymbol{\Phi}_n \right) \bar{\mathbf{M}}_{n|n-1}, \qquad (38)$$

where the matrix $\bar{\boldsymbol{\Gamma}}_{n+1} \in \mathbb{C}^{2N_R N_T L \times 2N_R N_T L}$ is a block diagonal matrix with two block components, both of which are identical to $\frac{1}{2} \left(\mathbf{I}_{N_R N_T} \otimes \hat{\boldsymbol{\Gamma}}_n \right)$. The prediction error covariance matrix $\bar{\mathbf{M}}_{n+1|n}$ for the $(n+1)$st block can be related to the previous block $\bar{\mathbf{M}}_{n|n-1}$, in steady state, as $\bar{\mathbf{M}}_{n+1|n} = \bar{\mathbf{M}}_{n|n-1}$. Upon substituting the expression for $\bar{\mathbf{M}}_{n|n}$ from (38), Equation (36) for the $(n+1)$st block can be further simplified to

$$\bar{\mathbf{M}}_{n+1|n} = \rho^2 (\mathbf{I} - \bar{\mathbf{K}}_n \boldsymbol{\Phi}_n) \bar{\mathbf{M}}_{n|n-1} + \left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1}$$
$$= \rho^2 \bar{\mathbf{M}}_{n|n-1} - \rho^2 \bar{\mathbf{K}}_n \boldsymbol{\Phi}_n \bar{\mathbf{M}}_{n|n-1} + \left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1}. \qquad (39)$$

Substituting the value of $\bar{\mathbf{K}}_n$ from (37) into (39), the resultant expression for $\bar{\mathbf{M}}_{n+1|n}$ can be simplified to the form shown in (32). Employing the Woodbury Matrix Identity [25], the expression in (32) can be modified as

$$\bar{\mathbf{M}}_{n+1|n} = \rho^2 \left(\bar{\mathbf{M}}_{n|n-1}^{-1} + \sigma^{-2} \boldsymbol{\Phi}_n^T \boldsymbol{\Phi}_n \right)^{-1} + \left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1}. \qquad (40)$$

The above equation can be recast as the standard prediction Riccati equation of [33], using the notation defined therein for mathematical convenience, as

$$\bar{\mathbf{M}}_{n+1|n} = \mathbf{f}_p \left(\bar{\mathbf{M}}_{n|n-1}^{-1} + \mathbf{h}_p^T \mathbf{r}_p^{-1} \mathbf{h}_p \right) \mathbf{f}_p^T + \mathbf{q}_p, \qquad (41)$$

where the various quantities above can be seen to be given from (40) as

$$\mathbf{q}_p = \left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1}, \mathbf{h}_p = \boldsymbol{\Phi}_n, \mathbf{f}_p = \rho \mathbf{I}, \mathbf{r}_p = \sigma^2 \mathbf{I}.$$

Similarly, for the estimator at steady state, it follows that $\bar{\mathbf{M}}_{n+1|n+1} = \bar{\mathbf{M}}_{n|n}$. On substituting the Kalman gain matrix $\bar{\mathbf{K}}_{n+1}$ given by Equation (37) into Equation (38), the matrix $\bar{\mathbf{M}}_{n+1|n+1}$ can be simplified to the form shown in Equation (33). Using the Woodbury Identity [25] once again, Equation (33) can be mathematically manipulated to obtain the following equation

$$\bar{\mathbf{M}}_{n+1|n+1} = \left(\bar{\mathbf{M}}_{n+1|n}^{-1} + \sigma^{-2} \boldsymbol{\Phi}_{n+1}^T \boldsymbol{\Phi}_{n+1} \right)^{-1}. \qquad (42)$$

Further simplification of the above equation requires the evaluation of $\bar{\mathbf{M}}_{n+1|n}^{-1}$, which can be carried out as follows. Starting with (40), one can simplify the expression for $\bar{\mathbf{M}}_{n+1|n}^{-1}$ using the Woodbury matrix identity, to the form shown in (34). This can now be substituted in (42) followed by a simplification to obtain the expression for $\bar{\mathbf{M}}_{n+1|n+1}$ as shown in (43), which on further mathematical manipulation and rearrangement of the terms yields (44). Upon comparison to the standard Riccati equation of [33], the expression in (44) can also be represented in the form

$$\bar{\mathbf{M}}_{n+1|n+1} = \mathbf{q}_e + \mathbf{f}_e \left[\bar{\mathbf{M}}_{n|n}^{-1} + \mathbf{h}_e^T \mathbf{r}_e^{-1} \mathbf{h}_e \right]^{-1} \mathbf{f}_e^T, \qquad (45)$$

where the pertinent quantities above are defined as

$$\mathbf{q}_e = \left[\left(\left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1} \right)^{-1} + \boldsymbol{\Phi}_{n+1}^T \sigma^{-2} \boldsymbol{\Phi}_{n+1} \right]^{-1},$$

$$\mathbf{f}_e = \rho \mathbf{q}_e \left[\left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1} \right]^{-1}, \qquad (46)$$

$$\mathbf{r}_e = \sigma^2 \mathbf{I} + \boldsymbol{\Phi}_{n+1} \left(1 - \rho^2 \right) \bar{\boldsymbol{\Gamma}}_{n+1} \boldsymbol{\Phi}_{n+1}^T \text{ and } \mathbf{h}_e = \rho \boldsymbol{\Phi}_{n+1}. \qquad (47)$$

Finally, Equation (45) can now be solved using the standard Discrete-time Algebraic Riccati Equation (DARE) solver [34], after substituting the matrix $\boldsymbol{\Phi}_n$ corresponding the P-HBKF shown in (35), to obtain the numerical value of the steady

$$\bar{\mathbf{M}}_{n+1|n+1} = \Big[\big((1-\rho^2)\,\bar{\boldsymbol{\Gamma}}_{n+1} \big)^{-1} - \big((1-\rho^2)\,\bar{\boldsymbol{\Gamma}}_{n+1} \big)^{-1} \rho^2 \Big[\bar{\mathbf{M}}_{n|n}^{-1} + \rho^2 \big((1-\rho^2)\,\bar{\boldsymbol{\Gamma}}_{n+1} \big)^{-1} \Big]^{-1} \times$$
$$\big((1-\rho^2)\,\bar{\boldsymbol{\Gamma}}_{n+1} \big)^{-1} + \boldsymbol{\Phi}_{n+1}^T \sigma^{-2} \boldsymbol{\Phi}_{n+1} \Big]^{-1} \tag{43}$$

$$\bar{\mathbf{M}}_{n+1|n+1} = \mathbf{q}_e + \rho \mathbf{q}_e \big((1-\rho^2)\,\bar{\boldsymbol{\Gamma}}_{n+1} \big)^{-1} \times$$
$$\Big[\bar{\mathbf{M}}_{n|n}^{-1} + \rho \boldsymbol{\Phi}_{n+1}^T \big[\sigma^2 \mathbf{I} + \boldsymbol{\Phi}_{n+1} (1-\rho^2)\,\bar{\boldsymbol{\Gamma}}_{n+1} \boldsymbol{\Phi}_{n+1}^T \big]^{-1} \rho \boldsymbol{\Phi}_{n+1} \Big]^{-1} \rho \mathbf{q}_e \big((1-\rho^2)\,\bar{\boldsymbol{\Gamma}}_{n+1} \big)^{-1} \tag{44}$$
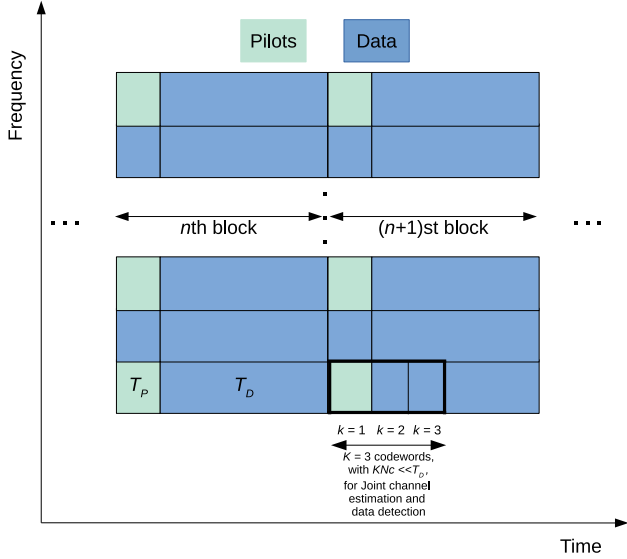
Fig. 3. Time-frequency block structure for data aided channel estimation

state error covariance matrix $\bar{\mathbf{M}}_\infty \triangleq \lim_{n\to\infty} \bar{\mathbf{M}}_{n|n}$. The asymptotic MSE is then expressed as

$$\text{MSE}_\infty = \text{Tr}\big(\bar{\mathbf{M}}_\infty \big). \tag{48}$$

To improve the performance further over that of the P-HBKF, which exclusively employs pilot symbols, a technique for HBKF based joint doubly-selective channel estimation and data detection is presented in the next section.

