Article:

You, Y and Zhang, L (2020) Bayesian Matching Pursuit Based Channel Estimation for Millimeter Wave Communication. IEEE Communications Letters, 24 (2). pp. 344-348. ISSN 1089-7798

https://doi.org/10.1109/LCOMM.2019.2953706

Bayesian Matching Pursuit Based Channel Estimation for Millimeter Wave Communication

You You, *Student Member, IEEE,* and Li Zhang, *Senior Member, IEEE*

***Abstract*—Hybrid precoding is considered as a solution to reduce the high power consumption caused by devices operating at radio frequency (RF) in millimeter wave (mmWave) communication. For hybrid precoding, the channel state information (CSI) is critical but hard to obtain because of the analog precoding at RF and the large number of antennas. mmWave channel has been proved to be sparse by real-world experiments. Compressive sensing (CS) methods can be applied to the channel estimation to decrease complexity. However, there is a distinct performance gap between the estimation of the existing CS methods with or without given sparsity pattern (SP). In this letter, a new method based on Bayesian matching pursuit(BMP) idea is proposed to improve sparse channel estimation performance. We make appropriate assumptions according to the characteristics of mmWave channel. We select a set of candidate SPs with high posterior probabilities to estimate CSI. Numerical simulation shows that our proposed method has significantly improved channel estimation performance with acceptable complexity compared to existing methods including orthogonal matching pursuit, sparse Bayesian learning and Bayesian compressive sensing.**

I. Introduction

THE large amount available spectrum at millimeter wave (mmWave) enables the 5G network to meet ever-growing data rate demands and tackle the exponential increase traffic volumes [1]. Thanks to the short wave length, massive MIMO can be equipped at both base station (BS) and mobile station (MS) to overcome the huge propagation loss in mmWave communication. However, full digital precoding as in microwave system requires big number of radio frequency (RF) chains , which is impractical for mmWave system because of their high power consumption [2]. Therefore, a hybrid MIMO architecture consisting of an analog beamformer cascaded with a digital processor is proposed. It reduces the amount of RF chains without compromising too much beamforming performance [3].

Channel state information (CSI) is crucial to the design of precoding and combining in mmWave system. The new constraints on the hardware of hybrid architecture and the huge number of antennas make channel estimation a challenging problem. Due to the sparsity of mmWave channel [4], CSI can be described by a limited number of angle of arrive (AoA), angle of departure (AoD) and path gains. Compressive sensing (CS) theory [5] and virtual angle representation [6] are widely used to solve channel estimation problem as a sparse signal recovery problem. In the sparse signal, the set of locations of nonzero elements is called sparsity pattern (SP) which represents the AoDs/AoAs of corresponding nonzero paths. The values of the nonzero elements represent the corresponding path gains.

There have been many works on the application of CS to mmWave channel estimation. They can be divided into close-loop and open-loop. [7] and [8] are close-loop beam training-based methods, which use multistage process to avoid exhaustive search. However, close-loop method is difficult to be applied to outdoor channel, because limited transmitted power prevents the use of wide beam. An alternative way is to apply the open-loop methods which can decrease the feedback overhead and use a fixed beam width. Open-loop methods include non Bayesian based algorithms such as orthogonal matching pursuit (OMP) [9] and Bayesian based algorithms such as sparse Bayesian learning (SBL) [10] and Bayesian compressive sensing (BCS) [11]. OMP is an iterative algorithm that finds the sub-optimal solution. The nonzero locations (SP) of CSI correspond to the columns of sensing matrix which are highly correlated with the received signal. Bayesian based method makes appropriate statistic assumption and apply estimation techniques to identify the desired sparse solution. Specifically, the SBL adopts a Bayesian framework with each element following independent, zero-mean, Gaussian distribution with unknown variance which are assigned the Gamma conjugate prior as hyperpriori. Expectation maximization (EM) method is utilized to computes a Maximum A Posteriori (MAP) estimate. BCS is another Bayesian method, instead of applying EM to calculate MAP estimate, a more efficient implementation has been derived by analyzing the properties of the marginal likelihood function. It estimates CSI through maximizing the marginal likelihood. All grid based CS algorithms have off-grid error. [12], [13] and [14] propose methods to mitigate this error for OMP, l_1-norm minimization and SBL respectively. However, for mmWave channel estimation, there still have a distinct gap compared with that with known SP especially at low SNRs as shown in [12], [13]. It indicates that, even without off-gird error, SP estimation method needs to be further enhanced. So this paper focuses on improving mmWave channel estimation performance using Bayesian matching pursuit (BMP) idea. Numerical simulations demonstrate that the estimation performance of the proposed method outperforms the existing methods with affordable complexity.

