

Bayesian Learning based Millimeter-Wave Sparse Channel Estimation with Hybrid Antenna Arrays

Mubarak Umar Aminu, Marian Codreanu and Markku Juntti

Centre for Wireless Communications, University of Oulu

P.O. Box 4500, 90014 University of Oulu, Finland

Email: mubarak.aminu@oulu.fi, marian.codreanu@oulu.fi, markku.juntti@oulu.fi

Abstract—We consider the problem of millimeter-wave (mmWave) channel estimation with a hybrid digital-analog two-stage beamforming structure. A radio frequency (RF) chain excites a dedicated set of antenna subarrays. To compensate for the severe path loss, known training signals are beamformed and swept to scan the angular space. Since the mmWave channels typically exhibit sparsity, the channel response can usually be expressed as a linear combination of a small number of scattering clusters. Thereby the number of angles of arrival (AoAs) and angles of departure (AoDs) with significant signal components is limited, and compressive sensing techniques can be leveraged for estimating the channel. In this paper, we investigate two sparse recovery algorithms: a Bayesian and non-Bayesian one. In the Bayesian approach, we invoke the sparse Bayesian learning (SBL) framework, which relies on a 2-layer hierarchical prior model for channel. A highly efficient and fast iterative Bayesian inference method is then applied to the proposed model. The non-Bayesian approach is a LASSO-based approach, where we devise a low complexity solution by adopting alternating directions method of multipliers (ADMM) technique to solve the problem. The efficacy of the proposed algorithms is demonstrated using numerical examples. The Bayesian approach shows improved estimation performance in relation to the non-Bayesian approach.

Index Terms—Millimeter-wave communications, sparse channel estimation, sparse Bayesian learning, compressive sensing.

I. INTRODUCTION

The abundant spectrum resources at the millimeter-wave (mmWave) bands, i.e., 10–100 GHz frequencies, is one core capacity increase enabler for the Fifth Generation (5G) wireless networks [1]. The tiny wavelength of signals at mmWave allows packing of large numbers of antenna elements, making it feasible to realize the needed high beamforming gains to overcome the severe pathloss. However, conventional fully-digital beamforming architectures cannot be directly applied in mmWave communications due to the radio frequency (RF) hardware constraints. Therefore, hybrid beamforming architectures [2], which is a two-stage digital-analog beamforming procedure, have been proposed. They allow the use of large antenna arrays with a limited number of RF chains.

The knowledge of the channel state information is challenging with the narrow beams. The CSI acquisition could be based on an exhaustive beamformer sweeping over all directions or by using a bisection search to estimate the mmWave channel in the angle of arrival (AoA) and angle of departure (AoD) domains. Since the mmWave channel is typically sparse in the angular domain (i.e., it can be

represented as a linear combination of a small number of AoAs and AoDs), compressive sensing (CS) and sparse signal recovery can be leveraged to provide accurate estimates of the corresponding AoAs and AoDs.

The application of CS to mmWave channel estimation has become a hot research topic. Specifically, the CS theory, which includes the ℓ_1 -norm convex constrained optimization [3], [4] and the sparse Bayesian learning (SBL) methods [5], [6] for estimating the wireless channel, has been studied recently. We refer here the sparse channel estimators based on the ℓ_1 -norm convex constrained optimization as the *non-Bayesian* and the SBL as the *Bayesian*. For example, the non-Bayesian estimator, which is based on ℓ_1 -norm constrained optimization, popularly referred to as Least Absolute Shrinkage and Selection Operator (LASSO) was proposed for estimating the channel in a mmWave-hybrid system in [7], [8]. A greedy orthogonal matching pursuit (OMP) algorithm was then proposed to find an approximate sparse solution. The LASSO-based estimator is popular due to its convexity. The Bayesian estimators for estimating the channel in a mmWave-hybrid system have also been recently proposed in [9], [10]. For the Bayesian estimators, a prior probability density function (pdf) of the unknown parameter is specified and the aim is to find a sparse maximum a posteriori (MAP) estimate.