VI. D-HBKF FOR JOINT DOUBLY-SELECTIVE CHANNEL ESTIMATION AND DATA DETECTION

As illustrated in Fig. 3, in D-HBKF, K consecutive MIMO-OFDM codewords are employed for channel estimation. In the first MIMO-OFDM codeword, pilot symbols are transmitted over N_P pilot subcarriers in the interval of duration T_P, while the remaining non-pilot subcarriers in the interval T_P and all the subcarriers in the subsequent interval of duration T_D are used for data transmission. Thus, the pilot overhead in D-HBKF is identical to that of P-HBKF. The proposed D-HBKF technique exploits all the available outputs pertaining to the pilot as well as the data subcarriers over the K MIMO-OFDM codewords for joint channel estimation and data detection, in turn leading to an improved estimate. Let $s_{k,n_s,n}(m)$, $1 \le n_s \le N_s$, denote the transmit symbols drawn from a suitable constellation in the kth, $1 \le k \le K$, MIMO-OFDM codeword over mth subcarrier. The OSTBC codeword matrix $\mathbf{X}_{k,n}(m) \in \mathbb{C}^{N_T \times N_c}$ corresponding to the

mth subcarrier in the kth codeword can be obtained using (1) as

$$\mathbf{X}_{k,n}(m) = \alpha \sum_{n_s=1}^{N_s} \mathbf{P}_{n_s} \Re\{s_{k,n_s,n}(m)\} + j\mathbf{Q}_{n_s} \Im\{s_{k,n_s,n}(m)\}. \tag{49}$$

Let the complex scalar quantity $X_{k,t,n_c,n}(m)$ denote the (t,n_c)th element of the OSTBC codeword matrix $\mathbf{X}_{k,n}(m)$ and let the vector $\mathbf{x}_{k,t,n_c,n} \in \mathbb{C}^{N \times 1}$ be defined as

$$\mathbf{x}_{k,t,n_c,n} = [X_{k,t,n_c,n}(1), X_{k,t,n_c,n}(2), \dots, X_{k,t,n_c,n}(N)]^T. \tag{50}$$

The signal $\mathbf{y}_{k,r,n_c,n} \in \mathbb{C}^{N \times 1}$ received by the rth RA after CP removal, followed by the FFT operation can be expressed as

$$\mathbf{y}_{k,r,n_c,n} = [\text{diag}(\mathbf{x}_{k,1,n_c,n}), \dots, \text{diag}(\mathbf{x}_{k,N_T,n_c,n})] \times$$
$$(\mathbf{I}_{N_T} \otimes \mathbf{F})\, \mathbf{h}_{r,n} + \mathbf{w}_{k,r,n_c,n}, \tag{51}$$

where the vector $\mathbf{w}_{k,r,n_c,n} \in \mathbb{C}^{N \times 1}$ denotes the additive noise, which is assumed to be comprised of symmetric zero-mean i.i.d. complex Gaussian noise samples of variance σ^2 and the vector $\mathbf{h}_{r,n}$ is defined in (4). Let the vectors $\mathbf{y}_{k,r,n} \in \mathbb{C}^{NN_c \times 1}$ and $\mathbf{w}_{k,r,n} \in \mathbb{C}^{NN_c \times 1}$ represent the concatenated receive and noise vectors corresponding to the rth RA across all the N_c time instants in the kth codeword, respectively, which are defined as

$$\mathbf{y}_{k,r,n} = \big[\mathbf{y}_{k,r,1,n}^T, \mathbf{y}_{k,r,2,n}^T, \dots, \mathbf{y}_{k,r,N_c,n}^T \big]^T,$$
$$\mathbf{w}_{k,r,n} = \big[\mathbf{w}_{k,r,1,n}^T, \mathbf{w}_{k,r,2,n}^T, \dots, \mathbf{w}_{k,r,N_c,n}^T \big]^T. \tag{52}$$

Let $\mathbf{Y}_{k,n} \in \mathbb{C}^{NN_c \times N_R}$ and $\mathbf{W}_{k,n} \in \mathbb{C}^{NN_c \times N_R}$ denote the overall concatenated received signal and noise matrices for the kth codeword defined as

$$\mathbf{Y}_{k,n} = [\mathbf{y}_{k,1,n}, \dots, \mathbf{y}_{k,N_R,n}],$$
$$\mathbf{W}_{k,n} = [\mathbf{w}_{k,1,n}, \dots, \mathbf{w}_{k,N_R,n}]. \tag{53}$$

The OSTBC-based MIMO-OFDM system model corresponding to the kth codeword in the nth TB can now be succinctly represented as

$$\mathbf{Y}_{k,n} = \boldsymbol{\Psi}_{k,n} \mathbf{H}_n + \mathbf{W}_{k,n}, \tag{54}$$

where the dictionary matrix $\boldsymbol{\Psi}_{k,n} \in \mathbb{C}^{NN_c \times LN_t}$ and the corresponding MIMO-OFDM codeword matrix $\widetilde{\mathbf{X}}_{k,n} \in$

$\mathbb{C}^{NN_c \times NN_T}$ are given by:

$$\boldsymbol{\Psi}_{k,n} = \widetilde{\mathbf{X}}_{k,n} \left(\mathbf{I}_{N_t} \otimes \mathbf{F}\right),$$

$$\widetilde{\mathbf{X}}_{k,n} = \begin{bmatrix} \mathrm{diag}\left(\mathbf{x}_{k,1,1,n}\right) & \ldots & \mathrm{diag}\left(\mathbf{x}_{k,N_T,1,n}\right) \\ \vdots & \ddots & \vdots \\ \mathrm{diag}\left(\mathbf{x}_{k,1,N_c,n}\right) & \ldots & \mathrm{diag}\left(\mathbf{x}_{k,N_T,N_c,n}\right) \end{bmatrix}.$$

(55)

The matrix $\mathbf{H}_n$ denotes the resultant channel matrix corresponding to the nth block as defined previously in (6). Finally, the concatenated receiver noise and dictionary matrices corresponding to nth block are denoted by $\mathbf{Y}_{\mathcal{B},n} \in \mathbb{C}^{KNN_c \times N_R}$, $\mathbf{W}_{\mathcal{B},n} \in \times\mathbb{C}^{KNN_c \times N_R}$ and $\boldsymbol{\Psi}_{\mathcal{B},n} \in \mathbb{C}^{KNN_c \times LN_T}$, respectively, which can be obtained as

$$\mathbf{Y}_{\mathcal{B},n} = \left[\mathbf{Y}_{1,n}^T, \ldots, \mathbf{Y}_{K,n}^T\right],$$
$$\mathbf{W}_{\mathcal{B},n} = \left[\mathbf{W}_{1,n}^T, \ldots, \mathbf{W}_{K,n}^T\right],$$
$$\boldsymbol{\Psi}_{\mathcal{B},n} = \left[\boldsymbol{\Psi}_{1,n}^T, \ldots, \boldsymbol{\Psi}_{K,n}^T\right].$$

(56)

Thus, the overall system model corresponding to the nth TB follows as

$$\mathbf{Y}_{\mathcal{B},n} = \boldsymbol{\Psi}_{\mathcal{B},n}\mathbf{H}_n + \mathbf{W}_{\mathcal{B},n}.$$

(57)

The vectorized form of the OSTBC-based MIMO-OFDM system model formulated above for joint channel estimation and data detection can be obtained as

$$\mathbf{y}_{\mathcal{B},n} = \mathrm{vec}\left(\mathbf{Y}_{\mathcal{B},n}\right) = \boldsymbol{\Phi}_{\mathcal{B},n}\mathbf{h}_n + \mathbf{w}_{\mathcal{B},n},$$

