

Sparse Bayesian Learning-Aided Joint Sparse Channel Estimation and ML Sequence Detection in Space-Time Trellis Coded MIMO-OFDM Systems

Amrita Mishra, *Member, IEEE*, Aditya K. Jagannatham, *Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

Abstract—Sparse Bayesian learning (SBL)-based approximately sparse channel estimation schemes are conceived for space-time trellis coded (STTC) multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) systems relying on trellis-based encoding and decoding over the data subcarriers. First, a pilot-aided channel estimation scheme is developed employing the multiple response extension of SBL (MSBL) framework. Subsequently, a novel data-aided joint channel estimation and data decoding framework relying on optimal maximum likelihood sequence detection (MLSD) is intrinsically amalgamated with our powerful EM-based MSBL algorithm. Explicitly, an MSBL-based MIMO channel estimate is gleaned in the E-step followed by a novel modified path-metric-based Viterbi decoder in the M-step. Our theoretical analysis characterizes the performance of the proposed schemes in terms of the associated frame error rate (FER) upper bounds by explicitly considering the effect of estimation errors along with evaluating the product measure of the STTC under consideration. Finally, our simulation results are complemented by the Bayesian Cramér-Rao bound (BCRB), the associated complexity analysis and the performance of the proposed schemes for validating the theoretical bounds.

Index Terms—Space-time trellis codes, MIMO-OFDM, sparse Bayesian learning, maximum likelihood sequence detection, frame error rate

I. INTRODUCTION

Space-time trellis codes (STTC) [1], [2] have garnered significant attention due to their attractive performance gains as a result of the inherent spatio-temporal diversity for multiple-input multiple-output (MIMO) systems [3]. Towards this end, the recent work in [4] employs STTC for vehicular communication while [5] combines STTC systems with space-time shift keying to further improve the associated performance gains. Further, orthogonal frequency division multiplexing (OFDM)

in multi-antenna systems has emerged as one of the key technologies such as 3GPP-LTE, Mobile WiMAX, IMT-Advanced as well as wireless LAN standards like IEEE 802.11a, IEEE 802.11n [6]. In this context, [7], [8] have explored the integration of STTC-based signal design for frequency selective MIMO-OFDM systems. Towards this end, the performance of trellis coded MIMO-OFDM wireless systems depends heavily on the quality of the channel state information (CSI) available at the receiver, which necessitates the development of schemes for improved accuracy of estimation. A brief review of related works in the existing literature is presented next.

1. Review of Existing Contributions

Existing works developed pilot-based techniques [9] and training-based signal designs [10], [11] for channel estimation in OFDM and MIMO-OFDM wireless systems respectively. The frameworks therein can be readily extended for channel estimation in STBC/STTC MIMO-OFDM systems. However, the accuracy of such schemes is highly dependent on the number of pilot subcarriers, thereby leading to a significant bandwidth overhead. This limitation has resulted in the need for the design of accurate channel estimation schemes in the presence of fewer information devoid pilot symbols. In this regard, joint channel estimation and symbol detection techniques were developed based on the expectation-maximization paradigm [12], [13] and evolutionary algorithms [14] which resulted in an improved performance owing to the iterative nature of channel estimation and symbol detection. Nevertheless, none of the above OFDM-based channel estimation schemes take into account the temporal sparsity inherent in multipath wireless channels.

To overcome the above shortcomings of conventional pilot-based estimation approaches, the sparse signal recovery framework of [15] has been widely applied for OFDM channel estimation [16]–[19]. To elaborate, the sparsity of OFDM channels arises since the number of significant channel taps of a wireless multipath channel is typically much lower than the total number of taps corresponding to the maximum delay spread of the channel [20]. In this regard, the sparse Bayesian learning framework [21] has been demonstrated to be less prone to converging to estimates other than the maximally sparse solution in comparison to popular sparse signal recovery algorithms such as basis pursuit [22] and focal underdetermined system solver FOCUSS [23]. These events were referred to as ‘structural errors’ in [21]. Another strength

A. Mishra is with the Department of Electronics and Communication Engineering, DSPM International Institute of Information Technology Naya Raipur, Atal Nagar, Chhattisgarh 493661, India (e-mail: amrita@iiitr.edu.in).

A. K. Jagannatham is with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur, UP 208016, India (e-mail: adityaj@iitk.ac.in).

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).

This research has been supported in part by the IIMA IDEA Telecom Centre of Excellence (IITCOE) and Qualcomm Innovation Fellowship. L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/N004558/1, EP/PO34284/1, COALESCE, of the Royal Society’s Global Challenges Research Fund Grant as well as of the European Research Council’s Advanced Fellow Grant QuantCom.

of the SBL framework is that instead of choosing a fixed sparsity promoting prior, it evaluates the Bayesian evidence corresponding to each parameterized model prior and then finally chooses the one which maximizes the Bayesian evidence. The diverse applications of SBL have been demonstrated in recent works such as [24] and [25] to perform target imaging in MIMO radars and channel estimation in millimeter wave hybrid MIMO systems, respectively. The authors of [26], [28] proposed schemes to exploit the simultaneous sparsity for the estimation of a MIMO-OFDM channel via the M-SBL framework in uncoded scenarios. Furthermore, in the external turbo encoder/decoder considered in [27] for simulations, the log-likelihood ratio-based bit detection was not integrated with the EM framework, while we developed the paradigm of joint channel estimation and sequence detection in this study. Our recent contribution in [27] overcomes the above shortcoming with the aid of our SBL-based sparse channel estimation in OSTBC MIMO-OFDM wireless systems. A brief summary of the existing and proposed works is presented in a tabular format in Table I.

2. Our Contributions

- Against this background we develop a new pilot-based MSBL technique for approximately sparse channel estimation in trellis coded MIMO-OFDM systems.
- Subsequently, joint sparse channel estimation and sequence detection is conceived for STTC MIMO-OFDM systems by intrinsically amalgamating an evolved Viterbi decoder with our EM framework. This potent arrangement yields an MSBL-based channel estimate in the E-step followed by a novel modified path-metric-based trellis decoder in the M-step.
- New theoretical performance bounds are presented in terms of the coded frame error rate (FER), which cannot be found in the SBL-based OFDM channel estimation literature.
- Bayesian Cramér-Rao bounds (BCRBs) are derived for benchmarking the MSE performance, followed by a comparison of the computational complexities of the proposed and existing schemes. The simulation results confirm the theory and they are also contrasted to the family of existing and genie-based benchmark schemes.

3. Outline and Notations

The rest of the paper is organized as follows. The system model and encoding/decoding aspects of the trellis coded MIMO-OFDM wireless system under consideration are described in Section II. Section III motivates the paradigm of approximately sparse channel estimation followed by a discussion of the MSBL-based pilot aided scheme designed for our STTC MIMO-OFDM system under consideration. Section IV develops the proposed data-aided joint estimation framework. The analytical pairwise error probability (PEP) upper bounds characterizing the FER performance of the MSBL-based MLSD in STTC MIMO-OFDM systems are developed in Section V followed by our theoretical BCRBs and our complexity analysis in Sections VI and VII-A respectively.

Our simulation results are presented in Section VII followed by our conclusions in Section VIII.

The following notations have been used. Small case letters (k), small case boldface letters (\mathbf{h}), and upper case boldface letters (\mathbf{H}) are used to represent scalars, vectors and matrices respectively. The quantities $\mathbf{A} \in \mathbb{R}^{L \times M}$ and $\mathbf{B} \in \mathbb{C}^{L \times M}$ denote a real and complex-valued $L \times M$ matrices \mathbf{A} and \mathbf{B} , respectively. The superscript notation such as $\mathbf{x}^{(r)}$ denotes the estimate of the variable \mathbf{x} in the r^{th} iteration while $\hat{\mathbf{x}}$ is used to denote the estimate of \mathbf{x} . The operators $(\cdot)^*$, \otimes , $(\cdot)^H$, and $\text{Tr}(\cdot)$ are used to denote the complex conjugate, matrix Kronecker product, conjugate transpose and trace of a matrix, respectively. The notation \mathbf{I}_N is used to represent a $N \times N$ identity matrix and $\mathbf{0}_{L \times M}$ represents a $L \times M$ matrix of zeros, respectively. For a vector \mathbf{a} , a_m denotes the m^{th} element of \mathbf{a} . Furthermore, $\text{diag}(a_1, \dots, a_Z)$ represents a diagonal matrix with principal diagonal elements a_1, \dots, a_Z . For a matrix \mathbf{A} , $\mathbf{A}_{m,n}$, $\text{vec}(\mathbf{A})$, $|\mathbf{A}|$, and $\mathbf{A}_{:,k}$ denote the $(m, n)^{th}$ element, vector obtained by stacking the columns of the matrix \mathbf{A} , determinant and k^{th} column of \mathbf{A} , respectively. A complex Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is represented as $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The notation \triangleq represents a definition while \equiv denotes equivalence upto a known constant term. Calligraphic subscript notations such as $(\cdot)_{\mathcal{P}}$, $(\cdot)_{\mathcal{F}}$, and $(\cdot)_{\mathcal{F}, \mathcal{P}}$ denote the pilot, frame and pilot-based equivalent of the frame-based nature of the input argument vector/matrix.

II. SPACE-TIME TRELLIS CODED MIMO-OFDM SYSTEM MODEL: AN OVERVIEW

Consider a space-time trellis coded MIMO system with N transmit antennas (TAs) and M receive antennas (RAs). The complex baseband multipath channel impulse response (MCIR) $\tilde{h}_{i,j}(t)$ between TA i and RA j is given by

$$\tilde{h}_{i,j}(t) = \sum_{l=1}^{L_s} \alpha_{i,j}(l) \delta(t - \tau(l)), \quad (1)$$

where the quantities $\alpha_{i,j}(l)$ and $\tau(l)$ represent the channel attenuation factor and the propagation delay corresponding to the l^{th} multipath component, while $\delta(t)$ and L_s denote the Dirac delta function and the number of multipath components. The channel vector $\mathbf{h}_{i,j} \in \mathbb{C}^{L \times 1}$ is obtained by sampling the channel $h_{i,j}(t) = g_i(t) * \tilde{h}_{i,j}(t) * g_j(t)$, at the baud rate $t = T$, where $g_i(t)$ and $g_j(t)$ represent the baseband transmit and receive filters respectively and $*$ denotes the convolution operator. Due to the finite bandwidth of the transmitted signal, the resulting sampled channel impulse response corresponds to a low-pass filtered of the sparse CIR $\tilde{h}_{i,j}(t)$, and thereby exhibits approximate sparsity [26], [29]. Let the L -tap equivalent channel vector corresponding to the discrete time approximately sparse channel be denoted by $\mathbf{h}_{i,j} = [h_{i,j}(1), \dots, h_{i,j}(L)]$ where the components $h_{i,j}(l)$, $1 \leq l \leq L$, are assumed to be independent of each other. Furthermore, similar to [26], [27], we assume the non-zero components of $\mathbf{h}_{i,j}$ to coincide for all the transmit receive antenna pairs (i, j) . This yields the MN channel vectors $\mathbf{h}_{i,j} \forall 1 \leq i \leq N, 1 \leq j \leq M$ exhibit similar sparse CIRs, which can be exploited for improving the

TABLE I: Review of Existing Works

	[10]-2003, [11]-2006	[12]-2008	[13]-2014	[14]-2014	[16]-2002	[17]-2008 [18]-2010	[26]-2015	[27]-2018	Proposed Work-2019
MIMO	✓	✓	✓	✓			✓	✓	✓
STBC/OSTBC		✓						✓	
Turbo				✓					
STTC			✓						✓
OFDM	✓	✓		✓	✓	✓	✓	✓	✓
Pilot-based channel estimation	✓				✓	✓			
Joint channel estimation and symbol detection		✓	✓	✓			✓	✓	✓
Channel sparsity					✓	✓	✓	✓	✓
Compressed sensing (CS)-based sparse channel estimation						✓			
Fixed sparsity promoting prior					✓				
SBL-based prior						✓	✓	✓	✓
Joint likelihood maximization incorporates bit detection			✓					✓	✓
Theoretical BER/FER bounds			✓					✓	✓

performance of channel estimation. Recent contributions such as [30], [31] also exploited the common sparsity for channel estimation in frequency division duplex massive MIMO systems. For brevity, the term *approximately sparse* is used in the rest of this work to refer to approximate simultaneous sparsity.