In this paper, we present Bayesian and non-Bayesian sparse channel estimation algorithms for mmWave-hybrid system. The Bayesian approach is a novel application of the fast SBL algorithm [11] for mmWave-hybrid channel estimation. Mishra *et al.* [10] applied the EM algorithm, which is known to suffer from high computational complexity, for the hyperparameter estimation. We use a 2-layer hierarchical prior model. We then introduce a highly accelerated algorithm, which exploits the properties of the marginal likelihood function. The non-Bayesian approach is a LASSO-based estimator where we propose a low complexity solution by adopting alternating directions method of multipliers (ADMM) technique for the reweighted ℓ_2 -norm minimization problem. We point out also that the SBL framework realizes the LASSO-based cost function when a certain parameter prior pdf is chosen. Finally, the simulation results show that both the Bayesian and non-Bayesian algorithms provide good estimates in terms of the normalized mean square error (MSE).

II. CHANNEL MODEL

We consider a multiple-input multiple-output (MIMO) system with one base station (BS) and a single user equipment (UE) as in Fig. 1. The BS is equipped with N_T^{BS} antennas and $N_{\text{RF}}^{\text{BS}}$ radio frequency (RF) chains. Similarly, the UE is equipped with M_R^{UE} antennas and $M_{\text{RF}}^{\text{UE}}$ radio frequency (RF) chains. We assume that $N_{\text{RF}}^{\text{BS}} \leq N_T^{\text{BS}}$, $M_{\text{RF}}^{\text{UE}} \leq M_R^{\text{UE}}$, and narrowband flat-fading channels with L scattering clusters in the environment. The channel matrix can be expressed as

$$\mathbf{H} = \sum_{\ell=1}^L \xi_{\ell} \mathbf{s}_R(\theta_{\ell}) \mathbf{s}_T^H(\phi_{\ell}), \quad (1)$$

where ξ_{ℓ} is the complex channel gain of ℓ th path. The vectors $\mathbf{s}_R(\theta_{\ell}) \in \mathbb{C}^{M_R^{\text{UE}} \times 1}$ and $\mathbf{s}_T(\phi_{\ell}) \in \mathbb{C}^{N_T^{\text{BS}} \times 1}$ denote the antenna array response at the UE and BS, respectively, where $\theta_{\ell} \in [0, 2\pi)$ and $\phi_{\ell} \in [0, 2\pi)$ are AoA and AoD, respectively. Assuming a uniform linear array (ULA), $\mathbf{s}_R(\theta_{\ell})$ and $\mathbf{s}_T(\phi_{\ell})$ are given as

$$\mathbf{s}_R(\theta_{\ell}) = [1, e^{j\mu(\theta_{\ell})}, e^{j2\mu(\theta_{\ell})}, \dots, e^{j(M-1)\mu(\theta_{\ell})}]^T \quad (2)$$

$$\mathbf{s}_T(\phi_{\ell}) = [1, e^{j\mu(\phi_{\ell})}, e^{j2\mu(\phi_{\ell})}, \dots, e^{j(N-1)\mu(\phi_{\ell})}]^T \quad (3)$$

where $\mu(\theta_{\ell}) = \frac{2\pi}{\lambda} d \sin(\theta_{\ell})$, λ is the carrier wavelength and d is the spacing between the antenna elements. Since both θ_{ℓ} and ϕ_{ℓ} are unknown, we use the quantized grid for both AoAs and AoDs by uniformly sampling $[0, 2\pi]$ with Q samples. Neglecting the grid quantization error, the transformed channel model can be expressed as [7]

$$\mathbf{H} = \mathbf{S}(\Theta) \mathbf{H}^v \mathbf{S}(\Phi)^H, \quad (4)$$

where $\mathbf{S}(\Theta)$ and $\mathbf{S}(\Phi)$ contain steering vectors $\mathbf{s}_R(\theta)$ and $\mathbf{s}_T(\phi)$, respectively, for Q resolvable directions ($Q \gg L$), i.e.,

$$\mathbf{S}(\Theta) = [\mathbf{s}_R(0), \mathbf{s}_R(\frac{2\pi}{Q}), \dots, \mathbf{s}_R(\frac{2\pi}{Q}(Q-1))]$$

$$\mathbf{S}(\Phi) = [\mathbf{s}_T(0), \mathbf{s}_T(\frac{2\pi}{Q}), \dots, \mathbf{s}_T(\frac{2\pi}{Q}(Q-1))].$$

Matrix $\mathbf{H}^v \in \mathbb{C}^{Q \times Q}$ contains the channel gains where the (i, j) th entry can be interpreted as the channel gain between the i -th angular bin for the AoA and the j -th angular bin for the AoD. Note that the channel exhibits L multipaths which implies that the channel matrix \mathbf{H}^v will have L dominant elements while the remaining elements are close to zero.