To the best of our knowledge, this letter is the first paper to apply the BMP [15] to mmWave channel estimation. We propose a method using 'virtual sparsity' to apply BMP without known sparsity at a low complexity. We make appropriate assumptions according to the characteristics of mmWave channel and select a set of candidate SPs with significant posterior probabilities for minimum mean square error (MMSE) channel estimation.

II. System Model

We consider a single user hybrid MIMO system shown in Fig. 1, where the BS and MS are equipped with N_T and N_R antennas. Both BS and MS have N_{RF} RF chains ($N_{RF} \le \min(N_T, N_R)$).

In the channel estimation stage, BS uses pilot beam training vectors $\{\mathbf{f}_m \in \mathbb{C}^{N_T \times 1} : m = 1, \ldots, N_T^{Beam}\}$ ($N_T^{Beam} \le N_T$) to scan N_T^{Beam} different directions successively. The pilot beams are received by N_R^{Beam} ($N_R^{Beam} \le N_R$) combining vectors $\{\mathbf{w}_n \in \mathbb{C}^{N_R \times 1} : n = 1, \ldots, N_R^{Beam}\}$ ($N_R^{Beam} \le N_R$) at MS. The received signal for the mth pilot beam is given by

$$\mathbf{y}_m = \mathbf{W}^H \mathbf{H} \mathbf{f}_m x_p + \mathbf{W}^H \mathbf{n}_m, \tag{1}$$

where x_p is the transmitted pilot symbol. $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_{N_R^{Beam}}] \in \mathbb{C}^{N_R \times N_R^{Beam}}$ is the combining matrix at MS. $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ represents the channel matrix, and $\mathbf{n}_m \in \mathbb{C}^{N_R \times 1}$ is the i.i.d Gaussian noise vector. Collecting $\mathbf{y}_m$ for $m \in \{1, \ldots, N_T^{Beam}\}$, we get

$$\begin{aligned} \mathbf{Y} &= \mathbf{W}^H \mathbf{H} \mathbf{F} \mathbf{X} + \mathbf{N} \\ &= \sqrt{P_t} \mathbf{W}^H \mathbf{H} \mathbf{F} + \mathbf{N} \end{aligned} \tag{2}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_{N_T^{Beam}}] \in \mathbb{C}^{N_R^{Beam} \times N_T^{Beam}}$, $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_{N_T^{Beam}}] \in \mathbb{C}^{N_T \times N_T^{Beam}}$ and $\mathbf{N} = [\mathbf{W}^H \mathbf{n}_1, \ldots, \mathbf{W}^H \mathbf{n}_{N_R^{Beam}}] \in \mathbb{C}^{N_R^{Beam} \times N_T^{Beam}}$ is the noise matrix. $\mathbf{X} \in \mathbb{C}^{N_T^{Beam} \times N_T^{Beam}}$ is a diagonal matrix with x_p on its diagonal. We assume identical pilot symbols so that $\mathbf{X} = \sqrt{P_t} \mathbf{I}_{N_T^{Beam}}$ where P_t is the pilot signal power.