(58)

where the effective block dictionary matrix can be expressed as $\boldsymbol{\Phi}_{\mathcal{B},n} = \left(\mathbf{I}_{N_R} \otimes \boldsymbol{\Psi}_{\mathcal{B},n}\right) \in \mathbb{C}^{KNN_cN_R \times LN_TN_R}$ and the stacked noise vector is given by $\mathbf{w}_{\mathcal{B},n} = \mathrm{vec}\left(\mathbf{W}_{\mathcal{B},n}\right) \in \mathbb{C}^{KNN_cN_R \times 1}$. The D-HBKF procedure of jointly estimating the doubly-selective sparse channel and the OSTBC data codewords can be formulated as follows. The measurement prediction error vector $\mathbf{y}_{\mathcal{B},e,n}$ in the HBKF framework can be defined as

$$\mathbf{y}_{\mathcal{B},e,n} = \mathbf{y}_{\mathcal{B},n} - \widehat{\mathbf{y}}_{\mathcal{B},n|n-1} \approx \sqrt{1-\rho^2}\boldsymbol{\Phi}_{\mathcal{B},n}\mathbf{u}_n + \mathbf{w}_{\mathcal{B},n},$$

(59)

where the predicted measurement $\widehat{\mathbf{y}}_{\mathcal{B},n|n-1}$ may be determined using the standard KF, as shown in the Appendix A. The D-HBKF begins its operation by assigning a Gaussian parameterized prior similar to (19) to the innovation noise vector $\mathbf{u}_n$ with the hyperparameter matrix $\boldsymbol{\Gamma}_n$. The complete data for this scenario is $\{\mathbf{y}_{\mathcal{B},e,n}, \mathbf{u}_n\}$, while the unknown parameter set is $\boldsymbol{\theta}_n = \{\boldsymbol{\Gamma}_n, \mathbf{X}_{\mathcal{D},n}\}$, where the matrix $\mathbf{X}_{\mathcal{D},n} \in \mathbb{C}^{N_T \times (KN-N_P)N_c}$ corresponding to the unknown data codewords is constructed as follows. Let $\mathcal{P}^c$ denote the set of non-pilot or data subcarriers. The concatenated matrix of OSTBC codewords corresponding to the non-pilot subcarriers for the first codeword duration can be defined as $\mathbf{X}_{\mathcal{D},1,n} = \left[\mathbf{X}_{1,n}(m_1), \ldots, \mathbf{X}_{1,n}(m_{N-N_P})\right] \in \mathbb{C}^{N_T \times (N-N_P)N_c}$, where $m_j \in \mathcal{P}^c$. Similarly, the concatenated matrix $\mathbf{X}_{\mathcal{D},k,n}, k = 2,3,\ldots,K$, is constructed as $\mathbf{X}_{\mathcal{D},k,n} = \left[\mathbf{X}_{k,n}(1), \ldots, \mathbf{X}_{k,n}(N)\right] \in \mathbb{C}^{N_T \times NN_c}$. The overall block data codeword matrix $\mathbf{X}_{\mathcal{D},n}$ can be further defined as $\mathbf{X}_{\mathcal{D},n} = \left[\mathbf{X}_{\mathcal{D},1,n}, \mathbf{X}_{\mathcal{D},2,n}, \ldots, \mathbf{X}_{\mathcal{D},K,n}\right]$. The D-HBKF jointly estimates the associated hyperparameters $\boldsymbol{\Gamma}_n$ and the data codeword matrix $\mathbf{X}_{\mathcal{D},n}$ using the EM procedure derived below.

Fig. 4. Block diagram representation of the D-HBKF scheme for joint doubly-selective sparse channel estimation and data detection in OSTBC MIMO-OFDM systems

Let $\widehat{\mathbf{X}}_{\mathcal{D},n}^{(i-1)}$ and $\widehat{\boldsymbol{\gamma}}_n^{(i-1)}$ denote the estimates of the overall data codeword matrix and the hyperparameter vector obtained after the completion of the $(i-1)$st EM iteration, with $\boldsymbol{\theta}_n^{(i-1)} = \left\{\widehat{\mathbf{X}}_{\mathcal{D},n}^{(i-1)}, \widehat{\boldsymbol{\gamma}}_n^{(i-1)}\right\}$. The E-step evaluates the average log-likelihood $\mathcal{L}\left(\boldsymbol{\theta}_n|\boldsymbol{\theta}_n^{(i-1)}\right)$ in the ith iteration as

$$\mathcal{L}\left(\boldsymbol{\theta}_n|\boldsymbol{\theta}_n^{(i-1)}\right) = \mathbb{E}_{\mathbf{u}_n|\mathbf{y}_{\mathcal{B},e,n};\boldsymbol{\theta}_n^{(i-1)}} \left\{\log p\left(\mathbf{y}_{\mathcal{B},e,n}, \mathbf{u}_n; \boldsymbol{\theta}_n\right)\right\}$$
$$= \mathbb{E}_{\mathbf{u}_n|\mathbf{y}_{\mathcal{B},e,n};\boldsymbol{\theta}_n^{(i-1)}} \left\{\log p\left(\mathbf{y}_{\mathcal{B},e,n}|\mathbf{u}_n; \mathbf{X}_{\mathcal{D},n}\right)\right\}$$
$$+ \mathbb{E}_{\mathbf{u}_n|\mathbf{y}_{\mathcal{B},e,n};\boldsymbol{\theta}_n^{(i-1)}} \left\{\log p\left(\mathbf{u}_n; \boldsymbol{\gamma}_n\right)\right\}.$$

The quantity $\log p\left(\mathbf{y}_{\mathcal{B},e,n}|\mathbf{u}_n; \mathbf{X}_{\mathcal{D},n}\right)$ is seen to be independent of the hyperparameter vector $\boldsymbol{\gamma}_n$. Furthermore, it can be observed from (19) that the term $\log p(\mathbf{u}_n; \boldsymbol{\gamma}_n)$ is independent of the block data codeword matrix $\mathbf{X}_{\mathcal{D},n}$. Therefore, the M-step of the D-HBKF scheme reduces to the decoupled maximization of the first and second terms of $\mathcal{L}\left(\boldsymbol{\theta}_n|\boldsymbol{\theta}_n^{(i-1)}\right)$ with respect to $\mathbf{X}_{\mathcal{D},n}$ and $\boldsymbol{\gamma}_n$, respectively. The log-likelihood maximization with respect to $\boldsymbol{\gamma}_n$ in the ith iteration of the M-step can be expressed as

$$\widehat{\boldsymbol{\gamma}}_n^{(i)} = \arg\max_{\boldsymbol{\gamma}_n} \mathbb{E}_{\mathbf{u}_n|\mathbf{y}_{\mathcal{B},e,n};\widehat{\boldsymbol{\gamma}}_n^{(i-1)}} \left\{\log p\left(\mathbf{u}_n; \boldsymbol{\gamma}_n\right)\right\}. \quad (60)$$

Similar to P-HBKF, the hyperparameter estimates $\widehat{\boldsymbol{\gamma}}_n^{(i)}$ for D-HBKF are obtained as in (23) with the *a posterior* mean vector $\boldsymbol{\mu}_{\mathcal{B},n}^{(i)} \in \mathbb{C}^{LN_RN_T \times 1}$ and the covariance matrix $\boldsymbol{\Sigma}_{\mathcal{B},n}^{(i)} \in \mathbb{C}^{LN_RN_T \times LN_RN_T}$ derived as

$$\boldsymbol{\mu}_{\mathcal{B},n}^{(i)} = \sigma^{-2}\sqrt{1-\rho^2}\boldsymbol{\Sigma}_{\mathcal{B},n}^{(i)} \left(\boldsymbol{\Phi}_{\mathcal{B},n}^{(i)}\right)^H \mathbf{y}_{\mathcal{B},e,n}, \quad (61)$$

$$\boldsymbol{\Sigma}_{\mathcal{B},n}^{(i)} = \left(\frac{(1-\rho^2)}{\sigma^{-2}}\left(\boldsymbol{\Phi}_{\mathcal{B},n}^{(i)}\right)^H \boldsymbol{\Phi}_{\mathcal{B},n}^{(i)} + \left(\mathbf{I} \otimes \widehat{\boldsymbol{\Gamma}}_n^{(i-1)}\right)^{-1}\right)^{-1},$$

