Learned ChEster

Sai Subramanyam Thoota*, Christo Kurisummoottil Thomas†, Rakesh Mundlamuri†,
Sameera Bharadwaja*, Chandra R Murthy*, and Marios Kountouris†
*Indian Institute of Science, Bangalore, India.
Email: {thoota, sameerah, cmurthy}@iisc.ac.in
†EURECOM, Sophia-Antipolis, France.
Email: {rakesh.mundlamuri, kurisumm, marios.kountouris}@eurecom.fr

A. Notation

The operator $(\cdot)^*$ represents the conjugate transpose or conjugate for a matrix or a scalar respectively. $\bar{\mathbf{A}}$ denotes conjugate of a matrix $\mathbf{A}$. The probability density function (pdf) of a complex Gaussian random variable x with mean μ and variance σ^2 is denoted by $\mathcal{CN}(x; \mu, \nu)$. blkdiag$(\cdot)$ represents blockdiagonal part of a matrix. diag$(\mathbf{X})$ or diag$(\boldsymbol{x})$ represents a vector obtained by the diagonal elements of the matrix $\mathbf{X}$ or the diagonal matrix obtained with the elements of $\boldsymbol{x}$ in the diagonal respectively. $\mathbf{A} \otimes \mathbf{B}$ denotes a Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$. Tx denotes the transmitter and Rx denotes the receiver.

I. System Model

In this section, we outline the problem formulation for the mmWave channel estimation challenge. We consider a single user (UE) hybrid multi-carrier mmWave MIMO OFDM uplink with N_t transmit antennas at the user and N_r receive antennas at the base station (BS). The hybrid architecture involves L_t RF chains at the transmit side and L_r RF chains at the receive side. We employ only RF precoding and combining during the channel estimation phase, and therefore the baseband precoding and combining matrices are set to $\mathbf{I}_{L_t}$ and $\mathbf{I}_{L_r}$, respectively. Since there is no knowledge of channel during the initial access phase, we use random phase shifts for the analog precoder (denoted $\mathbf{F}_{\text{tr}} \in \mathcal{C}^{N_t \times L_t}$) and combiner (denoted $\mathbf{W}_{\text{tr}} \in \mathcal{C}^{N_r \times L_r}$). We denote the number of data streams and subcarriers by N_s and K, respectively. The system operates with uniform linear arrays (ULAs) at both the UE and BS with half wavelength spacing between each antenna.

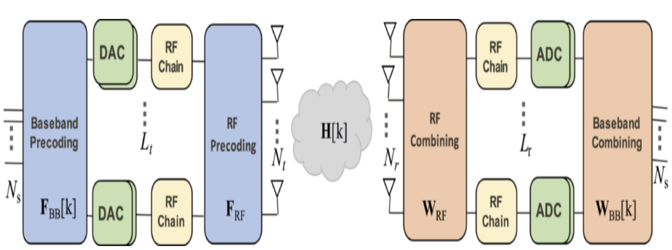

Fig. 1. Millimeter wave MIMO system based on a hybrid architecture. In the site-specific channel estimation challenge, the BS operates as receiver and the UE as transmitter.

After RF combining and downconversion, the received signal during the m^{th} pilot frame for the k^{th} subcarrier is given by

$$\mathbf{y}^{(m)}[k] = \mathbf{W}_{\text{tr}}^{(m)^*}(\mathbf{H}[k]\mathbf{F}_{\text{tr}}^{(m)}\mathbf{q}^{(m)}t^{(m)}[k] + \mathbf{n}^{(m)}[k]), \quad (1)$$

where $\mathbf{H}[k] \in \mathcal{C}^{N_r \times N_t}$ represents the frequency domain MIMO channel matrix for subcarrier k, $\mathbf{q}^{(m)}$ is the frequency flat pilot vector, and $t^{(m)}[k]$ is a frequency dependent pilot symbol. The noise vector $\mathbf{n}^{(m)}[k]$ is circularly symmetric complex Gaussian distributed with zero mean independent and identically distributed (i.i.d.) components of variance σ^2, denoted by $\mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{N_r})$. Each entry of the transmit pilot $\mathbf{q}^{(m)}$ is selected as $\frac{1}{\sqrt{2L_t}}(a + jb)$, where $a, b \in \{-1, 1\}$ and are uniformly distributed. Note that $\left\|\mathbf{q}^{(m)}\right\|^2 = 1$ and hence the transmit SNR is defined as $\rho = \frac{1}{\sigma_n^2}$. For this challenge, the transmit power is kept constant throughout. More details on the channel model is provided in [1].