The mmWave channel can be approximated by a geometric channel mode with L scatters due to its limited scattering feature. Each scatterer contributes only one path of propagation between BS and MS. The channel matrix can be written as

$$\mathbf{H} = \sqrt{\frac{N_T N_R}{L}} \sum_{\ell=1}^{L} \alpha_\ell \mathbf{a}_R(\theta_\ell^r) \mathbf{a}_T^H(\theta_\ell^t), \tag{3}$$

where α_ℓ is the complex gain of the l-th path, θ_l^r and θ_l^t are the AoA and AoD of the l-th path, respectively. $\mathbf{a}_T(\theta_l^t)$ and $\mathbf{a}_R(\theta_l^r)$ are array response vector for BS and MS. Assuming that we use N_T and N_R uniform linear array (ULA), $\mathbf{a}_T(\theta_l^t)$ and $\mathbf{a}_R(\theta_l^r)$ can be given by

$$\begin{aligned} \mathbf{a}_T(\theta_l^t) &= [1, e^{-j2\pi \frac{d}{\lambda} \cos\theta_l^t}, \ldots, e^{-j2\pi \frac{d}{\lambda} \cos\theta_l^t (N_T - 1)}]^T \\ \mathbf{a}_R(\theta_l^r) &= [1, e^{-j2\pi \frac{d}{\lambda} \cos\theta_l^r}, \ldots, e^{-j2\pi \frac{d}{\lambda} \cos\theta_l^r (N_R - 1)}]^T \end{aligned} \tag{4}$$

where d denotes the antenna spacing, λ denotes the wavelength of operation. In this letter, we consider $d = \frac{\lambda}{2}$. The channel gains $\{\alpha_\ell\}_{\ell=1}^L$ are modeled by i.i.d. random variables with distribution $\mathcal{CN}(0, \sigma^2)$. The AoAs and AoDs are modeled by a Laplacian distribution whose mean is uniformly distributed over $[0, \pi)$, and angular standard deviation is σ_{AS}.

To apply CS techniques to channel estimation, virtual channel representation is used. Specifically, we assume that all the angles fall onto a set of discrete angles called grid. In this letter, we choose uniform grid as $[0, \frac{\pi}{G-1}, \frac{2\pi}{G-1}, \ldots, \frac{\pi(G-1)}{G-1}]$, and $G \gg L$ to achieve the desired resolution. Using discrete angle grid, the channel matrix $\mathbf{H}$ in (3) can be approximated as

$$\mathbf{H} \cong \mathbf{A}_R \mathbf{H}_b \mathbf{A}_T^H, \tag{5}$$

where $\mathbf{A}_R = [\mathbf{a}_R(0), \ldots, \mathbf{a}_R(\frac{\pi}{G-1}), \ldots, \mathbf{a}_R(\frac{\pi(G-1)}{G-1})] \in \mathbb{C}^{N_R \times G}$, $\mathbf{A}_T = [\mathbf{a}_T(0), \ldots, \mathbf{a}_T(\frac{\pi}{G-1}), \ldots, \mathbf{a}_T(\frac{\pi(G-1)}{G-1})] \in \mathbb{C}^{N_T \times G}$ and $\mathbf{H}_b \in \mathbb{C}^{G \times G}$ is a L-sparse channel gain matrix. The virtual channel representation is not exactly equal to the real channel matrix $\mathbf{H}$ because of the quantized grid error as the simulation results in [12].

III. Formulation of mmWave Channel Estimation Problem

Considering the system model in (2) and channel model in (5), the mmWave channel estimation problem can be formulated as a sparse signal recovery problem by vectorizing $\mathbf{Y}$ in (2). Using property of Khatri-Rao product $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \cdot \text{vec}(\mathbf{B})$ for $\mathbf{Y}$ and $\mathbf{H}$, we can get

