mmWave Channel Estimation via Compressive Covariance Estimation: Role of Sparsity and Intra-vector Correlation

Dheeraj Prasanna, *Student Member, IEEE* and Chandra R. Murthy, *Senior Member, IEEE*

Abstract—In this work, we address the problem of multiple-input multiple-output mmWave channel estimation in a hybrid analog-digital architecture, by exploiting both the underlying spatial sparsity as well as the spatial correlation in the channel. We accomplish this via compressive covariance estimation, where we estimate the channel covariance matrix from noisy low dimensional projections of the channel obtained in the pilot transmission phase. We use the estimated covariance matrix as a plug-in to the linear minimum mean square estimator to obtain the channel estimate. We present a new Gaussian prior model, inspired by sparse Bayesian learning (SBL), which incorporates parameters to capture the channel correlation in addition to sparsity. Based on this prior, we develop the `Corr-SBL` algorithm, which uses an expectation maximization procedure to learn the parameters of the prior and update the posterior channel estimates. A closed form solution is obtained for the maximization step based on fixed-point iterations. To facilitate practical implementation, an online version of the algorithm is developed which significantly reduces the latency at a marginal loss in performance. The efficacy of the prior model is studied by analyzing the normalized mean squared error in the channel estimate. Our results show that, when compared to a genie-aided estimator and other existing sparse recovery algorithms, exploiting both sparsity and correlation results in significant performance gains, even under imperfect covariance estimates obtained using a limited number of samples.

Index Terms—Millimeter wave, channel estimation, sparsity, correlation, Bayesian learning.

I. INTRODUCTION

Millimeter wave (mmWave) communication has been investigated as a promising technology for the fifth generation (5G) cellular networks [2]–[4]. The large bandwidth available at mmWave frequencies can be utilized to obtain high data rates. However, signals at these frequencies experience high attenuation, leading to significant path loss [4]. To overcome the path loss, multiple antennas are used along with a hybrid analog-digital architecture to keep the hardware cost low [5]. The performance of these systems critically depends on the availability of accurate channel state information (CSI) at both the base station (BS) and the user equipments (UEs). In turn, this requires a large training or pilot overhead for channel estimation. In this work, our goal is to investigate the role of spatial sparsity and intra-vector correlation to obtain reliable

channel estimates with low pilot overhead. We start with a review of literature.

A. mmWave Channel Estimation

The hybrid analog-digital architecture used in mmWave communications precludes the use of traditional channel estimation techniques, as the channel is only observed through the lens of the analog beams used at the radio-frequency (RF) front-end. Analog beam sweeping based procedures have been proposed to sample the channel subspace and estimate the mmWave channel links [6]–[8], but these procedures typically incur large pilot overheads. An alternative approach is to exploit structure in the channel, and estimate it by solving an optimization problem. Measurement campaigns for mmWave channels [4], [9], [10] have revealed structures like sparsity and correlation, which can be incorporated into a statistical model to estimate channel using far fewer pilot transmissions compared to beam scanning based approaches.

Spatial sparsity arises in mmWave channels because the signals arrive at the receiver in a small number of path clusters [4], [9]. Different sparse representations of the mmWave channel have been studied in [11]–[13], and sparse recovery algorithms such as orthogonal matching pursuit (OMP) have been applied to estimate the channel using a reduced number of pilots. In [14], [15], different training strategies for sparse channel estimation have been discussed. In [16], a parameter-perturbation framework combined with a low-complexity simultaneous OMP algorithm is presented mmWave channel estimation, accounting for off-grid effects. These traditional sparse recovery algorithms (e.g., OMP) yield accurate point estimates when the measurement matrix satisfies strong requirements such as restricted isometric property (RIP), which are rarely met in highly measurement-constrained scenarios. Further, they do not reveal the posterior distribution which can add flexibility in dealing with additional structure.

In addition to sparsity, structures such as spatial correlation have been observed due to mutual coupling at the antennas [10], [17]. In [10], spatial fading models were provided to fit measured spatial correlation using a parameterized exponential model. In a general massive MIMO scenario, [18] developed a Toeplitz model for the covariance matrix under the assumption that the angle of arrivals (AoAs) and path gains are i.i.d. random variables. They also presented an algorithm to exploit the correlation for channel estimation.

To the best of our knowledge, very few of the existing studies capture sparsity (and the resulting spatial correlation)

as well as correlation among the nonzero entries of the sparse vector in the same statistical model. Such correlations can arise, for example, due to signal propagation and scattering in clusters (called block sparsity). Channel estimation techniques that exploit block sparsity have been presented in [19], [20]. While these studies consider a common support in the sparse representation of the channel from the user to the different antennas at the BS, correlation among the nonzero entries are not modeled or exploited. The block sparsity setup can be modified to include correlation among path gains in each block [21], but it does not generalize to correlation across blocks. Rain attenuation leads to multiple scattering and correlation at mmWave frequencies [22], [23]. Also, atmospheric turbulence and correlated shadowing [24] lead to correlation among paths that are further apart, which cannot be considered under the ambit of block sparsity.

From the above discussion, correlation among the nonzero entries of the sparse representation of the channel is unavoidable in general, which makes it necessary to incorporate correlation along with sparsity into statistical models to obtain accurate channel estimates. To this end, we start with a general statistical model for the channel, where all nonzero entries of the sparse representation are assumed to be correlated with each other, similar to the model considered by Park et al. [25]. The focus of Park et al. was on precoder design, and a modification of OMP was developed for recovering the covariance of the sparse channel. The goal of this paper is to understand the role of sparsity and correlation in facilitating accurate channel estimation, especially when the underlying covariance structure is unknown and has to be estimated from the received data itself. In the process, we also develop new algorithms for channel estimation, based on Bayesian inference, that exploit both correlation and sparsity.

Bayesian inference algorithms such as sparse Bayesian learning (SBL) [26] are well suited to exploit correlation, as they estimate the covariance as a key step within the algorithm. Their primary goal is to infer the best-fitting distribution from a parameterized class of distributions. Point estimates can then be obtained from the posterior distribution. A message passing algorithm for mmWave channel estimation was developed in [27], and SBL based algorithms are presented in [20], [28] for recovery of spatially uncorrelated sparse channels. The algorithm developed in this paper is also related to algorithms like TMSBL [29], BSBL [21], PCSBL [30]. However, to the best of our knowledge, there is no existing algorithm based on Bayesian inference, which considers possible correlation among all the nonzero entries of a sparse vector.

B. Contributions

In this work, our main contributions are as follows:

1) We develop a novel Bayesian sparse recovery algorithm called Corr-SBL for recovery of sparse vectors with intra-vector correlation, in Sec. IV. We formulate a zero mean hierarchical correlated complex Gaussian prior with covariance matrix that can incorporate known correlation structure while at the same time induce sparsity. Our algorithm performs Bayesian inference based on evidence maximization. A closed-form solution to update the hyper-parameters is obtained as a fixed point iteration. We also present a pragmatic approach for learning the correlation coefficient in the unknown correlation case.

2) We investigate the utility of exploiting both spatial sparsity and correlation in the multi-paths of a mmWave channel, for uplink channel estimation in a multi-user MIMO setup with the hybrid architecture. We present the application of Corr-SBL for this problem.

3) We extend Corr-SBL to jointly estimate the channels over multiple coherence blocks in Sec. IV-D. To reduce the overall latency in channel estimation, in Sec. IV-E, we develop an online version of the algorithm.

4) We derive an alternative representation of the output of the algorithm as a plug-in LMMSE estimator in Sec. V-A. In Sec. V-B, we derive an expression for the normalized mean squared error (NMSE) in channel estimation, and discuss the efficacy of Corr-SBL for channel estimation.

5) We present a hybrid scheme for combining the data signals using the channel estimates in Sec. V-C, and derive a lower bound on the per-user spectral efficiency.

We elucidate the utility of the Corr-SBL prior using the analytical expression for the NMSE, and present empirical comparisons against optimal genie-aided estimators in highly measurement-constrained scenarios. The Monte Carlo simulation results in Sec. VI illustrate the advantage of exploiting correlation as well as sparsity even under imperfect correlation information, depending on the correlation level and number of independent channel instantiations that are available to estimate the covariance. Finally, our framework for recovering sparse vectors with correlated nonzero entries could be of independent interest, as it can potentially be very useful in other applications besides mmWave channel estimation.

Notation: Bold lowercase letters denote vectors, bold uppercase letters denote matrices and script styled alphabets represent sets. The operator $|\cdot|$ when applied on a scalar denotes the absolute value, and on a set denotes the cardinality. The operator $\mathrm{diag}()$ when applied to a vector generates a diagonal matrix consisting entries of a vector, and when applied to a matrix returns the diagonal entries as a vector. The trace of a matrix is denoted by $\mathrm{Tr}()$. $\mathrm{Re}\{x\}$ denotes the real part of a complex number x. $\mathbb{E}_x$ denotes expectation with respect to the probability distribution $p(x)$. $\mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the complex Gaussian distribution on $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The notation $\mathbf{A} \odot \mathbf{B}$ denotes the Hadamard product of $\mathbf{A}$ and $\mathbf{B}$. $[n]$ denotes the set of integers from 1 to n and $a : b$ denotes the set of integers from a to b. Given a set of indices $\mathcal{S}$, $\mathbf{x}_{\mathcal{S}}$ denotes the subvector of $\mathbf{x}$ obtained by retaining only the values corresponding to the indices in $\mathcal{S}$. $[\mathbf{A}]_{(\mathcal{S}_r, \mathcal{S}_c)}$ denotes a submatrix of $\mathbf{A}$ containing its rows and columns indexed by $\mathcal{S}_r$ and $\mathcal{S}_c$, respectively.

II. SYSTEM MODEL

In this section, we present the wireless system setup and the channel model for single cell multi-user mmWave MIMO uplink communication with a hybrid analog-digital architecture.
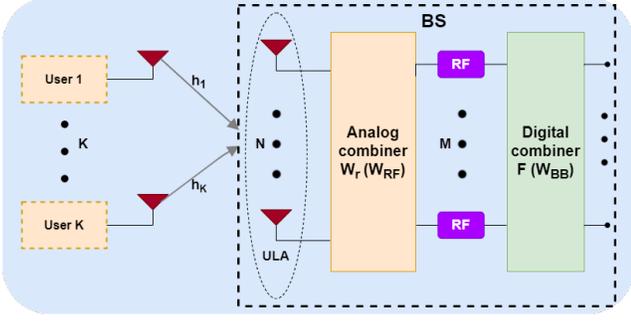
Fig. 1. Wireless uplink system with a single N antenna BS serving K single antenna users.

A. System and Channel Model

The mmWave wireless system consists of a single MIMO BS equipped with a uniform linear array (ULA) consisting of N equally spaced antennas, which serves K spatially distributed single antenna users. The BS employs a hybrid MIMO architecture with M (with $K \leq M \ll N$) RF chains.

The signal at the antenna is converted to the digital domain by a network of phase shifters in a fully connected structure, represented by an analog combining matrix $\mathbf{W}_r \in \mathbb{C}^{M \times N}$, with the phase-only control of the phase shifters satisfied by the constant modulus constraint $|[\mathbf{W}_r]_{(m,n)}| = \frac{1}{\sqrt{N}}$ $\forall m \in [M], n \in [N]$. The digital baseband combiner $\mathbf{F} \in \mathbb{C}^{K \times M}$ then converts the signal into K data streams.

The uplink mmWave channel from each user to the BS is assumed to have a block flat-fading structure, with coherence time T_c (in s) and coherence bandwidth B_c (in Hz), i.e., the channel is constant within time-frequency coherence blocks of $\tau_c = B_c T_c$ channel uses. Further, the channel from k^{th} user to the BS is comprised of $L_k \ll N$ multi-paths. The lth path at the rth coherence block is characterized using a complex baseband path gain $\bar{g}_{r,k,l}$ and the corresponding AoA $\psi_{r,k,l}$. The array response vector $\bar{\mathbf{a}}(\psi_{r,k,l}) \in \mathbb{C}^N$ for the ULA is

$$\bar{\mathbf{a}}(\psi_{r,k,l}) = \frac{1}{\sqrt{N}} \left[1, e^{-j2\pi v_{r,k,l}}, \ldots, e^{-j2\pi(N-1)v_{r,k,l}} \right]^T. \quad (1)$$

where $v_{r,k,l} = \frac{d}{\lambda} \cos(\psi_{r,k,l})$ is the directional cosine corresponding to the AoA $\psi_{r,k,l}$. The uplink mmWave channel can then be represented as [25]

$$\mathbf{h}_{r,k} = \sum_{l=1}^{L_k} \bar{g}_{r,k,l} \bar{\mathbf{a}}(\psi_{r,k,l}) = \bar{\mathbf{A}} \bar{\mathbf{g}}_{r,k} \in \mathbb{C}^N, \quad (2)$$

where columns of $\bar{\mathbf{A}} \in \mathbb{C}^{N \times L_k}$ are the array response vectors given by (1). If we grid the range of possible directional cosine values (-1 to 1) using $D \gg L_k$ points, (2) can be approximated as

$$\mathbf{h}_{r,k} = \mathbf{A} \mathbf{g}_{r,k}. \quad (3)$$

The matrix $\mathbf{A} \in \mathbb{C}^{N \times D}$ consists of the ULA response vectors for the AoAs corresponding to the grid points. The index of columns of $\bar{\mathbf{A}}$ in $\mathbf{A}$, denoted by the set $\mathcal{S}_k$ (with $|\mathcal{S}_k| = L_k$), are the grid points corresponding to AoAs given in (2), and the corresponding nonzero entries of $\mathbf{g}_{r,k} \in \mathbb{C}^D$ are the path gains $\bar{g}_{r,k,l}$. Since $D \gg L_k$, $\mathbf{g}_{r,k}$ is a sparse vector.