After vectorizing (1), and using the result $vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})vec(\mathbf{X})$, we obtain

$$vec(\mathbf{y}^{(m)}[k]) = \underbrace{\left(\mathbf{q}^{(m)\,T}\mathbf{F}_{\text{tr}}^{(m)\,T} \otimes \mathbf{W}_{\text{tr}}^{(m)^*}\right)}_{\mathbf{\Phi}^{(m)}} vec(\mathbf{H}[k]) + \mathbf{W}_{\text{tr}}^{(m)^*}\mathbf{n}^{(m)}[k]. \quad (2)$$

Our goal is to estimate the channels $\mathbf{H}[k], k = 1, 2, \ldots, K$ using the above received pilot signals.

A. Channel Model

The datasets provided for the challenge involve channels which are generated using the Raymobtime dataset available at https://www.lasse.ufpa.br/raymobtime/. The MIMO channel is assumed to be frequency selective, with delay tap length N_c in the time domain. The d^{th} delay tap of the channel is modeled as a pathwise channel model with L paths as follows:

$$\mathbf{H}_d = \sum_{l=1}^{L} \alpha_l p(dT_s - \tau_l) a_{\text{R}}(\theta_l) a_{\text{T}}(\phi_l)^*, \quad (3)$$

where α_l represents the complex path coefficient, θ_l is the angle of arrival (AoA) at the BS side, ϕ_l is the angle of departure (AoD) from the UE side and $p(\tau)$ represents the pulse shaping filter. τ_l denotes the delay of the l^{th} path and there are L multipaths. T_s denotes the sampling time. The channel $\mathbf{H}_d$ can be approximated using the virtual channel model [2]

$$\mathbf{H}_d = \mathbf{A}_{\mathrm{R}} \boldsymbol{\Delta}_d \mathbf{A}_{\mathrm{T}}^*. \tag{4}$$

The matrix $\mathbf{A}_{\mathrm{R}} \in \mathcal{C}^{N_r \times G_r}$ contains the Tx side antenna array response vectors evaluated at a grid of size G_r for the AoA and $\mathbf{A}_{\mathrm{T}} \in \mathcal{C}^{N_t \times G_t}$ contains the Rx side antenna array response vectors at a grid of size G_t for the AoD. $\boldsymbol{\Delta}_d \in \mathcal{C}^{G_r \times G_t}$ represents a sparse matrix with entries corresponding to the complex path coefficients $\alpha_l p(dT_s - \tau_l)$ at the locations where a path with corresponding AoA and AoD exists. Further, the frequency domain representation of the channel at subcarrier k can be written in terms of delay taps as follows.

$$\mathbf{H}[k] = \sum_{d=0}^{N_c-1} \mathbf{H}_d e^{-j2\pi \frac{kd}{K}} = \mathbf{A}_{\mathrm{R}} \boldsymbol{\Delta}[k] \mathbf{A}_{\mathrm{T}}^*, \tag{5}$$

where

$$\boldsymbol{\Delta}[k] = \sum_{d=1}^{N_c} \boldsymbol{\Delta}_d e^{-j2\pi \frac{kd}{K}}. \tag{6}$$

Vectorizing $\mathbf{H}[k]$

$$vec(\mathbf{H}[k]) = \left(\bar{\mathbf{A}}_{\mathrm{T}} \otimes \mathbf{A}_{\mathrm{R}} \right) vec(\boldsymbol{\Delta}[k]). \tag{7}$$

Further, defining $\boldsymbol{\Psi} = \bar{\mathbf{A}}_{\mathrm{T}} \otimes \mathbf{A}_{\mathrm{R}}$ and $\mathbf{h}[k] = vec(\boldsymbol{\Delta}[k])$ and substituting for $vec(\mathbf{H}[k])$ in (2), we get the received signal model as

$$vec(\mathbf{y}^{(m)}[k]) = \boldsymbol{\Phi}^{(m)} \boldsymbol{\Psi} \mathbf{h}[k] + \mathbf{W}_{\text{tr}}^{(m)^*} \mathbf{n}^{(m)}[k]. \tag{8}$$