$$\begin{aligned} \mathbf{y}_v =& \sqrt{P} (\mathbf{F}^T \otimes \mathbf{W}^H) \cdot \text{vec}(\mathbf{H}) + \text{vec}(\mathbf{N}) \\ =& \sqrt{P} (\mathbf{F}^T \otimes \mathbf{W}^H) \text{vec}(\mathbf{A}_R \mathbf{H}_b \mathbf{A}_T^H) + \mathbf{n}_Q \\ =& \sqrt{P} (\mathbf{F}^T \otimes \mathbf{W}^H) \mathbf{A}_D \mathbf{h} + \mathbf{n}_Q \\ =& \mathbf{Q} \cdot (\mathbf{h}) + \mathbf{n}_Q, \end{aligned} \tag{6}$$

where $\mathbf{y}_v \in \mathbb{C}^{M \times 1}$ is the vectorized received signal where $M = N_T^{Beam} N_R^{Beam}$ is the measurement dimension. $\mathbf{A}_D = \mathbf{A}_T^* \otimes \mathbf{A}_R$ is an $N_T N_R \times G^2$ dictionary matrix that consists of the G^2 column vectors of the form $\mathbf{a}_T^H(\theta_u) \otimes \mathbf{a}_R(\theta_v)$, with θ_u and θ_v, the uth and vth points, respectively, of the angle uniform grid. $\mathbf{h} = \text{vec}(\mathbf{H}_b)$ represents the path gains of the corresponding quantized directions. $\mathbf{h}$ is an $N \times 1$ vector where $N = G^2$ is the virtual channel dimension. $\mathbf{Q} = \sqrt{P} (\mathbf{F}^T \otimes \mathbf{W}^H) \mathbf{A}_D \in \mathbb{C}^{M \times N}$ is the sensing matrix. (6) is a sparse signal recovery problem as $\mathbf{h}$ has only L non-zero elements and $L \ll N$. Compressive sensing(CS) methods including OMP [9], SBL [10] and BCS [11] can be leveraged to recover $\mathbf{h}$ from noisy received signal $\mathbf{y}_v$.

As introduced in Section I. All theses algorithms aim to find the most likely SP, which may not be the most accurate one. In contrast to the MAP estimator, Minimum Mean-Squared-Error (MMSE) uses a fusion of SPs to form its result. Thus, In this letter, we propose to work with a mixture of chosen candidate SPs based on posterior possibility with appropriate assumption.

IV. Proposed Bayesian Matching Pursuit Method for mmWave Channel Estimation

A. Assumptions for mmWave channel

To apply the BMP to estimate the mmWave channel, we need to make appropriate statistic assumptions according to the characteristics of mmWave channel. The noise $\mathbf{n}_Q$ in (6) is assumed to be white circular Gaussian with variance σ^2, i.e., $\mathbf{n}_Q \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$. $\{h_n\}_{n=0}^N$ are the elements in sparse vector $\mathbf{h}$. We assume that $\{h_n\}_{n=0}^N$ are drawn from T specific Gaussian distribution. In this application, simulations shows that larger T provides the same performance but with higher complexity. Therefore, we chose $T = 2$. $s_n = t \in \{0, 1\}$

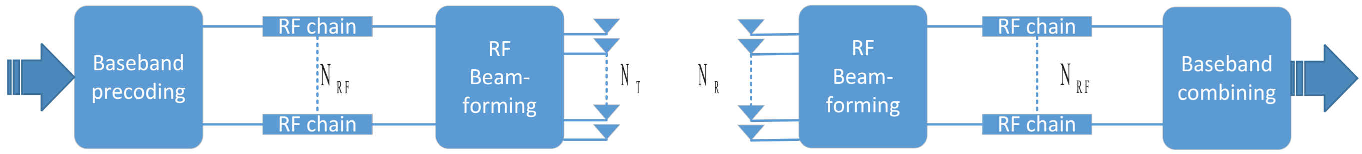

Fig. 1: Hybrid massive MIMO architecture for mmWave communication .