We assume that the non-zero entries in the complex baseband path gain vector $\mathbf{g}_{r,k}$ have zero mean and are correlated with each other. The $(i,j)^{\text{th}}$ entry of the covariance matrix $\mathbf{R}_{\mathbf{g}_{r,k}} = \mathbb{E}\left[\mathbf{g}_{r,k}\mathbf{g}_{r,k}^H\right] \in \mathbb{C}^{D \times D}$ can be obtained from the Pearson product-moment correlation coefficient definition [31] as $\left[\mathbf{R}_{\mathbf{g}_{r,k}}\right]_{(i,j)} = \rho_{ij} \sqrt{\gamma_i^*} \sqrt{\gamma_j^*}$ where γ_i^* denotes the variance of ith entry of $\mathbf{g}_{r,k}$ and ρ_{ij} is the correlation coefficient between the two entries, which is a function of the pair of indices (i,j) and is governed by a known correlation model (e.g., uniform correlation model, exponential model, Toeplitz model etc.). A matrix $\mathbf{U} \in \mathbb{C}^{D \times D}$ with ρ_{ij} as its $(i,j)^{\text{th}}$ entry is called the correlation matrix. With this notation, we have $\mathbf{R}_{\mathbf{g}_{r,k}} = (\mathbf{\Gamma}^*)^{1/2} \mathbf{U} (\mathbf{\Gamma}^*)^{1/2}$ where $\mathbf{\Gamma}^* = \text{diag}(\gamma_1^*, \gamma_2^*, \ldots, \gamma_D^*)$.

Note that, for a grid point j which does not correspond to any of the L_k paths, the variance of the jth entry, γ_j^*, is zero. This results in the covariance structure similar to [25], where $\mathbf{R}_{\mathbf{g}_{r,k}}$ is a $D \times D$ positive semi-definite (PSD) matrix which contains a nonzero $L_k \times L_k$ positive definite (PD) principal submatrix corresponding to the index set $\mathcal{S}_k$, with its other entries equal to 0. Using (3), the channel covariance matrix $\mathbf{R}_{\mathbf{h}_{r,k}} = \mathbb{E}\left[\mathbf{h}_{r,k}\mathbf{h}_{r,k}^H\right] = \mathbf{A}\mathbf{R}_{\mathbf{g}_{r,k}}\mathbf{A}^H \in \mathbb{C}^{N \times N}$. The channel statistics vary slowly compared to the channel instantiations [32], i.e., the channel statistics remain constant over a time interval $T_s > T_c$. Hence, the channel covariance and AoAs remain constant over $\tau_s = \frac{T_s}{T_c}$ coherence blocks. In the sequel, since we focus on estimating the covariance matrix within τ_s coherence blocks where the covariance matrix remains constant, we drop the subscript r in the covariance matrices $\mathbf{R}_{\mathbf{g}_{r,k}}$ and $\mathbf{R}_{\mathbf{h}_{r,k}}$.

B. Pilot Transmission

At the start of each coherence block, the kth user transmits a unique orthonormal pilot $\mathbf{p}_k \in \mathbb{C}^{\tau_p}$, $k = 1, 2, \ldots, K$, where τ_p is the pilot length satisfying $\tau_p = K < \tau_c$. This pilot signal is used by the BS to estimate the uplink channels from all the users. The pilot signal received at the BS is processed using an analog combining matrix $\mathbf{W}_r \in \mathbb{C}^{M \times N}$, resulting in the following signal at the digital front-end:

$$\mathbf{Y}_r^p = \mathbf{W}_r \left(\sum_{k=1}^K \mathbf{h}_{r,k} \mathbf{p}_k^H + \mathbf{N} \right) \in \mathbb{C}^{M \times \tau_p}, \quad (4)$$

where the entries of additive noise $\mathbf{N}$ are independent and identically distributed (i.i.d.) Gaussian with zero mean and variance σ_n^2.[1] Since the pilot signals are orthonormal, by post-multiplying $\mathbf{Y}_r^p$ with $\mathbf{p}_k$, we obtain the pilot signal for estimating the kth user's channel as

$$\mathbf{y}_{r,k} = \mathbf{Y}_r^p \mathbf{p}_k = \mathbf{W}_r \mathbf{h}_{r,k} + \mathbf{W}_r \mathbf{N} \mathbf{p}_k \in \mathbb{C}^M. \quad (5)$$

The covariance of the effective noise $\mathbf{n}_r = \mathbf{W}_r \mathbf{N} \mathbf{p}_k$ is $\sigma_n^2 \mathbf{W}_r \mathbf{W}_r^H$. In Sec. III-B, we present a choice of $\mathbf{W}_r$ such that $\mathbf{W}_r \mathbf{W}_r^H$ approaches $\mathbf{I}_M$ asymptotically ($N \to \infty$) [33]. Hence, $\mathbf{n}_r$ is assumed to comprise of i.i.d. $\mathcal{CN}\left(0, \sigma_n^2\right)$ entries.

---

[1]Without loss of generality, we consider the pilot signal to be of unit power, and include the effect of pilot transmission power in σ_n^2. Note that (4) assumes the users are time and frequency synchronized with the BS. In practice, this synchronization can be achieved using the primary and secondary synchronization signals transmitted by the BS.

From (5), it can be observed that the measurement $\mathbf{y}_{r,k}$ does not involve interference from other users. Also, since the pilot sequences are orthogonal, the noise in the post-processed received training signals are independent across the users, and we can perform channel estimation independently for each user. Hence, in the sequel, we drop the subscript k and consider the channel estimation for a single user. Estimating $\mathbf{h}_r$ from $\mathbf{y}_r$ in (5) constitutes the channel estimation problem. Specifically, using (3), the channel estimation problem (5) can be posed as a sparse recovery algorithm as:

$$\mathbf{y}_r = \boldsymbol{\Phi}_r \mathbf{g}_r + \mathbf{n}_r, \tag{6}$$

where $\boldsymbol{\Phi}_r = \mathbf{W}_r \mathbf{A} \in \mathbb{C}^{M \times D}$. Hence, our goal boils down to exploiting sparsity and correlation among the nonzero entries of the sparse vector to obtain the channel estimate.

Note that, in the above system model, we have considered a frequency-flat channel model and single antenna users. Extensions to frequency-selective channel models and multi-antenna users are discussed in Appendix B.

III. CHANNEL ESTIMATION: PRELIMINARIES

A. Channel Estimation Schemes

In this subsection, we present an overview of two existing classes of channel estimation techniques, which form the baseline for the comparisons in this paper. The details of these estimators are presented in Appendix A.

The first class of estimators are linear estimators, where the estimate of the channel in each coherence block can be represented as $\hat{\mathbf{h}}_r = \mathbf{M}_r \mathbf{y}_r$. This includes linear minimum mean squared error (LMMSE) and least squared (LS) estimation, among others. When the covariance matrix is not known, a plug-in LMMSE estimation can be formed using an estimate $\hat{\mathbf{R}}_\mathbf{h}$ for the covariance matrix $\mathbf{R}_\mathbf{h}$. When the sample covariance is used to estimate $\hat{\mathbf{R}}_\mathbf{h}$, the estimator exploits correlation structure, but ignores sparsity in the channel.

The second class of estimators utilize sparse signal recovery algorithms such as OMP [11] and SBL [26] to recover sparse representation of the channel $\mathbf{g}_r$ from the measurements using (6). We note that these algorithms exploit the sparsity structure, but neglect channel correlation.

Channel estimators based on covariance estimation require knowledge of the covariance matrix. Estimating the covariance entails computing the sample-averaged covariance using multiple channel instantiations. Since the channel statistics are constant over τ_s coherence blocks, pilots from multiple coherence blocks can be used to estimate the covariance matrix without additional pilot overhead. The support is also constant within a coherence block, and sparse recovery algorithms developed in the multiple measurement vector (MMV) paradigm such as SOMP [25] and MSBL [29] can be used to obtain better performance. In this work, compute the sample covariance by averaging over $T \leq \tau_s$ coherence blocks, with T being chosen to trade-off between performance and complexity.

B. Analog Combiner Matrix Design

The analog combining matrix $\mathbf{W}_r \in \mathbb{C}^{M \times N}$ plays an key role in the performance of channel estimation. In this subsection, we discuss two schemes used in the paper for the choice of combining matrices across coherence blocks.

The analog combining matrix can be represented as $\mathbf{W}_r = \frac{1}{\sqrt{N}} e^{i\boldsymbol{\Theta}_r}$ to satisfy the constant modulus constraint, where the entries of $\boldsymbol{\Theta}_r$ represent the phase of each entry of $\mathbf{W}_r$. The first of the two schemes uses the same combining matrix across coherence blocks and is referred to as "shared $\mathbf{W}_r$ scheme". In this scheme, $\boldsymbol{\Theta}_r = \boldsymbol{\Theta} \, \forall r$, with the entries of $\boldsymbol{\Theta}$ chosen independently from a uniform distribution in $[0, 2\pi]$.

In [34, Proposition 2], the authors proved that when a shared compression matrix is applied on every sample of a signal, all possible estimators are asymptotically biased. Instead, if independent compression matrices are applied on different samples, unbiased estimators can be designed. Using this for the channel estimation problem, better estimation performance can be obtained when a different combining matrix is chosen for each coherence block. Thus, in our second scheme, $\boldsymbol{\Theta}_r$ is chosen independently across r, with the distribution for each $\boldsymbol{\Theta}_r$ being the same as in the shared $\mathbf{W}_r$ scheme. This scheme is referred to as "i.i.d. $\mathbf{W}_r$ scheme". We note that the memory and computational complexity of the i.i.d. $\mathbf{W}_r$ scheme is higher than that of the shared $\mathbf{W}_r$ scheme.

In the sequel, we borrow terminology from the compressed sensing literature to refer to estimating the channel within a single coherence block as single measurement vector (SMV) channel estimation and estimation over multiple blocks by exploiting the joint sparsity structure as multiple measurement vector (MMV) channel estimation. In the next section, we present a novel Bayesian learning algorithm which exploits both sparsity and correlation.

IV. CORRELATED SPARSE BAYESIAN LEARNING

We recall that our goal for SMV channel estimation is to recover $\mathbf{g}_r \in \mathbb{C}^D$ from the measurements $\mathbf{y}_r = \boldsymbol{\Phi}_r \mathbf{g}_r + \mathbf{n}_r \in \mathbb{C}^M$ which, given $\mathbf{g}_r$, are distributed as $p(\mathbf{y}_r | \mathbf{g}_r) = \mathcal{CN}(\mathbf{y}_r; \boldsymbol{\Phi}_r \mathbf{g}_r, \sigma_n^2 \mathbf{I}_M)$. Following the Bayesian learning philosophy, we impose a parameterized prior on $\mathbf{g}_r$, and develop an algorithm to find its MAP estimate. The choice of the prior is crucial to the success of the algorithm. It must promote the structure – sparsity and correlation – in $\mathbf{g}_r$, and must facilitate computation of the posterior. The Bayesian inference procedure learns the prior parameters such that the resulting distribution best fits the observed data according to the underlying channel and measurement model. We present a choice for such a prior in the next subsection. We discuss the utility of the prior in obtaining low NMSE for channel estimation in Sec. V-B.

A. Prior Model

In SBL, a parameterized Gaussian prior with the covariance matrix $\boldsymbol{\Sigma}_\theta \in \mathbb{C}^{D \times D}$ is modeled as $\boldsymbol{\Sigma}_\theta = \text{diag}(\boldsymbol{\gamma})$, where the unknown hyperparameter θ consists of the entries of the vector $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_D]^T \in \mathbb{R}_+^D$ which denote the variances of the entries of $\mathbf{g}_r$. This choice of prior model is known to induce sparsity in the final channel estimate. The diagonal nature of the covariance matrix implicitly assumes

that the entries of $\mathbf{g}_r$ are uncorrelated. Hence, this prior cannot accommodate any knowledge of the correlation structure in $\mathbf{g}_r$.