The employment of STTC in MIMO-OFDM can lead to two different system models based on the specific choice of the encoding/decoding procedures. These can be classified as MIMO-OFDM SubCarrier (MO-SC) and MIMO-OFDM SYmbol (MO-SY) systems, wherein space-time trellis encoding/decoding is performed across the data subcarriers [7], [32] for a single symbol and consecutive OFDM symbols for each subcarrier [33], respectively. Note that the MO-SC system described in this work is similar in nature to the space-frequency trellis coded systems presented in earlier contributions such as [7], [32]. However, in order to distinguish between space-frequency trellis coded and MO-SY systems that perform coding across subcarriers and OFDM symbols, respectively, we adhere to the nomenclature of MO-SC throughout this work to refer to space-frequency trellis coded systems. Our previous work in [34] presented a preliminary study of MSBL-based approximately sparse channel estimation techniques designed for MO-SY systems. By contrast, this work presents a detailed description of SBL-based approximately sparse channel estimation schemes for MO-SC systems followed by comprehensive performance and complexity analyses. In the MO-SC system, the binary information bits generated by a source are initially mapped to a complex modulated symbol drawn from a constellation \mathcal{S} such as Q -ary phase shift keying (PSK), similar to [7], [32] where the quantity Q denotes the constellation size. These are subsequently fed into a single space-time trellis encoder that produces code vectors $\mathbf{c}_k \in \mathbb{C}^{N \times 1}$ as, $\mathbf{c}_k = [c_{1,k}, \dots, c_{N,k}]^T$, which satisfies $\mathbf{E}\{|c_{i,k}|^2\} = 1$. Let the codeword matrix be defined as

$$\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_K] \in \mathbb{C}^{N \times K}, \quad (2)$$

where K denotes the block length of the STTC under consideration. Without loss of generality, this work also assumes the block length K to be equal to the total number of subcarriers in the MO-SC system. The codeword \mathbf{C} is subsequently passed

through a bank of demultiplexers corresponding to each TA followed by the IFFT and multiplexing operations, as shown in Fig. 1. Furthermore, prior to transmission, a cyclic prefix (CP) of sufficient length is appended to each of the OFDM symbols formed from the N IFFT outputs. The complex baseband signal $\mathbf{y}_j \in \mathbb{C}^{K \times 1}$ at the j^{th} receive antenna after CP removal and FFT can be represented as

$$\mathbf{y}_j = \sum_{i=1}^N \bar{\mathbf{C}}_i \mathbf{F} \mathbf{h}_{i,j} + \mathbf{n}_j, \quad (3)$$

where each diagonal matrix $\bar{\mathbf{C}}_i \in \mathbb{C}^{K \times K}$ is constructed from the i^{th} row of the codeword matrix \mathbf{C} and denotes the coded MIMO-OFDM symbol of the i^{th} TA. The matrix $\mathbf{F} \in \mathbb{C}^{K \times L}$ denotes the submatrix comprised of the first L columns of the standard discrete Fourier transform (DFT) matrix. The vector $\mathbf{n}_j \in \mathbb{C}^{K \times 1}$ represents the complex symmetric zero mean additive white Gaussian noise (AWGN) vector with independent and identically distributed (i.i.d) samples and covariance matrix $\mathbf{R}_n = \sigma_n^2 \mathbf{I}_K$. Let $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_M] \in \mathbb{C}^{K \times M}$ represent a concatenation of the receive vectors \mathbf{y}_j for all RAs and $\mathbf{N} \in \mathbb{C}^{M \times K}$ denote a similar concatenation of noise vectors \mathbf{n}_j . Let $\mathbf{Y}_{\cdot k}^T$ denote the k^{th} column of \mathbf{Y}^T , which corresponds to the received signal over the k^{th} subcarrier. It can be expressed as

$$\mathbf{Y}_{\cdot k}^T = \mathcal{H}_k \mathbf{c}_k + \mathbf{N}_{\cdot k}, \quad (4)$$

where $\mathcal{H}_k \in \mathbb{C}^{M \times N}$ denotes the effective MIMO channel matrix corresponding to the k^{th} subcarrier and is defined as

$$\mathcal{H}_k = \begin{bmatrix} (\mathbf{F} \mathbf{h}_{1,1})_k & \dots & (\mathbf{F} \mathbf{h}_{1,N})_k \\ \vdots & \ddots & \vdots \\ (\mathbf{F} \mathbf{h}_{M,1})_k & \dots & (\mathbf{F} \mathbf{h}_{M,N})_k \end{bmatrix}, \quad (5)$$

with each element $(\mathbf{F} \mathbf{h}_{i,j})_k$ representing the k^{th} DFT component of the channel vector $\mathbf{h}_{i,j}$. The decoded codeword $\hat{\mathbf{C}} = [\hat{\mathbf{c}}_1, \dots, \hat{\mathbf{c}}_K] \in \mathbb{C}^{N \times K}$ can now be obtained via Viterbi decoding, which minimizes the cost function given by [1]

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{C} \in \mathcal{S}^{N \times K}} \sum_{k=1}^K \|\mathbf{Y}_{\cdot k}^T - \mathcal{H}_k \mathbf{c}_k\|^2. \quad (6)$$

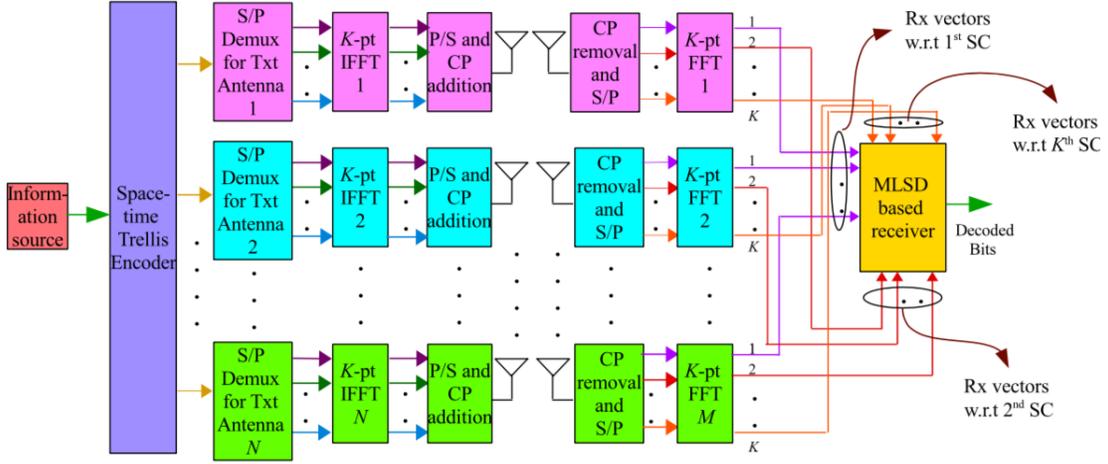


Fig. 1: Block diagram representation of a trellis coded MIMO-OFDM system with coding over subcarriers (MO-SC)

The performance of the ML decoder above naturally depends on the accuracy of CSI $\widehat{\mathcal{H}}_k$ at the receiver. This motivates the SBL-based schemes described next to obtain improved CSI estimates by exploiting sparsity of the underlying channel in STTC MIMO-OFDM systems. In contrast to the MO-SC system described above and shown in Fig. 1 that has a single space-time trellis encoder/decoder across all the subcarriers, the MO-SY system employed in our previous work in [34] that operates across multiple OFDM symbols, has a total of K independent space-time trellis encoder/decoder pairs, one for each of the K subcarriers. Thus, the coded block of the MO-SC system considered in this work spans a single OFDM symbol, which leads to a significantly lower processing delay in comparison to the MO-SY system that has a block duration of K OFDM symbols per subcarrier. This also renders the MO-SC system well-suited for practical implementation.

III. PILOT-BASED MSBL (P-MSBL) FOR APPROXIMATELY SPARSE CHANNEL ESTIMATION

Similar to previous MIMO-OFDM works such as [26], [27] based on the multiple measurement vectors (MMV) \mathbf{y}_j corresponding to all M RAs, the MO-SC system can be expressed as

$$\mathbf{Y} = \Phi \mathbf{H} + \mathbf{N}, \quad (7)$$

where $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_M] \in \mathbb{C}^{K \times M}$, $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_M] \in \mathbb{C}^{K \times M}$, $\Phi = \overline{\mathbf{C}}(\mathbf{I}_N \otimes \mathbf{F}) \in \mathbb{C}^{K \times LN}$, and $\mathbf{H} = \begin{bmatrix} \mathbf{h}_{1,1} & \dots & \mathbf{h}_{1,M} \\ \mathbf{h}_{N,1} & \dots & \mathbf{h}_{N,M} \end{bmatrix} \in \mathbb{C}^{LN \times M}$ denote the receive, noise, and channel matrices respectively with $\overline{\mathbf{C}} = [\overline{\mathbf{C}}_1, \overline{\mathbf{C}}_2, \dots, \overline{\mathbf{C}}_N] \in \mathbb{C}^{K \times KN}$ denoting the transmit codeword. Based on the terminology used in the popular works of CS literature such as [35], [36], Φ is termed as the ‘measurement matrix’ since it directly relates the receive signal matrix \mathbf{Y} to the unknown sparse MIMO-OFDM channel matrix \mathbf{H} as seen in (7). For the purpose of channel estimation, consider P pilot symbols placed uniformly across the K subcarriers such that the pilot equivalent $\overline{\mathbf{C}}_{\mathcal{P}}$ of the transmit codeword $\overline{\mathbf{C}}$ is given by, $\overline{\mathbf{C}}_{\mathcal{P}} = [\overline{\mathbf{C}}_{\mathcal{P},1}, \overline{\mathbf{C}}_{\mathcal{P},2}, \dots, \overline{\mathbf{C}}_{\mathcal{P},N}] \in \mathbb{C}^{P \times PN}$ where the diagonal codeword matrix $\overline{\mathbf{C}}_{\mathcal{P},i}$ corresponding to

the transmit antenna i comprises of the P pilot symbols as its diagonal entries. Let $\mathbf{y}_{\mathcal{P},j} \in \mathbb{C}^{P \times 1}$ represent the received signal at the P pilot subcarrier locations on the j^{th} receive antenna and $\mathbf{Y}_{\mathcal{P}} = [\mathbf{y}_{\mathcal{P},1}, \mathbf{y}_{\mathcal{P},2}, \dots, \mathbf{y}_{\mathcal{P},M}] \in \mathbb{C}^{P \times M}$ denote the matrix of concatenated received vectors $\mathbf{y}_{\mathcal{P},j}$ across all the M RAs. The system model corresponding to the pilot observations can be expressed as

$$\mathbf{Y}_{\mathcal{P}} = \Phi_{\mathcal{P}} \mathbf{H} + \mathbf{N}_{\mathcal{P}}, \quad (8)$$

where the measurement matrix $\Phi_{\mathcal{P}} = \overline{\mathbf{C}}_{\mathcal{P}}(\mathbf{I}_N \otimes \mathbf{F}_{\mathcal{P}}) \in \mathbb{C}^{P \times LN}$ and $\mathbf{F}_{\mathcal{P}} \in \mathbb{C}^{P \times L}$ denotes a submatrix of the DFT matrix \mathbf{F} , comprising only the P rows corresponding to the pilot subcarrier positions. Let us assume the quasi-static block fading MIMO channel matrix \mathbf{H} to remain constant over a frame comprising V consecutive OFDM symbols. Let $\mathbf{Y}_{\mathcal{P},v} \in \mathbb{C}^{P \times M}$, $1 \leq v \leq V$ denote the pilot-based receive matrix corresponding to the v^{th} frame. The equivalent system model for the block fading scenario above is formulated as

$$\mathbf{Y}_{\mathcal{F},\mathcal{P}} = \Phi_{\mathcal{F},\mathcal{P}} \mathbf{H} + \mathbf{N}_{\mathcal{F},\mathcal{P}}, \quad (9)$$

where $\mathbf{Y}_{\mathcal{F},\mathcal{P}} = [\mathbf{Y}_{\mathcal{P},1}^T, \dots, \mathbf{Y}_{\mathcal{P},V}^T]^T \in \mathbb{C}^{PV \times M}$, $\Phi_{\mathcal{F},\mathcal{P}} = [\Phi_{\mathcal{P},1}^T, \dots, \Phi_{\mathcal{P},V}^T] \in \mathbb{C}^{PV \times LN}$ and $\mathbf{N}_{\mathcal{F},\mathcal{P}} = [\mathbf{N}_{\mathcal{P},1}^T, \dots, \mathbf{N}_{\mathcal{P},V}^T] \in \mathbb{C}^{PV \times M}$ denote the corresponding receive, measurement and noise matrices concatenated over the MO-SC frame, with $\Phi_{\mathcal{P},v}$ and $\mathbf{N}_{\mathcal{P},v}$ similarly corresponding to the v^{th} frame. Note that when $PV < LN$, the resultant channel estimation problem above is ill-posed, rendering popular conventional techniques such as least-squares ineffective. In such scenarios, SBL-based schemes similar to the ones proposed in [20], [28] that exploit the sparsity of the underlying channel are ideally suited for channel estimation. As a part of the SBL framework, the parameterized prior $\mathcal{CN}(\mathbf{0}, \Gamma)$ is assigned to each of the channel vectors $\mathbf{h}_{i,j}$ associated with $\Gamma = \text{diag}(\boldsymbol{\gamma})$ where $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_L]^T$ denotes the hyperparameter vector. Since the vectors $\mathbf{h}_{i,j}$ are assumed to be simultaneously sparse, the matrix Γ is identical for all the transmit receive antenna pairs (i, j) . Additionally, it has also been established in [21], [37] that for hyperparameter estimation based on Bayesian evidence maximization, the SBL-based

hyperparameter estimates $\hat{\gamma}_l$ corresponding to non-significant multipath components approach zero. Thus, the associated channel coefficient estimates $\hat{\mathbf{h}}_{i,j}(l)$ corresponding to the non-significant paths l approach $0 \forall (i, j)$. The parametric form of the multiple SBL (MSBL) prior assignment for the channel matrix \mathbf{H} of the MO-SC space-time trellis coded system under consideration is given by

$$\begin{aligned} p(\mathbf{H}; \Gamma) &= \prod_{j=1}^M p(\mathbf{h}_j; \Gamma), \\ &= \prod_{j=1}^M \prod_{i=1}^N \prod_{l=1}^L \frac{1}{\pi \gamma_l} \exp\left(-\frac{|h_{i,j}(l)|^2}{\gamma_l}\right) \end{aligned} \quad (10)$$

where $\mathbf{h}_j = [\mathbf{h}_{1,j}^T, \mathbf{h}_{2,j}^T, \dots, \mathbf{h}_{N,j}^T]^T \in \mathbb{C}^{LN \times 1}$ denotes the j^{th} column of the channel matrix \mathbf{H} . The P-MSBL framework for approximately sparse channel estimation is outlined next.