is used as a mixture parameter to index the component distribution. When $s_n = 0$, $(\mu_0, \sigma_0^2) = (0, 0)$ is set to make sure that $h_n = 0$. When $s_n = 1$, $(\mu_1 = 0, \sigma_1^2 = 100P)$ is set to indicate an active non-zero coefficient where P is the power of the received signal. It is the simplest way to represent a sparse signal when we don't know the mean of the nonzero values. We set 100P as the variance for nonzero element. Because we assume zero mean for non-zero elements, it is hard to distinguish them from non-active elements. A relative large variance can improve the accuracy. Our simulation based analysis shows that variance larger than 100P would not improve performance further in our application. So we set 100P as the variance for nonzero elements. $\{s_n\}_{n=0}^{N-1}$ are treated as i.i.d random variables as $\Pr\{s_n = t\} = \lambda_t$ $(0 < \lambda_t \leq 1)$. λ_t is the probability that the value follows Gaussian distribution indexed by $s_n = t$. We make $\sum_{t=1} \lambda_t \ll 1$ to ensure the sparsity. Considering $\mathbf{h} = [h_0, \ldots, h_{N-1}]^T$ and $\mathbf{s} = [s_0, \ldots, s_{N-1}]^T$, the priors can be written as

$$\mathbf{h} \mid \mathbf{s} \sim \mathcal{CN}(\boldsymbol{\mu}(\mathbf{s}), \mathbf{R}(\mathbf{s})), \tag{7}$$

where $[\boldsymbol{\mu}(\mathbf{s})]_n = \mu_{s_n}$ and $\mathbf{R}(\mathbf{s})$ has diagonal $[\mathbf{R}(\mathbf{s})]_{n,n} = \sigma_{s_n}^2$. Considering (6), the channel vector $\mathbf{h}$ and the received signal $\mathbf{y}_v$ are joint Gaussian conditioned on the mixture parameters $\mathbf{s}$ as

$$\begin{bmatrix} \mathbf{y}_v \\ \mathbf{h} \end{bmatrix} \Bigg| \mathbf{s} \sim \mathcal{CN}\left(\begin{bmatrix} \mathbf{Q}\boldsymbol{\mu}(\mathbf{s}) \\ \boldsymbol{\mu}(\mathbf{s}) \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Phi}(\mathbf{s}) & \mathbf{Q}\mathbf{R}(\mathbf{s}) \\ \mathbf{R}(\mathbf{s})\mathbf{Q}^H & \mathbf{R}(\mathbf{s}) \end{bmatrix} \right), \tag{8}$$

where

$$\boldsymbol{\Phi}(\mathbf{s}) \triangleq \mathbf{Q}\mathbf{R}(\mathbf{s})\mathbf{Q}^H + \sigma^2 \mathbf{I}_M. \tag{9}$$

B. MMSE Coefficient Estimation

For channel estimation, MMSE estimate of $\mathbf{h}$ from $\mathbf{y}_v$ is

$$\hat{\mathbf{h}}_{\text{mmse}} \triangleq \mathrm{E}\{\mathbf{h}|\mathbf{y}_v\} = \sum_{\mathbf{s}\in\mathbf{S}} p(\mathbf{s}|\mathbf{y}_v)\mathrm{E}\{\mathbf{h}|\mathbf{y}_v, \mathbf{s}\}. \tag{10}$$

From (8) it is straightforward [16] to obtain

$$\mathrm{E}\{\mathbf{h}|\mathbf{y}_v, \mathbf{s}\} = \boldsymbol{\mu}(\mathbf{s}) + \mathbf{R}(\mathbf{s})\mathbf{Q}^H \boldsymbol{\Phi}(\mathbf{s})^{-1}\big(\mathbf{y}_v - \mathbf{Q}\boldsymbol{\mu}(\mathbf{s})\big). \tag{11}$$