Similar to SBL, in this paper, we consider a parameterized Gaussian prior for $\mathbf{g}_r$ with hyperparameter set θ as

$$p(\mathbf{g}_r; \theta) = \mathcal{CN}(\mathbf{g}_r; \mathbf{0}, \boldsymbol{\Sigma}_\theta). \tag{7}$$

Recall that the covariance matrix $\mathbf{R_g}$ is a $D \times D$ positive semidefinite matrix with an $L \times L$ positive definite principal submatrix corresponding to L active paths, and other entries equal to zero [25]. In order to incorporate this structure, we model $\boldsymbol{\Sigma}_\theta$ using two hyperparameter sets: $\{\boldsymbol{\gamma} \in \mathbb{R}_+^D\}$ denoting the vector containing unknown variances of the entries of $\mathbf{g}_r$, and $\mathbf{U} \in \mathbb{C}^{D \times D}$ denoting a positive definite matrix with correlation coefficients between the entries of $\mathbf{g}_r$. For ease of exposition, we deal with the hyperparameter $\mathbf{U}$ in two cases: (i) when $\mathbf{U}$ is known; and (ii) when $\mathbf{U}$ is an unknown matrix parameterized based on the underlying channel model.[2] Using the definition for Pearson product-moment correlation coefficient [31], $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}, \mathbf{U}} \triangleq \boldsymbol{\Sigma}_\theta$ is modeled as

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}, \mathbf{U}} = \boldsymbol{\Gamma}^{\frac{1}{2}} \mathbf{U} \boldsymbol{\Gamma}^{\frac{1}{2}}, \tag{8}$$

where $\boldsymbol{\Gamma} \triangleq \mathrm{diag}(\boldsymbol{\gamma})$. To see how the above model incorporates the structure in $\mathbf{R_g}$, consider the case where i^{th} element of $\boldsymbol{\gamma}$ is zero. Since it denotes the variance of a zero mean random variable, i^{th} entry of $\mathbf{g}_r$ is zero, and the ith row and column of $\boldsymbol{\Sigma}_\theta$ are zero vectors. If *all but L entries* are zero, then $\boldsymbol{\Sigma}_\theta$ has a structure similar to $\mathbf{R_g}$. Consequently, the MAP estimate for $\mathbf{g}_r$ is also an L-sparse estimate, with the same support as that of $\boldsymbol{\gamma}$. Also, the prior model is a generalization to SBL prior, as it reduces to the SBL prior $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \boldsymbol{\Gamma}$ when $\mathbf{U} = \mathbf{I}_D$.

In the sequel, it will be more convenient to work with the precision matrix $\boldsymbol{\Omega}_{\mathbf{c}, \mathbf{U}} \triangleq \boldsymbol{\Sigma}_{\boldsymbol{\gamma}, \mathbf{U}}^{-1}$. Note that $\boldsymbol{\Omega}_{\mathbf{c}, \mathbf{U}} = \mathbf{C} \mathbf{U}^{-1} \mathbf{C}$, where $\mathbf{C} \triangleq \mathrm{diag}(\mathbf{c})$ and $\mathbf{c} \in \mathbb{R}_+^D$ has $1/\sqrt{\gamma_i}$ as its ith entry. In the following section, we present an iterative algorithm to estimate the value of c_i. When an index i is not in the support, the value of c_i goes to infinity as the iterations proceed. To counter the numerical instability, in any iteration, if the value of c_i exceeds a threshold ε, we remove the index i and the corresponding column of $\boldsymbol{\Phi}_r$ from the support. This also speeds up the algorithm. In the Algorithms 1 and 2 presented later in this section, the thresholding operation described above is performed at the start of each EM iteration (while loop), and indexes i for which $c_i > \varepsilon$ are removed from the support.

B. Case I: Algorithm Development for Known $\mathbf{U}$

We now proceed with developing a Bayesian algorithm for the proposed choice of prior when $\mathbf{U}$ is known, and drop the subscript $\mathbf{U}$ in $\boldsymbol{\Omega}_{\mathbf{c}, \mathbf{U}}$. This involves two steps: obtaining the optimal value for the hyperparameters $\mathbf{c}$, and computing the posterior distribution.

We use type-II ML estimation to obtain the optimal value of $\mathbf{c}$. This is based on the evidence maximization framework, where the cost function is the marginal likelihood of $\mathbf{y}$. By marginalizing the joint density $p(\mathbf{g}_r, \mathbf{y}_r; \mathbf{c}, \sigma_n^2)$ with respect to

---

$\mathbf{g}_r$, it is straightforward to show that the marginal likelihood is given by $p(\mathbf{y}_r; \mathbf{c}, \sigma_n^2) = \mathcal{CN}(\mathbf{y}_r; \mathbf{0}, \boldsymbol{\Omega}_{\mathbf{y}}^{-1})$, where $\boldsymbol{\Omega}_{\mathbf{y}} \triangleq \left[\sigma_n^2 \mathbf{I}_M + \boldsymbol{\Phi}_r \boldsymbol{\Omega}_{\mathbf{c}}^{-1} \boldsymbol{\Phi}_r^H\right]^{-1}$ denotes the precision matrix of $\mathbf{y}_r$. Thus, the cost function that needs to be maximized for finding $\mathbf{c}$ is obtained from the log likelihood $\log(p(\mathbf{y}_r; \mathbf{c}, \sigma_n^2))$ as

$$L(\mathbf{c}) \triangleq \mathrm{logdet}(\boldsymbol{\Omega}_{\mathbf{y}}) - \mathbf{y}_r^H \boldsymbol{\Omega}_{\mathbf{y}} \mathbf{y}_r. \tag{9}$$

The optimal $\mathbf{c}$ is then used to compute the posterior distribution and the channel estimate using the following Lemma.

Lemma 1. *Let the prior distribution on $\mathbf{g}_r$ be modeled as $p(\mathbf{g}_r; \mathbf{c}) = \mathcal{CN}(\mathbf{g}_r; \mathbf{0}, \boldsymbol{\Omega}_{\mathbf{c}}^{-1})$. Then, the posterior distribution of $\mathbf{g}_r$ given the observation $\mathbf{y}_r$ and hyperparameter $\mathbf{c}$, is $p(\mathbf{g}_r | \mathbf{y}_r; \mathbf{c}) = \mathcal{CN}(\mathbf{g}_r; \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}, \boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}})$, where*

$$\boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}} = \frac{1}{\sigma_n^2} \boldsymbol{\Phi}_r^H \boldsymbol{\Phi}_r + \boldsymbol{\Omega}_{\mathbf{c}}; \quad \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} = \frac{1}{\sigma_n^2} \boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} \boldsymbol{\Phi}_r^H \mathbf{y}. \tag{10}$$

Proof. See Appendix C. ∎

Since the posterior distribution of $\mathbf{g}_r$ is Gaussian, its mode (i.e., the MAP estimate) is the same as its mean. Hence, the posterior mean $\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}$ computed using the optimal value of the hyperparameters $\mathbf{c}$ is our channel estimate.

The problem of maximizing the cost function (9) is non-convex and does not admit a closed form solution. Hence, we use the expectation-maximization (EM) procedure to maximize (9) by treating $\mathbf{g}_r$ as a hidden (latent) variable. The EM procedure involves iterating between an expectation step (E-step) and a maximization step (M-step) [35, Section 9.3].

1) Expectation step: This step involves computing the expected value of the complete-data log likelihood with respect to the posterior distribution for $\mathbf{g}_r$ computed at the hyperparameter value $\mathbf{c}_{\text{old}}$ obtained from previous iteration of the EM algorithm. The expected value is denoted by the so-called Q function, which is defined as follows:

$$Q(\mathbf{c}, \mathbf{c}_{\text{old}}) \triangleq \mathbb{E}_{\mathbf{g}_r | \mathbf{y}_r; \mathbf{c}_{\text{old}}, \sigma_n^2} \left[\log(p(\mathbf{g}_r, \mathbf{y}_r; \mathbf{c}, \sigma_n^2))\right]. \tag{11}$$

The posterior distribution $p(\mathbf{g}_r | \mathbf{y}_r; \mathbf{c}_{\text{old}}, \sigma_n^2)$ is computed using Lemma 1. The Q function is given by the following theorem.

Theorem 1. *The expected value of complete-data log likelihood evaluated using the hyperparameter value $\mathbf{c}_{old}$ corresponding to the cost function $L(\mathbf{c}) = \mathrm{logdet}(\boldsymbol{\Omega}_{\mathbf{y}}) - \mathbf{y}_r^H \boldsymbol{\Omega}_{\mathbf{y}} \mathbf{y}_r$ is given by*

$$Q(\mathbf{c}, \mathbf{c}_{old}) = k' + (\mathrm{logdet}(\boldsymbol{\Omega}_{\mathbf{c}})) - Tr\left[\boldsymbol{\Omega}_{\mathbf{c}} \hat{\mathbf{R}}_{\mathbf{g}}\right], \tag{12}$$

where k' is a constant independent of $\mathbf{c}$, and $\hat{\mathbf{R}}_{\mathbf{g}} \triangleq \left[\boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} + \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}^H\right]$.

Proof. See Appendix D. ∎

From Theorem 1, we observe that the Q function can be characterized by the matrix $\hat{\mathbf{R}}_{\mathbf{g}}$. Thus, the E-step involves computing $\hat{\mathbf{R}}_{\mathbf{g}}$ given by the theorem, which in turn requires the computation of the covariance matrix $\boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1}$ and the mean $\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}$ of the posterior distribution given by (10). Note that (10) involves inverting the $D \times D$ matrix $\boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}$. The complexity can

---

[2]An algorithm is derived for the first case in Sec. IV-B, while a pragmatic procedure to estimate the parameters of $\mathbf{U}$ is presented in Sec. IV-C.

be reduced from $\mathcal{O}(D^3)$ to $\mathcal{O}(M^3)$ by using the Woodbury matrix identity, which speeds up the algorithm.

2) Maximization step: In this step, the hyperparameter $\mathbf{c}$ is updated by maximizing the Q function. The first order optimality condition for the stationary points of $Q(\mathbf{c}, \mathbf{c}_{old})$ is given by the following theorem.

Theorem 2. *The first order optimality condition for the optimization problem $\mathbf{c}^* = \arg\max_{\mathbf{c}} Q(\mathbf{c}, \mathbf{c}_{old})$ is given by*

$$\frac{1}{c_i} = \mathrm{Re} \left\{ \sum_{k=1}^{D} c_k (\mathbf{U}^{-1})_{ik} \left[\hat{\mathbf{R}}_{\mathbf{g}} \right]_{(k,i)} \right\}, \ i \in [D], \qquad (13)$$

Proof. See Appendix E. ∎

In the conventional sparse Bayesian learning algorithm, the optimality condition above decouples into separate equations in each hyperparameter c_i, and the update for c_i can be obtained independent of c_j, $j \neq i$. However, for the prior considered here, the optimality condition in (13) is a coupled quadratic equation, which cannot be solved in closed form. Gradient based methods can be used to search for the solution, but the computational complexity involved is large. Instead, we draw from the generalized EM theory [36]: any hyperparameter update rule which ensures that $Q(\mathbf{c}, \mathbf{c}_{old})$ is non-decreasing in each EM iteration will ensure convergence of the EM iterations to a local maximum or saddle point of $L(\mathbf{c})$.

To this end, we consider a vector representation for (13) as $\mathbf{c} = (\mathrm{Re}\{\mathbf{K}\})^{-1} \frac{1}{\mathbf{c}}$, where $\mathbf{K} \triangleq \mathbf{U}^{-1} \odot \hat{\mathbf{R}}_{\mathbf{g}}^T \in \mathbb{C}^{D \times D}$ and $\frac{1}{\mathbf{c}}$ denotes the element-wise inverse. A single iteration of this fixed point equation results in a non-decreasing cost function value, as asserted by the following proposition.

Proposition 1. *Consider the update for the hyperparameter given by*

$$\mathbf{c}_{new} = (\mathrm{Re}\{\mathbf{K}\})^{-1} \frac{1}{\mathbf{c}_{old}}. \qquad (14)$$

This satisfies the condition $\sum_i \left(\frac{dQ}{dc_i} \left((c_{new})_i - (c_{old})_i \right) \right) \geq 0$. Consequently, by the generalized EM theory, the cost function in (12) does not decrease after the update. In turn, this guarantees the convergence of the overall algorithm.

Proof. See Appendix F. ∎

In the preceding derivation, it was assumed that the noise variance σ_n^2 was known. In some applications, it is desirable for the algorithm to learn the noise variance also. This can be incorporated by learning a hyperparameter $\lambda \triangleq \sigma_n^2$. The E-step of the algorithm remains unchanged, while the M-step decouples into independent updates for $\mathbf{c}$ (given by (14)) and λ. The update for λ is given by

$$\lambda = \frac{1}{M} \left[\mathbf{y}_r^H \mathbf{y}_r - \mathbf{y}_r^H \boldsymbol{\Phi}_r \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} - \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}^H \boldsymbol{\Phi}_r^H \mathbf{y}_r + Tr\left[\boldsymbol{\Phi}_r^H \boldsymbol{\Phi}_r \boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} \right] \right].$$

C. Case II: A Pragmatic Approach for Learning the Parameters of the Correlation Matrix

The Bayesian inference discussed above assumed that correlation matrix $\mathbf{U}$ is known. One approach for estimating $\mathbf{U}$ could be to consider it as an additional hidden parameter in the EM algorithm, and use the optimality condition to obtain an update. However, due to the matrix derivatives involved, deriving a closed form update for $\mathbf{U}$ not straightforward. Instead, we present a pragmatic approach for learning the correlation in this subsection.