1. P-MSBL for MO-SC Systems

The SBL framework aims to estimate the hyperparameters by maximizing the likelihood function $p(\mathbf{Y}_{\mathcal{F}, \mathcal{P}}; \Gamma)$ given by

$$p(\mathbf{Y}_{\mathcal{F}, \mathcal{P}}; \Gamma) = \prod_{j=1}^M (\pi)^{-PV} |\Sigma_y|^{-1} \exp(-\mathbf{y}_j^H \Sigma_y^{-1} \mathbf{y}_j) \quad (11)$$

with respect to the hyperparameters Γ where $\Sigma_y = \sigma^2 \mathbf{I}_{PV} + \Phi_{\mathcal{P}} (\mathbf{I}_N \otimes \Gamma) \Phi_{\mathcal{P}}^H$. However, owing to the intractability of the conventional ML approach above, the iterative expectation maximization (EM) algorithm can be employed for hyperparameter estimation [21]. Let the channel matrix \mathbf{H} denote the hidden parameter for the EM algorithm and $\Gamma^{(r)}$ represent the estimate of the hyperparameter matrix in the r^{th} iteration. The expectation (E-step) in the r^{th} iteration computes the log-likelihood function $\mathcal{L}(\Gamma | \Gamma^{(r)})$ as [21]

$$\begin{aligned} \mathcal{L}(\Gamma | \Gamma^{(r)}) &= \mathbf{E}_{\mathbf{H} | \mathbf{Y}_{\mathcal{F}, \mathcal{P}}; \Gamma^{(r)}} \{\log p(\mathbf{Y}_{\mathcal{F}, \mathcal{P}}, \mathbf{H}; \Gamma)\} \\ &= \mathbf{E}_{\mathbf{H} | \mathbf{Y}_{\mathcal{F}, \mathcal{P}}; \Gamma^{(r)}} \{\log p(\mathbf{Y}_{\mathcal{F}, \mathcal{P}} | \mathbf{H}; \Gamma)\} + \\ &\quad \mathbf{E}_{\mathbf{H} | \mathbf{Y}_{\mathcal{F}, \mathcal{P}}; \Gamma^{(r)}} \{\log p(\mathbf{H}; \Gamma)\}, \end{aligned} \quad (13)$$

where $\log p(\mathbf{Y}_{\mathcal{F}, \mathcal{P}} | \mathbf{H}; \Gamma) \equiv -\frac{\|\mathbf{Y}_{\mathcal{F}, \mathcal{P}} - \Phi_{\mathcal{F}, \mathcal{P}} \mathbf{H}\|^2}{\sigma^2}$ and $\log p(\mathbf{H}; \Gamma)$ can be computed from (10). Further, the *a posteriori* probability density function (PDF) corresponding to each \mathbf{h}_j can be obtained as $p(\mathbf{h}_j | \mathbf{Y}_{\mathcal{F}, \mathcal{P}}; \Gamma^{(r)}) \sim \mathcal{CN}(\boldsymbol{\mu}_j^{(r)}, \Sigma^{(r)})$, with the *a posteriori* mean vector $\boldsymbol{\mu}_j^{(r)} \in \mathbb{C}^{LN \times 1}$ given as [37] $\boldsymbol{\mu}_j^{(r)} = \sigma^{-2} \Sigma^{(r)} \Phi_{\mathcal{F}, \mathcal{P}}^H (\mathbf{Y}_{\mathcal{F}, \mathcal{P}})_{\cdot, j}$ and covariance matrix $\Sigma^{(r)} \in \mathbb{C}^{LN \times LN}$ given as [37]

$$\Sigma^{(r)} = \left(\sigma^{-2} \Phi_{\mathcal{F}, \mathcal{P}}^H \Phi_{\mathcal{F}, \mathcal{P}} + \left(\Gamma_{\mathcal{K}}^{(r)} \right)^{-1} \right)^{-1}, \quad (14)$$

where $\Gamma_{\mathcal{K}}^{(r)} = \mathbf{I}_N \otimes \Gamma^{(r)}$. Finally, the posterior density of the concatenated MIMO-OFDM channel matrix \mathbf{H} can be evaluated as [37] $p(\mathbf{H} | \mathbf{Y}_{\mathcal{F}, \mathcal{P}}; \Gamma^{(r)}) \sim \mathcal{CN}(\mathbf{M}^{(r)}, \Sigma^{(r)})$, with the mean $\mathbf{M}^{(r)} = [\boldsymbol{\mu}_1^{(r)}, \boldsymbol{\mu}_2^{(r)}, \dots, \boldsymbol{\mu}_M^{(r)}] \in \mathbb{C}^{LN \times M}$ and

Algorithm 1 P-MSBL-based Channel Estimation in MO-SC Trellis Coded Systems

Input: $\mathbf{Y}_{\mathcal{F}, \mathcal{P}}, \Phi_{\mathcal{F}, \mathcal{P}}, \sigma^2, I$

Initialization: $\gamma_l^{(0)} = 1, \forall 1 \leq l \leq L \implies \Gamma^{(0)} = \mathbf{I}_L$

Set counter $r = -1$

while $r < I$ **do**

$r \leftarrow r + 1$

E-step: Evaluate *a posteriori* covariance matrix $\Sigma^{(r)}$ as per (14) and mean matrix $\mathbf{M}^{(r)}$ as

$$\mathbf{M}^{(r)} = [\boldsymbol{\mu}_1^{(r)}, \dots, \boldsymbol{\mu}_M^{(r)}], \boldsymbol{\mu}_j^{(r)} = \sigma^{-2} \Sigma^{(r)} \Phi_{\mathcal{F}, \mathcal{P}}^H (\mathbf{Y}_{\mathcal{F}, \mathcal{P}})_{\cdot, j}$$

M-step: Evaluate hyperparameter estimates $\gamma_l^{(r+1)}, \forall 1 \leq l \leq L$ as per (18)

end while

Output: $\hat{\mathbf{H}}_{\mathcal{P}} = \mathbf{M}^{(I)}$

$\Sigma^{(r)}$ as defined in (14). The subsequent maximization (M-step) obtains the hyperparameter matrix estimate $\Gamma^{(r+1)}$ by maximizing $\mathcal{L}(\Gamma | \Gamma^{(r)})$ as

$$\begin{aligned} \Gamma^{(r+1)} &= \arg \max_{\Gamma} \left(\mathbf{E}_{\mathbf{H} | \mathbf{Y}_{\mathcal{F}, \mathcal{P}}; \Gamma^{(r)}} \{\log p(\mathbf{Y}_{\mathcal{F}, \mathcal{P}} | \mathbf{H}; \Gamma)\} + \right. \\ &\quad \left. \mathbf{E}_{\mathbf{H} | \mathbf{Y}_{\mathcal{F}, \mathcal{P}}; \Gamma^{(r)}} \{\log p(\mathbf{H}; \Gamma)\} \right) \end{aligned} \quad (15)$$

$$\equiv \arg \max_{\Gamma} \mathbf{E}_{\mathbf{H} | \mathbf{Y}_{\mathcal{F}, \mathcal{P}}; \Gamma^{(r)}} \{\log p(\mathbf{H}; \Gamma)\}. \quad (16)$$

The equivalence of the optimization problem in (16) follows from the fact that $\log p(\mathbf{Y}_{\mathcal{F}, \mathcal{P}} | \mathbf{H}; \Gamma) \equiv -\frac{\|\mathbf{Y}_{\mathcal{F}, \mathcal{P}} - \Phi_{\mathcal{F}, \mathcal{P}} \mathbf{H}\|^2}{\sigma^2}$ is independent of the hyperparameter vector γ as shown above. Employing the MSBL-based prior assignment for the MIMO channel matrix of (10), the maximization problem in (16) can be recast and solved to obtain the final expression for hyperparameter update as shown in (17) and (18). The E and M-steps are iterated for a fixed number of iterations I to obtain the P-MSBL-based MIMO-OFDM channel estimate $\hat{\mathbf{H}}_{\mathcal{P}} = \mathbf{M}^{(I)}$. In the next section we propose a framework for joint estimation of the hyperparameters Γ and ML decoding of the space-time trellis codewords $\mathbf{C}_v, 1 \leq v \leq V$ in order to further enhance the performance of channel estimation.

IV. DATA-AIDED MSBL (DA-MSBL) FOR JOINT CHANNEL ESTIMATION AND SEQUENCE DETECTION

The DA-MSBL framework proposed for the MO-SC system begins by assigning the parameter to be estimated as $\Theta = \{\mathbf{C}_{1:V}, \Gamma\}$ where the coded frame $\mathbf{C}_{1:V} = [\mathbf{C}_1, \dots, \mathbf{C}_V] \in \mathbb{C}^{N \times KV}$ represents the concatenation of space-time trellis codewords \mathbf{C}_v for V consecutive OFDM symbols. Let us consider the parameter set $\{\mathbf{Y}_{\mathcal{F}}, \mathbf{H}\}$ as the complete data associated with $\mathbf{Y}_{\mathcal{F}} = [\mathbf{Y}_1^T, \mathbf{Y}_2^T, \dots, \mathbf{Y}_V^T]^T \in \mathbb{C}^{KV \times M}$ denoting a concatenation of the receive matrices $\mathbf{Y}_v, 1 \leq v \leq V$. The log-likelihood function $\mathcal{L}(\mathbf{C}_{1:V}, \Gamma | \mathbf{C}_{1:V}^{(r)}, \Gamma_{\mathcal{K}}^{(r)})$ in the r^{th} iteration of the E-step invoked for the joint estimation of the MO-SC code frame $\mathbf{C}_{1:V}$ and the hyperparameter matrix Γ

$$\mathbf{\Gamma}^{(r+1)} = \arg \max_{\gamma} \mathbf{E}_{\mathbf{H}|\mathbf{Y}_{\mathcal{F}}, \mathbf{P}; \mathbf{\Gamma}^{(r)}} \left\{ \log \left(\prod_{j=1}^M \prod_{i=1}^N \prod_{l=1}^L \frac{1}{\pi \gamma_l} \exp \left(-\frac{|h_{i,j}(l)|^2}{\gamma_l} \right) \right) \right\} \quad (17)$$

$$\Rightarrow \gamma_l^{(r+1)} = \frac{1}{MN} \sum_{j=1}^M \sum_{i=0}^{N-1} \left(|\mathbf{M}_{(l+iL, j)}^{(r)}|^2 + \Sigma_{(l+iL, l+iL)}^{(r)} \right). \quad (18)$$

$$\mathbf{C}_{1:V}^{(r+1)} = \arg \max_{\mathbf{C}_{1:V} \in \mathcal{S}^{N \times V \times K}} \mathbf{E}_{\mathcal{H}|\mathbf{Y}_{\mathcal{F}}; \mathbf{C}_{1:V}^{(r)}, \mathbf{\Gamma}^{(r)}} \left\{ \sum_{v=1}^V \log p(\mathbf{Y}_v^T | \mathcal{H}; \mathbf{C}_v) \right\} \quad (19)$$

$$\mathbf{R}_k^{(r)} = \begin{bmatrix} \tilde{\mathbf{R}}_{(k,k)} & \tilde{\mathbf{R}}_{(k, k+K)}^* & \cdots & \tilde{\mathbf{R}}_{(k, k+(N-1)K)}^* \\ \vdots & \ddots & \ddots & \vdots \\ \tilde{\mathbf{R}}_{(k+(N-1)K, k)}^* & \tilde{\mathbf{R}}_{(k+(N-1)K, k+K)}^* & \cdots & \tilde{\mathbf{R}}_{(k+(N-1)K, k+(N-1)K)}^* \end{bmatrix} \quad (20)$$

$$\mathbf{C}_v^{(r+1)} = \min_{\mathbf{C}_v \in \mathcal{S}^{N \times K}} \sum_{k=1}^K \left\| (\mathbf{Y}_v^T)_{\cdot, k} - \mathcal{H}_k^{(r)} \mathbf{c}_v(k) \right\|^2 + \mathbf{c}_{v,k}^H \mathbf{R}_k^{(r)} \mathbf{c}_{v,k}. \quad (21)$$

can be simplified as

$$\mathcal{L}(\mathbf{C}_{1:V}, \mathbf{\Gamma} | \mathbf{C}_{1:V}^{(r)}, \mathbf{\Gamma}^{(r)}) = \mathbf{E}_{\mathbf{H}|\mathbf{Y}_{\mathcal{F}}; \mathbf{C}_{1:V}^{(r)}, \mathbf{\Gamma}^{(r)}} \{ \log p(\mathbf{Y}_{\mathcal{F}}, \mathbf{H}; \mathbf{C}_{1:V}, \mathbf{\Gamma}) \} \quad (22)$$

$$= \mathbf{E}_{\mathbf{H}|\mathbf{Y}_{\mathcal{F}}; \mathbf{C}_{1:V}^{(r)}, \mathbf{\Gamma}^{(r)}} \{ \log p(\mathbf{Y}_{\mathcal{F}} | \mathbf{H}; \mathbf{C}_{1:V}, \mathbf{\Gamma}) + \log p(\mathbf{H}; \mathbf{C}_{1:V}, \mathbf{\Gamma}) \} \quad (23)$$

$$= \mathbf{E}_{\mathbf{H}|\mathbf{Y}_{\mathcal{F}}; \mathbf{C}_{1:V}^{(r)}, \mathbf{\Gamma}^{(r)}} \{ \log p(\mathbf{Y}_{\mathcal{F}} | \mathbf{H}; \mathbf{C}_{1:V}, \mathbf{\Gamma}) + \log p(\mathbf{H}; \mathbf{\Gamma}) \}, \quad (24)$$

where $\log p(\mathbf{Y}_{\mathcal{F}} | \mathbf{H}; \mathbf{C}_{1:V}, \mathbf{\Gamma}) \equiv -\frac{\sum_{v=1}^V \|\mathbf{Y}_v - \Phi_v \mathbf{H}\|^2}{\sigma^2}$ and the simplification in (24) results from the fact that the channel likelihood $p(\mathbf{H}; \mathbf{C}_{1:V}, \mathbf{\Gamma})$ is independent of the codewords $\mathbf{C}_{1:V}$. The *a posteriori* density $p(\mathbf{H}|\mathbf{Y}_{\mathcal{F}}; \mathbf{C}_{1:V}^{(r)}, \mathbf{\Gamma}^{(r)})$ is derived similar to that of the P-MSBL scenario with mean $\mathbf{M}^{(r)} \in \mathbb{C}^{LN \times M}$ and covariance $\Sigma^{(r)} \in \mathbb{C}^{LN \times LN}$ given as

$$\mathbf{M}^{(r)} = [\boldsymbol{\mu}_1^{(r)}, \dots, \boldsymbol{\mu}_M^{(r)}] \quad (25)$$

$$\Sigma^{(r)} = \left(\sigma^{-2} (\Phi_{\mathcal{F}}^{(r)})^H \Phi_{\mathcal{F}}^{(r)} + (\Gamma_{\mathcal{K}}^{(r)})^{-1} \right)^{-1}. \quad (26)$$