We store the set of all possible SPs as $\mathbf{S}$. If we know all possible 2^N $(\{0,1\}^N)$ posterior probability $p(\mathbf{s}|\mathbf{y}_v)_{\mathbf{s}\in\mathbf{S}}$, (10) can be calculated. But it is impractical to compute all possible 2^N posterior probability $p(\mathbf{s}|\mathbf{y}_v)_{\mathbf{s}\in\mathbf{S}}$. Note that, the size of $\mathbf{S}_\Omega$ which includes the SPs with non-negligible posterior probability $p(\mathbf{s}|\mathbf{y}_v)_{\mathbf{s}\in\mathbf{S}_\Omega}$ can be small and practical to compute because of the sparsity. Using only the dominant SPs in $\mathbf{S}_\Omega$ yields the approximate MMSE estimate

$$\hat{\mathbf{h}}_{\text{ammse}} \triangleq \mathrm{E}\{\mathbf{h}|\mathbf{y}_v\} = \sum_{\mathbf{s}\in\mathbf{S}_\Omega} p(\mathbf{s}|\mathbf{y}_v)\mathrm{E}\{\mathbf{h}|\mathbf{y}_v, \mathbf{s}\}. \tag{12}$$

The primary challenge in the computation of (12) is to obtain $\mathbf{S}_\Omega$ to calculate $p(\mathbf{s}|\mathbf{y}_v)$ and $\boldsymbol{\Phi}(\mathbf{s})^{-1}$. So, we first leverage a fast method to search for $\mathbf{S}_\Omega$.

C. Search for Dominant SPs

We search for $\mathbf{S}_\Omega$ by selecting $\mathbf{s} \in \mathbf{S}$ with the significant posterior probability $p(\mathbf{s}|\mathbf{y}_v)$. According to Bayesian rule, the posterior probability can be written as

$$p(\mathbf{s}|\mathbf{y}_v) = \frac{p(\mathbf{y}_v|\mathbf{s})p(\mathbf{s})}{\sum_{\mathbf{s}'\in\mathbf{S}} p(\mathbf{y}_v|\mathbf{s}')p(\mathbf{s}')}, \tag{13}$$

where $p(\mathbf{s}|\mathbf{y}_v)$ are equal to $p(\mathbf{y}_v|\mathbf{s})p(\mathbf{s})$ up to a scale. For convenience, we work in logarithm domain and define $\alpha(\mathbf{s}, \mathbf{y}_v)$ as SP selection metric:

$$\begin{aligned} \alpha(\mathbf{s}, \mathbf{y}_v) &\triangleq \ln p(\mathbf{y}_v|\mathbf{s})p(\mathbf{s}) \\ &= \ln p(\mathbf{y}_v|\mathbf{s}) + \sum_{n=0}^{N-1} \ln p(s) \\ &= -\big(\mathbf{y}_v - \mathbf{Q}\boldsymbol{\mu}(\mathbf{s})\big)^H \boldsymbol{\Phi}(\mathbf{s})^{-1}\big(\mathbf{y}_v - \mathbf{Q}\boldsymbol{\mu}(\mathbf{s})\big) \\ &\quad - \ln\det\big(\boldsymbol{\Phi}(\mathbf{s})\big) - M \ln \pi + \sum_{n=0}^{N-1} \ln \lambda_{s_n}. \end{aligned} \tag{14}$$

The significant $p(\mathbf{s}|\mathbf{y}_v)$ corresponds to significant value of $\alpha(\mathbf{s}, \mathbf{y}_v)$. So we search $\mathbf{S}_\Omega$ based on metric $\alpha(\mathbf{s}, \mathbf{y}_v)$ using non-exhaustive tree search method.

The search starts with $\mathbf{s} = \mathbf{0}$. In the first stage, we change only one element to non-zero in $\mathbf{s}$ which corresponds to N different 'one element active' SPs. We store all these possible SPs as $\mathbf{S}^{(1)}$ and calculate the metric $\alpha(\mathbf{s})$ for them. We choose D SPs with largest metrics and store them as $\mathbf{S}_\Omega^{(1)}$. In the second step, we activate one more element from the D chosen SPs in $\mathbf{S}_\Omega^{(1)}$ so that we have $(N-1) + (N-2) + ... + (N-D)$ possible 'two element active' SPs in $\mathbf{S}^{(2)}$. Then D 'two element active' SPs with largest metrics among these $(ND - \frac{(1+D)D}{2})$ possible SPs are chosen and stored in $\mathbf{S}_\Omega^{(2)}$. We do this procedure J times to get D 'J element active' SPs with largest posterior possibility as candidate SPs.