In the previous section, we saw that the mean $\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}$ of the posterior distribution is the MAP estimate for $\mathbf{g}_r$ upon convergence of the algorithm. Similarly, the matrix $\hat{\mathbf{R}}_{\mathbf{g}} = \boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} + \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}^H$ can be interpreted as an estimate for the covariance matrix $\mathbf{R}_{\mathbf{g}}$. In fact, SBL uses the diagonal entries of this estimate to update the variance hyperparameters in each iteration. In the correlated case, a similar update for the inverse variance entries $\mathbf{c}$ does not satisfy the first order optimality condition (13), and hence is not a viable choice for an update rule for $\mathbf{c}$ in the EM algorithm. Instead, an update for $\mathbf{U}$ can be obtained by using (8) to project $\hat{\mathbf{R}}_{\mathbf{g}}$ onto the space of correlation matrices. To this end, the estimate for $\boldsymbol{\Gamma}$ in (8), denoted by $\hat{\boldsymbol{\Gamma}}$, is obtained by considering the diagonal entries of $\hat{\mathbf{R}}_{\mathbf{g}}$. An estimate for the correlation matrix, denoted by $\hat{\mathbf{U}}$, is then obtained by projecting onto space of correlation matrices with unit diagonal entries, given by

$$\hat{\mathbf{U}} = \hat{\boldsymbol{\Gamma}}^{-1/2} \hat{\mathbf{R}}_{\mathbf{g}} \hat{\boldsymbol{\Gamma}}^{-1/2}. \qquad (15)$$

In case a parameterized model for $\mathbf{U}$ is available, e.g., if $\mathbf{U}$ is determined by a scalar parameter ρ, the above estimate can be used as a sufficient statistic to estimate ρ. In particular, for the uniform correlation model, $\rho \in [0, 1)$ and can be obtained by averaging the off-diagonal entries of $\hat{\mathbf{U}}$. In Sec. VI, we empirically show that the performance of this update is close to the original algorithm that has knowledge of $\mathbf{U}$.

D. Multiple Measurement Vector Setup

The performance of correlation-aware algorithms depends on accuracy of the estimates for the channel covariance. Since the channel covariance is constant for T coherence blocks, multiple measurements can be used to estimate the channel statistics, and improve the performance. The extension of the algorithm derived above to the MMV case can be summarized as follows. The E-step updates are averaged over all coherence blocks under consideration. The resulting algorithm for known $\mathbf{U}$ case is called `Corr-SBL`, and the algorithm in the i.i.d. $\mathbf{W}_r$ scheme is presented as Algorithm 1. When the approach to learn the correlation outlined in Sec. IV-C is incorporated, the algorithm is called `Corr-SBL-learn`. In the shared $\mathbf{W}_r$ scheme, the updates in the inner for-loop are independent of the loop index, and thus a computationally simpler algorithm can be obtained by vectorizing the updates.

E. Online Algorithm

In the MMV estimation problem, the channel is estimated at the end of T coherence blocks, which could be impractical in terms of latency when T is large. In this subsection, we present an online version of Corr-SBL, which also incorporates the advantages of MMV estimation.

We first consider the shared $\mathbf{W}_r$ scheme. The only update which involves averaging over multiple coherence blocks is in computing $\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}^H$. A running sum $\sum_{s=1}^{r-1} \left(\boldsymbol{\Phi}^H \mathbf{y}_s \mathbf{y}_s^H \boldsymbol{\Phi} \right)$

Algorithm 1: `Corr-SBL`

Input: $\{\boldsymbol{\Phi_r}\}_{r=1}^T, \{\mathbf{y}_r\}_{r=1}^T, \mathbf{U}, \sigma_n^2$

Initialize: $k \leftarrow 0, \mathbf{c} \leftarrow \mathbf{1}$

while $k \leq k_{max}$ *and* $\| \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_k - \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_{k-1} \|_F < \epsilon$ **do**

$\quad \hat{\mathbf{R}}_{\mathbf{g}} = \mathbf{0}_{D \times D}$

$\quad$ **for** r = 1 to T **do**

$\quad\quad \boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}} = \frac{1}{\sigma_n^2} \boldsymbol{\Phi}_r^H \boldsymbol{\Phi}_r + \boldsymbol{\Omega}_{\mathbf{c}}$

$\quad\quad \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_{([D],r)} = \frac{1}{\sigma_n^2} \boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} \boldsymbol{\Phi}_r^H \mathbf{y}_r$

$\quad\quad \hat{\mathbf{R}}_{\mathbf{g}} \leftarrow \hat{\mathbf{R}}_{\mathbf{g}} + \frac{\boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} + \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_{([D],r)} \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_{([D],r)}^H}{T}$

$\quad$ **end**

$\quad \mathbf{c} \leftarrow \left(\mathrm{Re} \left\{ \mathbf{U}^{-1} \odot \hat{\mathbf{R}}_{\mathbf{g}}^T \right\} \right)^{-1} \frac{1}{\mathbf{c}}$

$\quad \boldsymbol{\Omega}_{\mathbf{c}} = \mathrm{diag}(\mathbf{c}) \mathbf{U}^{-1} \mathrm{diag}(\mathbf{c})$

$\quad k \leftarrow k + 1$

end

Output: $\{\hat{\mathbf{g}}_r\}_{r=1}^T = \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}, \mathbf{c}$.

Algorithm 2: `Online Corr-SBL`

Input: $\boldsymbol{\Phi}, \mathbf{Y}_{r,\Delta} = \left[\mathbf{y}_r \ \mathbf{y}_{r+1} \ \cdots \ \mathbf{y}_{r+\Delta-1}\right], \mathbf{U}, \sigma_n^2, r, \Delta$

Prior: $\mathbf{R}_{\mathrm{old}} = \sum_{s=1}^{r-1} \left(\boldsymbol{\Phi}^H \mathbf{y}_s \mathbf{y}_s^H \boldsymbol{\Phi}\right)$

Initialize: $k \leftarrow 0, \mathbf{c} \leftarrow \mathbf{1}, \hat{\mathbf{U}} = \mathbf{U}$

while $k \leq k_{max}$ *and* $\| \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_k - \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_{k-1} \|_F < \epsilon$ **do**

$\quad \boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}} = \frac{1}{\sigma_n^2} \boldsymbol{\Phi}^H \boldsymbol{\Phi} + \boldsymbol{\Omega}_{\mathbf{c}}$

$\quad \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} = \frac{1}{\sigma_n^2} \boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} \boldsymbol{\Phi}^H \mathbf{Y}_{r,\Delta}$

$\quad \hat{\mathbf{R}}_{\mathbf{g}} = \boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} + \frac{1}{r+\Delta} \left(\boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} \mathbf{R}_{\mathrm{old}} \boldsymbol{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} + \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}^H \right)$

$\quad \mathbf{c} \leftarrow \left(\mathrm{Re} \left\{ \hat{\mathbf{U}}^{-1} \odot \hat{\mathbf{R}}_{\mathbf{g}}^T \right\} \right)^{-1} \frac{1}{\mathbf{c}}$

$\quad \hat{\boldsymbol{\Gamma}} = \mathrm{diag}(\mathrm{diag}(\hat{\mathbf{R}}_{\mathbf{g}}))$

$\quad \hat{\mathbf{U}} = \hat{\boldsymbol{\Gamma}}^{-\frac{1}{2}} \hat{\mathbf{R}}_{\mathbf{g}} \hat{\boldsymbol{\Gamma}}^{-\frac{1}{2}}$

$\quad \boldsymbol{\Omega}_{\mathbf{c}} = \mathrm{diag}(\mathbf{c}) \hat{\mathbf{U}}^{-1} \mathrm{diag}(\mathbf{c})$

$\quad k \leftarrow k + 1$

end

Output: $\{\hat{\mathbf{g}}_t\}_{s=r}^{r+\Delta-1} = \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}, \mathbf{c},$

$\mathbf{R}_{\mathrm{old}} \leftarrow \mathbf{R}_{\mathrm{old}} + \sum_{s=r}^{r+\Delta-1} \left(\boldsymbol{\Phi}^H \mathbf{Y}_{r,\Delta} \mathbf{Y}_{r,\Delta}^H \boldsymbol{\Phi}\right)$

---

for the $r-1$ preceding coherence blocks can be used as prior knowledge for estimating the channel in the rth coherence block. Building on this idea, we perform joint channel estimation over $\Delta \ll T$ blocks starting from the rth coherence block. We present this solution in Algorithm 2. For the i.i.d. $\mathbf{W}_r$ scheme, a running sum of $\hat{\mathbf{R}}_{\mathbf{g}}$ computed in Theorem 1 is required for each iteration of EM algorithm, resulting in a large memory overhead. Instead, an approximation for the MMV setup can be obtained by storing the running sum of $\hat{\mathbf{R}}_{\mathbf{g}}$ at the end of the final EM iteration of the preceding block.

We note that, at the Tth coherence block, the online al-

gorithm is the same as the MMV algorithm for the shared $\mathbf{W}_r$ scheme, and is an approximation of the MMV algorithm for the i.i.d. $\mathbf{W}_r$ scheme. Thus, for both schemes, the overall latency is reduced at the cost of a slightly higher mean squared error in the initial coherence blocks due to the approximation.

F. Complexity Analysis

Table I compares the per-iteration memory and computational complexity of Corr-SBL, its online version, and MSBL. Here, the computational complexity is measured in terms of the number of floating point operations. The overall complexity of Corr-SBL and Corr-SBL-learn (and online versions) are the same since the additional computation of $\hat{\mathbf{U}}$ given by $\mathcal{O}(D^2)$ is included in $\mathcal{O}(D^3)$. The additional complexity of $\mathcal{O}(D^3)$ of Corr-SBL over MSBL is due to the **c**-update step in Algorithm 1. The complexity is higher for the i.i.d. $\mathbf{W}_r$ scheme due to the additional inner-loop. The online version of Corr-SBL has lower computational complexity, in addition to lower latency. From our experiments, Corr-SBL converges in a similar number of iterations as MSBL, both with and without learning the correlation.

The additional complexity of $\mathcal{O}(D^3)$ can be computationally expensive, especially when large grid sizes are used. An approach to reduce complexity is to use lower grid sizes, but this can result in performance loss due to grid mismatch. A recent approach[3] has considered integrating a greedy search procedure to obtain coarse estimates of the AoAs with a statistical interference model (based on MSBL), followed by dictionary refinement with smaller grid sizes. Another approach to reduce complexity is to consider the algorithm unrolling framework [37], where model based approaches can be used to develop deep learning techniques that perform similar to or better than the optimization based approaches while reducing the complexity [38].

The memory complexity is measured by the storage required for input information. Storing the correlation matrix information leads to higher memory complexity in Corr-SBL compared to MSBL. However, this can be significantly reduced for parameterized models by storing only the parameters. The memory complexity is higher for the i.i.d. $\mathbf{W}_r$ scheme compared to the shared $\mathbf{W}_r$ scheme, as expected.

V. PERFORMANCE ANALYSIS

Corr-SBL algorithm is designed based on the maximum likelihood principle and was shown to attain a stationary point of the cost function. However, the performance of channel estimation is usually measured using other cost functions such as NMSE or spectral efficiency of the system. To analyze the performance of Corr-SBL under these measures, we present a unified framework by considering the class of plug-in LMMSE estimators, which includes the LMMSE, E-LMMSE and IPCI estimators described in Appendix A. We also show that the output of Bayesian learning algorithms, MSBL and Corr-SBL, can be represented as a plug-in LMMSE estimator.

---

[3]https://ece.iisc.ac.in/~cmurthy/Learned_Chester_AI5GPHY_Challenge.pdf

Table I. Memory and Computational Complexity

Algorithm	Mode	Memory complexity	Computational complexity
MSBL	i.i.d. $\mathbf{W}_r$	$\mathcal{O}\left(DMT\right)$	$\mathcal{O}\left(DM^2T\right)$
	Shared $\mathbf{W}_r$	$\mathcal{O}\left(M(D+T)\right)$	$\mathcal{O}\left(DM(M+T)\right)$
Corr-SBL	i.i.d. $\mathbf{W}_r$	$\mathcal{O}\left(DMT + D^2\right)$	$\mathcal{O}\left(DM^2T + D^3\right)$
	Shared $\mathbf{W}_r$	$\mathcal{O}\left(M(D+T) + D^2\right)$	$\mathcal{O}\left(DM(M+T) + D^3\right)$
Online-Corr-SBL-learn	i.i.d. $\mathbf{W}_r$	$\mathcal{O}\left(DM\Delta + D^2\right)$	$\mathcal{O}\left(DM^2\Delta + D^3\right)$
	Shared $\mathbf{W}_r$	$\mathcal{O}\left(M(D+\Delta) + D^2\right)$	$\mathcal{O}\left(DM(M+\Delta) + D^3\right)$

A. Plug-in LMMSE Estimators

The following result unifies the channel estimation schemes considered in this work, and shows that they can be represented as plug-in LMMSE estimators.

Proposition 2. *The class of plug-in LMMSE estimators can be represented as*

$$\hat{\mathbf{h}}_r = \mathbf{M}_r\mathbf{y}_r = \hat{\mathbf{R}}_\mathbf{h}\mathbf{W}_r^H\left(\mathbf{W}_r\hat{\mathbf{R}}_\mathbf{h}\mathbf{W}_r^H + \sigma_n^2\mathbf{I}_M\right)^{-1}\mathbf{y}_r, \quad (16)$$

where the covariance matrix $\hat{\mathbf{R}}_\mathbf{h}$ depends on the estimator:

$$\hat{\mathbf{R}}_\mathbf{h} = \begin{cases} \mathbf{R}_\mathbf{h} = \mathbf{A}\Gamma^{\frac{1}{2}}\mathbf{U}\Gamma^{\frac{1}{2}}\mathbf{A}^H & \textit{LMMSE} \\ \mathbf{A}\Gamma\mathbf{A}^H & \textit{E-LMMSE} \\ \mathbf{A}\Gamma_{Corr\text{-}SBL}^{\frac{1}{2}}\mathbf{U}\Gamma_{Corr\text{-}SBL}^{\frac{1}{2}}\mathbf{A}^H & \textit{Corr-SBL} \\ \mathbf{A}\Gamma_{MSBL}\mathbf{A}^H & \textit{MSBL} \\ \hat{\mathbf{R}}_\mathbf{h}^{IPCI} & \textit{IPCI} \end{cases}$$

where Γ and $\mathbf{U}$ denote the variance matrix and the correlation matrix, and $\hat{\mathbf{R}}_\mathbf{h}^{IPCI} \triangleq \mathbf{W}_r^\dagger\left(\frac{1}{T}\sum_{r=1}^{T}\mathbf{y}_r\mathbf{y}_r^H\right)\left(\mathbf{W}_r^\dagger\right)^H$ with $\mathbf{W}_r^\dagger$ denoting the pseudoinverse of $\mathbf{W}$.