The vector $\boldsymbol{\mu}_j^{(r)}$ is evaluated as $\boldsymbol{\mu}_j^{(r)} = \sigma^{-2} \Sigma^{(r)} (\Phi_{\mathcal{F}}^{(r)})^H (\mathbf{Y}_{\mathcal{F}})_{\cdot, j}$ and corresponds to column-wise concatenation of the mean vectors $(\boldsymbol{\mu}_{j,i}^{(r)})$ for all $1 \leq i \leq N$ transmit antennas i.e. $\boldsymbol{\mu}_j^{(r)} = \left[(\boldsymbol{\mu}_{j,1}^{(r)})^T, \dots, (\boldsymbol{\mu}_{j,N}^{(r)})^T \right]^T$. The concatenated measurement

$$\text{matrix } \Phi_{\mathcal{F}}^{(r)} = \left[(\Phi_1^{(r)})^T, \dots, (\Phi_V^{(r)})^T \right]^T \in \mathbb{C}^{KV \times LN}.$$

The M-step aims at jointly maximizing the cost function $\mathcal{L}(\mathbf{C}_{1:V}, \mathbf{\Gamma} | \mathbf{C}_{1:V}^{(r)}, \mathbf{\Gamma}^{(r)})$ with respect to both the unknown hyperparameter matrix $\mathbf{\Gamma}$ and codewords $\mathbf{C}_{1:V}$. Interestingly, the joint maximization problem in (24) reduces to decoupled optimization problems with respect to the hyperparameters $\mathbf{\Gamma}$ and the codewords $\mathbf{C}_{1:V}$. This arises since the first term in (23) is independent of the hyperparameters γ . The second term $\log p(\mathbf{H}; \mathbf{C}_{1:V}, \mathbf{\Gamma})$ corresponding to the channel

likelihood is thus independent of the block codeword $\mathbf{C}_{1:V}$. The resulting expression for hyperparameter estimation is similar to that of the P-MSBL scheme and is given as

$$\gamma_l^{(r+1)} = \frac{1}{MN} \sum_{j=1}^M \sum_{i=0}^{N-1} \left(|\mathbf{M}_{(l+iL, j)}^{(r)}|^2 + \Sigma_{(l+iL, l+iL)}^{(r)} \right), \quad (27)$$

where the *a posteriori* mean and covariance matrices $\mathbf{M}^{(r)}$ and $\Sigma^{(r)}$ are given in (25) and (26) respectively. For optimal codeword detection, one can now express maximization of the log-likelihood function $\mathcal{L}(\mathbf{C}_{1:V}, \mathbf{\Gamma} | \mathbf{C}_{1:V}^{(r)}, \mathbf{\Gamma}^{(r)})$ in (23) with respect to the block code matrix $\mathbf{C}_{1:V}$ as shown in (29). Upon employing the equivalent MIMO-OFDM system model of (4), the codeword detection optimization problem above can be recast as shown in (19), where $\mathcal{H} = [\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_K]$ denotes a concatenation of the effective MIMO channel matrix \mathcal{H}_k corresponding to $1 \leq k \leq K$ subcarriers. The decoupled optimization problem constructed for each codeword \mathbf{C}_v is shown in (30) and can be further simplified as demonstrated in (31). One then can further simplify the maximization problem by the addition and subtraction of the term $\mathbf{c}_{v,k}^H \mathbf{E}_{\mathcal{H}|\mathbf{Y}_{\mathcal{F}}; \mathbf{C}_v^{(r)}, \mathbf{\Gamma}^{(r)}} \{ \mathcal{H}_k^H \} \mathbf{E}_{\mathcal{H}|\mathbf{Y}_{\mathcal{F}}; \mathbf{C}_v^{(r)}, \mathbf{\Gamma}^{(r)}} \{ \mathcal{H}_k \} \mathbf{c}_{v,k}$ in the summation above to obtain $\mathbf{C}_v^{(r+1)}$ as shown in (21).

The matrix $\mathcal{H}_k^{(r)} \triangleq \mathbf{E}_{\mathcal{H}|\mathbf{Y}_{\mathcal{F}}; \mathbf{C}_v^{(r)}, \mathbf{\Gamma}^{(r)}} \{ \mathcal{H}_k \} \in \mathbb{C}^{M \times N}$ denotes the *a posteriori* estimate of the frequency domain MIMO channel matrix \mathcal{H}_k and is determined employing updates of the previous r^{th} iteration of the E-step as

$$\mathcal{H}_k^{(r)} = \begin{bmatrix} (\mathbf{F}\boldsymbol{\mu}_{1,1}^{(r)})_k & \cdots & (\mathbf{F}\boldsymbol{\mu}_{1,N}^{(r)})_k \\ \vdots & \ddots & \vdots \\ (\mathbf{F}\boldsymbol{\mu}_{M,1}^{(r)})_k & \cdots & (\mathbf{F}\boldsymbol{\mu}_{M,N}^{(r)})_k \end{bmatrix} \quad (28)$$

and $\mathbf{R}_k^{(r)}$ denotes the covariance of $(\mathcal{H}_k^{(r)})^H$ as depicted in (21) with $\tilde{\mathbf{R}} = (\mathbf{I}_N \otimes \mathbf{F}) \Sigma^{(r)} (\mathbf{I}_N \otimes \mathbf{F})^H \in \mathbb{C}^{NK \times NK}$. The optimal codeword estimation problem for the DA-MSBL

$$\begin{aligned} \mathbf{C}_{1:V}^{(r+1)} &= \arg \max_{\mathbf{C}_{1:V} \in \mathcal{S}^{N \times V \times K}} \mathbf{E}_{\mathbf{H} | \mathbf{Y}_{\mathcal{F}}; \mathbf{C}_{1:V}^{(r)}, \Gamma^{(r)}} \{ \log p(\mathbf{Y}_{\mathcal{F}} | \mathbf{H}; \mathbf{C}_{1:V}, \Gamma) \} \\ &= \arg \max_{\mathbf{C}_{1:V} \in \mathcal{S}^{N \times V \times K}} \mathbf{E}_{\mathbf{H} | \mathbf{Y}_{\mathcal{F}}; \mathbf{C}_{1:V}^{(r)}, \Gamma^{(r)}} \left\{ \sum_{v=1}^V \log p(\mathbf{Y}_v | \mathbf{H}; \mathbf{C}_v, \Gamma) \right\} \end{aligned} \quad (29)$$

$$\mathbf{C}_v^{(r+1)} = - \arg \max_{\mathbf{C}_v \in \mathcal{S}^{N \times K}} \mathbf{E}_{\mathcal{H} | \mathbf{Y}_{\mathcal{F}}; \mathbf{C}_v^{(r)}, \Gamma^{(r)}} \left\{ \sum_{k=1}^K \left\| (\mathbf{Y}_v^T)_{\cdot k} - \mathcal{H}_k \mathbf{c}_{v,k} \right\|^2 \right\}. \quad (30)$$

$$\begin{aligned} \mathbf{C}_{1:V}^{(r+1)} &= - \arg \max_{\mathbf{C}_v \in \mathcal{S}^{N \times K}} \sum_{k=1}^K (\mathbf{Y}_v^T)_{\cdot k}^H (\mathbf{Y}_v^T)_{\cdot k} - \mathbf{E}_{\mathcal{H} | \mathbf{Y}_{\mathcal{F}}; \mathbf{C}_v^{(r)}, \Gamma^{(r)}} \left\{ \mathcal{H}_k^H \right\} \mathbf{c}_{v,k}^H (\mathbf{Y}_v^T)_{\cdot k} - \\ &\quad (\mathbf{Y}_v^T)_{\cdot k}^H \mathbf{E}_{\mathcal{H} | \mathbf{Y}_{\mathcal{F}}; \mathbf{C}_v^{(r)}, \Gamma^{(r)}} \left\{ \mathcal{H}_k \right\} \mathbf{c}_{v,k} + \mathbf{c}_{v,k}^H \mathbf{E}_{\mathcal{H} | \mathbf{Y}_{\mathcal{F}}; \mathbf{C}_v^{(r)}, \Gamma^{(r)}} \left\{ \mathcal{H}_k^H \mathcal{H}_k \right\} \mathbf{c}_{v,k} \end{aligned} \quad (31)$$

approach in MO-SC wireless systems takes the final form of

$$\mathbf{C}_v^{(r+1)} = \arg \min_{\mathbf{C}_v \in \mathcal{S}^{N \times K}} \sum_{k=1}^K \left\| \begin{bmatrix} (\mathbf{Y}_v^T)_{\cdot k} \\ \mathbf{0}_{n \times 1} \end{bmatrix} - \begin{bmatrix} \mathcal{H}_k^{(r)} \\ (\mathcal{R}_k^{(r)})^{1/2} \end{bmatrix} \mathbf{c}_{v,k} \right\|^2 \quad (32)$$

As can be seen, a unique aspect of the procedure for codeword estimation described above is that it directly integrates the paradigm for ML sequence detection in the M-step of the EM framework thereby yielding the optimal bit estimates, unlike existing approaches in [20], [26]. Finally, the MIMO channel and codeword estimates after I iterations are given by, $\hat{\mathbf{H}}_{\mathcal{D}, \mathcal{A}} = \mathbf{M}^I$ and $\hat{\mathbf{C}}_{1:V} = \mathbf{C}_{1:V}^I$ respectively. As demonstrated in the seminal contributions on SBL by Wipf and Rao [21], [37], the proposed MSBL schemes are globally convergent, i.e. each iteration leads to an increase in the ML cost function until a fixed point is reached. Furthermore, similar to the criterion employed in other related SBL treatises [24], [26], the proposed schemes are said to converge, when the difference between the successive hyperparameter estimates becomes sufficiently small, i.e. $\|\gamma^{(r+1)} - \gamma^{(r)}\|^2 \leq \epsilon$ and the threshold ϵ is a suitably chosen small value. For instance, this has been set as $\epsilon = 10^{-8}$ in the simulations.

V. PEP UPPER BOUND FOR MSBL-BASED MLSD

Similar to existing works [38], [39], the $(i, j)^{th}$ MIMO channel matrix coefficient estimate $[\hat{\mathcal{H}}_k]_{(i,j)}$ is modeled as

$$[\hat{\mathcal{H}}_k]_{(i,j)} = [\mathcal{H}_k]_{(i,j)} + [\Delta_k]_{(i,j)}. \quad (33)$$

The variance $\sigma_{\mathcal{H},k}^2 = \mathbf{E} \left\{ \left| [\mathcal{H}_k]_{(i,j)} \right|^2 \right\}$ of $[\mathcal{H}_k]_{(i,j)}$ can be evaluated as

$$\sigma_{\mathcal{H},k}^2 = \mathbf{E} \left\{ \left| \sum_{l=1}^L \mathbf{F}_{\cdot k}(l) h_{i,j}(l) \right|^2 \right\} \quad (34)$$

$$= \sum_{l=1}^L \mathbf{E} \left\{ \left| \mathbf{F}_{\cdot k}(l) h_{i,j}(l) \right|^2 \right\} \quad (35)$$

$$= \sum_{l=1}^L \mathbf{E} \left\{ \left| h_{i,j}(l) \right|^2 \right\} \quad (36)$$

where the simplification from (34) to (35) exploits the fact that the channel coefficients $h_{i,j}(l)$ are independent $\forall 1 \leq l \leq L$

Algorithm 2 DA-MSBL-based Joint Channel Estimation and Sequence Detection in MO-SC Trellis Coded Systems

Input: $\mathbf{Y}_{\mathcal{F}}$, σ^2 , V , I

Initialization: $\gamma_l^{(0)} = 1$, $\forall 1 \leq l \leq L \implies \Gamma^{(0)} = \mathbf{I}_L$, $\mathbf{C}_{1:V}^{(0)}$: Obtained via (6) using $\hat{\mathbf{H}}_{\mathcal{D}}$,

Set counter $r = -1$

while $r < I$ **do**

$r \leftarrow r + 1$

E-step:

- 1) $\Phi_{\mathcal{F}}^{(r)} = \left[\left(\Phi_1^{(r)} \right)^T, \dots, \left(\Phi_V^{(r)} \right)^T \right]^T$; $\Phi_v^{(r)}$, $1 \leq v \leq V$ using $\mathbf{C}_{1:V}^{(r)}$ as per (7)
- 2) Evaluate *a posteriori* mean $\mathbf{M}^{(r)}$ and covariance matrix $\Sigma^{(r)}$ as per (25) and (26) respectively
- 3) Construct $\mathcal{H}_k^{(r)}$ and $\mathcal{R}_k^{(r)}$ as per (28) and (21) respectively

M-step:

- 1) Evaluate $\gamma_l^{(r+1)}$, $\forall 1 \leq l \leq L$ as per (27)
- 2) Evaluate $\mathbf{C}_v^{(r+1)}$ $\forall 1 \leq v \leq V$ as per (32)

end while

Output: $\hat{\mathbf{H}}_{\mathcal{D}, \mathcal{A}} = \mathbf{M}^{(I)}$, $\hat{\mathbf{C}}_{1:V} = \mathbf{C}_{1:V}^{(I)}$

with $\mathbf{E}\{h_{i,j}(l)h_{i,j}(l')\} = \mathbf{E}\{h_{i,j}(l)\}\mathbf{E}\{h_{i,j}(l')\} = 0 \forall l \neq l'$. The final result in (36) can be obtained by invoking the following simplifications in (35). Since we have $\mathbf{E}\{|ax|^2\} = |a|^2\mathbf{E}\{|x|^2\}$ where a is a constant and $|\mathbf{F}_{\cdot k}(l)|^2 = 1$, $\forall 1 \leq k \leq K, 1 \leq l \leq L$, the variance $\sigma_{\mathcal{H},k}^2$ is the same for all $1 \leq k \leq K$ and the subscript k is dropped henceforth for notational convenience.