(62)

where the quantity $\boldsymbol{\Phi}_{\mathcal{B},n}^{(i)}$ is obtained similar to $\boldsymbol{\Phi}_{\mathcal{B},n}$ using the estimated data codeword matrix $\widehat{\mathbf{X}}_{\mathcal{D},n}^{(i-1)}$ after the $(i-1)$st EM iteration. The estimate $\widehat{\mathbf{X}}_{\mathcal{D},n}^{(i)}$ at the ith iteration is obtained as

$$\widehat{\mathbf{X}}_{\mathcal{D},n}^{(i)} = \arg\max_{\mathbf{X}_{\mathcal{D},n}} \mathbb{E}_{\mathbf{h}_n|\mathbf{y}_{\mathcal{B},n};\boldsymbol{\theta}_n^{(i-1)}} \left\{\log p\left(\mathbf{y}_{\mathcal{B},n}|\mathbf{h}_n; \mathbf{X}_{\mathcal{D},n}\right)\right\}.$$

(63)

The above maximization leads to the decoupled optimization problem in (68) for the evaluation of the individual estimates

$\widehat{\mathbf{X}}_{k,n}^{(i)}(m)$. The quantity $\widetilde{\mathbf{Y}}_{k,n}(m) \in \mathbb{C}^{N_R \times N_c}$ denotes the output matrix for the kth codeword over the mth subcarrier in the nth TB. The quantity $\boldsymbol{\mathcal{H}}_n(m) \in \mathbb{C}^{N_R \times N_T}$ represents the MIMO channel matrix corresponding to the mth subcarrier in the nth TB, which is defined as

$$\boldsymbol{\mathcal{H}}_n(m) = \begin{bmatrix} (\mathbf{Fh}_{1,1,n})_m & (\mathbf{Fh}_{2,1,n})_m & \cdots & (\mathbf{Fh}_{N_T,1,n})_m \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{Fh}_{1,N_R,n})_m & (\mathbf{Fh}_{2,N_R,n})_m \cdots & (\mathbf{Fh}_{N_T,N_R,n})_m \end{bmatrix},$$

(64)

where $(\mathbf{Fh}_{t,r,n})_m$ denotes the mth DFT coefficient of the channel vector $\mathbf{h}_{t,r,n}$. For decoding the transmitted symbols, one can define the quantities $\mathbf{z}_{k,n}(m) = \mathrm{vec}\,(\widetilde{\mathbf{Y}}_{k,n}(m)) \in \mathbb{C}^{N_R N_c \times 1}$ and the symbol vector estimate $\widehat{\mathbf{a}}_{k,n}^{(i)}(m) = \left[\left(\Re\left\{ \widehat{\mathbf{s}}_{k,n}^{(i)}(m) \right\} \right)^T \left(\Im\left\{ \widehat{\mathbf{s}}_{k,n}^{(i)}(m) \right\} \right)^T \right]^T$. The optimization problem of evaluating the symbol estimate $\widehat{\mathbf{a}}_{k,n}^{(i)}(m)$ corresponding to $\widehat{\mathbf{X}}_{k,n}^{(i)}(m)$ is now formulated in (69). This can be further simplified as follows. Let the effective channel matrix $\boldsymbol{\mathcal{C}}_n(m) \in \mathbb{C}^{N_R N_c \times 2N_s}$ for the nth TB be constructed as

$$\boldsymbol{\mathcal{C}}_n(m) = [\boldsymbol{\mathcal{C}}_{p,n}(m), \boldsymbol{\mathcal{C}}_{q,n}(m)],$$

(65)

where the matrices $\boldsymbol{\mathcal{C}}_{p,n}(m) \in \mathbb{C}^{N_R N_c \times N_s}, \boldsymbol{\mathcal{C}}_{q,n}(m) \in \mathbb{C}^{N_R N_c \times N_s}$ above are defined as [26]

$$\boldsymbol{\mathcal{C}}_{p,n}(m) = [\mathrm{vec}\,(\boldsymbol{\mathcal{H}}_n(m)\,\mathbf{P}_1),\ldots,\mathrm{vec}\,(\boldsymbol{\mathcal{H}}_n(m)\,\mathbf{P}_{N_s})],$$

(66)

$$\boldsymbol{\mathcal{C}}_{q,n}(m) = [\mathrm{vec}\,(\boldsymbol{\mathcal{H}}_n(m)\,\mathbf{Q}_1),\ldots,\mathrm{vec}\,(\boldsymbol{\mathcal{H}}_n(m)\,\mathbf{Q}_{N_s})].$$

(67)

Using the above quantities, together with the fact that the noise is additive white Gaussian in nature, the log-likelihood and also consequently the detector formulated in (69) can be further simplified as shown in (70). Upon using the result $\Re\left\{ \boldsymbol{\mathcal{C}}_n^H(m)\boldsymbol{\mathcal{C}}_n(m) \right\} = \|\boldsymbol{\mathcal{H}}_n(m)\|_F^2\,\mathbf{I}_{2N_s}$ for OSTBCs [26] followed by further simplification, the D-HBKF based symbol vector estimate $\widehat{\mathbf{a}}_{k,n}^{(i)}(m)$ can be obtained as

$$\widehat{\mathbf{a}}_{k,n}^{(i)}(m) = \frac{\Re\left\{ \left(\widehat{\boldsymbol{\mathcal{C}}}_n^{(i)}(m) \right)^H \mathbf{z}_{k,n}(m) \right\}}{\zeta_n^{(i)}(m)},$$

(71)

where the estimate of the effective channel matrix $\widehat{\boldsymbol{\mathcal{C}}}_n^{(i)}(m)$, similar to (65), is obtained upon replacing $\boldsymbol{\mathcal{H}}_n(m)$ by its estimate $\widehat{\boldsymbol{\mathcal{H}}}_n^{(i)}(m)$ obtained at the ith iteration in (66) and (67). The *a posteriori* MIMO channel estimate $\widehat{\boldsymbol{\mathcal{H}}}_n^{(i)}(m)$ corresponding to the mth subcarrier can be constructed by substituting the estimate $\widehat{\mathbf{h}}_{n|n}^{(i)}$, obtained using the KF equations outlined in Appendix A, with $\boldsymbol{\Phi}_{\mathcal{P}}$ replaced by $\boldsymbol{\Phi}_{\mathcal{B},n}$. Furthermore, the quantity $\zeta_n^{(i)}(m)$ in the denominator of (71) can be shown to be given as

$$\zeta_n^{(i)}(m) = \left\| \widehat{\boldsymbol{\mathcal{H}}}_n^{(i)}(m) \right\|_F^2 + \sum_{r=1}^{N_R}\sum_{t=1}^{N_T} \left(\widetilde{\mathbf{M}}_n^{(i)}\,(\bar{d},\bar{d}) \right),$$

(72)

where the index $\bar{d}$ obeys $\bar{d} = [(r-1)N_T + (t-1)]\,N + m$, $\forall\ 1 \le r \le N_R, 1 \le t \le N_T$ and $\widetilde{\mathbf{M}}_n^{(i)} = (\mathbf{I}_{N_T N_R} \otimes$

$\mathbf{F})\mathbf{M}_{n|n}^{(i)}(\mathbf{I}_{N_T N_R}\otimes\mathbf{F})^H \in \mathbb{C}^{NN_T N_R \times NN_T N_R}$ is obtained using $\mathbf{M}_{n|n}^{(i)}$ computed as one of the outputs of the KF procedure in Appendix A.

Upon convergence, the D-HBKF based joint estimates of the MIMO channel and data symbols are given by $\widehat{\mathbf{h}}_{\mathrm{D},n} = \widehat{\mathbf{h}}_{n|n}^{(i)}$ and $\widehat{\mathbf{s}}_{k,n}^{(i)}(m)$, respectively. The BCRB and the asymptotic MSE for the D-HBKF based channel estimate can be derived along similar lines to that of the P-HBKF upon replacing $\mathbf{y}_{\mathcal{P},n}$ and $\boldsymbol{\Phi}_{\mathcal{P}}$ by $\mathbf{y}_{\mathcal{B},n}$ and $\boldsymbol{\Phi}_{\mathcal{B},n}$, respectively, in (29), (35). We note that the Kalman filtering time-evolution model, also used in [21], [35], requires the knowledge of the temporal correlation ρ, which in turn depends on the Doppler shift of the channel, and on the noise variance σ^2. However, as shown in the results of Section-VII(A), the performance of the proposed HBKF-based techniques is robust to mismatches in the values of the parameters ρ, σ^2, thus making them eminently suited for practical implementation.