Proof. See Appendix G. ∎

B. Normalized Mean Squared Error (NMSE)

The NMSE in the channel estimate is defined as NMSE $= \mathbb{E}\left[\|\hat{\mathbf{h}}_r - \mathbf{h}\|_2^2\right]/\mathbb{E}\left[\|\mathbf{h}\|_2^2\right]$. The following theorem provides the NMSE for the plug-in LMMSE estimators. Its proof follows from direct computation and is omitted.

Theorem 3. *Consider the estimator $\hat{\mathbf{h}}_r = \mathbf{M}_r\mathbf{y}_r$, where $\mathbf{M}_r = \hat{\mathbf{R}}_\mathbf{h}\mathbf{W}_r^H\left(\mathbf{W}_r\hat{\mathbf{R}}_\mathbf{h}\mathbf{W}_r^H + \sigma_n^2\mathbf{I}_M\right)^{-1}$ is the plug-in LMMSE matrix. Assuming that the estimate $\hat{\mathbf{R}}_\mathbf{h}$ is independent of the measurement $\mathbf{y}_r$, i.e., $\hat{\mathbf{R}}_\mathbf{h}$ is computed using measurements from other coherence blocks, the NMSE in the channel estimate is given by*

$$\frac{1}{\mathrm{Tr}\left[\mathbf{R}_\mathbf{h}\right]} \times \mathbb{E}\left[\mathrm{Re}\left\{\mathrm{Tr}\left[\mathbf{M}_r\left(\mathbf{W}_r\mathbf{R}_\mathbf{h}\mathbf{W}_r^H + \sigma_n^2\mathbf{I}_M\right)\mathbf{M}_r^H\right.\right.\right.$$
$$\left.\left.\left. +\ \mathbf{R}_\mathbf{h} - 2\mathbf{M}_r\mathbf{W}_r\mathbf{R}_\mathbf{h}\right]\right\}\right], \quad (17)$$

where the expectation is over the randomness in $\mathbf{M}_r$ and $\mathbf{W}_r$.

In the above, we assumed that $\hat{\mathbf{R}}_\mathbf{h}$ is independent of the measurement $\mathbf{y}_r$. If this does not hold, the expression for the MSE becomes complicated due to the coupling between the two, making the analysis more involved. In any case,

one typically uses multiple previous channel instantiations to estimate $\hat{\mathbf{R}}_\mathbf{h}$, hence, this is not unduly restrictive.

We note that the error in estimating the covariance matrix using performance measures like Frobenius and spectral norm have been studied in literature [34]. However, very few works consider the *error in signal recovery* using the noisy covariance estimates, especially for the compressed sensing case. Theorem 3 presents the NMSE performance for the class of plug-in LMMSE estimators. For a given set of system parameters $\mathbf{W}, \mathbf{R}_\mathbf{h}$ and σ_n^2, the NMSE depends on $\mathbf{M}_r$, which in turn depends on the covariance estimate. From the theorem, it is straightforward to verify that the least NMSE is obtained by the LMMSE estimator, which assumes perfect knowledge of the covariance matrix $\mathbf{R}_\mathbf{h}$. It can be verified that the genie-aided LMMSE achieves the Cramér-Rao bound for the above problem. The E-LMMSE estimator and MSBL force a diagonal structure correlation, and therefore do not exploit the full covariance structure. The IPCI estimator uses the sample covariance, which requires large number of samples to learn the structure of $\mathbf{R}_\mathbf{h}$ because the underlying sparsity is not exploited. The covariance estimate of Corr-SBL follows has a structure similar to actual covariance, and has the potential to learn $\mathbf{R}_\mathbf{h}$ accurately using only a small number of samples. We illustrate this by comparing the NMSE value computed using Theorem 3 with the simulated NMSE in Fig. 5 in Sec. VI.

C. Spectral Efficiency (SE) Analysis

Let $\hat{\mathbf{H}} = \left[\hat{\mathbf{h}}_1\hat{\mathbf{h}}_2\ldots\hat{\mathbf{h}}_K\right] \in \mathbb{C}^{N\times K}$ denote the channel estimates for all users obtained using (16) in a single coherence block. If the kth user transmits a symbol $x_k\ \forall k$ with zero mean and power P, the received combined vector at the BS is

$$\mathbf{y} = \mathbf{F}\mathbf{W}_{RF}\left(\sum_{k=1}^{K}\mathbf{h}_kx_k + \mathbf{n}\right). \quad (18)$$

where we use $\mathbf{W}_{RF} \in \mathbb{C}^{M\times N}$ as the analog combiner matrix to distinguish it from the matrix $\mathbf{W}_r$ used in the pilot transmission phase. Various designs for the analog combiner $\mathbf{W}_{RF}$ and digital combiner $\mathbf{F} \in \mathbb{C}^{M\times M}$ matrices have been studied in the literature when perfect CSI is available at the BS; one such design aims at designing them independently [39]. In this work, we use the same methodology, with channel estimates $\hat{\mathbf{H}}$ replacing the unknown CSI $\mathbf{H}$. We let the number of users served by the BS to be equal to the number of RF chains, i.e., $K = M$. Then, the phase-only control of the analog combining matrix $\mathbf{W}_{RF}$ is satisfied by constraining the amplitude of

all entries to $\frac{1}{\sqrt{N}}$, and the phase of $(m, k)^{\text{th}}$ entry of $\mathbf{W}_{RF}$, denoted by $\hat{\theta}_{m,k}$, is set using the channel estimate as

$$\hat{\theta}_{m,k} = \hat{\phi}_{m,k}, \ m \in [M], k \in [K], \quad (19)$$

where $\hat{\phi}_{m,k}$ is the phase of $\left[\hat{\mathbf{H}}^H\right]_{(m,k)}$. The effective baseband channel $\hat{\mathbf{H}}_B = \mathbf{W}_{RF}\hat{\mathbf{H}}$ is used to design the digital combiner using the regularized zero forcing (RZF) combining with regularization parameter $\eta = \frac{M\sigma_n^2}{P}$:

$$\mathbf{F} = \left(\hat{\mathbf{H}}_B^H \hat{\mathbf{H}}_B + \eta \mathbf{I}_M\right)^{-1} \hat{\mathbf{H}}_B^H. \quad (20)$$

In order to compute the SE, we write the kth row of (18) as

$$y_k = \underbrace{\mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k\right] x_k}_{\text{signal}} + \underbrace{\left(\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k x_k - \mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k\right] x_k\right)}_{\text{self-interference}}$$
$$+ \underbrace{\sum_{q=1, q\neq k}^{K} \mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_q x_q}_{\text{inter-user interference}} + \underbrace{\mathbf{f}_k \mathbf{W}_{RF} \mathbf{n}}_{\text{AWGN}}, \quad (21)$$

where $\mathbf{f}_k$ is the kth row of $\mathbf{F}$. The term $\mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k\right]$ is treated as a known deterministic channel. Noting that it is uncorrelated with other terms in the summation, we obtain the following result. The proof follows from applying the use-and-then forget bound [40, Section 3.2] and is omitted.

Theorem 4. *The spectral efficiency (SE) of user k is lower-bounded by*

$$SE_k \geq \left(1 - \frac{K}{\tau_c}\right) \log_2\left(1 + \tilde{\gamma}_k\right) \ \text{bits/s/Hz}, \quad (22)$$

where the pre-log factor accounts for pilot overhead, and $\tilde{\gamma}_k$ is an effective SINR term given by

$$\tilde{\gamma}_k = \frac{P|\mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k\right]|^2}{P \sum_{q=1}^{K} \mathbb{E}\left[|\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_q|^2\right] - P|\mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k\right]|^2 + \alpha_k} \quad (23)$$

where $\alpha_k = \sigma_n^2 \mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{W}_{RF}^H \mathbf{f}_k^H\right]$ is the effective noise power in the combined signal.

The expectations in Theorem 4 cannot be obtained in closed form. They are computed using simulations in the next section.

VI. SIMULATION RESULTS

A. Simulation Setup

We now present simulation results to elucidate the role of correlation and sparsity in mmWave channel estimation. We compare the NMSE performance of Corr-SBL with the genie-aided LMMSE, LS, IPCI, E-LMMSE estimators given by (24), (25), (27), (28) in Appendix A, respectively, and sparse recovery algorithms SOMP [25], CovOMP [25] and MSBL [41]. We also compare the average sum rate achieved by the hybrid MIMO architecture with channel estimates obtained using Corr-SBL against that of a system equipped with a fully digital architecture (where there is no need for an analog beamforming stage) and genie-aided LMMSE channel estimates. We consider the uniform correlation model, where $\mathbf{U}_{ij}$

Table II. Simulation Parameters

Parameters	Values
Number of antenna N	256
Grid size D	256
Number of RF chains M	16
Number of users K	16
Number of multipaths L_k	16
Number of snapshots T	512
SNR	10 dB
Correlation coefficient ρ	0.5
Coherence interval τ_c	2000
Hyperparameter pruning threshold ε	10^5
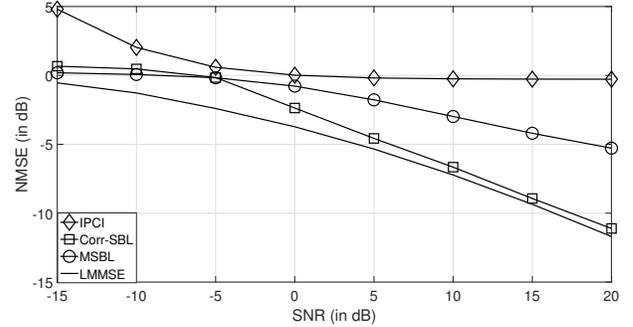
Fig. 2. Comparison of NMSE against SNR for the class of plug-in LMMSE estimators.

equals unity when $i = j$ and equals $\rho \in [0, 1)$ otherwise [42]. The support of the sparse vector is obtained by drawing L_k samples from the D grid points uniformly at random without replacement. The path gain vector is obtained from a complex normal distribution with zero mean and covariance matrix of the form described in Section. II. We use the command `zadoffChuSeq(19,17)` (in MATLAB) to generate a base sequence of length 17. Then, we consider 15 successive cyclically permuted sequences of the base sequence, along with the base sequence, as pilot sequences for the $K = 16$ users. The other parameter values are as listed in Table II, unless specified otherwise.

B. Effect of SNR

In the first experiment, we study the NMSE performance of the class of plug-in LMMSE estimators as a function of the SNR. From Fig. 2, we see that for SNR less than 0 dB, the advantage of exploiting correlation is not significant since the noise overwhelms the signal. However, as the SNR increases beyond 0 dB, performance of all algorithms exploiting sparsity increases, with the gain being higher for Corr-SBL and genie aided LMMSE estimator compared to MSBL. In the further simulations, we fix the SNR to 10 dB, and study the effect of other parameters of the system to elucidate the role of correlation in sparse signal recovery.

C. mmWave Channel Estimation

In the second set of experiments, we present the cumulative distribution function (CDF) of the NMSE of different algo-

Fig. 3. Comparison of NMSE for different algorithms for both shared $\mathbf{W}_r$ and i.i.d. $\mathbf{W}_r$ schemes.

Fig. 5. NMSE performance against sparsity level.

Fig. 4. Loss in NMSE performance with and without learning the correlation due to imperfect correlation information.

D. Corr-SBL Performance

Here, we demonstrate the robustness of the Corr-SBL algorithm to imperfect correlation information and high sparsity levels. First, we demonstrate the advantage of learning the correlation. Let ρ be the true correlation coefficient. The initial value of the correlation coefficient $\hat{\rho}$ for both Corr-SBL and Corr-SBL-learn is chosen uniformly at random in the range $[(1-\epsilon)\rho, (1+\epsilon)\rho]$. The performance loss due to imperfect correlation is presented in Fig. 4, where we plot the difference between the NMSE of each of the algorithms initialized with the imperfect correlation $\hat{\rho}$ and the NMSE of genie-Corr-SBL which has the exact knowledge of ρ, as a function of ϵ. Since Corr-SBL continues with the initial value of the correlation coefficient, its performance deteriorates with increasing ϵ. Corr-SBL-learn learns ρ from the data, and exhibits relatively stable performance irrespective of ϵ, and outperforms Corr-SBL. The NMSE increases slightly with the size of the uncertainty interval, but the loss in NMSE is small even for high values of ϵ. This shows that Corr-SBL-learn does not require a very accurate knowledge of ρ in order to exploit the underlying channel correlation.