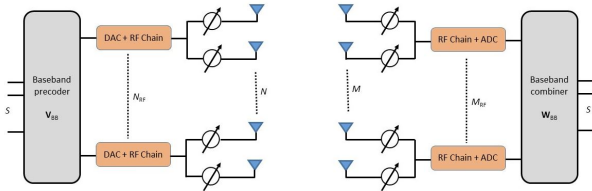


Fig. 1. Hybrid analog-digital transmitter and receiver.

A. Channel Sensing Method

In the channel acquisition phase, the BS uses beamforming vectors $\mathbf{v}_1, \dots, \mathbf{v}_D \in \mathbb{C}^{N_T^{\text{BS}} \times 1}$ to scan D different directions. The UE in turn applies the combining vectors $\mathbf{w}_1, \dots, \mathbf{w}_D \in \mathbb{C}^{M_R^{\text{UE}} \times 1}$ to the received signals. After the UE uses the combining vectors $\mathbf{w}_1, \dots, \mathbf{w}_D$ to detect the signal, the detector output \mathbf{z}_i becomes

$$\mathbf{z}_i = \mathbf{W}^H \mathbf{H} \mathbf{v}_i s_i + \mathbf{W}^H \mathbf{n}_i, \quad (5)$$

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_D] \in \mathbb{C}^{M_R^{\text{UE}} \times D}$ is the combining matrix and $\mathbf{n}_i \in \mathbb{C}^{M_R^{\text{UE}} \times 1}$ is the i.i.d Gaussian noise vector and s_i is the transmitted symbol. Since s_i is known symbol, we can let $s_i = 1$. By stacking all D received vectors $\mathbf{z}_1, \dots, \mathbf{z}_D$, we have

$$\mathbf{y} = [\mathbf{z}_1^T, \dots, \mathbf{z}_D^T]^T = \text{vec}(\mathbf{W}^H \mathbf{H} \mathbf{V}) + \mathbf{n} \quad (6)$$

$$= (\mathbf{V}^T \text{conj}(\mathbf{S}(\Phi)) \otimes \mathbf{W}^H \mathbf{S}(\Theta)) \mathbf{h} + \mathbf{n} \quad (7)$$

where $\text{vec}(\cdot)$ and $\text{conj}(\cdot)$ refer to vectorization and complex conjugate operation respectively, $\mathbf{h} = \text{vec}(\mathbf{H}^v)$, and $\mathbf{n} = [\mathbf{n}_1^T, \dots, \mathbf{n}_D^T]^T$. A more compact representation of (7) is expressed as

$$\mathbf{y} = \mathbf{\Psi} \mathbf{h} + \mathbf{n} \quad (8)$$

where $\mathbf{y} \in \mathbb{C}^{D^2 \times 1}$ is the received vector and $\mathbf{\Psi} = (\mathbf{V}^T \text{conj}(\mathbf{S}(\Phi)) \otimes \mathbf{W}^H \mathbf{S}(\Theta)) \in \mathbb{C}^{D^2 \times Q^2}$ is the sensing matrix. The vector $\mathbf{h} \in \mathbb{C}^{Q^2}$ contains the channel gains and most of the entries are close to zero.