A. Computational Complexity Analysis

Due to lack of space, the detailed derivations of the computational complexity had to be relegated to our technical report in [36]. The key implications of the results in Table I and II therein are discussed below. It is evident that the joint channel estimation and data detection of the K MIMO-OFDM codewords in the D-HBKF scheme leads to a higher complexity of $\mathcal{O}(L^3 N_R^3 N_T^3 + N_R^3 N_T^3 N^2 L + K^2 N^2 N_c^2 N_R^3 L N_T)$, which arises due to the fact that it necessitates the processing of the $KNN_c N_R \times 1$ vector $\mathbf{y}_{\mathcal{B},n}$ followed by the subsequent data detection procedure using the EM framework. By contrast, the P-HBKF technique described in Section III employs the $N_P N_c N_R \times 1$ vector $\mathbf{y}_{\mathcal{P},n}$ for channel estimation, which is extracted from the first MIMO-OFDM pilot codeword of each TB. Note that the number of pilot subcarriers and the total number of subcarriers obey $N_P << N$. Moreover, since the data detection procedure in P-HBKF is decoupled from the EM framework, this leads to a lower computational complexity of $\mathcal{O}\left(L^3 N_R^3 N_T^3 + N_P^2 N_c^2 N_R^3 L N_T \right)$. Naturally, while the overall complexity of the P-HBKF scheme is lower in comparison to D-HBKF, as shown in the simulation results of Section VII, the performance of the former is also poor in comparison to that of the D-HBKF algorithm.

VII. SIMULATION RESULTS

An OSTBC coded MIMO-OFDM system having $N_R = 2$ RAs and $N_T \in \{2,4\}$ TAs is considered. The number of subcarriers is set to $N = 64$ with a bandwidth of 5 MHz. The OFDM symbols have a duration of 12.8 μs and a cyclic prefix of duration 6.4 μs. Quadrature phase shift keying (QPSK) modulation is considered for the symbols followed by OSTBC encoding, with the performance demonstrated both for the full-rate Alamouti code [3], as well as for the rate-$\frac{3}{4}$ non-square OSTBCs described in [26]. The $\{N_s, N_c, N_T\}$ parameter set is fixed as $\{2,2,2\}$ and $\{3,4,4\}$, respectively. In each OFDM symbol, a total of $N_P = 22$ pilots are loaded across the uniformly-spaced subcarriers, for each TA. The transmission frame is comprised of $R = 10$ TBs. For P-HBKF, the pilot

$$\widehat{\mathbf{X}}_{k,n}^{(i)}(m) = \arg \max_{\mathbf{X}_{k,n}(m)} \mathbb{E}_{\boldsymbol{\mathcal{H}}_n(m)|\widetilde{\mathbf{Y}}_{k,n}(m);\widehat{\mathbf{X}}_{k,n}^{(i-1)}(m)} \left\{ \log p \left(\widetilde{\mathbf{Y}}_{k,n}(m) | \boldsymbol{\mathcal{H}}_n(m) ; \mathbf{X}_{k,n}(m) \right) \right\} \tag{68}$$

$$\widehat{\mathbf{a}}_{k,n}^{(i)}(m) = \arg \max_{\mathbf{a}_{k,n}(m)} \mathbb{E}_{\boldsymbol{\mathcal{H}}_n(m)|\mathbf{z}_{k,n}(m);\widehat{\mathbf{a}}_{k,n}^{(i-1)}(m)} \left\{ \log p \left(\mathbf{z}_{k,n}(m) | \boldsymbol{\mathcal{H}}_n(m) ; \mathbf{a}_{k,n}(m) \right) \right\} \tag{69}$$

$$\widehat{\mathbf{a}}_{k,n}^{(i)}(m) = \arg \min_{\mathbf{a}_{k,n}(m)} \mathbb{E}_{\boldsymbol{\mathcal{H}}_n(m)|\mathbf{z}_{k,n}(m);\widehat{\mathbf{a}}_{k,n}^{(i-1)}(m)} \left\{ \| \mathbf{z}_{k,n}(m) - \boldsymbol{\mathcal{C}}_n(m) \mathbf{a}_{k,n}(m) \|^2 \right\} \tag{70}$$

(a)   (b)
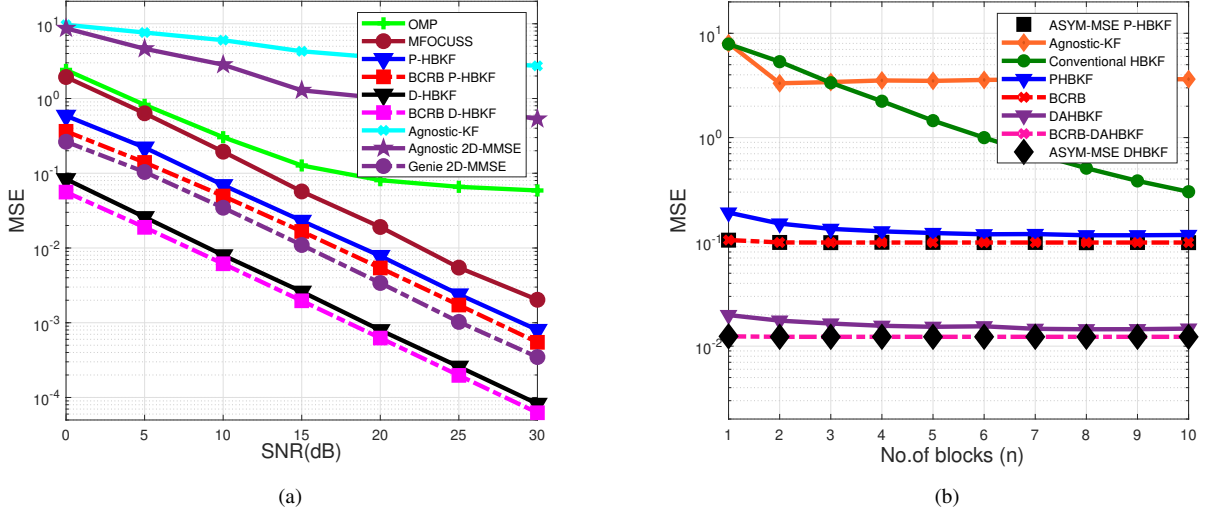
Fig. 5. (a) MSE versus SNR performance comparison (b) MSE performance comparison versus number of blocks n for Alamouti coded 2×2 MIMO-OFDM for the following techniques: Agnostic-KF [25], Agnostic 2D-MMSE [36], Genie 2D-MMSE [36], OMP [37], FOCUSS [12], SBL [18], P-HBKF, D-HBKF along with their respective BCRBs (31) and asymptotic MSE (48).

(a)   (b)
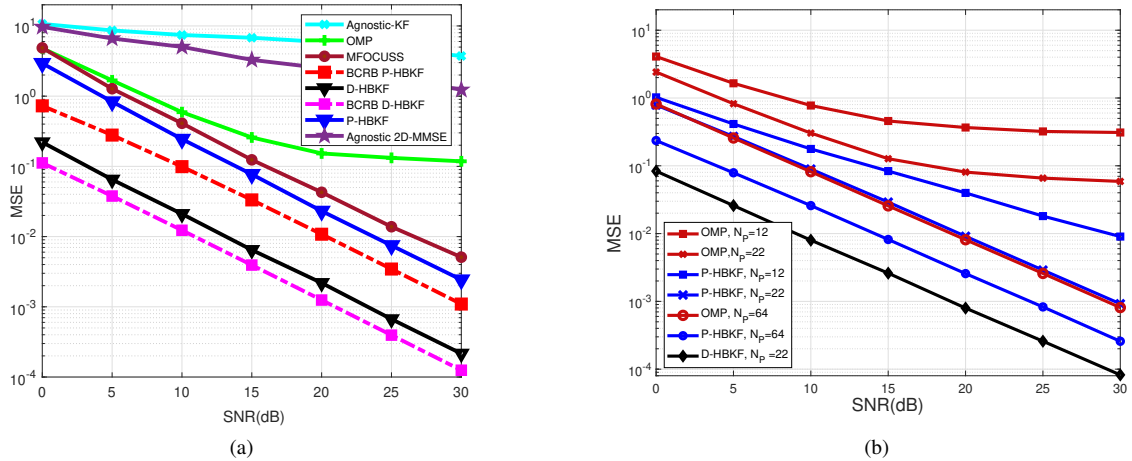
Fig. 6. (a) MSE versus SNR performance comparison for rate $\frac{3}{4}$ non-square space time block coded [26] 8×4 MIMO-OFDM (b) MSE versus SNR performance comparison for Alamouti coded 2×2 MIMO-OFDM by varying number of subcarriers used as pilots, for the following techniques: Agnostic 2D-MMSE [36], OMP [37], FOCUSS [12], SBL [18], P-HBKF, D-HBKF along with their respective BCRBs (31).