The NMSE performance of different algorithms is compared as a function of the number of multi-path components, i.e., the sparsity level of the channel, in Fig. 5. The performance of all sparse recovery algorithms are similar at low sparsity levels, hence exploiting the correlation is not crucial. The performance of SOMP, CovOMP and MSBL deteriorate quickly with increasing sparsity. In contrast, both Corr-SBL and Corr-SBL-learn continue to perform close to the optimal genie-aided estimator even at high sparsity levels, and significantly outperform the other methods. Thus, Bayesian methods can offer significant advantages especially in highly measurement-constrained scenarios, when the prior model is chosen to best-fit the underlying model. Also, the NMSE values computed using Theorem 3 overlap perfectly with the simulated NMSE values, illustrating the accuracy of the theoretical expressions.

rithms, run for 1000 independent realizations of the channel and T coherence blocks. The curves for both i.i.d. $\mathbf{W}_r$ and shared $\mathbf{W}_r$ schemes are presented in Fig. 3. The genie based LMMSE estimator (curve labeled LMMSE), sets the best achievable benchmark for all algorithms and the performance of Corr-SBL (curve labeled Corr-SBL) is only marginally worse. For the case where the correlation coefficient ρ is not known, we estimate it by averaging the off-diagonal entries of $\hat{\mathbf{U}}$ as explained in Sec. IV-C. This curve, labeled Corr-SBL-learn, matches the performance of Corr-SBL, showing that the algorithm can learn ρ without appreciable loss in performance. Exploiting only correlation (IPCI) or only sparsity (MSBL) performs worse than Corr-SBL.

The NMSE with the i.i.d. $\mathbf{W}_r$ scheme is lower than that of the shared $\mathbf{W}_r$ scheme, showing that it is better to use independent measurement matrices across coherence blocks. The genie-based estimator E-LMMSE only exploits sparsity and is a lower bound on the performance of sparse recovery algorithms that do not exploit correlation. It performs worse than Corr-SBL, which illustrates the importance of exploiting correlation in addition to sparsity. However, although it does not exploit correlation, the hierarchical Bayesian prior used in MSBL results in better performance than CovOMP (which does exploit both correlation and sparsity). In the shared $\mathbf{W}_r$ scheme, the algorithms CovOMP and SOMP in fact perform worse than a trivial estimator which outputs all zeros (and yields an NMSE of 0 dB).

E. Shared $\mathbf{W}_r$ vs. i.i.d. $\mathbf{W}_r$ schemes

Fig. 6 compares the channel estimation performance of the i.i.d. $\mathbf{W}_r$ and shared $\mathbf{W}_r$ schemes. As the number of coherence blocks used to estimate the channel covariance (T) increases, the performance of Corr-SBL and Corr-SBL-learn converge to the LMMSE estimator and that of MSBL

(a) Shared $\mathbf{W}_r$ scheme.

(b) i.i.d. $\mathbf{W}_r$ scheme.
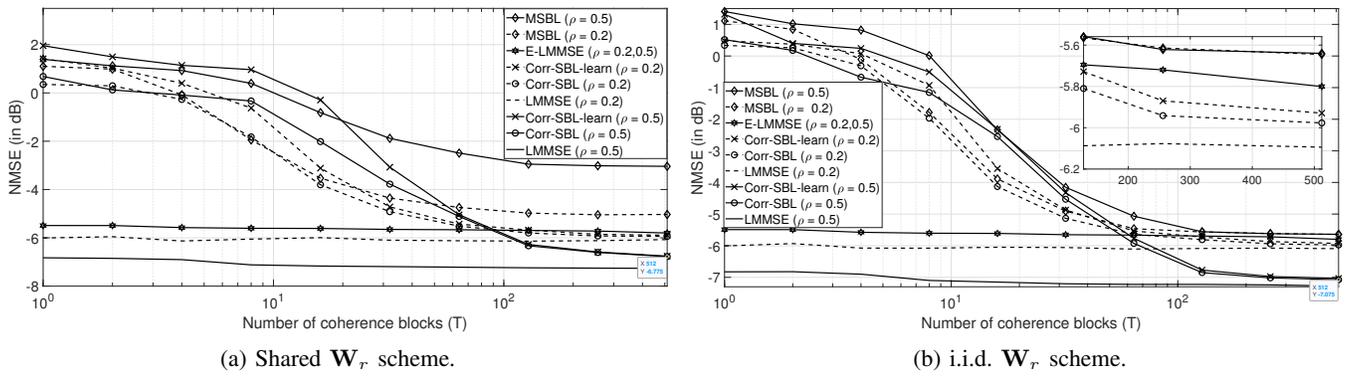
Fig. 6. NMSE performance with averaging over multiple coherence blocks.
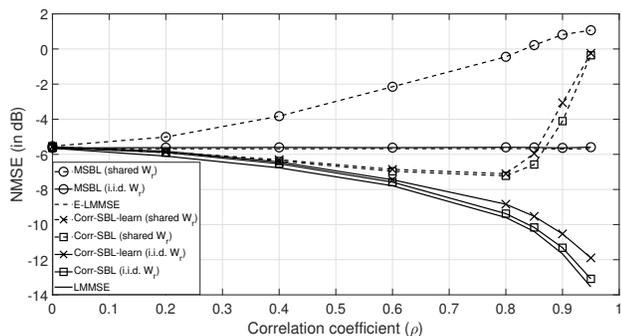
Fig. 7. NMSE performance against the correlation level.

Fig. 9. Sum-rate performance of the system with channel estimates obtained by different algorithms.

Fig. 8. Expected percentage of coherence blocks with NMSE value more than x, where x is the x-axis coordinate value.
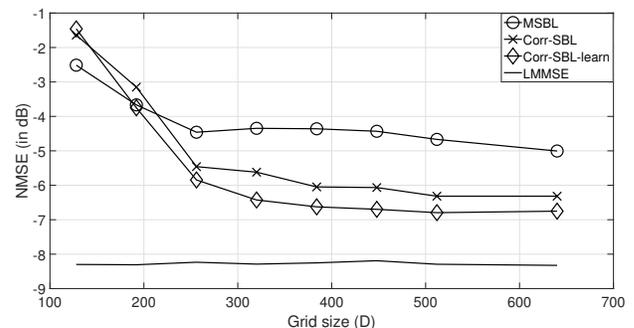
Fig. 10. NMSE performance under a practical clustered channel model (NYUSIM) with grid mismatch.

converges to the E-LMMSE estimator, for both schemes. The NMSE value of the Corr-SBL algorithm for $T = 512$ is lower for the i.i.d. $\mathbf{W}_r$ scheme compared to the shared $\mathbf{W}_r$ scheme. The i.i.d. $\mathbf{W}_r$ scheme offers better performance compared to the shared $\mathbf{W}_r$ scheme at high correlation, as shown in Fig. 7. In the shared $\mathbf{W}_r$ scheme, the performance of MSBL deteriorates with increasing correlation, and Corr-SBL follows the performance of the genie aided LMMSE estimator until a correlation threshold, after which the high correlation overwhelms the covariance estimation procedure, possibly due to the larger condition number of $\mathbf{U}$. However, as seen for both MSBL and Corr-SBL, in the i.i.d. $\mathbf{W}_r$ scheme, better covariance estimation leads to better performance even at high correlation levels.

F. Online Estimation

Here, we compare the performance of the online version of Corr-SBL for different block lengths Δ. We consider the scenario of estimating the channel over $T = 512$ channel coherence blocks. In Fig. 8, we plot the percentage of the 512 coherence blocks where the NMSE is greater than a given value, say x, as a function of x. With $\Delta = 64$, which offers a significant reduction in latency compared to using all 512 coherence blocks, less than 0.2% of the blocks have an NMSE higher than -5 dB, while the MMV solution that uses all the 512 coherence blocks achieves an NMSE slightly lower than -6 dB. Even with $\Delta = 1$, less than 10% of the blocks have an NMSE higher than -3 dB. The higher NMSE at

smaller Δ is due to poor estimates in the initial few blocks; the NMSE in fact rapidly decreases in the later blocks. This small performance loss can be insignificant when compared to the latency reduction advantage in practical applications.

G. Spectral Efficiency

In Fig. 9, we compare the sum-rate performance of the system when the channel estimates are obtained using the different algorithms, calculated using (22). The hybrid architecture in Sec. V-C with genie-aided LMMSE channel estimates performs close to the fully digital architecture with the optimal genie aided LMMSE channel estimates. The sum-rate performance using Corr-SBL channel estimates is close to the genie-aided estimator. It offers nearly 50% higher sum rate compared to MSBL across the range of number of antennas. The performance of IPCI is limited by the accuracy of estimating the covariance itself, and as a consequence, it does not improve with the number of antennas.

H. NYUSIM Channel Model

Lastly, we compare the performance of Corr-SBL and MSBL with the genie-aided LMMSE estimator for the channel model obtained from the NYUSIM mmWave channel simulator [43], [44], which considers a practical clustering model. Since the NYUSIM model does not consider correlation among the path gains, we premultiply the obtained path gain vector with a matrix that results in the covariance matrix of the final vector to be similar to the model considered in this paper. The parameters of the simulation are same as given in Table II, but with the number of multipaths obtained from the NYUSIM varying between 12 and 30 across different Monte Carlo simulations. In Fig. 10, we plot the 95th percentile NMSE (NMSE of top 95% runs, to remove outliers) as a function of the grid size. We can observe that as the grid size increases, the approximation of the channel using (2) becomes more accurate, and, as a result, the performance of the algorithm improves before it saturates beyond a grid size of about 350. Also, unlike the previous simulations, Corr-SBL-learn performs slightly better than Corr-SBL. This is because, with the practical channel model, the covariance matrix may not exactly match the $\mathbf{U}$ assumed by Corr-SBL. This shows that learning the correlation from data, rather than fixing it based on a model, can result in better performance in practice.

VII. CONCLUSION

In this paper, we explored the role of sparsity and intra-vector correlation in the mmWave channel estimation problem. We presented a novel prior model for exploiting both structures and developed a Bayesian inference algorithm named Corr-SBL for channel estimation. Using the optimal prior estimate obtained from an EM-type procedure, we derived the MAP estimate of the sparse mmWave channel with intra-vector correlation. For the case with imperfect correlation information, we presented an approach for learning the correlation. Experimental results showed that Corr-SBL outperforms existing approaches and achieves close to genie-aided optimal performance over a wide range of scenarios. The algorithm is also robust to imperfect correlation information. In practical implementations, the online version reduces the latency at a slight loss in NMSE performance. Future extensions of this work could consider exploiting inter-vector correlation across coherence blocks, in addition to sparsity and intra-vector correlation (see, e.g., [45], [46]). Also, while the paper presented a framework for exploiting correlation, measurement campaigns to estimate the correlation seen in practice will be useful to fine-tune the algorithm for practical use-cases.

APPENDIX

A. Channel Estimation Schemes

In this subsection, the different channel estimation schemes considered for comparison in the simulation results are presented. We recall that the goal is to estimate $\mathbf{h}_r \in \mathbb{C}^N$ given the pilot signal $\mathbf{y}_r \in \mathbb{C}^M$, where $\mathbf{y}_r = \mathbf{W}_r \mathbf{h}_r + \mathbf{n}_r$.

1) Linear Estimation: Here, we restrict the attention to estimation schemes of the form $\hat{\mathbf{h}}_r = \mathbf{M}_r \mathbf{y}_r$. When the channel covariance matrix $\mathbf{R_h}$ is known, LMMSE is the optimal linear estimator.

$$\hat{\mathbf{h}}_r^{\text{LMMSE}} = \mathbf{R_h} \mathbf{W}_r^H \left(\mathbf{W}_r \mathbf{R_h} \mathbf{W}_r^H + \sigma_n^2 \mathbf{I}_M \right)^{-1} \mathbf{y}_r. \quad (24)$$

However, in practice $\mathbf{R_h}$ is not known at the receiver. In this case, the simplest estimator is the least squares (LS) estimator:

$$\hat{\mathbf{h}}_r^{\text{LS}} = \mathbf{W}_r^{\dagger} \mathbf{y}_r, \quad (25)$$

where $\mathbf{W}_r^{\dagger}$ denotes the pseudo-inverse of $\mathbf{W}_r$. The LS estimator does not exploit the correlation $\mathbf{R_h}$ in the channel. To exploit correlation, LS estimates of the channel in T coherence blocks are computed, and the sample covariance obtained from the estimates is used as an estimate for $\mathbf{R_h}$ as

$$\hat{\mathbf{R}}_{\mathbf{h}} = \mathbf{W}_r^{\dagger} \left(\frac{1}{T} \sum_{r=1}^{T} \mathbf{y}_r \mathbf{y}_r^H \right) \left(\mathbf{W}_r^{\dagger} \right)^H. \quad (26)$$

$\hat{\mathbf{R}}_{\mathbf{h}}$ is plugged into the LMMSE estimator resulting in an imperfect channel covariance information (IPCI) based estimator:

$$\hat{\mathbf{h}}_r^{\text{IPCI}} = \hat{\mathbf{R}}_{\mathbf{h}} \mathbf{W}_r^H \left(\mathbf{W}_r \hat{\mathbf{R}}_{\mathbf{h}} \mathbf{W}_r^H + \sigma_n^2 \mathbf{I}_M \right)^{-1} \mathbf{y}_r. \quad (27)$$

2) Compressed Sensing based Estimation: The estimators discussed in the previous subsection do not exploit the spatial sparsity in the channel. Using (3) in (5), and considering $\boldsymbol{\Phi}_r = \mathbf{W}_r \mathbf{A} \in \mathbb{C}^{M \times D}$ as the measurement matrix, the received pilot can be written in the compressed sensing framework as $\mathbf{y}_r = \boldsymbol{\Phi}_r \mathbf{g}_r + \mathbf{n}_r$. Any of the well known sparse recovery algorithms can be used to recover an estimate $\hat{\mathbf{g}}_r$ for $\mathbf{g}_r$ from the pilot signal $\mathbf{y}_r$. Using $\hat{\mathbf{g}}_r$, the channel is estimated as $\hat{\mathbf{h}}_r = \mathbf{A} \hat{\mathbf{g}}_r$. For performance comparison, we consider two algorithms, OMP [11] and SBL [26]. For estimating the channel over multiple coherence blocks, since the channel support remains constant, OMP and SBL can be replaced with their MMV counterparts, SOMP [25] and MSBL [41], respectively. These algorithms exploit the sparse structure in $\mathbf{g}_r$, but do not use the intra-vector correlation.