III. BAYESIAN CHANNEL ESTIMATION

Given the signal model in (8), and the noise term \mathbf{n} modeled as an i.i.d Gaussian random vector with some variance σ^2 , we can express the pdf of the signal vector \mathbf{y} , given noise variance σ^2 and \mathbf{h} as

$$p(\mathbf{y} | \mathbf{h}, \sigma^2) = (\pi \sigma^2)^{-N} e^{(-\sigma^{-2} \|\mathbf{y} - \mathbf{\Psi} \mathbf{h}\|_2^2)}, \quad (9)$$

where we set $N = D^2$. A key feature of the SBL framework is the incorporation of a parameterized prior that encourages sparsity in the representation. Invoking a two-layer hierarchical model, we consider the channel \mathbf{h} to be zero-mean Gaussian with pdf specified as

$$p(\mathbf{h} | \alpha) = (\pi)^{-M} \prod_{m=1}^M \alpha_m^{-1} e^{(-\alpha_m^{-1} |h_m|^2)}, \quad (10)$$

where we define $M = Q^2$ and $\alpha = [\alpha_1, \dots, \alpha_M]^T$ is a vector of M hyperparameters. The sparseness of \mathbf{h} is moderated by the hyperparameters (notice that $h_m = 0$ when $\alpha_m = \infty$) whereby individual α_m is independently assigned to each h_m . To complete the specification of the second layer of the hierarchical model, hyperpriors are assigned to the hyperparameters and suitable priors are the Gamma distributions [6]:

$$p(\alpha) = \prod_{i=1}^M \text{Gamma}(\alpha_i | a, b), \quad (11)$$

where $\text{Gamma}(\alpha_i|a, b) = \Gamma(a)^{-1} b^a \alpha^{a-1} e^{-b\alpha}$ with $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ being the gamma function. Choosing the Gamma distribution allows for a closed form Bayesian analysis, since it is the conjugate prior of the Gaussian likelihood function [12, Chapter 2.3.7]. The parameters of the hyperpriors are chosen so that it is uninformative (by setting $a, b \rightarrow 0$ in (11)), i.e., so as to express no preference over any value for the corresponding h_m . Given the priors, the Bayesian inference proceeds by computing the posterior distribution of the channel vector \mathbf{h} conditioned on the received signal vector \mathbf{y} as

$$p(\mathbf{h}, \alpha, \sigma^2 | \mathbf{y}) = p(\mathbf{h} | \mathbf{y}, \alpha, \sigma^2) p(\alpha, \sigma^2 | \mathbf{y}), \quad (12)$$

where $p(\mathbf{h} | \mathbf{y}, \alpha, \sigma^2)$ can be computed analytically as

$$\begin{aligned} p(\mathbf{h} | \mathbf{y}, \alpha, \sigma^2) &= \frac{p(\mathbf{y} | \mathbf{h}, \sigma^2) p(\mathbf{h} | \alpha)}{p(\mathbf{y} | \alpha, \sigma^2)} \\ &= (\pi)^{-M} |\Sigma|^{-1} e^{-(\mathbf{h} - \mu)^H \Sigma^{-1} (\mathbf{h} - \mu)}, \end{aligned} \quad (13)$$

and the covariance and the mean are given respectively as

$$\Sigma = (\sigma^{-2} \Psi^H \Psi + \mathbf{A})^{-1} \quad (15)$$

$$\mu = \sigma^{-2} \Sigma \Psi^H \mathbf{y}, \quad (16)$$

with $\mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_M)$. The focus now turns to the second term on the right-hand side of (12), the hyperparameters posterior $p(\alpha, \sigma^2 | \mathbf{y})$, which is analytically intractable. The Bayesian learning thus becomes the search for the hyperparameter posterior mode via a *type-II maximum likelihood* procedure [6], [11]. Therefore, the estimate of α is sought by maximizing the likelihood function, or equivalently, its logarithm:

$$\begin{aligned} \mathcal{L}(\alpha) &= \log(p(\mathbf{y} | \alpha, \sigma^2)) \\ &= -[N \log \pi + \log |\mathbf{C}| + \mathbf{y}^H \mathbf{C}^{-1} \mathbf{y}], \end{aligned} \quad (17)$$

where $\mathbf{C} = \sigma^2 \mathbf{I} + \Psi \mathbf{A}^{-1} \Psi^H$. Typically, the learning algorithm can proceed by iteratively maximizing (17), i.e.,

$$\arg\max_{\alpha} \mathcal{L}(\alpha) = \arg\min_{\alpha} [\log |\mathbf{C}| + \mathbf{y}^H \mathbf{C}^{-1} \mathbf{y}],$$