duration T_P in each TB is set to the number of slots N_c occupied by a single OSTBC codeword, while for D-HBKF, it is assumed to be equal to $K = 3$ OSTBC codeword slots, i.e., $T_P = 3N_c$. The hyperparameters for the P-HBKF and D-HBKF schemes are initialized as indicated in Algorithm 1. The initial estimates of the data symbols on the non-pilot subcarriers $\widehat{\mathbf{a}}_{k,n}^{(0)}(m)$, $\forall \ m \in \mathcal{P}^c, 1 \leq k \leq K, 1 \leq n \leq R$, for the D-HBKF algorithm, are obtained by using the P-

HBKF channel estimate for detection. The stopping parameters ϵ and $i_{\max}$ are set as 10^{-5} and 50, respectively, for both the schemes. The frequency-selective channel's CIR vector $\mathbf{h}_{t,r,n}$ is assumed to have $L = 32$ taps, which is equal to the CP length. The approximately sparse channel is generated using the standard Rayleigh fading vehicular-A channel model of [38] with $L_s = 6$ non-zero channel taps until specified otherwise. The MIMO-OFDM system is considered

to be deployed in the 2.5 GHz band at a mobile velocity of $v = 30$ km/h, leading to a Doppler shift of $f_D = 69$ Hz. The coherence time is set to $T_c = 2.1$ ms with a block length of $T_B = T_c$. Substituting the values of f_d and T_B in Jake's model, the temporal correlation parameter ρ of the time varying channel model in (10) is $\rho = J_0(2\pi f_d T_B) \approx 0.8$. For the parameters N_c, T_P, T_D set as above, considering the transmission of 39 OSTBC MIMO-OFDM data codewords, it follows that $T_B = T_P + T_D = 40 \times 51.2 \ \mu s = 2 \ ms$. This is well within the coherence time $T_c = 2.1 \ ms$. Thus, for this scenario, the wireless channel can be assumed to be approximately constant in each duration of length T_B.

The performance of the proposed P-HBKF and D-HBKF techniques for doubly-selective MIMO-OFDM channel estimation is also compared with that of the other popular sparse channel estimation techniques such as SBL [18], OMP [37] and FOCUSS [12]. The stopping criterion for the OMP is set for ensuring that the algorithm terminates when the residual error $E(i)$ between the consecutive iterations obeys $E(i + 1) - E(i) < \sigma^2$. The regularization parameter for MFOCUSS is set as the noise variance σ^2, the norm parameter $p = 0.8$ and the stopping threshold is set as 10^{-5} with the maximum number of iterations set to 800. The stopping parameters for SBL are set identical to that of the P-HBKF and D-HBKF schemes. The performance of the proposed schemes is also compared to 2D-MMSE [6], [36], which is an MMSE approach that estimates the channel's frequency response (CFR) using time-frequency correlation.

A. MSE Performance Comparison

The MSE performance of the channel estimates is presented next. The popular full-rate Alamouti OSTBC [3] associated with $N_T = N_R = 2$ is considered first. Fig. 5(a) plots the MSE performance versus SNR of the proposed P-HBKF and D-HBKF techniques along with that of the OMP, SBL, FOCUSS and 2D-MMSE. Furthermore, the BCRB corresponding to the doubly-selective sparse channel estimate obtained in (31) is also included in the figure. It can be observed that the proposed P-HBKF scheme achieves a lower MSE in comparison to the existing sparse approaches, viz. OMP, SBL and FOCUSS, which do not utilize the temporal correlation across multiple measurements. The Agnostic 2D-MMSE, which has no knowledge of the channel power delay profile (PDP) and hence assumes a typical exponential PDP, can be seen to perform poorly in terms of MSE of channel estimation. In contrast, the P-HBKF that also has no knowledge of the PDP, i.e., either of the non-zero tap locations or their respective powers, yields a significantly improved performance. The Genie 2D-MMSE, which assumes knowledge of the true channel PDP, has an MSE that is close to the BCRB. However, due to its ideal nature, it can act only as a performance benchmark to lower bound the performance of the proposed schemes. Finally, it is also interesting to note that the Genie 2D-MMSE, which is a block estimation scheme has a slightly improved performance than the BCRB, since the BCRB as derived in the paper is evaluated only for the forward Kalman iteration and lacks a smoothing step, in order to maintain the online nature

of the P-HBKF technique. Moreover, the MSE performance of the P-HBKF and D-HBKF techniques is close to that of the corresponding BCRB bounds, which is worth noting since the proposed P-HBKF and D-HBKF techniques do not require prior knowledge of the PDP, while the respective benchmarks do. Fig. 5(b) plots the MSE of the proposed and existing schemes versus number of blocks (n) at an SNR of 10dB, for an Alamouti-coded system. The performance is benchmarked with respect to the sparsity-agnostic KF, conventional-HBKF and BCRB. From the figure, it can be seen that the Agnostic-KF has the worst MSE performance that naturally arises owing to the fact that it ignores the channel-sparsity. It can be observed that the proposed P-HBKF and D-HBKF schemes achieve a significantly lower MSE than the conventional HBKF and also require significantly fewer blocks for convergence owing to the sparse initialization of the error covariance matrix and hyperparameter vector, as described in Subsection III-B. This also validates the assumption employed in the simplification of the HBKF in (16), which is met by the P-HBKF and D-HBKF due to the sparse initialization. By contrast, the conventional HBKF satisfies this criterion only after several blocks, thus leading to its slower convergence and to a significant performance gap with respect to the proposed techniques. Both the P-HBKF and D-HBKF yield performance curves that are significantly improved in comparison to the existing approaches, while also being close to the corresponding BCRBs. Fig. 5(b) also shows the asymptotic MSE performance, the analysis of which has been presented in Section V based on the solution of the discrete-time Riccati equation. It can be seen that the P-HBKF and D-HBKF schemes closely approach their asymptotic MSE benchmarks determined in (48) within as few as $n = 10$ blocks, thus demonstrating the rapid convergence of the proposed online estimation schemes.

In order to study the performance with respect to general non-square OSTBCs using a practical channel model, Fig. 6(a) demonstrates the performance of a rate-$\frac{3}{4}$ non-square OSTBC using $N_T = 4$ TAs and $N_R = 8$ RAs [26] for the 3GPP spatial channel model (SCM) described in [38]. A power delay profile with $L_s = 6$ non-zero channel taps is considered as stipulated by the vehicular-A model provided in the SCM link level parameters (Table 4.1 of [38]). The MSEs of the proposed and existing schemes are compared in Fig. 6(a) along with their respective BCRBs with the time domain samples of the channel generated using the SCM Matlab-package [39]. The P-HBKF and D-HBKF can once again be seen to outperform the existing schemes, which demonstrates their viability for practical applications. Fig. 6(b) depicts the MSE performance of the Alamouti OSTBC with $N_T = N_R = 2$, as the number of pilot subcarriers N_P is varied. The MSE performance of both the OMP and P-HBKF techniques is seen to improve upon increasing the number of pilot subcarriers, which can be attributed to the increased number of measurements. Furthermore, it can be observed that the performance of the P-HBKF scheme with $N_P = 22$ subcarriers is comparable to that of the OMP scheme with all the $N = N_P = 64$ subcarriers designated as pilot subcarriers, which demonstrates a significant reduction of around 65%
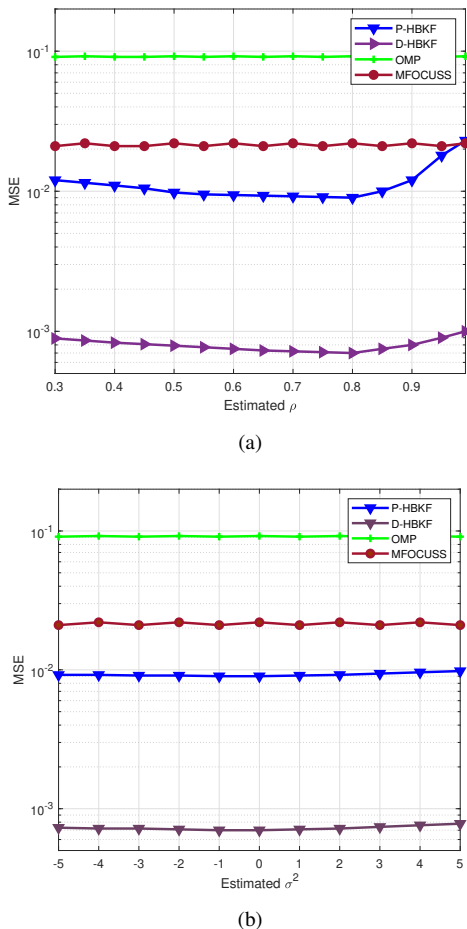
(a)