We denote $\mathbf{W}_{\text{tr}}^{(m)^*} \mathbf{n}^{(m)}[k]$ as $\mathbf{n}^{(m)}[k]$ for convenience. By concatenating the RF combined signal of M training frames, we get

$$\underbrace{\begin{bmatrix} \mathbf{y}^{(1)}[k] \\ \vdots \\ \mathbf{y}^{(M)}[k] \end{bmatrix}}_{\mathbf{y}[k]} = \underbrace{\begin{bmatrix} \boldsymbol{\Phi}^{(1)} \\ \vdots \\ \boldsymbol{\Phi}^{(M)} \end{bmatrix}}_{\boldsymbol{\Phi}} \boldsymbol{\Psi} \mathbf{h}^{\mathrm{v}}[k] + \underbrace{\begin{bmatrix} \mathbf{n}^{(1)}[k] \\ \vdots \\ \mathbf{n}^{(M)}[k] \end{bmatrix}}_{\mathbf{n}[k]}. \tag{9}$$

Now, by stacking the received signal of K subcarriers, we get the final system equation

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} \mathbf{y}[1] & \dots & \mathbf{y}[K] \end{bmatrix} \\ &= \boldsymbol{\Phi} \boldsymbol{\Psi} \begin{bmatrix} \mathbf{h}^{\mathrm{v}}[1] & \dots & \mathbf{h}^{\mathrm{v}}[K] \end{bmatrix} + \begin{bmatrix} \mathbf{n}[1] & \dots & \mathbf{n}[K] \end{bmatrix} \\ &= \boldsymbol{\Phi} \boldsymbol{\Psi} \mathbf{H}^{\mathrm{v}} + \mathbf{N}. \end{aligned} \tag{10}$$

Now, our goal is to estimate $\mathbf{H}^{\mathrm{v}}$ given $\mathbf{Y}$ and $\boldsymbol{\Phi}$. As the AoD/AoA are the same for all the subcarriers, $\mathbf{H}^{\mathrm{v}}$ has a joint row sparse structure, which we will utilize in our proposed solution.

II. Proposed Approach

In this section, we propose a model based approach using a compressed sensing (CS) framework to estimate the mmWave channel given the received pilot measurements and the frequency flat transmit vector. We integrate a greedy search procedure and a statistical inference method to estimate the channel. The algorithm consists of the following steps:

1) Preconditioning
2) Multi-level greedy search to obtain initial channel and dictionary estimates
3) Noise variance estimation
4) Bayesian learning to refine the channel and noise variance estimates
5) Channel estimate denoising

We provide a detailed description of each step below. We summarize the ML challenge problem concisely here [1]. The received signal after RF combining is

A. Preconditioning

Sparse signal recovery using greedy algorithms such as orthogonal matching pursuit (OMP) assumes that the noise covariance matrix is diagonal, without which the algorithm may not select the atom that has the highest correlation with the received signal. Due to the RF combining using $\mathbf{W}_{\text{tr}}$ at the front end of the receiver, the noise becomes correlated. To nullify the effect of $\mathbf{W}_{\text{tr}}$, we use a noise whitening filter using an approach provided in [1] that diagonalizes the noise covariance matrix.

The noise covariance matrix before whitening is

$$\mathbb{E}\left[\mathbf{n}^{(i)}[k]\mathbf{n}^{(j)^*}[k]\right] = \mathbf{W}_{\text{tr}}^{(i)^*} \mathbf{W}_{\text{tr}}^{(j)} \delta[i-j] \tag{11}$$

$$\begin{aligned} \mathbf{C}_{\mathrm{w}} &= \mathbb{E}\left[\mathbf{n}[k]\mathbf{n}^*[k]\right] \\ &= \text{blkdiag}\{\mathbf{W}_{\text{tr}}^{(1)^*} \mathbf{W}_{\text{tr}}^{(1)}, \dots, \mathbf{W}_{\text{tr}}^{(M)^*} \mathbf{W}_{\text{tr}}^{(M)}\}. \end{aligned} \tag{12}$$