We also consider two genie-aided estimators. When the covariance matrix $\mathbf{R_g}$ is known, the LMMSE estimate results

in the same estimator as in (24), with $\mathbf{R_h}$ replaced with $\mathbf{A}\mathbf{R_g}\mathbf{A}^H$. To characterize performance of estimators that neglect the correlation information, using the support and individual variances as the diagonal matrix $\mathbf{\Gamma_g}$, a genie-aided element-wise plug-in LMMSE estimator is

$$\hat{\mathbf{g}}_r^{\text{E-LMMSE}} = \mathbf{\Gamma}_g \mathbf{\Phi}_r^H \left(\mathbf{\Phi}_r \mathbf{\Gamma}_g \mathbf{\Phi}_r^H + \sigma_n^2 \mathbf{I}_M \right)^{-1} \mathbf{y}_r. \quad (28)$$

An OMP-based sparse recovery algorithm, CovOMP, was proposed in [25], which exploits both sparsity and correlation. Simulation results show that our solution outperforms Cov-OMP, especially in highly measurement-constrained scenarios.

B. Extensions

The main goal of this paper is to elucidate the role of spatial sparsity and correlation in mmWave channel estimation. For simplicity of exposition, we considered a frequency-flat channel model and single-antenna users in developing our solution. In this section, we briefly discuss extensions of our approach to the cases where the channels are frequency-selective and the users are equipped with multiple antennas.

1) Multiple Antennas at the Users: In this subsection, we present a measurement model for the correlated sparse recovery problem when the users are equipped with multiple antennas. Suppose the BS and the K users employ ULAs with N_B and N_U antennas, and have $M_B < N_B$ and $M_U < N_U$ RF chains, respectively. For this setup, the channel in (2) can be written as,

$$\mathbf{H}_{r,k} = \sum_{l=1}^{L_k} \bar{g}_{r,k,l} \bar{\mathbf{a}}_B(\psi_{r,k,l}) \bar{\mathbf{a}}_U^H(\theta_{r,k,l}) = \bar{\mathbf{A}}_B \bar{\mathbf{g}}_{r,k} \bar{\mathbf{A}}_U^H \in \mathbb{C}^{N_B \times N_U},$$

where $\bar{\mathbf{A}}_B \in \mathbb{C}^{N_B \times L_k}$ and $\bar{\mathbf{A}}_U \in \mathbb{C}^{N_U \times L_k}$ denote the array response vectors at the BS and the kth user, respectively. Using grids of size D_B and D_U, $\mathbf{H}_{r,k}$ can be approximated as

$$\mathbf{H}_{r,k} = \mathbf{A}_B \mathbf{G}_{r,k} \mathbf{A}_U^H$$

where $\mathbf{G}_{r,k} \in \mathbb{C}^{D_B \times D_U}$ is a sparse matrix with L_k nonzero path gains corresponding to the AoAs and AoDs. If the BS uses a combiner $\mathbf{W}_r \in \mathbb{C}^{M_B \times N_B}$ and the user employs a precoder $\mathbf{F}_r \in \mathbb{C}^{N_U \times M_U}$, the received signal at the BS for estimating the kth user's channel (corresponding to (4)) is

$$\mathbf{Y}_{r,k} = \mathbf{W}_r \mathbf{A}_B \mathbf{G}_{r,k} \mathbf{A}_U^H \mathbf{F}_r + \mathbf{N}_{r,k}.$$

By vectorizing $\mathbf{Y}_{r,k}$, we obtain the linear system given by

$$\mathbf{y}_{r,k} = \left(\mathbf{F}^T \otimes \mathbf{W} \right) \left(\mathbf{A}_U^c \otimes \mathbf{A}_B \right) \mathbf{g}_{r,k} + \mathbf{n}_{r,k} \in \mathbb{C}^{M_B M_U \times 1},$$

where $\mathbf{A}_U^c$ denotes the element wise conjugate of $\mathbf{A}_U$, and $\mathbf{g}_{r,k} \in \mathbb{C}^{D_B D_U \times 1}$ is vectorized form of $\mathbf{G}_{r,k}$. This is now in the same form as the sparse recovery framework developed in this paper, albeit with a larger dimensional measurement matrix and sparse vector. At the cost of higher computational complexity, the Corr-SBL can now be directly used for learning the channel $\mathbf{g}_{r,k}$ from the measurements $\mathbf{y}_{r,k}$.

2) Frequency-selective Channel Models: In OFDM systems, the use of subcarriers for data transmission allows one to convert frequency selective channels into multiple parallel frequency flat channels. In this case, the algorithm described

in the paper can be applied independently over the different sub-carriers. However, this does not exploit the correlation in the channels across sub-carriers. The work in [25] considers a combination of time-domain and frequency-domain algorithms. A challenge in extending Corr-SBL to this case is in estimating the parameters of the channel correlation across subcarriers. This is an interesting direction for future work.

C. Proof of Lemma 1

The posterior distribution of $\mathbf{g}_r$ given the observations $\mathbf{y}_r$ and hyperparameter value $\mathbf{c}_{\text{old}}$ is given by

$$p(\mathbf{g}_r | \mathbf{y}_r; \mathbf{c}_{\text{old}}, \sigma_n^2) = \frac{p(\mathbf{y}_r | \mathbf{g}_r; \sigma_n^2) p(\mathbf{g}_r; \mathbf{c}_{\text{old}})}{p(\mathbf{y}_r; \mathbf{c}_{\text{old}}, \sigma_n^2)}$$

Using the prior on $\mathbf{g}_r$: $\mathcal{CN}(\mathbf{g}_r; \mathbf{0}, \mathbf{\Omega}_{\mathbf{c}_{\text{old}}}^{-1})$, we get

$$p(\mathbf{g}_r | \mathbf{y}_r; \mathbf{c}_{\text{old}}, \sigma_n^2) = \frac{\mathcal{CN}(\mathbf{y}_r; \mathbf{\Phi}_r \mathbf{g}_r, \sigma_n^2 \mathbf{I}_M) \, \mathcal{CN}(\mathbf{g}_r; \mathbf{0}, \mathbf{\Omega}_{\mathbf{c}_{\text{old}}}^{-1})}{\mathcal{CN}(\mathbf{y}_r; \mathbf{0}, \mathbf{\Omega}_y^{-1})}$$

$$= k \exp\left(-\mathbf{g}_r^H (\mathbf{\Omega}_{\mathbf{c}_{\text{old}}} + \frac{\mathbf{\Phi}_r^H \mathbf{\Phi}_r}{\sigma_n^2}) \mathbf{g}_r + \frac{\mathbf{g}_r^H \mathbf{\Phi}_r^H \mathbf{y}_r + \mathbf{y}_r^H \mathbf{\Phi}_r \mathbf{g}_r}{\sigma_n^2} \right),$$

where k is a normalization constant. Using $\mathbf{\Omega}_{\mathbf{g}|\mathbf{y}} = \mathbf{\Omega}_{\mathbf{c}_{\text{old}}} + \frac{\mathbf{\Phi}_r^H \mathbf{\Phi}_r}{\sigma_n^2}$ and $\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} = \frac{1}{\sigma_n^2} \mathbf{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} \mathbf{\Phi}_r^H \mathbf{y}_r$ and completing the squares, the posterior distribution can be written as

$$p(\mathbf{g}_r | \mathbf{y}_r; \mathbf{c}_{\text{old}}, \sigma_n^2) = k \exp\left(-(\mathbf{g}_r - \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}})^H \mathbf{\Omega}_{\mathbf{g}|\mathbf{y}} (\mathbf{g}_r - \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}) \right),$$

which is the Gaussian distribution with mean $\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}$ and covariance $\mathbf{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1}$, as given in the statement of the Lemma.

D. Proof of Theorem 1

The lower bound Q on the cost function L using the EM framework is given as

$$Q(\mathbf{c}, \mathbf{c}_{\text{old}}) = \mathbb{E}_{\mathbf{g}_r | \mathbf{y}_r; \mathbf{c}_{\text{old}}, \sigma_n^2} \left[\log(p(\mathbf{g}_r, \mathbf{y}_r; \mathbf{c}, \sigma_n^2)) \right]$$

$$= \mathbb{E}_{\mathbf{g}_r | \mathbf{y}_r, \mathbf{c}_{\text{old}}} \left[\log(p(\mathbf{y}_r | \mathbf{g}_r; \sigma_n^2 \mathbf{I}_M)) \right] + \mathbb{E}_{\mathbf{g}_r | \mathbf{y}_r, \mathbf{c}_{\text{old}}} \left[\log(p(\mathbf{g}_r; \mathbf{c})) \right]. \quad (29)$$

The first expectation is a constant with respect to $\mathbf{c}$, which does not affect the M-step. The second expectation is computed as

$$\mathbb{E}_{\mathbf{g}_r | \mathbf{y}_r, \mathbf{c}_{\text{old}}} [\log(p(\mathbf{g}_r; \mathbf{c}))] = \mathbb{E}_{\mathbf{g}_r | \mathbf{y}_r, \mathbf{c}_{\text{old}}} \left[\log(\mathcal{CN}(\mathbf{g}_r; \mathbf{0}, \mathbf{\Omega}_{\mathbf{c}}^{-1})) \right]$$

$$= -D \log(\pi) + \log(\det(\mathbf{\Omega}_{\mathbf{c}})) - \mathbb{E}_{\mathbf{g}_r | \mathbf{y}_r, \mathbf{c}_{\text{old}}} \left[\mathbf{g}_r^H \mathbf{\Omega}_{\mathbf{c}} \mathbf{g}_r \right].$$

The scalar term $\mathbf{g}_r^H \mathbf{\Omega}_{\mathbf{c}} \mathbf{g}_r$ can be rewritten using the trace operator. Using the product property of trace ($\text{Tr}[\mathbf{AB}] = \text{Tr}[\mathbf{BA}]$), the expectation term can be rewritten as

$$\mathbb{E}_{\mathbf{g}_r | \mathbf{y}_r, \mathbf{c}_{\text{old}}} \left[\mathbf{g}_r^H \mathbf{\Omega}_{\mathbf{c}} \mathbf{g}_r \right] = \text{Tr} \left[\mathbf{\Omega}_{\mathbf{c}} \mathbb{E}_{\mathbf{g}_r | \mathbf{y}_r, \mathbf{c}_{\text{old}}} \left[\mathbf{g}_r \mathbf{g}_r^H \right] \right].$$

The expectation term inside the trace is the second moment matrix with respect to the posterior probability distribution. Using Lemma 1, the second moment matrix is obtained as $\hat{\mathbf{R}}_{\mathbf{g}} \triangleq \mathbb{E}_{\mathbf{g}_r | \mathbf{y}_r, \mathbf{c}_{\text{old}}} \left[\mathbf{g}_r \mathbf{g}_r^H \right] = \mathbf{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} + \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}^H$. Substituting this into (29), the Q function is given by

$$Q(\mathbf{c}, \mathbf{c}_{\text{old}}) = \text{constant} + \log(\det(\mathbf{\Omega}_{\mathbf{c}})) - \text{Tr} \left[\mathbf{\Omega}_{\mathbf{c}} \hat{\mathbf{R}}_{\mathbf{g}} \right].$$

E. Proof of Theorem 2

Using first order optimality condition for maximizing $Q(\mathbf{c}, \mathbf{c}_{\text{old}})$, $\frac{\partial Q}{\partial c_i} = 0$, $i \in \{1, 2, \ldots D\}$, and from Theorem 1,

$$\frac{\partial \left(\log(\det(\mathbf{\Omega_c})) - \text{Tr}\left[\mathbf{\Omega_c} \hat{\mathbf{R}}_{\mathbf{g}} \right] \right)}{\partial c_i} = 0. \tag{30}$$

The first term is

$$\frac{\partial \log(\det(\mathbf{\Omega_c}))}{\partial c_i} = \frac{\partial \log(\det(\mathbf{C U}^{-1} \mathbf{C}))}{\partial c_i} = \frac{2}{c_i}. \tag{31}$$

Considering the second term,

$$\frac{\partial \left(\text{Tr}\left[\mathbf{\Omega_c} \hat{\mathbf{R}}_{\mathbf{g}} \right] \right)}{\partial c_i} = \frac{\partial \left(\text{Tr}\left[\mathbf{C U}^{-1} \mathbf{C} \hat{\mathbf{R}}_{\mathbf{g}} \right] \right)}{\partial c_i}$$
$$= \text{Tr}\left[\mathbf{J}_{ii} \mathbf{U}^{-1} \mathbf{C} \hat{\mathbf{R}}_{\mathbf{g}} + \hat{\mathbf{R}}_{\mathbf{g}} \mathbf{C} \mathbf{U}^{-1} \mathbf{J}_{ii} \right], \tag{32}$$

where $\mathbf{J}_{ii}$ an $N \times N$ matrix with a single 1 in the ith diagonal entry and zeros elsewhere. Using the fact that the two matrices in the trace are Hermitian transpose of each other and using property of single entry matrices, the term is simplified as,

$$\frac{\partial \left(\text{Tr}\left[\mathbf{\Omega_c} \hat{\mathbf{R}}_{\mathbf{g}} \right] \right)}{\partial c_i} = 2 \,\text{Re}\left\{ \left[\mathbf{U}^{-1} \mathbf{C} \hat{\mathbf{R}}_{\mathbf{g}} \right]_{(i,i)} \right\}. \tag{33}$$

Substituting (31) and (33) in (30), and simplifying expressions, the optimality condition is obtained as

$$\frac{1}{c_i} = \text{Re}\left\{ \sum_{k=1}^{D} c_k \left[\mathbf{U}^{-1} \right]_{(i,k)} \left[\hat{\mathbf{R}}_{\mathbf{g}} \right]_{(k,i)} \right\}. \tag{34}$$

F. Proof of Proposition 1

For ease of presentation of the proof, we change our notation, with $\mathbf{c}$ replaced with $\mathbf{c}_{\text{new}}$ and $\mathbf{c}_{\text{old}}$ replaced with $\mathbf{c}$. Let $s_{i,j}$ and $t_{i,j}$ denote the $\{i, j\}^{\text{th}}$ entry of the real symmetric matrix $\mathbf{S} \triangleq \text{Re}\{\mathbf{K}\}$ and its inverse $\mathbf{T}$, respectively. These follow the condition $\sum_{i=1}^{D} s_{i,k} t_{i,l} = \sum_{i=1}^{D} s_{i,k} t_{l,i} = 1$ if $k = l$ and 0 otherwise. Also, since $\mathbf{U}^{-1}$ and $\hat{\mathbf{R}}_{\mathbf{g}}^T$ are positive definite, $\text{Re}\{\mathbf{K}\}$ is a positive definite matrix.