Let the channel estimation error $[\Delta_k]_{(i,j)}$ in (33) be assumed to be independent of $[\mathcal{H}_k]_{(i,j)}$ and modeled as a symmetric complex Gaussian random variable with zero mean and variance σ_{Δ}^2 . Similar to $\sigma_{\mathcal{H}}^2$, one can deduce that the variance $\sigma_{\hat{\mathcal{H}}}^2$ is independent of k and can be evaluated as

$$\sigma_{\hat{\mathcal{H}}}^2 = \sum_{l=1}^L \mathbf{E} \left\{ \left| \hat{h}_{i,j}(l) \right|^2 \right\}. \quad (37)$$

Hence, it follows that $[\hat{\mathcal{H}}_k]_{(i,j)}$ is a zero mean complex Gaussian random variable with variance $\sigma_{\hat{\mathcal{H}}}^2$ given by, $\sigma_{\hat{\mathcal{H}}}^2 = \sigma_{\mathcal{H}}^2 + \sigma_{\Delta}^2$. The correlation coefficient ρ between $[\mathcal{H}_k]_{(i,j)}$ and

$$d^2(\mathbf{C}, \hat{\mathbf{C}}) = \frac{1}{\sigma_{\hat{\mathcal{H}}}^2} \sum_{k=1}^K \sum_{j=1}^M \left| \sum_{i=1}^N [\hat{\mathcal{H}}_k]_{(i,j)} \Delta c_{i,k} \right|^2, \quad (38)$$

$$= \frac{1}{\sigma_{\hat{\mathcal{H}}}^2} \sum_{j=1}^M \sum_{k=1}^K \left[[\hat{\mathcal{H}}_k]_{(1,j)}, \dots, [\hat{\mathcal{H}}_k]_{(N,j)} \right] \underbrace{\begin{bmatrix} \Delta c_{1,k} \\ \vdots \\ \Delta c_{N,k} \end{bmatrix}}_{\Delta \mathbf{c}_k} [\Delta c_{1,k}^*, \dots, \Delta c_{N,k}^*] \begin{bmatrix} [\hat{\mathcal{H}}_k^*]_{(1,j)} \\ \vdots \\ [\hat{\mathcal{H}}_k]_{(N,j)} \end{bmatrix} \quad (39)$$

$$= \frac{1}{\sigma_{\hat{\mathcal{H}}}^2} \sum_{j=1}^M [\hat{\mathbf{h}}_{1,j}^T, \dots, \hat{\mathbf{h}}_{N,j}^T] \left(\sum_{k=1}^K \mathbf{W}_k \Delta \mathbf{c}_k \Delta \mathbf{c}_k^H \mathbf{W}_k^H \right) [\hat{\mathbf{h}}_{1,j}^T, \dots, \hat{\mathbf{h}}_{N,j}^T]^H, \quad (40)$$

$[\hat{\mathcal{H}}_k]_{(i,j)}$ is given by [40], $\rho = \frac{\sqrt{\sigma_{\hat{\mathcal{H}}}^2}}{\sqrt{\sigma_{\hat{\mathcal{H}}}^2 + \sigma_{\Delta}^2}}$. Although the evaluation of the true posterior distribution of the MSBL-based channel estimates is intractable owing to the complex nature of the iterative EM algorithm, the above model of channel estimation errors in (33) directly determines the frame error rate of the STTC MO-SC system as seen later. Consider the transmission of any codeword \mathbf{C} defined in (2). The cost function given by the path metric in the trellis allows us to determine the ML estimate $\hat{\mathbf{C}}$ of the MO-SC codeword given by [38]

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{C}} \sum_{k=1}^K \sum_{j=1}^M \left| [\mathbf{Y}^T]_{(j,k)} - \frac{\rho}{\sigma_{\hat{\mathcal{H}}}} \sum_{i=1}^N [\hat{\mathcal{H}}_k]_{(i,j)} c_{i,k} \right|^2. \quad (41)$$

The conditional PEP $\Pr(\mathbf{C} \rightarrow \hat{\mathbf{C}} \mid [\hat{\mathcal{H}}_k]_{(i,j)})$ of the transmitted codeword \mathbf{C} being erroneously decoded as $\hat{\mathbf{C}}$ can be expressed as [38]

$$\Pr(\mathbf{C} \rightarrow \hat{\mathbf{C}} \mid [\hat{\mathcal{H}}_k]_{(i,j)}) \leq \exp \left(\frac{-\nu |\rho|^2 d^2(\mathbf{C}, \hat{\mathbf{C}})}{4(1 + N\nu(\sigma_{\hat{\mathcal{H}}}^2 - |\rho|^2))} \right). \quad (42)$$

The quantity $\nu = \frac{SNR}{N}$ denotes the total signal to noise power ratio (SNR) per TA. Let the codeword difference vector be defined as $\Delta c_{i,k} = (c_{i,k} - \hat{c}_{i,k})$. Then the squared distance metric $d^2(\mathbf{C}, \hat{\mathbf{C}})$ can be simplified as shown in (38)-(40), where we have $\mathbf{W}_k = (\mathbf{I}_N \otimes \mathbf{w}_k) \in \mathbb{C}^{NL \times N}$ associated with each Fourier vector $\mathbf{w}_k = [1, e^{-j2\pi \frac{k}{K}}, \dots, e^{-j2\pi(L-1)\frac{k}{K}}]^T \in \mathbb{C}^{L \times 1}$. Let the vector $\hat{\mathbf{h}}_{i,j} \in \mathbb{C}^{L_s \times 1}$ denote the sub-vector corresponding to the non-zero coefficients $\mathbf{h}_{i,j}$ with $\hat{\mathbf{h}}_{i,j}$ representing its M-SBL estimate. Furthermore, let $\hat{\mathbf{h}}_j$ corresponding to the j^{th} RA be defined as $\hat{\mathbf{h}}_j = [\hat{\mathbf{h}}_{1,j}^H, \dots, \hat{\mathbf{h}}_{N,j}^H]^T \in \mathbb{C}^{NL_s \times 1}$. Note that the underlying assumption is that the M-SBL estimate considers perfect recovery of the channel support associated with the L_s significant channel components and can be justified as follows. At moderate to high SNRs, the M-SBL-based estimates are sufficiently accurate to guarantee perfect recovery of the channel support associated with the dominant channel components. Furthermore, the error variances at the non-significant tap locations can be neglected, since they are

substantially lower than those of the significant taps. Thus, the squared distance metric can be closely approximated as

$$d^2(\mathbf{C}, \hat{\mathbf{C}}) \approx \sum_{j=1}^M \frac{\hat{\mathbf{h}}_j^H \mathbf{D} \hat{\mathbf{h}}_j}{\sigma_{\hat{\mathcal{H}}}^2}, \quad (43)$$

where the effective codeword difference matrix $\mathbf{D} \in \mathbb{C}^{NL_s \times NL_s}$ in (43) is defined as

$$\mathbf{D} = \sum_{k=1}^K \mathbf{W}_k \Delta \mathbf{c}_k \Delta \mathbf{c}_k^H \mathbf{W}_k^H. \quad (44)$$

Then we have $\mathbf{W}_k = (\mathbf{I}_N \otimes \mathbf{w}_k) \in \mathbb{C}^{NL_s \times N}$ and similar to $\hat{\mathbf{h}}_{i,j}$, \mathbf{w}_k represents a sub-vector of \mathbf{w}_k defined earlier with entries corresponding to only the significant channel tap locations of $\mathbf{h}_{i,j}$. Let $q = \text{rank}(\mathbf{D})$. The positive semi-definite matrix $\mathbf{D} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^H$, where $\mathbf{V} \in \mathbb{C}^{NL_s \times NL_s}$ is a unitary matrix and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_q, 0, \dots, 0)$ is a $NL_s \times NL_s$ diagonal matrix comprised of the eigenvalues of \mathbf{D} . Since the MO-SC system performs encoding/decoding across subcarriers, one has to additionally account for the correlated nature of the subcarrier MIMO channel matrices \mathcal{H}_k which renders the ensuing FER analysis quite challenging. In particular, the rank and eigenvalues associated with the codeword difference matrix \mathbf{D} cannot be obtained for these systems in a straightforward manner from the existing literature on STTC MIMO systems [41]. Thus, these results have been derived in Appendix A and subsequently employed for obtaining the FER bounds, which makes the analysis in this work novel. In contrast, since the MO-SY system described in [34] encodes/decodes across the symbols, the pertinent analysis to determine the FER bounds is much simplified.

Upon substituting the expression $\mathbf{D} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^H$ into (43) we can further simplify $d^2(\mathbf{C}, \hat{\mathbf{C}})$ as

$$d^2(\mathbf{C}, \hat{\mathbf{C}}) = \sum_{j=1}^M \sum_{\tilde{q}=1}^q \lambda_{\tilde{q}} \left| \hat{\beta}_{j,\tilde{q}} \right|^2, \quad (45)$$

where $\hat{\beta}_{j,\tilde{q}}$ denotes the \tilde{q}^{th} component of the vector $\hat{\beta}_j = \frac{1}{\sigma_{\hat{\mathcal{H}}}} \mathbf{V}^H \hat{\mathbf{h}}_j \in \mathbb{C}^{NL_s \times 1}$. Furthermore, owing to the previous assumption that the channel's estimate vector $\hat{\mathbf{h}}_{i,j}$ is comprised of independent components $\forall 1 \leq i \leq N, 1 \leq j \leq M$, all the NL_s components of the vector $\frac{\hat{\mathbf{h}}_j}{\sigma_{\hat{\mathcal{H}}}}$ are also independent. We

can also observe that each element of the vector $\widehat{\beta}_j$ is symmetric zero mean complex Gaussian with its component-wise variance given as $\mathbf{E} \left\{ \left| \widehat{\beta}_j(l) \right|^2 \right\} = \frac{1}{\sigma_{\mathcal{H}}^2} \mathbf{E} \left\{ \left| \widehat{h}_j(l) \right|^2 \right\} \forall 1 \leq l \leq NL_s$. Let $\mathbf{E} \left\{ \left| \widehat{h}_{i,j}(\tilde{l}) \right|^2 \right\} \approx \phi, 1 \leq i \leq N, 1 \leq j \leq M$ for all L_s entries of $\widehat{h}_{i,j}$. On substituting the value of $\sigma_{\mathcal{H}}^2$ from (37), one can write $\mathbf{E} \left\{ \left| \widehat{\beta}_j(l) \right|^2 \right\} = \frac{\mathbf{E} \left\{ \left| \widehat{h}_j(l) \right|^2 \right\}}{\sum_{\tilde{l}=1}^{L_s} \mathbf{E} \left\{ \left| \widehat{h}_{i,j}(\tilde{l}) \right|^2 \right\}}$ which can be further simplified as

$$\mathbf{E} \left\{ \left| \widehat{\beta}_j(l) \right|^2 \right\} \approx \frac{\mathbf{E} \left\{ \left| \widehat{h}_j(l) \right|^2 \right\}}{\sum_{\tilde{l}=1}^{L_s} \mathbf{E} \left\{ \left| \widehat{h}_{i,j}(\tilde{l}) \right|^2 \right\}}, \quad (46)$$

$$= \frac{\phi}{\sum_{\tilde{l}=1}^{L_s} \phi} = \frac{1}{L_s}. \quad (47)$$

Furthermore, since \mathbf{V} is a unitary matrix, the NL_s components of $\frac{\widehat{h}_j(l)}{\sigma_{\mathcal{H}}}$ are distributed as i.i.d $\mathcal{CN} \left(0, \frac{1}{L_s} \right)$. The quantity $\left| \widehat{\beta}_{j,\bar{q}} \right|$ is Rayleigh distributed hence has the PDF of $p \left(\left| \widehat{\beta}_{j,\bar{q}} \right| \right) = 2L_s \left| \widehat{\beta}_{j,\bar{q}} \right| e^{-L_s \left| \widehat{\beta}_{j,\bar{q}} \right|^2}$. Upon substituting the expression of $d^2 \left(\mathbf{C}, \widehat{\mathbf{C}} \right)$ from (45) into (42), the resultant expression of PEP can be expressed in terms of $\widehat{\beta}_{j,\bar{q}}$ as

$$\begin{aligned} & \Pr \left(\mathbf{C} \rightarrow \widehat{\mathbf{C}} \mid \widehat{\beta}_{j,\bar{q}}, 1 \leq \bar{q} \leq q, 1 \leq j \leq M \right) \\ & \leq \prod_{j=1}^M \exp \left(-\bar{\nu} \sum_{\bar{q}=1}^q \lambda_{\bar{q}} \left| \widehat{\beta}_{j,\bar{q}} \right|^2 \right), \end{aligned} \quad (48)$$

with $\bar{\nu} = \frac{\nu |\rho|^2}{4(1+N\nu(\sigma_{\mathcal{H}}^2 - |\rho|^2))}$. The PEP can be obtained by averaging the expression on the right hand side of (48) with respect to the i.i.d Rayleigh PDFs of $\left| \widehat{\beta}_{j,\bar{q}} \right|$

$$\overline{\Pr} \left(\mathbf{C} \rightarrow \widehat{\mathbf{C}} \right) \leq \mathbf{E} \left\{ \prod_{j=1}^M \exp \left(-\bar{\nu} \sum_{\bar{q}=1}^q \lambda_{\bar{q}} \left| \widehat{\beta}_{j,\bar{q}} \right|^2 \right) \right\} \quad (49)$$

$$= \prod_{j=1}^M \prod_{\bar{q}=1}^q \mathbf{E} \left\{ e^{-\bar{\nu} \lambda_{\bar{q}} \left| \widehat{\beta}_{j,\bar{q}} \right|^2} \right\} \quad (50)$$

$$= \left(\prod_{\bar{q}=1}^q \left(1 + \frac{\lambda_{\bar{q}} |\rho|^2 \frac{\nu}{4L_s}}{1 + N\nu(\sigma_{\mathcal{H}}^2 - |\rho|^2)} \right) \right)^{-M} \quad (51)$$

where (51) follows by substituting $\mathbf{E} \left\{ e^{-\bar{\nu} \lambda_{\bar{q}} \left| \widehat{\beta}_{j,\bar{q}} \right|^2} \right\} = 2L_s \int_0^\infty e^{-\bar{\nu} \lambda_{\bar{q}} \left| \widehat{\beta}_{j,\bar{q}} \right|^2} \left| \widehat{\beta}_{j,\bar{q}} \right| e^{-L_s \left| \widehat{\beta}_{j,\bar{q}} \right|^2} d \left| \widehat{\beta}_{j,\bar{q}} \right| = \frac{1}{1 + \frac{\bar{\nu} \lambda_{\bar{q}}}{L_s}}$ and $\bar{\nu}$ in (50). Finally, the resultant average PEP of the STTC MO-SC system in the presence of channel estimation errors can be obtained as

$$\overline{\Pr} \left(\mathbf{C} \rightarrow \widehat{\mathbf{C}} \right) \leq \underbrace{(\Delta_q)^{-M} \left(\frac{|\rho|^2 \frac{\nu}{4L_s}}{1 + N\nu(\sigma_{\mathcal{H}}^2 - |\rho|^2)} \right)^{-qM}}_{P(\Delta_q)}, \quad (52)$$

where we have $\Delta_q = \prod_{\bar{q}=1}^q \lambda_{\bar{q}}$. The detailed procedure of evaluating of Δ_q and the rank q of the effective codeword

difference matrix \mathbf{D} has been described in Appendix A. Furthermore, for the P-MSBL and DA-MSBL channel estimation approaches presented in Sections III and IV respectively, by employing the *a posteriori* covariance matrices $\boldsymbol{\Sigma}^{(l)}$ obtained in the final EM iteration, we can evaluate the respective values of σ_{Δ}^2 as

$$\sigma_{\Delta}^2 = \frac{1}{MNK} \text{Tr} \left\{ (\mathbf{I}_N \otimes \mathbf{F}) \boldsymbol{\Sigma}^{(l)} (\mathbf{I}_N \otimes \mathbf{F})^H \right\}. \quad (53)$$