We do a Cholesky decomposition of $\mathbf{C}_{\mathrm{w}} = \mathbf{D}_{\mathrm{w}}^* \mathbf{D}_{\mathrm{w}}$, where $\mathbf{D}_{\mathrm{w}} \in \mathbb{C}^{ML_r \times ML_r}$ is upper triangular. Now we multiply the RF combined received signal (9) by $\mathbf{D}_{\mathrm{w}}^{-*}$ to obtain the noise whitened received signal.

$$\mathbf{y}_{\mathrm{w}}[k] = \mathbf{D}_{\mathrm{w}}^{-*} \mathbf{y}[k] = \mathbf{D}_{\mathrm{w}}^{-*} \boldsymbol{\Upsilon} \mathbf{h}^{\mathrm{v}}[k] + \mathbf{D}_{\mathrm{w}}^{-*} \mathbf{n}[k] \tag{13}$$

$$= \boldsymbol{\Upsilon}_{\mathrm{w}} \mathbf{h}^{\mathrm{v}}[k] + \mathbf{D}_{\mathrm{w}}^{-*} \mathbf{n}[k]. \tag{14}$$

Concatenating the noise whitened received signal of all the K subcarriers, we get

$$\begin{aligned} \mathbf{Y}_{\mathrm{w}} &= \begin{bmatrix} \mathbf{y}_{\mathrm{w}}[1] & \dots & \mathbf{y}_{\mathrm{w}}[K] \end{bmatrix} \\ &= \mathbf{D}_{\mathrm{w}}^{-*} \boldsymbol{\Upsilon} \begin{bmatrix} \mathbf{h}^{\mathrm{v}}[1] & \dots & \mathbf{h}^{\mathrm{v}}[K] \end{bmatrix} + \mathbf{D}_{\mathrm{w}}^{-*} \begin{bmatrix} \mathbf{n}[1] & \dots & \mathbf{n}[K] \end{bmatrix} \\ &= \boldsymbol{\Upsilon}_{\mathrm{w}} \mathbf{H}^{\mathrm{v}} + \mathbf{D}_{\mathrm{w}}^{-*} \mathbf{N} \\ &= \boldsymbol{\Phi}_{\mathrm{w}} \boldsymbol{\Psi} \mathbf{H}^{\mathrm{v}} + \mathbf{N}_{\mathrm{w}}\,, \end{aligned} \tag{15}$$

where $\mathbf{Y}_{\mathrm{w}} \in \mathbb{C}^{ML_r \times K}$, $\boldsymbol{\Upsilon}_{\mathrm{w}} \in \mathbb{C}^{ML_r \times G_t G_r}$, and $\mathbf{H}^{\mathrm{v}} \in \mathbb{C}^{G_t G_r \times K}$. Now, our goal is to estimate the row sparse matrix $\mathbf{H}^{\mathrm{v}}$ given $\mathbf{Y}_{\mathrm{w}}$ and $\boldsymbol{\Phi}_{\mathrm{w}}$.

B. Multi-level Greedy Search

We obtain an initial channel estimate using multi-level greedy search procedure (MLGS) with a coarsely quantized beamspace dictionary. We adopt simultaneously weighted orthogonal matching pursuit (SW-OMP) as our base algorithm to estimate the channel [1]. As the sparsifying dictionary is unknown a priori, we use row-truncated discrete Fourier transform matrices of size $N_t \times G_t$ and $N_r \times G_r$ as the transmit and receive array steering matrices, respectively. Let $\widetilde{\boldsymbol{\Psi}}$ be the initial sparsifying dictionary.