(b)

Fig. 7. (a) MSE versus mismatched correlation coefficient ρ (b) MSE versus mismatched noise variance σ^2 for an Alamouti coded 2×2 MIMO-OFDM system for the following techniques: OMP [37], FOCUSS [12], P-HBKF, D-HBKF.

in pilot overheads that can potentially be achieved by the proposed approach.

Fig. 7(a) and 7(b) present the MSE performance of the proposed P-HBKF and D-HBKF techniques in the presence of mismatches in the values of the correlation coefficient ρ and noise variance σ^2, respectively, at the receiver, for the Alamouti OSTBC using $N_T = N_R = 2$ and SNR= 20 dB. The true underlying values of these quantities are set as $\rho = 0.8$ and $\sigma^2 = 0$ dB. One can once again see that the HBKF-based schemes are robust to the effects of mismatch in the correlation coefficient ρ. The proposed D-HBKF technique exploits all the available KNN_c received measurements corresponding to the pilot as well as the data subcarriers over the K MIMO-OFDM codewords for channel estimation, whereas the P-HBKF technique is only able to leverage the $N_P N_c$ measurements corresponding to the pilot subcarriers. Thus, the large number of measurements makes D-HBKF robust against the parameter mismatches.

B. BER Performance Comparison

Fig. 8(a) compares the BER performance of the proposed and existing schemes for the full-rate Alamouti OSTBC with

$N_T = N_R = 2$ in a MIMO-OFDM system. This is benchmarked against a hypothetical genie receiver having perfect CSI. The sparse channel estimation accuracy improvement of the proposed HBKF based schemes is reflected in the BER reduction achieved by the P-HBKF in comparison to the existing schemes. The D-HBKF approach can be seen to lead to a further reduction in the BER, which can be attributed to its improved accuracy of channel estimation coupled with the modified ML detector derived in (71) that also incorporates the posteriori covariance matrix $\mathbf{M}_{n|n}^{(i)}$ of the channel estimation error. The theoretical BER values are computed utilizing the analytical expression in [22] by substituting the asymptotic covariance matrices obtained on solving Equation (45) for the P-HBKF and D-HBKF schemes. Fig. 8(b) considers the rate-$\frac{3}{4}$ non-square OSBTC MIMO-OFDM system using $N_T = 4$ TAs and $N_R = 2$ RAs for our BER performance comparison. It shows a trend similar to that of Fig. 8(a), with the D-HBKF yielding the best overall performance. The theoretical BER values for the square as well as non-square OSTBCs seen in Fig. 8(a), Fig. 8(b), respectively, are observed to be close to the corresponding values obtained via simulation. Thus, the improved estimation/ BER performance of the proposed P-HBKF and D-HBKF techniques coupled with the low-complexity online sparse channel estimation procedure makes them ideally suited for practical implementation in OSTBC-based MIMO-OFDM systems.

C. Extension of the Proposed Schemes to Next-Generation Systems

It is worth mentioning that the sparse channel estimation schemes developed in this work are not limited to the existing sub-6GHz MIMO-OFDM systems, and can be readily extended to large antenna array based next-generation millimeter wave (mmWave) and massive MIMO systems. As described in [40], [41], we can consider a millimeter wave hybrid MIMO architecture having $N_{RF}^T = 4$ radio frequency (RF) chains at the transmitter and $N_{RF}^R = 2$ RF chains at the receiver, i.e., a 2×4 MIMO in the baseband, and $N_T >> 4$ and $N_R >> 2$ TAs and RAs. In such systems, the RF precoder $\mathbf{F}_{RF} \in \mathbb{C}^{N_T \times N_{RF}^T}$ maps the output of the N_{RF}^T transmit RF chains to N_T transmit antennas, while its counterpart at the receiver, namely the RF combiner $\mathbf{W}_{RF}^H \in \mathbb{C}^{N_{RF}^R \times N_R}$ similarly maps the output of the N_R receive antennas to the N_{RF}^R receive RF chains. The pertinent channel estimation schemes for such systems have been proposed in [40]–[42], which exploit the sparsity inherent in the angular domain representation of the mmWave MIMO channel. Thus the proposed P-HBKF and D-HBKF techniques can also be suitably extended for time-selective channel estimation in mmWave hybrid MIMO systems. Following the channel estimation process, one can consider the equivalent baseband channel $\mathbf{H}_{\text{eq}} = \mathbf{W}_{RF}^H \mathbf{H} \mathbf{F}_{RF}$, so that the end-to-end baseband channel is of size $N_{RF}^R \times N_{RF}^T$. Subsequently, the conventional MMSE/ LS-based baseband combiner can be employed for data detection.

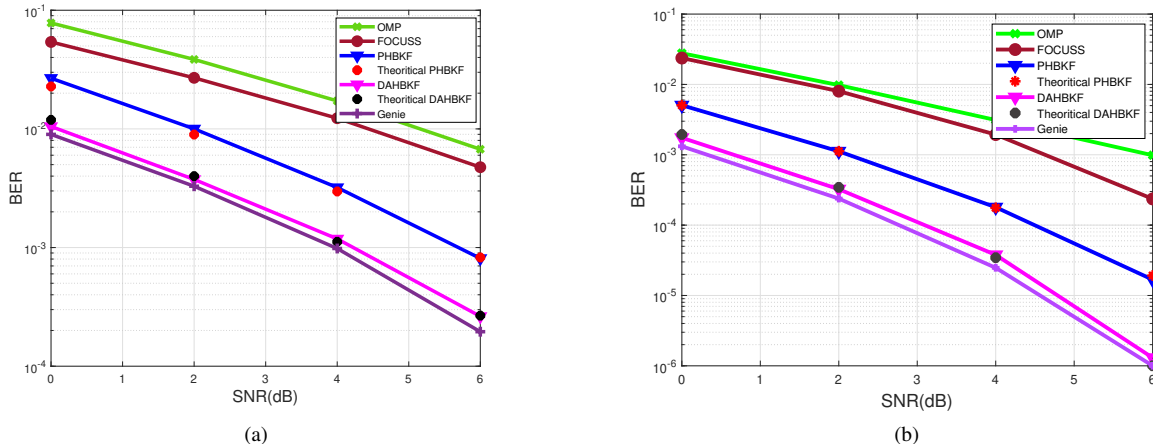(a)   (b)

Fig. 8. (a) BER versus SNR performance comparison for a 4-PSK Alamouti coded 2×2 MIMO-OFDM wireless system (b) BER versus SNR performance comparison for a 4-PSK rate $\frac{3}{4}$ non-square space time block coded 2×4 MIMO-OFDM wireless system for the following techniques: OMP [37], FOCUSS [12], P-HBKF, D-HBKF and Genie receiver.

VIII. Conclusions

We conceived HBKF-based schemes for sparse channel estimation in doubly-selective OSTBC MIMO-OFDM systems. The proposed P-HBKF technique successfully amalgamated the conventional KF with the SBL framework for improved tracking of the sparse MIMO-OFDM channel. The D-HBKF developed next was shown to further improve the performance by intrinsically integrating data detection with sparse channel estimation in the HBKF framework. The BCRBs and asymptotic MSE expressions were also derived for analytically characterize the MSE lower bound of the proposed techniques. Our simulation results demonstrated the performance improvement attained over the existing techniques, both in terms of the MSE and BER, while also validating the analytical bounds.