The expurgated union bound of the coded frame error rate (FER) P_{FER} of the STTC MO-SC system is given by [42]

$$P_{FER} \leq L_B \sum_{w=1}^W A(w) P[\Delta_q(w)]. \quad (54)$$

The quantities W , $L_B = K \log_2 |\mathcal{S}|$, and $A(w)$ denote the order of the expurgated union bound, the total number of symbols mapped to the codeword matrix \mathbf{C} and the normalized weighting coefficient corresponding to the w^{th} product metric. The PEP $P[\Delta_q(w)]$ is a function of the w^{th} product metric $\Delta_q(w)$, which can be readily evaluated from (52). The Bayesian Cramér-Rao bounds are derived next in order to benchmark the performance of the proposed schemes.

VI. BAYESIAN CRAMÉR-RAO BOUNDS (BCRB) FOR P-MSBL AND DA-MSBL

The analytical BCRBs for the MSE of the proposed schemes for the MO-SC systems are given below. Applying the $\text{vec}(\cdot)$ operator of (9), the observation model for the pilot observations $\mathbf{y}_{\mathcal{F},\mathcal{P}} = \text{vec}(\mathbf{Y}_{\mathcal{F},\mathcal{P}})$ can be expressed as

$$\mathbf{y}_{\mathcal{F},\mathcal{P}} = (\mathbf{I}_{MN} \otimes \boldsymbol{\Phi}_{\mathcal{F},\mathcal{P}}) \mathbf{h} + \mathbf{n}_{\mathcal{F},\mathcal{P}}, \quad (55)$$

where the effective parameter vector obeys $\mathbf{h} = \text{vec}(\mathbf{H}) \in \mathbb{C}^{LMN \times 1}$ and similarly $\mathbf{n}_{\mathcal{F},\mathcal{P}} = \text{vec}(\mathbf{N}_{\mathcal{F},\mathcal{P}}) \in \mathbb{C}^{PVM \times 1}$. The Bayesian Fisher information matrix (BFIM) $\mathbf{J}_B \in \mathbb{C}^{LMN \times LMN}$ is defined as [40]

$$\mathbf{J}_B = \underbrace{-\mathbf{E}_{(\mathbf{y}_{\mathcal{F},\mathcal{P}}, \mathbf{h})} \left\{ \frac{\partial^2 \mathcal{L}(\mathbf{y}_{\mathcal{F},\mathcal{P}} | \mathbf{h}; \boldsymbol{\Gamma})}{\partial \mathbf{h} \partial \mathbf{h}^H} \right\}}_{\mathbf{J}_{\mathcal{D}}} \underbrace{-\mathbf{E}_{\mathbf{h}} \left\{ \frac{\partial^2 \mathcal{L}(\mathbf{h}; \boldsymbol{\Gamma})}{\partial \mathbf{h} \partial \mathbf{h}^H} \right\}}_{\mathbf{J}_{\mathcal{P}}}, \quad (56)$$

where the matrices $\mathbf{J}_{\mathcal{D}}$ and $\mathbf{J}_{\mathcal{P}}$ denote the FIMs with respect to the receive vector $\mathbf{y}_{\mathcal{F},\mathcal{P}}$ and the prior density of the parameter vector \mathbf{h} , respectively. The quantity $\mathcal{L}(\mathbf{y}_{\mathcal{F},\mathcal{P}} | \mathbf{h}; \boldsymbol{\Gamma}) \equiv \frac{1}{\sigma^2} \|\mathbf{y}_{\mathcal{F},\mathcal{P}} - \boldsymbol{\Phi}_{\mathcal{F},\mathcal{P}} \mathbf{h}\|^2$ represents the equivalent log-likelihood function of the receive vector $\mathbf{y}_{\mathcal{F},\mathcal{P}}$, ignoring the constants that have no bearing on the BCRB. Upon employing the results in [46], the FIM becomes $\mathbf{J}_{\mathcal{D}} = \sigma^{-2} \boldsymbol{\Phi}_{\mathcal{F},\mathcal{P}}^H \boldsymbol{\Phi}_{\mathcal{F},\mathcal{P}}$. The quantity $\mathcal{L}(\mathbf{h}; \boldsymbol{\Gamma})$ denotes the channel likelihood parameterized by the hyperparameter matrix $\boldsymbol{\Gamma}$ corresponding to the distribution $\mathcal{CN}(\mathbf{0}_{LMN \times 1}, \mathbf{I}_{MN} \otimes \boldsymbol{\Gamma})$. Thus, the FIM component $\mathbf{J}_{\mathcal{P}}$ becomes [40] $\mathbf{J}_{\mathcal{P}} = \mathbf{I}_{MN} \otimes \boldsymbol{\Gamma}^{-1}$. The CRB of the mean squared error of the P-MSBL-based channel estimate $\widehat{\mathbf{h}}_{\mathcal{P}}$ is given by

$$\begin{aligned} & \mathbf{E} \left\{ \left\| \mathbf{h} - \widehat{\mathbf{h}}_{\mathcal{P}} \right\|^2 \right\} \\ & \geq \text{Tr} \left\{ \mathbf{J}_B^{-1} \right\} \quad (57) \\ & = \text{Tr} \left\{ \left((\mathbf{I}_{MN} \otimes \boldsymbol{\Gamma}^{-1})^{-1} + \sigma^{-2} \boldsymbol{\Phi}_{\mathcal{F},\mathcal{P}}^H \boldsymbol{\Phi}_{\mathcal{F},\mathcal{P}} \right)^{-1} \right\} \quad (58) \end{aligned}$$

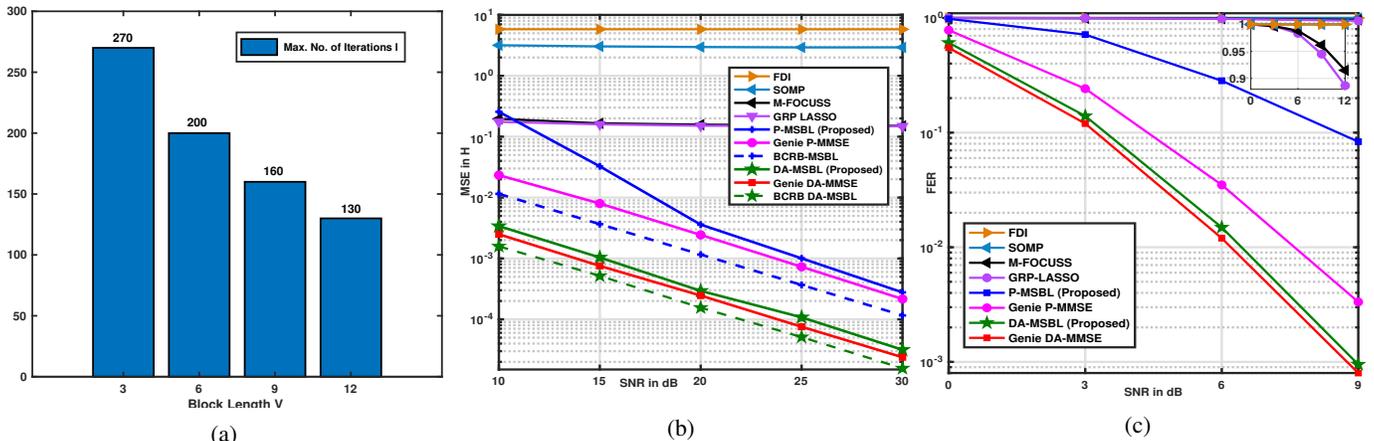


Fig. 2: (a) No. of EM iterations I versus block length V for $\epsilon = 10^{-8}$ (b) MSE and (c) BER performance of 2×2 , 4-state QPSK MO-SC wireless system for proposed SBL, existing sparse [43]–[45], and non-sparse [9] channel estimation schemes.

TABLE II: Power Delay Profile of Pedestrian B Channel Model [47]

Tap (l)	Rel. Delay $\tau(l)$ (in ns)	Avg. Power $\mathbf{E}\{ \alpha(l) ^2\}$ (in dB)
1	0	0
2	200	-0.9
3	800	-4.9
4	1200	-8.0
5	2300	-7.8
6	3700	-23.9

In order to evaluate the BCRB corresponding to the DA-MSBL scheme, since the *a priori* information related to the channel vector \mathbf{h} is identical for both P-MSBL and DA-MSBL, the FIM $\mathbf{J}_{\mathcal{P}}$ remains the same for both the cases. Similarly, the FIM $\mathbf{J}_{\mathcal{D}}$ for the DA-MSBL scheme evaluates to $\mathbf{J}_{\mathcal{D}} = \sigma^{-2} \Phi_{\mathcal{F}}^H \Phi_{\mathcal{F}}$. Finally, the corresponding BCRB can be evaluated as

$$\begin{aligned} & \mathbf{E} \left\{ \left\| \mathbf{h} - \hat{\mathbf{h}}_{\mathcal{D}} \right\|^2 \right\} \\ & \geq \text{Tr} \left\{ \left((\mathbf{I}_{MN} \otimes \Gamma^{-1})^{-1} + \sigma^{-2} \Phi_{\mathcal{F}}^H \Phi_{\mathcal{F}} \right)^{-1} \right\}. \end{aligned} \quad (59)$$

It can be observed that the BCRB expression in (58) corresponds to that of an equivalent pilot-based minimum mean squared error estimator (MMSE) with known channel sparsity, while (59) corresponds to an ideal MMSE, which assumes perfect knowledge of both the channel sparsity and of the transmitted codewords.

VII. SIMULATION RESULTS

An $M \times N$ space-time trellis coded MIMO-OFDM system with $M \in \{2, 4\}$ RAs and $N \in \{2, 4\}$ TAs is considered with $K = 64$ equispaced subcarriers spread over a transmission bandwidth of $B = 5$ MHz. A CP of duration $6.4 \mu\text{s}$ is appended at the beginning of each OFDM symbol. The number of subcarriers is $K = 64$ which equals the block length of the MO-SC system with $P = 10$ pilot subcarriers in each OFDM symbol. The information bits are mapped to a QPSK constellation having $|S| = 4$ symbols followed by the four-state STTC corresponding to $N = 2$ [1] and $N = 4$ TAs

[48]. For the purpose of simulation, the standard Rayleigh fading Pedestrian B channel model of [47] is employed. Table II presents the power delay profile of the Pedestrian B channel, i.e. the propagation delay $\tau(l)$ and average attenuation power $\mathbf{E}\{|\alpha(l)|^2\}$ of each channel tap l , where $1 \leq l \leq L_s$ and $L_s = 6$ denotes the number of non-zero channel taps. The transmit and receive filters $g_i(t)$ and $g_j(t)$ are assumed to be practical raised cosine filters with a roll-off factor of 0.5 [49] for the generation of the baseband channel impulse response $\hat{h}_{i,j}(t)$ in (1). The approximately sparse channel vectors are generated by sampling $\hat{h}_{i,j}(t)$ at the sampling rate of $f_s = 5.1$ MHz similar to the approach employed in [26]. As stated therein, these parameter values are chosen to better enable a performance comparison between the different schemes, while being as close as possible to a realistic settings. The STTC frame is comprised of $V = 6$ consecutive OFDM symbols. The hyperparameters are heuristically initialized as $\gamma_l^{(0)} = 1$ prior to the start of the EM algorithm. Since the proposed schemes do not assume any knowledge of the underlying multipath delay profile, the length of each channel vector $\mathbf{h}_{i,j}$ is set to the CP length of $L = 32$. The stopping criteria ϵ is set as $\epsilon = 10^{-8}$.

Fig. 2(a) demonstrates the number of EM iterations required for convergence of the proposed MSBL-based hyperparameter estimates for different values of V , with the stopping parameter $\epsilon = 10^{-8}$. It can be observed that for lower values of the block length V , the number of iterations required for convergence of the algorithms is higher and vice-versa. Further, for the value of $V = 6$ set in our simulations, the number of EM iterations is observed to be $I = 200$.