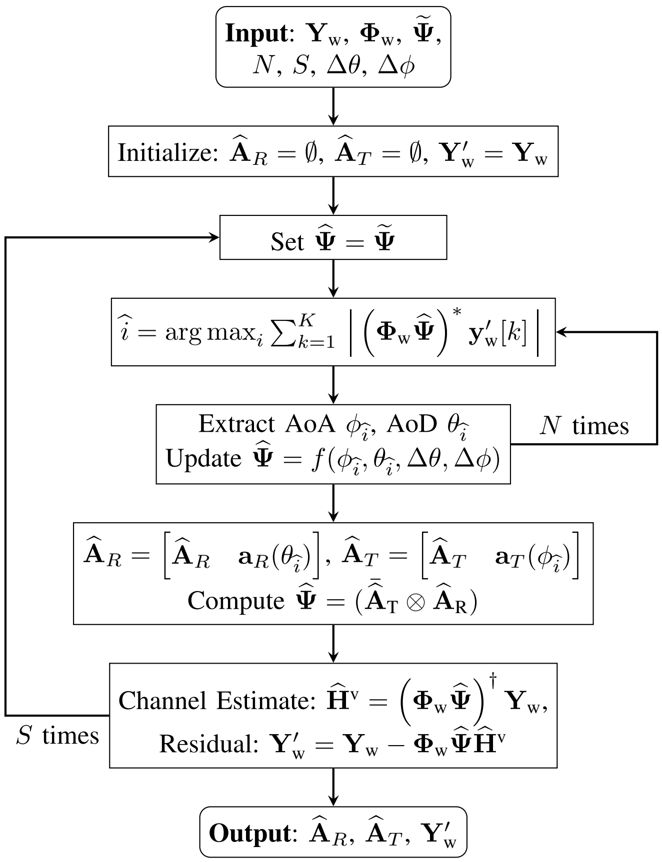

Fig. 2. Flow diagram of MLGS.

In the first step of MLGS, we select a column from $\widetilde{\mathbf{\Psi}}$ that is maximally correlated with the received signal. Mathematically,

$$\widehat{i} = \arg\max_i \sum_{k=1}^{K} \left| \left(\mathbf{\Phi}_{\text{w}}\widetilde{\mathbf{\Psi}}\right)^* \mathbf{y}_{\text{w}}[k] \right|^2, \tag{16}$$

where $|\cdot|$ denotes an element-wise modulus operation. Once we select $\widehat{i}$, we extract AoD $\theta_{\widehat{i}}$ and AoA $\phi_{\widehat{i}}$ using the structure of $\widetilde{\mathbf{\Psi}}$, and form a finely spaced dictionary of range $(\theta_{\widehat{i}} - \Delta\theta, \theta_{\widehat{i}} + \Delta\theta)$ and $(\phi_{\widehat{i}} - \Delta\phi, \phi_{\widehat{i}} + \Delta\phi)$, where $\Delta\theta$ and $\Delta\phi$ are appropriately chosen based on the spatial quantization of the previously chosen dictionary. We repeat (16) with $\widetilde{\mathbf{\Psi}}$ replaced by the newly formed dictionary, and choose a new $\{$AoD,AoA$\}$ pair. We repeat this process for a fixed number of times (say N) and select one set of AoD and AoA. Then, we compute

$$\widehat{\mathbf{H}}^{\text{v}} = \left(\mathbf{\Phi}_{\text{w}}\widehat{\mathbf{\Psi}}\right)^{\dagger} \mathbf{Y}_{\text{w}}, \tag{17}$$

where $\widehat{\mathbf{\Psi}}$ is formed using the currently chosen AoD and AoA. This whole procedure constitutes the first out of S iterations of the MLGS algorithm in which we recover a single tap.

In the s^{th} iteration of MLGS, we recover s channel taps by following the same steps as above, but with the residual $\mathbf{Y}'_{\text{w}} = \mathbf{Y}_{\text{w}} - \mathbf{\Phi}_{\text{w}}\widehat{\mathbf{\Psi}}\widehat{\mathbf{H}}^{\text{v}}$ as observations, where $\widehat{\mathbf{\Psi}}$ comprises the set of $\{$AoD, AoA$\}$ pairs chosen till $s-1$ iterations. Therefore, after S iterations, we recover S virtual beamspace channel taps. We summarize MLGS as a flow diagram in Fig. 2.

C. Noise Variance Estimation

We estimate the noise variance $\widehat{\sigma}_n^2$ using the residual output from MLGS. The noise variance is computed as

$$\widehat{\sigma}_n^2 = \frac{1}{MKL_r}\|\mathbf{Y}'_{\text{w}}\|_F^2, \tag{18}$$

where $\|\mathbf{A}\|_F$ denotes the Frobenius norm of a matrix $\mathbf{A}$.