We need to show that the update in (14) results in the quantity $\sum_i \left(\frac{dQ}{dc_i} ((c_{\text{new}})_i - c_i) \right)$ being greater than or equal to zero. This would in turn imply that the cost function is nondecreasing in each iteration. Using the expression for $\frac{dQ}{dc_i}$ derived in Appendix E and the expression for $\mathbf{c}_{\text{new}}$ from (14), the above quantity can be written as

$$\sum_{i=1}^{D} \left(\left(\frac{2}{c_i} - 2 \sum_{k=1}^{D} c_k s_{i,k} \right) \left(\sum_{k=1}^{D} \frac{t_{i,k}}{c_k} - c_i \right) \right)$$
$$= 2 \sum_{i=1}^{D} \left(\sum_{k=1}^{D} \frac{t_{i,k}}{c_i c_k} + s_{i,k} c_i c_k - 1 \right) - 2 \sum_{k,l=1}^{D} \frac{c_k}{c_l} \sum_{i=1}^{D} s_{i,k} t_{i,l}.$$

Using the properties of $\mathbf{S}$ and $\mathbf{T}$, the last term is equal to 1. Further, by considering $\mathbf{B} = \mathbf{CSC}$, where $\mathbf{C}$ is a diagonal matrix with c_i as its entries and using the eigen decomposition of $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, the above can be simplified as

$$2\mathbf{1}^T \left(\mathbf{B} + \mathbf{B}^{-1} - 2\mathbf{I} \right) \mathbf{1} = 2\mathbf{1}^T \mathbf{U} \left(\mathbf{\Lambda} + \mathbf{\Lambda}^{-1} - 2\mathbf{I} \right) \mathbf{U}^T \mathbf{1} \geq 0,$$

where $\mathbf{1}$ is the all ones vector of length D. The last inequality is because entries of $\mathbf{\Lambda}$ are positive since $\mathbf{B}$ is positive definite, and as a consequence $\mathbf{\Lambda} + \mathbf{\Lambda}^{-1} - 2\mathbf{I}$ is a diagonal matrix with non-negative entries, hence is positive semi-definite.

G. Proof of Proposition 2

The LMMSE, E-LMMSE and IPCI estimators presented in Appendix A are in the form of (16) as given in statement of the proposition. For Corr-SBL, we provide an alternative representation of the final estimate as a plug-in LMMSE estimator. The proof for MSBL follows similarly and is omitted.

Let $\mathbf{C}_{opt}$ denote the value of the hyperparameters obtained upon termination of the Corr-SBL algorithm, and let $\mathbf{\Omega_c} = \mathbf{C}_{opt} \mathbf{U}^{-1} \mathbf{C}_{opt}$. The posterior mean estimate $\hat{\mathbf{g}}_r$ is obtained as

$$\hat{\mathbf{g}}_r = \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} = \frac{1}{\sigma_n^2} \mathbf{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} \mathbf{y}_r = \frac{1}{\sigma_n^2} \left(\frac{\mathbf{\Phi}_r^H \mathbf{\Phi}_r}{\sigma_n^2} + \mathbf{\Omega_c} \right)^{-1} \mathbf{\Phi}_r^H \mathbf{y}_r.$$

Using the matrix identity

$$\left(\frac{\mathbf{\Phi}_r^H \mathbf{\Phi}_r}{\sigma_n^2} + \mathbf{\Omega_c} \right)^{-1} \mathbf{\Phi}_r^H = \sigma_n^2 \mathbf{\Omega_c}^{-1} \mathbf{\Phi}_r^H \left(\mathbf{\Phi}_r \mathbf{\Omega_c}^{-1} \mathbf{\Phi}_r + \sigma_n^2 \mathbf{I}_M \right)^{-1},$$

we obtain $\hat{\mathbf{g}}_r = \mathbf{\Sigma}_{opt} \mathbf{\Phi}_r^H \left(\mathbf{\Phi}_r \mathbf{\Sigma}_{opt} \mathbf{\Phi}_r + \sigma_n^2 \mathbf{I}_M \right)^{-1} \mathbf{y}_r$, where $\mathbf{\Sigma}_{opt} \triangleq \mathbf{C}_{opt}^{-1} \mathbf{U} \mathbf{C}_{opt}^{-1}$. From Sec. IV-A, we have $\mathbf{C}_{opt}^{-1} = \mathbf{\Gamma}_{\text{Corr-SBL}}^{\frac{1}{2}}$. Finally, using $\hat{\mathbf{h}}_r = \mathbf{A} \hat{\mathbf{g}}_r$ and substituting $\mathbf{\Phi}_r = \mathbf{W}_r \mathbf{A}$, the output of Corr-SBL can be written in the form (16), with $\hat{\mathbf{R}}_{\mathbf{h}}$ as given in the proposition.

REFERENCES

[1] D. Prasanna and C. R. Murthy, "On the role of sparsity and intra-vector correlation in mmwave channel estimation," in *Proc. SPAWC*, May 2020, pp. 1–5.

[2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[3] M. Cudak, T. Kovarik, T. A. Thomas, A. Ghosh, Y. Kishiyama, and T. Nakamura, "Experimental mmWave 5G cellular system," in *Proc. Globecom*, 2014, pp. 377–381.

[4] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[5] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, 2014.

[6] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[7] J. Palacios, D. De Donno, and J. Widmer, "Tracking mm-Wave channel dynamics: Fast beam training strategies under mobility," in *IEEE Conf. on Comput. Commun. (INFOCOM)*, 2017, pp. 1–9.

[8] J. Singh and S. Ramakrishna, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2670–2683, 2015.

[9] H. Zhang, S. Venkateswaran, and U. Madhow, "Channel modeling and MIMO capacity for outdoor millimeter wave links," in *Proc. WCNC*, 2010, pp. 1–6.

[10] M. K. Samimi, G. R. MacCartney, S. Sun, and T. S. Rappaport, "28 GHz millimeter-wave ultrawideband small-scale fading models in wireless channels," in *IEEE Veh. Technol. Conf.*, May 2016, pp. 1–6.

[11] J. Lee, G. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, June 2016.

[12] K. Venugopal, A. Alkhateeb, N. González Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, 2017.

[13] M. S. Dastgahian and H. Khoshbin, "Rank-defective millimeter-wave channel estimation based on subspace-compressive sensing," *Digit. Commun. Networks*, vol. 2, no. 4, pp. 206 – 217, 2016.

[14] A. Manoj and A. P. Kannu, "Multi-user millimeter wave channel estimation using generalized block OMP algorithm," in *Proc. SPAWC*, 2017, pp. 1–5.

[15] J. P. González-Coma et al., "FDD channel estimation via covariance estimation in wideband massive MIMO systems," *Sensors (Basel)*, vol. 20, no. 930, pp. 1–20, Feb. 2020.

[16] C. K. Anjinappa, A. C. Gürbüz, Y. Yapıcı, and I. Güvenç, "Off-grid aware channel and covariance estimation in mmwave networks," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3908–3921, Jun. 2020.

[17] C. T. Neil, A. Garcia-Rodriguez, P. J. Smith, P. A. Dmochowski, C. Masouros, and M. Shafi, "On the performance of spatially correlated large antenna arrays for millimeter-wave frequencies," *IEEE Trans. Antennas Propag.*, vol. 66, no. 1, pp. 132–148, 2018.

[18] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, 2020.

[19] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, 2018.

[20] 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. Signal Process.*, vol. 67, no. 5, pp. 1251–1266, Mar. 2019.

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

[22] M. Cheffena, L. E. Braten, and T. Ekman, "On the space-time variations of rain attenuation," *IEEE Trans. Antennas Propag.*, vol. 57, no. 6, pp. 1771–1782, Jun. 2009.

[23] V. K. Sakarellos, D. Skraparlis, A. D. Panagopoulos, and J. D. Kanellopoulos, "Cooperative diversity performance in millimeter wave radio systems," *IEEE Trans. Commun.*, vol. 60, pp. 3641–3649, Dec. 2012.

[24] Y. W. E. Chan, Cam Chung La, and B. Soong, "Comparative study of correlated shadowing loss model for wireless sensor networks," in *Proc. Int. Conf. Inform. Commun., and Signal Process.*, Dec. 2013, pp. 1–5.

[25] S. Park and R. W. Heath, "Spatial channel covariance estimation for the hybrid MIMO architecture: A compressive sensing-based approach," *IEEE Trans. Wireless Commun.*, vol. 17, pp. 8047–8062, Dec. 2018.

[26] 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.

[27] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using lse and sparse message passing for mmwave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, 2019.

[28] H. Tang, J. Wang, and L. He, "Off-grid sparse Bayesian learning-based channel estimation for mmWave massive MIMO uplink," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 45–48, 2019.

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

[30] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, 2014.

[31] J. Lee Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The Amer. Statistician*, vol. 42, no. 1, pp. 59–66, 1988.

[32] I. Viering, H. Hofstetter, and W. Utschick, "Spatial long-term variations in urban, rural and indoor environments," in *Proc. COST*, vol. 273. Citeseer, 2002.

[33] D. Kudathanthirige and G. Amarasuriya, "Sum rate analysis of massive MIMO downlink with hybrid beamforming," in *Proc. Globecom*, 2017, pp. 1–7.

[35] C. M. Bishop, *Pattern recognition and machine learning*. New York, NY: Springer, 2006.

[34] M. Azizyan, A. Krishnamurthy, and A. Singh, "Extreme compressive sampling for covariance estimation," *IEEE Trans. Inf. Theory*, vol. 64, no. 12, pp. 7613–7635, Dec. 2018.

[36] R. M. Neal and G. E. Hinton, *A View of the EM Algorithm that Justifies Incremental, Sparse, and other Variants*. Dordrecht: Springer Netherlands, 1998, pp. 355–368.

[37] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *arXiv preprint arXiv:1912.10557*, 2019.

[38] R. J. Peter and C. R. Murthy, "Learned-sbl: A deep learning architecture for sparse signal recovery," *arXiv preprint arXiv:1909.08185*, 2019.

[39] A. A. Nasir, H. D. Tuan, T. Q. Duong, H. V. Poor, and L. Hanzo, "Hybrid beamforming for multi-user millimeter-wave networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2943–2956, 2020.

[40] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge: Cambridge University Press, 2016.

[41] 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.

[42] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.

[43] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. on Microw. Theory Techn.*, vol. 64, no. 7, pp. 2207–2225, 2016.

[44] S. Ju, O. Kanhere, Y. Xing, and T. S. Rappaport, "A millimeter-wave channel simulator NYUSIM with spatial consistency and human blockage," in *Proc. Globecom*, 2019, pp. 1–6.

[45] 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, 2015.

[46] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, 2019.
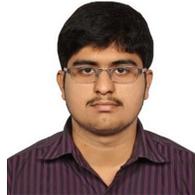
Dheeraj Prasanna (S'16) received the B. E. degree in Electronics and Communication Engineering from Rashtreeya Vidyalaya College of Engineering (RVCE), Bangalore. Currently, he is an M. Tech. (Research) student in the Department of Electrical Communication Engineering at the Indian Institute of Science, Bangalore, India. His research interests are in the areas of compressed sensing, Bayesian learning and wireless communications.

Chandra R. Murthy (S'03–M'06–SM'11) received the B. Tech. degree in Electrical Engineering from the Indian Institute of Technology, Madras, the M. S. and Ph. D. degrees in Electrical and Computer Engineering from Purdue University and the University of California, San Diego. Currently, he is a Professor in the Department of Electrical Communication Engineering at the Indian Institute of Science, Bangalore, India. His research interests are in the areas of energy harvesting communications, 5G/6G technologies and compressed sensing. He is currently serving as a senior area editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and as an associate editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON INFORMATION THEORY.