The proposed P-MSBL and DA-MSBL schemes are employed to obtain the channel estimates $\hat{\mathbf{H}}_{\mathcal{P}}$ and $\hat{\mathbf{H}}_{\mathcal{D},\mathcal{A}}$ respectively for the MO-SC system. The MSE of the proposed techniques is compared to that of the popular non-sparse frequency domain interpolation (FDI) [9] approach conceived for OFDM systems and also to several other existing sparse estimation approaches, namely the SOMP [43], multiple FOCUSS (M-FOCUSS) [44], group least absolute shrinkage and selection operator (GRP-LASSO) [45]. The performances of the various

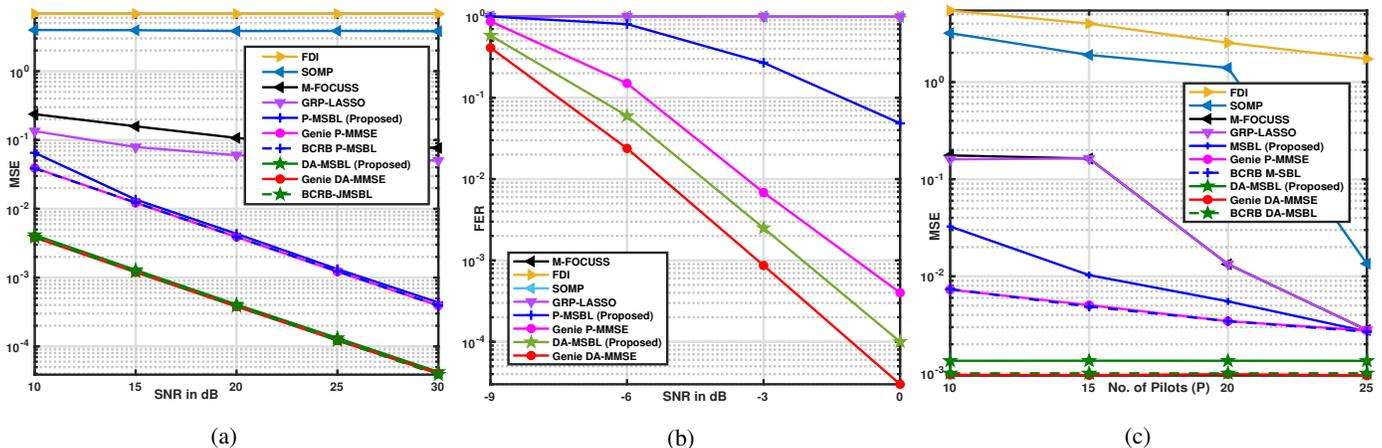


Fig. 3: (a) MSE and (b) BER performance of 4×4 , 4-state QPSK STTC MO-SC wireless system for proposed SBL, existing sparse [43]–[45], and non-sparse [9] channel estimation schemes (c) MSE versus P performance of 2×2 , 4-state QPSK STTC MO-SC wireless system for existing and proposed channel estimation schemes at SNR= 15dB.

schemes are benchmarked against that of the Genie P-MMSE and Genie DA-MMSE approaches, which correspond to ideal MMSE estimators. The Genie P-MMSE estimator employs only the known pilot symbols, whereas the Genie DA-MMSE approach considers all the transmitted symbols to be perfectly known at the receiver for channel estimation. These estimators are ideal, but impractical, owing to their known sparsity assumption i.e. these schemes assume complete knowledge of the location of the significant and non-significant channel taps. However, they serve as valuable benchmarks to compare the performance of the proposed and existing schemes.

Fig. 2(b) plots the MSEs of the approximately sparse channel estimates $\hat{\mathbf{H}}$ obtained using the proposed MSBL approaches, while also comparing them to those of the existing techniques, such as FDI [9], SOMP [43], M-FOCUSS [44], and GRP-LASSO [45]. It can be readily seen that the P-MSBL scheme has a significantly lower MSE than the non-sparse FDI as well as than the sparse SOMP, M-FOCUSS and GRP-LASSO schemes. Additionally, the DA-MSBL technique of joint approximately sparse channel estimation and Viterbi-based MLSD leads to a significant MSE reduction in the estimated channel coefficients. The MSE performance of the DA-MSBL scheme is also seen to almost coincide with that of the data-aided genie MMSE estimator having perfect prior knowledge of the sparse multi-path power delay profile. This demonstrates the effectiveness of the DA-MSBL framework, which does not assume any prior knowledge of the sparse multipath channel, and yet achieves a performance close to that of the genie scheme. Furthermore, the performance of the P-MSBL and DA-MSBL schemes is seen to be in close conformance with their respective BCRBs derived in Section VI, which further ascertains the robustness of the proposed MSBL techniques. This is because the BCRB of P-MSBL corresponds to a pilot-based MMSE having known channel sparsity while the BCRB for DA-MSBL represents an ideal MMSE which assumes perfect knowledge of both the channel sparsity and of the transmitted data symbols. By contrast, neither of our P-MSBL and DA-MSBL do not have any a

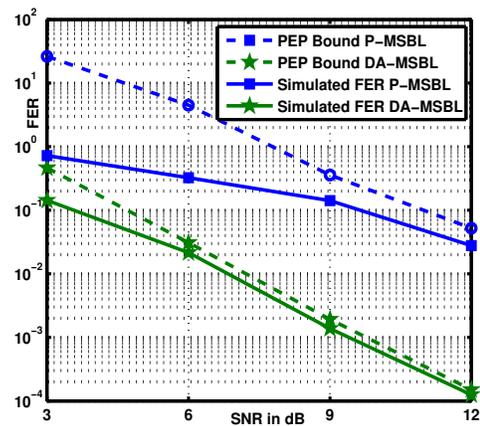


Fig. 4: Comparison of Theoretical FER bounds derived in (54) with their practical counterparts for the proposed MSBL-based MLSD techniques in MO-SC wireless system.

priori knowledge of the channel sparsity associated with DA-MSBL additionally performing data detection. Fig. 2(c) shows a comparison of the frame error rate (FER) of the STTC MIMO-OFDM system relying on the proposed and on the existing schemes. We can observe a trend similar to that of the MSE performance of the various competing schemes to that of the proposed P-MSBL techniques resulting in a significant performance improvement.

Fig. 3(a) and 3(b) depict the MSE and FER, respectively for a 4×4 four-state QPSK STTC MO-SC wireless system. The performance of the various existing and proposed schemes shows a trend similar to that of the 2×2 MO-SC system, thus demonstrating the improved performance of the proposed schemes for MIMO systems of various orders. Fig. 3(c) illustrates the MSE of the various schemes with respect to the number of pilots P with the total number of transmitted symbols fixed as K . Interestingly, it can be observed that the MSE performance of the pilot-based techniques improves upon increasing the number of pilot symbols, while that of

the data-aided DA-MSBL and the Genie DA-MMSE remains unaffected, since they employ all the pilot and data symbols for channel estimation, the sum total of which is constant. Thus, the DA-MSBL delivers equivalent performance for fewer pilot symbols, which demonstrates its efficiency. Fig. 4 demonstrates our comparison between the FER plots of P-MSBL and DA-MSBL obtained via simulation and the corresponding theoretical bounds developed in Section V. The various parameters used for computing the theoretical bounds are as follows. The standard four-state STTC of [1] is considered for the 2×2 STTC MIMO system with symbols drawn from a $|\mathcal{S}| = 4$ QPSK constellation. For this system, the order of the expurgated union bound W , the product metric $\tilde{\Delta}_q$, and the weighing coefficients $A(w)$ are obtained from [41], yielding $W = 2$, $\tilde{\Delta}_q(1, 2) = \{4, 16\}$, and $A(1, 2) = \{2, 1\}$ respectively. A detailed procedure of evaluating Δ_q corresponding to $\tilde{\Delta}_q(1, 2) = \{4, 16\}$ has been derived in the Appendix A. The gap between the theoretical and practical FER curves decreases upon increasing SNR since the PEP approximation in (43) incorporating NL_s dominant components and their squared distance metric $d^2(\mathbf{C}, \hat{\mathbf{C}})$ in (43) becomes progressively tight at high SNR. It is observed that the theoretical bounds approach the simulated curves at high SNR, thereby validating the derived analytical results.

1. Complexity analysis and comparison

This subsection presents a brief complexity comparison of the proposed MSBL schemes with that of the existing and ideal (albeit impractical) schemes employed for comparison purposes. As seen in Algorithms 1 and 2, the channel estimation complexities of the P-MSBL and DA-MSBL schemes are dominated by the matrix inverse computations associated with the evaluation of the *a posteriori* covariance matrix $\Sigma^{(r)}$ in (14) and (26) respectively. Employing the Woodbury matrix identity [50], it can be deduced that these inverse computations for the P-MSBL and DA-MSBL approaches have complexities of $\mathcal{O}(P^3V^3)$ and $\mathcal{O}(L^3N^3)$, respectively. The former is naturally lower than the latter. This is because the P-MSBL approach employs only pilots for channel estimation while the DA-MSBL incurs an additional complexity associated with the codeword estimation procedure in the M-step. A brief discussion on the MLSD complexity associated with the proposed DA-MSBL scheme is presented along with the comparison of the existing MSBL-based symbol detection technique of [26]. The exact step-wise details of the MLSD complexities for both the schemes is presented in the technical report in [51]. It is interesting to note that the scheme in [26] has an exponentially increasing complexity with respect to the number of TAs. Furthermore, for the outer turbo code in Fig. 1 of [26], the bit detection is performed disjointly with respect to the MSBL-based channel estimation scheme, hence resulting in an additional computational overhead and suboptimal bit estimates. By contrast, the proposed DA-MSBL scheme naturally amalgamates the ML decoding paradigm with the EM framework and performs joint channel estimation and sequence detection using a modified path metric-based Viterbi decoder. Coming now to the existing schemes, the

sparsity agnostic FDI scheme is observed to have the lowest computational complexity of order $\mathcal{O}(N^3M^3)$ imposed by the matrix inversion of least squares estimation followed by a simple linear interpolation. However, it yields an extremely poor MSE and FER performance compared to the remaining schemes. Furthermore, the complexity of the SOMP scheme is $\mathcal{O}(L^3N^3M^3)$, which is significantly higher than that of its pilot-based counterpart P-MSBL. This is owing to the interior point method-based convex solver employed to obtain the corresponding estimates [52]. The complexity order of the Genie - Pilot and Genie - Data Aided schemes having a known multipath delay profile is $\mathcal{O}(L_s^3N^3)$, where L_s denotes the number of significant channel taps, with an additional complexity imposed by STTC decoding in the genie DA scheme. Finally, albeit the complexities of these approaches are the lowest, they simply serve as performance benchmarks and are impractical. Table III summarizes the decoding complexities associated with the proposed and existing approaches. As illustrated therein, the complexity order of $\mathcal{O}(KVN^2)$ of MLSD employing DA-MSBL in MO-SC systems is lower than that of its analogous counterparts given by $\mathcal{O}[(K - P)VK^2N^2]$ for the existing approach in [26].

VIII. CONCLUSIONS

We have developed MSBL-based schemes for approximately sparse channel estimation in MO-SC STTC MIMO wireless systems. Starting with the parameterized MSBL prior for the MIMO-OFDM channel matrix, the pilot-based P-MSBL scheme estimates the channel via Bayesian evidence maximization. Then, the DA-MSBL framework was developed for joint maximization of the ensuing likelihood function with respect to both the hyperparameters and the unknown space-time trellis codewords via a novel modified path metric-based Viterbi decoder. Theoretical PEP bounds were also developed for the proposed Viterbi-based MLSD in the presence of channel estimation errors including a procedure of deriving the product metric for MO-SC STTC MIMO systems. Our simulation results demonstrated the enhanced MSE and FER performance of the proposed schemes and also validated the analytical bounds derived.

APPENDIX A

EVALUATION OF THE PRODUCT METRIC Δ_q FOR A MO-SC SYSTEM FOR \mathbf{D} IN (44)

For a conventional STTC MIMO system, the distance matrix $\tilde{\mathbf{D}}$ is given by [48]

$$\tilde{\mathbf{D}} = \sum_{k=1}^K \tilde{\Delta}_{\mathbf{c}_k} \tilde{\Delta}_{\mathbf{c}_k}^H, \quad (60)$$

where $\tilde{\Delta}_{\mathbf{c}_k} \in \mathbb{C}^{N \times 1}$ denotes the k^{th} independent error vector for a K block length STTC. For the 2×2 , $|\mathcal{S}| = 4$, and four-state trellis coded MIMO system [1] used in the simulation results, the product metric $\tilde{\Delta}_q$ and rank \tilde{q} corresponding to the distance matrix \mathbf{D}' of the 2×2 , four-state STTC is given by $\tilde{\Delta}_q = \{4, 16\}$ and $q = 2$ [41]. Furthermore, for the $w = 1$ product metric, the eigenvalues $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ are given by

TABLE III: Order of complexity for MLSD using proposed DA-MSBL scheme and symbol detection in [26]

DA-MSBL in MO-SC systems	$\mathcal{O}(AKV S ^2N^2 + KLN^3(K+L))$
Symbol detection in [26]	$\mathcal{O}(S ^N(K-P)VK^2N^2 + KLN^3(K+L))$