D. Modelling Off-Grid Effects

MLGS provides reasonable AoDs and AoA estimates, but the virtual beamspace quantized sparsifying dictionary used may not be able to obtain the exact AoD and AoA that lie in the off grid regions of the dictionary. To combat this, we adopt a statistical inference procedure to obtain near accurate AoDs and AoAs which improves the NMSE performance of our proposed algorithm [3]–[5]. We model the off grid effects as follows:

Suppose the original and estimated steering matrices are $\{\mathbf{A}_R, \mathbf{A}_T\}$ and $\{\widehat{\mathbf{A}}_R, \widehat{\mathbf{A}}_T\}$, respectively.

$$\begin{aligned}
\mathbf{A}_R &= \left[\mathbf{a}_R(\theta_1) \quad \dots \quad \mathbf{a}_R(\theta_S)\right], \\
\widehat{\mathbf{A}}_R &= \left[\widehat{\mathbf{a}}_R(\theta_1 + \Delta\theta_1) \quad \dots \quad \widehat{\mathbf{a}}_R(\theta_S + \Delta\theta_S)\right] \\
\mathbf{A}_T &= \left[\mathbf{a}_T(\phi_1) \quad \dots \quad \mathbf{a}_T(\phi_S)\right], \\
\widehat{\mathbf{A}}_T &= \left[\widehat{\mathbf{a}}_T(\phi_1 + \Delta\phi_1) \quad \dots \quad \widehat{\mathbf{a}}_T(\phi_S + \Delta\phi_S)\right]
\end{aligned}$$

It is straightforward to prove that, for appropriately chosen S, there exists bijective transformations $\mathbf{T}_R : \widehat{\mathbf{A}}_R \to \mathbf{A}_R$ and $\mathbf{T}_T : \widehat{\mathbf{A}}_T \to \mathbf{A}_T$. Therefore, for any $\mathbf{P} \in \mathbb{C}^{S\times S}$,

$$\begin{aligned}
\mathbf{A}_R\mathbf{P}\mathbf{A}_T^* &= \widehat{\mathbf{A}}_R\mathbf{T}_R\mathbf{P}\mathbf{T}_T^*\widehat{\mathbf{A}}_T^* \\
&\triangleq \widehat{\mathbf{A}}_R\mathbf{Q}\widehat{\mathbf{A}}_T^*.
\end{aligned} \tag{19}$$

From (19), we can see that the off-grid effects got absorbed into the matrix $\mathbf{Q}$, and for sufficiently small $\{\Delta\theta_1, \dots, \Delta\theta_S\}$ and $\{\Delta\phi_1, \dots, \Delta\phi_S\}$, $\mathbf{Q}$ is approximately sparse for a sparse $\mathbf{P}$. Note that, in the context of our channel estimation problem, our aim is to estimate the product $\mathbf{A}_R\mathbf{P}\mathbf{A}_T^*$ (or $\widehat{\mathbf{A}}_R\mathbf{Q}\widehat{\mathbf{A}}_T^*$), and not $\mathbf{P}$ (or $\mathbf{Q}$).

E. Bayesian Learning

Now, our goal is to refine the channel estimates output by MLGS procedure. We provide the measurement equation here for reference.

$$\mathbf{Y}_{\text{w}} = \mathbf{\Phi}_{\text{w}}\widehat{\mathbf{\Psi}}\mathbf{H}^{\text{v}} + \mathbf{N}_{\text{w}}\,, \tag{20}$$

where $\widehat{\mathbf{\Psi}} = (\bar{\widehat{\mathbf{A}}}_{\text{T}} \otimes \widehat{\mathbf{A}}_{\text{R}})$ is the dictionary output by MLGS.

We adopt a statistical inference approach to infer the posterior distribution of $\mathbf{H}^{\text{v}}$ given the measurements $\mathbf{Y}_{\text{w}}$, measurement matrix $\mathbf{\Phi}\widehat{\mathbf{\Psi}}$, and noise variance $\widehat{\sigma}_n^2$. We use sparse Bayesian learning (SBL) which is a type II maximum likelihood estimation procedure to obtain the channel estimate [3], [4]. We treat $\mathbf{H}^{\text{v}}$ as a hidden variable, and impose a complex Gaussian prior on it with mean $\mathbf{0}$. Joint sparsity is imposed by treating the variance of each entry of a row of $\mathbf{H}^{\text{v}}$ as a hyperparameter. We use expectation maximization (EM) procedure to obtain the posterior distribution of the channel. EM works on the principle of lower bounding the objective