concurrent with updating Σ and μ from (15) and (16) hoping it converges to a MAP solution [6]. This iterative process, however, suffers high complexity due to the fact that at each iteration, the covariance matrix (15) has to be computed. It requires a matrix inverse operation of size $M \times M$. An accelerated learning algorithm can instead be derived by exploiting the properties of the maximum likelihood function (17). The central idea is by isolating the effect of each α_m on the marginal likelihood function. In that regards, matrix \mathbf{C} in (17) can be decomposed into two parts

$$\mathbf{C} = \mathbf{C}_{-m} + \alpha_m \psi_m \psi_m^H, \quad (18)$$

where ψ_m is the m -th column vector of the matrix Ψ . Applying the Woodbury's matrix inversion lemma and matrix determinant lemma to \mathbf{C}^{-1} and $|\mathbf{C}|$ in (17), respectively (see [12, Appendix C, eq. C.7 and eq. C.15]), $\mathcal{L}(\alpha)$ can be expressed as

$$\mathcal{L}(\alpha) = \mathcal{L}(\alpha_{-m}) + \ell(\alpha_m), \quad (19)$$

where term $\ell(\alpha_m)$ only involves hyperparameter α_m , given by

$$\ell(\alpha_m) = \log \alpha_m - \log(\alpha_m + s_m) + \frac{|q_m|^2}{\alpha_m + s_m}, \quad (20)$$

with $s_m = \psi_m^H \mathbf{C}_{-m}^{-1} \psi_m$ and $q_m = \psi_m^H \mathbf{C}_{-m}^{-1} \mathbf{y}$. Thus far, the maximization of (17) is reduced to maximization of a simpler function (20). By differentiating (20) and setting it to zero, a unique maximum of $\mathcal{L}(\alpha)$ with respect to α_m is expressed as

$$\alpha_m = \begin{cases} \frac{s_m}{(|q_m|^2 - s_m)} & \text{if } |q_m|^2 > s_m, \\ \infty & \text{if } |q_m|^2 \leq s_m \end{cases} \quad (21)$$

It is relatively straightforward now to learn the channel efficiently by sequentially adding and deleting the columns of the sensing matrix Ψ . The sequential learning algorithm is provided in Algorithm 1. Upon completion of the learning, the channel vector estimate is given as $\hat{\mathbf{h}} = \mu$.

Algorithm 1 Fast Sparse Bayesian Learning.

- 1: Initialize with a single column ψ_m of Ψ and calculate α_m from (21).
 - 2: Compute Σ and μ from (15) and (16) (which are scalars initially), along with q_i and s_i for all columns ψ_i of Ψ .
 - 3: **repeat**
 - 4: Select any candidate column vector ψ_i
 - 5: If $|q_i|^2 > s_i$, and $\alpha_i < \infty$, ψ_i is in the model, update α_i using (21).
 - 6: If $|q_i|^2 > s_i$, and $\alpha_i = \infty$, **add** ψ_i to the model, update α_i using (21).
 - 7: If $|q_i|^2 \leq s_i$, and $\alpha_i < \infty$, **delete** ψ_i from the model, set $\alpha_i = \infty$.
 - 8: Recompute Σ and μ from (15) and (16) (with Ψ contains only those columns currently included in the model), along with q_i and s_i for all columns ψ_i of Ψ .
 - 9: **until** Desired level of convergence has been reached.
-

IV. NON-BAYESIAN CHANNEL ESTIMATION

In this section, we formulate the sparse channel estimation problem from the perspective of the popular LASSO regression. The LASSO approach is attractive, because it is a convex relaxation of the ℓ_0 and it promotes sparse solutions. We begin by casting the sparse channel reconstruction problem as ℓ_2 norm minimization problem [3], [4], [13]:

$$\underset{\mathbf{h}}{\text{minimize}} \quad \|\mathbf{y} - \Psi \mathbf{h}\|_2^2 + \lambda \sum_{m=1}^M \log(|h_m| + \epsilon), \quad (22)$$