APPENDIX A
P-HBKF STEPS

Prediction:

$$\widehat{\mathbf{h}}_{n|n-1} = \rho \, \widehat{\mathbf{h}}_{n-1|n-1} \text{ and } \widehat{\mathbf{y}}_{\mathcal{P},n|n-1} = \mathbf{\Phi}_{\mathcal{P}} \widehat{\mathbf{h}}_{n|n-1} \quad (73)$$

Prediction MSE:

$$\mathbf{M}_{n|n-1} = \rho^2 \mathbf{M}_{n-1|n-1} + \left(1 - \rho^2\right) \left(\mathbf{I}_{N_R N_T} \otimes \widehat{\mathbf{\Gamma}}_n^c\right) \quad (74)$$

Kalman Gain Matrix:

$$\mathbf{K}_n = \mathbf{M}_{n|n-1} \mathbf{\Phi}_{\mathcal{P}}^H \left(\sigma^2 \mathbf{I}_{N_P N_c N_R} + \mathbf{\Phi}_{\mathcal{P}} \mathbf{M}_{n|n-1} \mathbf{\Phi}_{\mathcal{P}}^H\right)^{-1} \quad (75)$$

Measurement Error:

$$\mathbf{y}_{\mathcal{P},e,n} = \mathbf{y}_{\mathcal{P},n} - \widehat{\mathbf{y}}_{\mathcal{P},n|n-1} \quad (76)$$

Correction:

$$\widehat{\mathbf{h}}_{n|n} = \widehat{\mathbf{h}}_{n|n-1} + \mathbf{K}_n \mathbf{y}_{\mathcal{P},e,n} \quad (77)$$

MSE Matrix:

$$\mathbf{M}_{n|n} = \left(\mathbf{I}_{L N_R N_T} - \mathbf{K}_n \mathbf{\Phi}_{\mathcal{P}}\right) \mathbf{M}_{n|n-1} \quad (78)$$

REFERENCES

[1] L. Hanzo, M. Münster, B. Choi, and T. Keller, *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting.* John Wiley & Sons, 2005.

[2] H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: design, performance, and field trial results," *IEEE Communications Magazine*, vol. 40, no. 9, pp. 143–149, 2002.

[3] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[4] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Transactions on Communications*, vol. 47, no. 2, pp. 199–207, 1999.

[5] E. ETSI, "300 744 v1. 6.1 (2009-01), Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television," *Standards Inst.(ETSI), Valbonne, France*, 2001.

[6] W. G. Jeon, K. H. Paik, and Y. S. Cho, "Two-dimensional MMSE channel estimation for OFDM systems with transmitter diversity," in *IEEE 54th Vehicular Technology Conference. VTC Fall 2001. Proceedings (Cat. No. 01CH37211)*, vol. 3. IEEE, 2001, pp. 1682–1685.

[7] A. Gilbert and J. Tropp, "Applications of sparse approximation in communications," in *International Symposium on Information Theory 2005. ISIT 2005.* IEEE, 2005, pp. 1000–1004.

[8] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Transactions on Communications*, vol. 60, no. 12, pp. 3705–3716, 2012.

[9] Y. Peng, X. Yang, X. Zhang, W. Wang, and B. Wu, "Compressed MIMO-OFDM channel estimation," in *12th IEEE International Conference on Communication Technology (ICCT), 2010.* IEEE, 2010, pp. 1291–1294.

[10] M. A. Khojastepour, K. Gomadam, and X. Wang, "Pilot-assisted channel estimation for MIMO OFDM systems using theory of sparse signal recovery," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009.* IEEE, 2009, pp. 2693–2696.

[11] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[12] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Transactions on signal processing*, vol. 45, no. 3, pp. 600–616, 1997.

[13] X. Lv, G. Bi, and C. Wan, "The group Lasso for stable recovery of block-sparse signal representations," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1371–1382, 2011.

[14] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[15] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4658–4672, 2013.

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

[17] F. Caltagirone, L. Zdeborová, and F. Krzakala, "On convergence of approximate message passing," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 1812–1816.

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

[19] ——, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.

[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 Processing*, vol. 62, no. 14, pp. 3591–3603, 2014.

[21] ——, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5369–5382, 2015.

[22] A. Mishra, N. Yashaswini, and A. K. Jagannatham, "SBL-based joint sparse channel estimation and maximum likelihood symbol detection in OSTBC MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4220–4232, 2018.

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

[24] E. Karseras, K. Leung, and W. Dai, "Tracking dynamic sparse signals using hierarchical Bayesian Kalman filters," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013*. IEEE, 2013, pp. 6546–6550.

[25] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Prentice Hall PTR, 1993.

[26] E. G. Larsson and P. Stocia, *Space-time block coding for wireless communications*. Cambridge university press, 2008.

[27] T. Feng, T. R. Field, and S. Haykin, "Stochastic differential equation theory applied to wireless channels," *IEEE Transactions on Communications*, vol. 55, no. 8, pp. 1478–1483, 2007.

[28] Z. Gao, L. Dai, Z. Lu, C. Yuen, and Z. Wang, "Super-resolution sparse MIMO-OFDM channel estimation based on spatial and temporal correlations," *IEEE communications letters*, vol. 18, no. 7, pp. 1266–1269, 2014.

[29] J. Ziniel, L. C. Potter, and P. Schniter, "Tracking and smoothing of time-varying sparse signals via approximate belief propagation," in *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*. IEEE, 2010, pp. 808–812.

[30] G. Joseph and C. R. Murthy, "A noniterative online Bayesian algorithm for the recovery of temporally correlated sparse vectors," *IEEE Transactions on Signal Processing*, vol. 65, no. 20, pp. 5510–5525, 2017.

[31] S. Srivastava, A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, "Quasi-static and time-selective channel estimation for block-sparse millimeter wave hybrid MIMO systems: Sparse Bayesian learning (SBL) based approaches," *IEEE Transactions on Signal Processing*, vol. 67, no. 5, pp. 1251–1266, 2018.

[32] H. L. Van Trees and K. L. Bell, "Bayesian bounds for parameter estimation and nonlinear filtering/tracking," *AMC*, vol. 10, p. 12, 2007.

[33] N. Assimakis and M. Adam, "Kalman filter Riccati equation for the prediction, estimation, and smoothing error covariance matrices," *ISRN Computational Mathematics*, vol. 2013, 2013.

[34] A. Laub, "A Schur method for solving algebraic Riccati equations," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 913–921, 1979.

[35] T. Y. Al-Naffouri, "An EM-based forward-backward Kalman filter for the estimation of time-variant channels in OFDM," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3924–3930, 2007.

[36] S. Srivastava, A. K. Jagannatham, and L. Hanzo, "Technical report: Sparse doubly-selective channel estimation techniques for OSTBC MIMO-OFDM systems: A hierarchical Bayesian Kalman filter based approach." IIT Kanpur, Tech. Rep., 2020. [Online]. Available: http://www.iitk.ac.in/mwn/documents/MWNLab_TR_HBKF_MIMO_OFDM.pdf.

[37] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, 1993. 1993*. IEEE, 1993, pp. 40–44.

[38] T. ETSI, "125 996 v. 6.1.0, Universal mobile telecommunications system (UMTS); spatial channel model for multiple input multiple output (MIMO) simulations," *3GPP TR 25.996 version 6.1.0), Release 6*, 2003.

[39] J. Salo, G. Del Galdo, J. Salmi, P. Kyösti, M. Milojevic, D. Laselva, and C. Schneider, "MATLAB implementation of the 3GPP spatial channel model (3GPP TR 25.996)," *on-line, Jan*, 2005.

[40] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.

[41] Y. Ding, S.-E. Chiu, and B. D. Rao, "Bayesian channel estimation algorithms for massive MIMO systems with hybrid analog-digital processing and low-resolution ADCs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 3, pp. 499–513, 2018.

[42] S.-E. Chiu, N. Ronquillo, and T. Javidi, "Active learning and CSI acquisition for mmwave initial alignment," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 11, pp. 2474–2489, 2019.