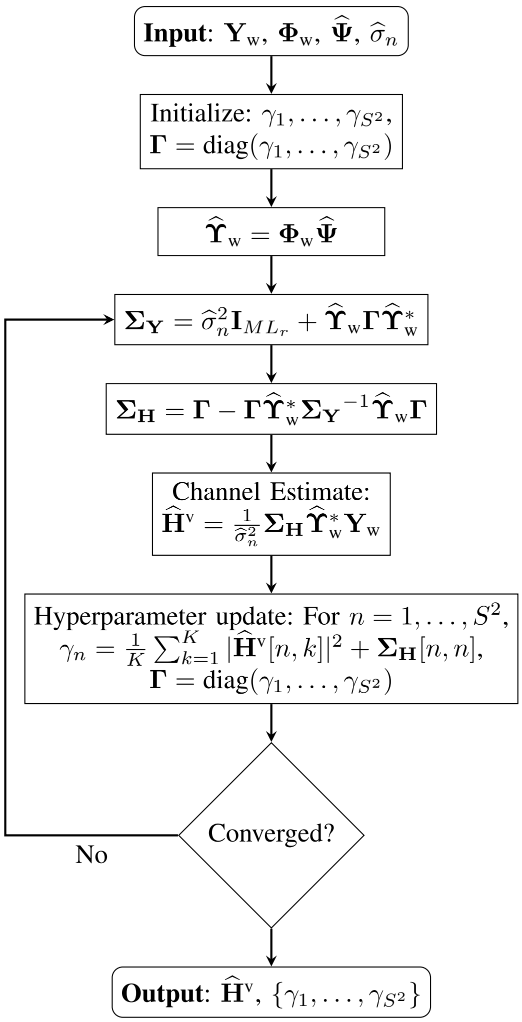

Fig. 3. Flow diagram of MSBL.

function of a maximization problem, and finds a local optimum in an iterative manner. More details of SBL and type-II ML estimation can be found in [5]. We provide a flow diagram to compute the posterior mean and covariance of the channel, and the hyperparameters in Fig. 3.

Once we obtain the frequency domain channel estimate $\widehat{\mathbf{H}}^{\mathrm{v}}$, we estimate the support of the sparse vector and the channel coefficients using the hyperparameters obtained using SBL. We refine the noise estimate using the Frobenius norm of the residual $\widetilde{\mathbf{Y}}_{\mathrm{w}} = \mathbf{Y}_{\mathrm{w}} - \mathbf{\Phi}_{\mathrm{w}} \widehat{\mathbf{\Psi}} \widehat{\mathbf{H}}^{\mathrm{v}}$.

F. Denoising

By analyzing the training dataset, we observed that the channel is sparse both in the virtual beamspace and delay domains. We exploited the beamspace sparsity, and obtained the frequency domain channel estimates using MLGS and SBL. And finally, we utilize the delay domain sparsity, and denoise the channel in the delay domain that reduces the mean square error between the original and estimated channels further.

For each subcarrier k, we compute $(\bar{\widehat{\mathbf{A}}}_T \otimes \widehat{\mathbf{A}}_R)\mathbf{H}^{\mathrm{v}}[:,k]$, and reshape it to form k^{th} subcarrier's channel matrix of size $N_r \times N_t$. Then, for each transmit and receive antenna pair, we compute a K-point inverse discrete Fourier transform to obtain a delay-domain channel estimate. We retain the P dominant taps in the delay-domain channel estimate, and set the other $K-P$ taps to 0. We fix P based on the estimated noise variance, and the number of training frames M. The value of P is inversely proportional to $\widehat{\sigma}_n^2$, and the training dataset is used to choose an appropriate P.

REFERENCES

[1] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, 2018.

[2] R. W. Heath Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Areas Commun*, vol. 10, no. 3, pp. 436–453, April 2016.

[3] 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.

[4] D. P. Wipf and B. D. Rao, "Sparse Bayesian Learning for Basis Selection ," *IEEE Trans. on Sig. Proc.*, vol. 52, no. 8, Aug. 2004.

[5] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *J. Mach. Learn. Res.*, vol. 1, 2001.