where λ is a *non-negative parameter*. Viewing (22) through the lenses of the Bayesian framework, we can see that the same objective function is arrived at with a proper choice of the prior pdf of the channel vector \mathbf{h} . Recall that, according to the Bayesian framework, a prior distribution is designed for \mathbf{h} and then a sparsity promoting maximum a posteriori (MAP) is sought for. Now from (14), we have defined the posteriori

distribution for the channel \mathbf{h} which is proportional to the product of the prior distribution and the likelihood function or

$$p(\mathbf{h}|\mathbf{y}, \alpha, \sigma^2) \propto p(\mathbf{y}|\mathbf{h}, \sigma^2)p(\mathbf{h}|\alpha),$$

where $p(\mathbf{y}|\mathbf{h}, \sigma^2)$ is given in (9). When the prior $p(\mathbf{h})$ is chosen such that $p(\mathbf{h}) \propto \prod_m^M |h_m|^{-2}$ [6], and we compute the logarithm of the posteriori $p(\mathbf{h}|\mathbf{y}, \alpha, \sigma^2)$, we obtain

$$\mathcal{L}(\mathbf{h}) = -\|\mathbf{y} - \Psi\mathbf{h}\|_2^2 - \sigma^2 \sum_m^M \log(|h_m|). \quad (23)$$

It is now clear that maximizing (23) is equivalent to the ℓ_2 norm minimization problem (22). We will now propose an iterative method to approximate the solution to (22) which admits closed form solutions at each step based on the successive convex approximation (SCA), wherein we approximate the concave term in the objective by its first order linear approximation. Note that due to the iterative nature of the relaxed problem a matrix inversion is computed in each iteration, which is computationally expensive. Therefore, we invoke the ADMM algorithm to efficiently solve the problem. By introducing an auxiliary variable $\mathbf{z} \in \mathbb{C}^M$, we recast the new problem as

$$\begin{aligned} \underset{\mathbf{h}, \mathbf{z}}{\text{minimize}} \quad & \|\mathbf{y} - \Psi\mathbf{h}\|_2^2 + \lambda \sum_m \log(|z_m| + \epsilon) \\ \text{subject to} \quad & \mathbf{h} = \mathbf{z}, \end{aligned} \quad (24)$$

where \mathbf{z} and \mathbf{h} are the optimization variables. With the reformulation (24), we derive the augmented Lagrangian in (25) [14], [15] as

$$\begin{aligned} L_\rho(\mathbf{h}, \mathbf{z}, \mathbf{u}) = & \lambda \sum_i \log(|z_i| + \epsilon) + \|\mathbf{y} - \Psi\mathbf{h}\|_2^2 \\ & + \rho \|\mathbf{z} - \mathbf{h} + \mathbf{u}\|_2^2 - \rho \|\mathbf{u}\|_2^2. \end{aligned} \quad (25)$$

The ADMM algorithm thus consists of the iterations

$$\mathbf{h}^{k+1} = \underset{\mathbf{h}}{\text{argmin}} L_\rho(\mathbf{h}, \mathbf{z}^k, \mathbf{u}^k) \quad (26)$$

$$\mathbf{z}^{k+1} = \underset{\mathbf{z}}{\text{argmin}} L_\rho(\mathbf{h}^{k+1}, \mathbf{z}, \mathbf{u}^k) \quad (27)$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{z}^{k+1} - \mathbf{h}^{k+1}, \quad (28)$$

where superscript k is the iteration counter and $\rho > 0$ is a *penalty factor*. Now we derive closed form solutions to the above iterative steps. The update \mathbf{h}^{k+1} in expression (26) is the solution to the following unconstrained optimization problem:

$$\underset{\mathbf{h}}{\text{minimize}} \quad \|\mathbf{y} - \Psi\mathbf{h}\|_2^2 + \rho \|\mathbf{z} - \mathbf{h} + \mathbf{u}\|_2^2, \quad (29)$$

where the solution is derived by taking the gradient of the objective function with respect to \mathbf{h} and equating to zero. The solution is given by the following closed form expression

$$\mathbf{h}^{k+1} = (\Psi^H \Psi + \rho \mathbf{I})^{-1} [\Psi^H \mathbf{y} + \rho (\mathbf{z}^k + \mathbf{u}^k)]. \quad (30)$$