$\tilde{\lambda}_1 = 2, \tilde{\lambda}_2 = 2$ such that $\tilde{\Delta}_q(1) = \tilde{\lambda}_1 \tilde{\lambda}_2 = 4$. One can express $\tilde{\mathbf{D}} = \tilde{\mathbf{D}}_1 + \tilde{\mathbf{D}}_2$, where $\tilde{\mathbf{D}}_1$ and $\tilde{\mathbf{D}}_2$ are the two constituent rank one outer product matrices corresponding to the linearly independent error vectors $\tilde{\Delta}\mathbf{c}_{k_1}$ and $\tilde{\Delta}\mathbf{c}_{k_2}$ for k_1, k_2 particular realizations of $k : 1 \leq k \leq K$. Using Weyl's inequality, the eigenvalues $\lambda(\mathbf{Y} + \mathbf{Z})$ of the sum of two matrices \mathbf{Y}, \mathbf{Z} satisfy [53] $\lambda_{u+v-1}(\mathbf{Y} + \mathbf{Z}) \leq \lambda_u(\mathbf{Y}) + \lambda_v(\mathbf{Z})$, where the indices $u, v \geq 1$ and $u+v-1 \leq n_{tot}$ with n_{tot} denoting the total number of positive eigenvalues of \mathbf{Y} and \mathbf{Z} . Using the above result, one can obtain the following relationships corresponding to the eigenvalues of the component matrices $\tilde{\mathbf{D}}_1$ and $\tilde{\mathbf{D}}_2$, $\lambda_1(\tilde{\mathbf{D}}_1) + \lambda_1(\tilde{\mathbf{D}}_2) \geq 2$, $\lambda_2(\tilde{\mathbf{D}}_1) + \lambda_1(\tilde{\mathbf{D}}_2) \geq 2$, and $\lambda_1(\tilde{\mathbf{D}}_1) + \lambda_2(\tilde{\mathbf{D}}_2) \geq 2$. Additionally, since $\tilde{\mathbf{D}}_1$ and $\tilde{\mathbf{D}}_2$ are both rank one matrices, one of $\lambda_1(\tilde{\mathbf{D}}_1)$ or $\lambda_2(\tilde{\mathbf{D}}_1)$ and $\lambda_1(\tilde{\mathbf{D}}_2)$ or $\lambda_2(\tilde{\mathbf{D}}_2)$ is zero. Without any loss of generality, one of the solutions is given by, $\lambda_1(\tilde{\mathbf{D}}_1) = 2, \lambda_2(\tilde{\mathbf{D}}_1) = 0, \lambda_1(\tilde{\mathbf{D}}_2) = 2$, and $\lambda_2(\tilde{\mathbf{D}}_2) = 0$. Corresponding to the dominant error events as in (60), the distance matrix \mathbf{D} for the MO-SC system in (44) can be expressed as

$$\mathbf{D} = \mathbf{D}_1 + \mathbf{D}_2, \quad (61)$$

where $\mathbf{D}_1 = \mathbf{W}_{k_1} \tilde{\mathbf{D}}_1 \mathbf{W}_{k_1}^H$, $\mathbf{D}_2 = \mathbf{W}_{k_2} \tilde{\mathbf{D}}_2 \mathbf{W}_{k_2}^H$, $\mathbf{W}_{k_i} = (\mathbf{I}_N \otimes \mathbf{w}_{k_i}) \in \mathbb{C}^{NL_s \times N}$ and $k_i, 1 \leq i \leq 2$ corresponds to two realizations of $\{k : 1 \leq k \leq K\}$ for which the codeword difference matrices $\tilde{\mathbf{D}}_1$ and $\tilde{\mathbf{D}}_2$ do not vanish. Employing the result, $\text{rank}(\mathbf{Y}\mathbf{Z}\mathbf{Y}^H) = \text{rank}(\mathbf{Z})$ for $\mathbf{Y} \in \mathbb{C}^{p \times q}$ and $\mathbf{Z} \in \mathbb{C}^{q \times q}$ provided \mathbf{Y} is a full column rank matrix, it follows that the ranks of the matrices \mathbf{D}_1 and \mathbf{D}_2 are $\text{rank}(\mathbf{W}_{k_1} \tilde{\mathbf{D}}_1 \mathbf{W}_{k_1}^H) = \text{rank}(\tilde{\mathbf{D}}_1)$ and $\text{rank}(\mathbf{W}_{k_2} \tilde{\mathbf{D}}_2 \mathbf{W}_{k_2}^H) = \text{rank}(\tilde{\mathbf{D}}_2)$ since \mathbf{W}_{k_1} and \mathbf{W}_{k_2} are full column rank matrices. Additionally, it can also be verified that the vectors $\Delta\mathbf{c}_{k_1} \triangleq \mathbf{W}_{k_1} \tilde{\Delta}\mathbf{c}_{k_1}$ and $\Delta\mathbf{c}_{k_2} \triangleq \mathbf{W}_{k_2} \tilde{\Delta}\mathbf{c}_{k_2}$ are linearly independent. Since $\mathbf{D}_1 = \Delta\mathbf{c}_{k_1} \Delta\mathbf{c}_{k_1}^H$ and $\mathbf{D}_2 = \Delta\mathbf{c}_{k_2} \Delta\mathbf{c}_{k_2}^H$, the rank q of the matrix \mathbf{D} which is a sum of the matrices \mathbf{D}_1 and \mathbf{D}_2 as shown in (61), evaluates as $q = 2$. Using the result for $\mathbf{Y} \in \mathbb{C}^{u \times v}$ and $\mathbf{Z} \in \mathbb{C}^{v \times u}$ with $v \geq u$, the v eigenvalues of the matrix $\mathbf{Z}\mathbf{Y}$ are the u eigenvalues of matrix $\mathbf{Y}\mathbf{Z}$ with the remaining eigenvalues being zero. Additionally, since $\text{rank}(\mathbf{D}_1) = \text{rank}(\tilde{\mathbf{D}}_1)$, we can evaluate the eigenvalues $\lambda(\mathbf{D}_1)$ of the matrix \mathbf{D}_1 as,

$$\lambda(\mathbf{D}_1) = \lambda(\mathbf{W}_{k_1} \tilde{\mathbf{D}}_1 \mathbf{W}_{k_1}^H) \quad (62)$$

$$= \lambda(\mathbf{W}_{k_1}^H \mathbf{W}_{k_1} \tilde{\mathbf{D}}_1) \quad (63)$$

$$= \lambda(L_s \mathbf{I}_n \tilde{\mathbf{D}}_1) \quad (64)$$

$$= L_s \lambda(\tilde{\mathbf{D}}_1). \quad (65)$$

Note that the result $\mathbf{W}_k^H \mathbf{W}_k = L_s \mathbf{I}_N \forall 1 \leq k \leq K$ is employed for the simplification of (63) to (65). For the channel's power delay profile considered in this work, one obtains $L_s = 6$ corresponding to the dominant channel taps from the channel PDP characteristics, and on substitution the eigenvalues of \mathbf{D}_1 and \mathbf{D}_2 can be seen to be $\lambda(\mathbf{D}_1) = [12, 0]^T$ and $\lambda(\mathbf{D}_2) = [12, 0]^T$ respectively. The Weyl's identity is employed to compute the eigenvalues of \mathbf{D}_1 and \mathbf{D}_2 that are used to obtain a bound on the eigenvalues of \mathbf{D} as $\lambda_1(\mathbf{D}) \leq 24$ and $\lambda_2(\mathbf{D}) \leq 12$ respectively. Using these upper bounds, the product measure Δ_q corresponding to $w = 1$ for the distance matrix \mathbf{D} can be evaluated as $\Delta_q(1) = 288$. For the $w = 2$ product metric, one can begin with $\tilde{\lambda}_1 = \tilde{\lambda}_2 = 4$ for the given value of $\tilde{\Delta}_q(2) = 16$. Following a similar approach as seen above $\Delta_q(2) = 1152$. Thus, for the MO-SC system under consideration, the product measure corresponding to the expurgated FER bound can be finally obtained as $\Delta_q(1, 2) = \{288, 1152\}$.

REFERENCES

- [1] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
- [2] L. Hanzo, T. H. Liew, and B. L. Yeap, *Turbo Coding, Turbo Equalisation and Space-time Coding: For Transmission over Fading Channels*. John Wiley & Sons, Ltd, 2011.
- [3] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-capacity Multifunctional MIMO Systems: Sphere-packing, Iterative Detection and Cooperation*. John Wiley & Sons, Ltd, 2009.
- [4] S. Ö. Ata and I. Altunbaş, "STTC design for vehicular communication systems employing fixed-gain af plnc over cascaded fading channels," *IET Commun.*, vol. 12, no. 11, pp. 1283–1289, Nov. 2018.
- [5] H. Elfadil, M. Maleki, N. Behdad, and H. R. Bahrami, "Trellis-coded space-time shift keying," *IEEE Trans. on Commun.*, vol. 66, no. 12, pp. 5888–5901, Dec. 2018.
- [6] L. Hanzo, M. Munster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-user communications, WLANs and Broadcasting*. John Wiley & Sons, Inc., 2003.
- [7] D. Agarwal, V. Tarokh, A. Naguib, and N. Seshadri, "Space-time coded OFDM for high data rate wireless communication over wideband channels," in *Proc. IEEE Vehicular Technology Conference (VTC)*, May 1998, pp. 2232–2236.
- [8] T.-H. Liew and L. Hanzo, "Space-time trellis and space-time block coding versus adaptive modulation and coding aided OFDM for wideband channels," *IEEE Trans. on Vehicul. Technol.*, vol. 55, no. 1, pp. 173–187, Jan. 2006.
- [9] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. on Broadcast.*, vol. 48, no. 3, pp. 223–229, Sep. 2002.
- [10] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.
- [11] H. Minn and N. A. Dahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.
- [12] T. Y. Al-Naffouri and A. A. Quadeer, "A forward-backward Kalman filter-based STBC MIMO OFDM receiver," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. Article ID 158037, p. 14, Jan. 2008.
- [13] A. Mishra, R. Gayathri, and A. K. Jagannatham, "Random parameter EM-based Kalman filter (REKF) for joint symbol detection and channel estimation in fast fading STTC MIMO systems," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 766–770, Jun. 2014.

- [14] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Evolutionary-algorithm-assisted joint channel estimation and turbo multiuser detection/decoding for OFDM/SDMA," *IEEE Trans. on Vehicul. Technol.*, vol. 63, no. 3, pp. 1204–1222, Mar. 2014.
- [15] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [16] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.
- [17] G. Tauböck and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr. 2008, pp. 2885–2888.
- [18] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
- [19] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Trans. on Commun.*, vol. 60, no. 12, pp. 3705–3716, Dec. 2012.
- [20] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591–3603, Jul. 2014.
- [21] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
- [22] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Aug. 1998.
- [23] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
- [24] A. Mishra, V. Gupta, S. Dwivedi, A. K. Jagannatham, and P. K. Varshney, "Sparse bayesian learning-based target imaging and parameter estimation for monostatic mimo radar systems," *IEEE Access*, vol. 6, pp. 68 545–68 559, Dec. 2018.
- [25] 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 Trans on Sig. Process.*, vol. 67, no. 5, pp. 1251–1266, Mar. 2019.
- [26] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: a sparse Bayesian learning approach," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5369–5382, Oct. 2015.
- [27] A. Mishra, N. S. Yashaswini, and A. K. Jagannatham, "SBL-based joint sparse channel estimation and maximum likelihood symbol detection in OSTBC MIMO-OFDM systems," *IEEE Trans. on Vehicul. Technol.*, vol. 67, no. 5, pp. 4220–4232, May 2018.
- [28] R. Prasad and C. R. Murthy, "Joint approximately group sparse channel estimation and data detection in MIMO-OFDM systems using sparse Bayesian learning," in *Proc. National Conference on Communications (NCC)*, Mar. 2014.
- [29] C. R. Berger, S. Zhou, W. Chen, and P. Willett, "Sparse channel estimation for OFDM: Over-complete dictionaries and super-resolution," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2009, pp. 196–200.
- [30] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for fdd massive mimo," *IEEE Trans. on Commun.*, vol. 64, no. 2, pp. 601–617, Feb. 2016.
- [31] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for fdd massive mimo," *IEEE Trans on Sig. Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.
- [32] H. Bölcskei and A. J. Paulraj, "Space-frequency coded broadband OFDM systems," in *Proc. IEEE Wireless Communication and Networking Conference (WCNC)*, Sep. 2000, pp. 1–6.
- [33] R. S. Blum, Y. Li, J. Winters, and Q. Yan, "Improved space-time coding for MIMO-OFDM wireless communications," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1832–1878, Nov. 2001.
- [34] A. Mishra and A. K. Jagannatham, "SBL-based joint channel estimation and ML sequence detection in STTC MIMO-OFDM systems," in *Proc. IEEE Global Communications Conference (GLOBECOMM)*, Dec. 2016, pp. 1–6.
- [35] Z. Gao, L. Dai, S. Han, I. Chih-Lin, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Feb. 2018.
- [36] Z. Qin, J. Fan, Y. Liu, Y. Gao, and G. Y. Li, "Sparse representation for wireless communications: A compressive sensing approach," *IEEE Sig. Process. Mag.*, vol. 35, no. 3, pp. 40–58, Apr. 2018.
- [37] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.
- [38] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: performance criteria in the presence of channel estimation errors, mobility and multiple paths," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 199–207, Feb. 1999.
- [39] Y. Li and P. Y. Kam, "Space-time trellis codes over rapid Rayleigh fading channels with channel estimation-Part I: Receiver design and performance analysis," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1640–1644, Aug. 2007.
- [40] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.
- [41] M. Byun and B. G. Lee, "New bounds of pairwise error probability for space-time codes in Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1484–1493, Aug. 2007.
- [42] D. Aktas and M. P. Fitz, "Distance spectrum analysis of space-time trellis-coded modulations in quasi-static Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3335–3344, Dec. 2003.
- [43] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation Part I: Greedy pursuit," *Elsevier Signal Processing*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
- [44] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 1371–1382, Jul. 2005.
- [45] X. Lv, G. Bi, and C. Wan, "The group LASSO for stable recovery of block-sparse signal representations," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1371–1382, Apr. 2011.
- [46] H. Hijazi and L. Ros, "Bayesian Cramer-Rao bounds for complex gain parameters estimation of slowly varying Rayleigh channel in OFDM systems," *ELSEVIER Signal Process. FAST Commun.*, vol. 89, pp. 111–115, Jan. 2009.
- [47] "Recommended ITU-R M.1225: Guidelines for evaluation of radio transmission technologies for IMT-2000," Tech. Rep., 1997.
- [48] B. Vucetic and J. Yuan, *Space-time coding*. John Wiley & Sons, 2003.
- [49] J. G. Proakis and M. Salehi, *Digital communications*. McGraw-hill New York, 2001, vol. 4.
- [50] T. Moon and W. Sterling, *Mathematical Methods And Algorithms For Signal Processing*. Prentice Hall, 2000.
- [51] A. Mishra, A. K. Jagannatham, and L. Hanzo, "Technical report I : SBL-based joint sparse channel estimation and ML sequence detection in space-time trellis coded MIMO-OFDM systems," IIT Kanpur, Tech. Rep., 2019. [Online]. Available: http://www.iitk.ac.in/mwn/documents/TR_SBL_MO-SC_2019.pdf
- [52] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [53] J. N. Franklin, *Matrix Theory*. Prentice Hall, 1978.