We now derive the update \mathbf{z}^{k+1} which is the solution to the optimization problem

$$\underset{\mathbf{z}}{\text{minimize}} \quad \lambda \sum_m \log(|z_m| + \epsilon) + \rho \|\mathbf{z} - \mathbf{h} + \mathbf{u}\|_2^2. \quad (31)$$

We replace the logarithm term $\sum_m \log(|z_m| + \epsilon)$ with its first order linear approximation. So, let \hat{z}_m be an arbitrary point, the

first order linear approximation of $\sum_m \log(|z_m| + \epsilon)$ around the point \hat{z}_m is given as

$$\sum_m \log(|\hat{z}_m| + \epsilon) + \sum_m (z_m - \hat{z}_m) \frac{1}{(|\hat{z}_m| + \epsilon)}. \quad (32)$$

We can therefore, rewrite (31) as

$$\underset{\mathbf{z}}{\text{minimize}} \quad \lambda \sum_m z_m \left(\frac{1}{(|\hat{z}_m| + \epsilon)} \right) + \rho \|\mathbf{z} - \mathbf{h} + \mathbf{u}\|_2^2 \quad (33)$$

where we have dropped the constant term $\lambda \left(\sum_m \log(|\hat{z}_m| + \epsilon) + \sum_m \hat{z}_m \left(\frac{1}{(|\hat{z}_m| + \epsilon)} \right) \right)$, since it does not affect the solution to the problem. The update \mathbf{z}^{k+1} can be derived by taking the gradient of the objective function in (33) with respect to \mathbf{z} and equating to zero. The closed form solution can be expressed as

$$\mathbf{z}^{k+1} = \mathbf{h}^{k+1} - \mathbf{u}^k - 2\lambda/\rho \left(\frac{1}{(|\hat{\mathbf{z}}|_m + \epsilon)} \right). \quad (34)$$

We now summarize the ADMM algorithm to solve problem (22) in Algorithm 2.

Algorithm 2 ADMM algorithm.

- 1: Initialization $\mathbf{z}^k, \mathbf{u}^k$ and $k = 0$.
 - 2: **repeat**
 - 3: Update \mathbf{h}^{k+1} by using expression (30)
 - 4: Update \mathbf{z}^{k+1} by using expression (34).
 - 5: Update \mathbf{u}^{k+1} by using expression (28).
 - 6: Set $k = k + 1$.
 - 7: **until** Desired level of convergence has been reached.
-

V. NUMERICAL RESULTS

We test the efficacy of the proposed algorithms using numerical examples considering a hybrid partially connected MIMO structure as depicted in Fig. 1, where each RF chain is connected to a dedicated antenna subarray. The training beamformers are designed similar to the approach in [7] taking into consideration of the partial connection and the power constraint, i.e., the transmit beamformer $\|\mathbf{v}\|^2 = \|\mathbf{v}_{\text{RF}}\mathbf{v}_{\text{BB}}\|^2 = 1$. We quantify the channel estimation performance using normalized MSE (NMSE), defined as $\text{NMSE} = \frac{\|\hat{\mathbf{h}} - \mathbf{h}\|_2^2}{\|\mathbf{h}\|_2^2}$. For each signal-to-noise ratio (SNR) value, we generated 100 channel and noise realizations. The AoA/AoD quantization grid is set to $Q = 2^5$, and the BS and the UE each search in $D = 20$ directions. The sensing matrix Ψ is of size 400×1024 and we set $L = 4$ scattering clusters (i.e., the expected channel sparsity of 0.0039). Fig. 2 shows the NMSE vs. SNR plot for the Bayesian and non-Bayesian algorithms. It can be seen that the Bayesian estimator shows an improved performance gain of around 2 dB. It is important to mention that the choice of penalty parameter ρ in (30) affects the performance of the non-Bayesian estimator. For our simulations, we set $\rho = 10$ and $\lambda = \sigma^2$.

Fig. 3 shows the angle domain representation of the channel for a single realization at SNR = 3 dB with the same setting as

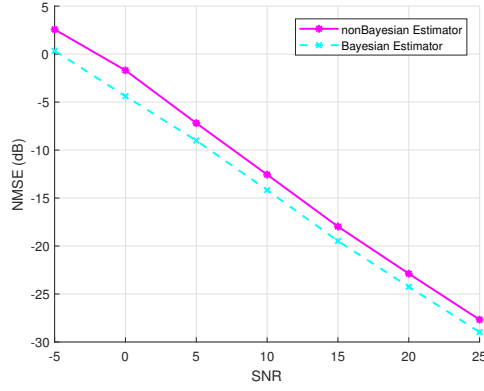


Fig. 2. NMSE estimation performance of the proposed algorithms for a system with $N_T^{BS} = M_R^{UE} = 8$, and $N_{RF}^{BS} = M_{RF}^{UE} = 2$.

that in Fig. 2. As seen in the figure, both the Bayesian and non-Bayesian algorithms can estimate the channel quite accurately: the 4 clusters can be easily identified. The approximate nature of the sparsity of the mmWave channel is also visible in Fig. 3.

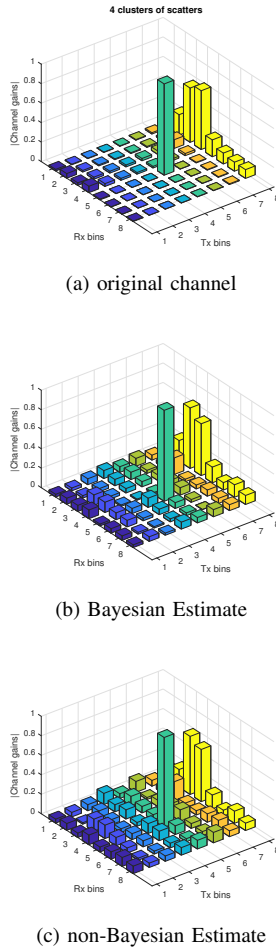


Fig. 3. Angle domain representation of the channel for single trial at SNR = 3 dB. Subfigure (a) shows the actual channel, (b) shows the Bayesian estimate, and (c) shows the non-Bayesian estimate.

VI. CONCLUSION

We have considered sparse channel estimation for millimeter-wave systems using partially connected hybrid digital-analog hybrid transceiver architecture. We presented two novel channel estimation algorithms. The first is a Bayesian based estimator, which relies on a 2-layer hierarchical prior model. A highly efficient and accelerated Bayesian inference method was then introduced and applied to the model. The second one is a non-Bayesian based estimator which was formulated as a reweighted ℓ_2 norm minimization problem. An efficient and low complexity algorithm based on the alternating directions method of multipliers was proposed to solve the problem. The proposed algorithms were validated using numeral examples. The results show good estimation performance. In our future work, we plan to study the performance also with larger arrays structures and hardware accelerated computation.

REFERENCES

- [1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [2] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO Precoding and Combining Solutions for Millimeter-Wave Systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, 2014.
- [3] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing Sparsity by Reweighted ℓ_1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.
- [4] S. J. Wright, R. D. Nowak, and M. A. Figueiredo, "Sparse Reconstruction by Separable Approximation," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, 2009.
- [5] D. P. Wipf and B. D. Rao, "Sparse Bayesian Learning for Basis Selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.
- [6] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, vol. 1, no. Jun, pp. 211–244, 2001.
- [7] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.
- [8] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370–2386, 2016.
- [9] R. T. Suryaprakash, M. Pajovic, K. J. Kim, and P. Orlik, "Millimeter Wave Communications Channel estimation via Bayesian Group Sparse Recovery," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3406–3410, 2016.
- [10] A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, "Sparse Bayesian Learning-based Channel Estimation in Millimeter Wave Hybrid MIMO Systems," *IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2017.
- [11] M. E. Tipping and A. C. Faul, "Fast Marginal Likelihood Maximisation for Sparse Bayesian Models." Citeseer, 2003.
- [12] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [13] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.
- [14] L. Li, X. Wang, and G. Wang, "Alternating Direction Method of Multipliers for Separable Convex Optimization of Real Functions in Complex Variables," *Mathematical Problems in Engineering*, 2015.
- [15] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.