The value of D is fixed and chosen as 5, because our simulation shows the benefits of increasing D diminish quickly for $D > 5$. The value of J is determined by the sparsity of the channel. However, we don't know the real sparsity of mmWave channel. So we define a virtual sparsity $L^{'}$. We choose an arbitrary small integer from 2 to 5 as the virtual sparsity because the real sparsity for mmWave channel is generally less than 10. And we calculate λ_1 as: $L^{'}/N$. $L^{'}$ follows Binomial (N, λ_1) distribution. It is common to use the approximation $L^{'} \sim \mathcal{N}(N\lambda_1, N\lambda_1(1-\lambda_1))$, in which case $Pr(L^{'} > J) = \frac{1}{2} erfc(\frac{J - N\lambda_1}{\sqrt{2N\lambda_1(1-\lambda_1)}})$. We choose

$J = \lceil \text{erfc}^{-1}(2J_0)\sqrt{2N\lambda_1(1-\lambda_1)} + N\lambda_1 \rceil$ where J_0 is a very small target value of $\Pr\{L^{'} > J\}$. The use of pre-determined virtual sparsity provide superior performance with low complexity without the need to know real sparsity.

Algorithm 1 Search via Bayesian Matching Pursuit

$\alpha^{\text{root}} = -\frac{1}{\sigma^2}\|\mathbf{y}_v\|_2^2 - M\ln\sigma^2 - M\ln\pi + N\ln\lambda_0$
for $n = 0 : N-1$ **do**
  $\mathbf{c}_n^{\text{root}} = \frac{1}{\sigma^2}\mathbf{q}_n$, $\beta_n^{\text{root}} = \sigma_1^2(1+\sigma_1^2\mathbf{q}_n^H\mathbf{c}_n^{\text{root}})^{-1}$
  for $t = 1 : T-1$ **do**
    $\alpha_{n,t}^{\text{root}} = \alpha^{\text{root}} + \ln\frac{\beta_n^{\text{root}}}{\sigma_1^2} + \beta_n^{\text{root}}|\mathbf{c}_n^{\text{root}H}\mathbf{y}_v + \frac{\mu_t}{\sigma_1^2}|^2 - \frac{|\mu_t|^2}{\sigma_1^2} + \ln\frac{\lambda_1}{\lambda_0}$
  end for
end for
for $d = 1 : D$ **do**
  n=[], **p**=[], $\hat{\mathbf{s}}^{(d,0)} = \mathbf{0}$, $\mathbf{z} = \mathbf{y}_v$
  for $n = 0 : N-1$ **do**
    $\mathbf{c}_n = \mathbf{c}_n^{\text{root}}$, $\beta_n = \beta_n^{\text{root}}$
    for $t = 1 : T-1$ **do**
      $\alpha_{n,t} = \alpha_{n,t}^{\text{root}}$
    end for
  end for
  for $j = 1 : J$ **do**
    $(n_\Omega, t_\Omega) = (n,t)$ indexing the largest element in $\{\alpha_{n,t}\}_{n=0:N-1}^{t=1:T-1}$ which leads to an as-of-yet unexplored node.
    $\alpha^{(d,j)} = \alpha_{n_\Omega,t_\Omega}$, $\hat{\mathbf{s}}^{(d,j)} = \hat{\mathbf{s}}^{(d,j-1)} + t_\Omega\boldsymbol{\delta}_\Omega$
    $\mathbf{n} = [\mathbf{n}, n_\Omega]$, $\mathbf{t} = [\mathbf{t}, t_\Omega]$, $\mathbf{z} = \mathbf{z} - \mathbf{q}_{n_\Omega}\mu_\Omega$
    for $n = 0 : N-1$ **do**
      $\mathbf{c}_n = \mathbf{c}_n - \beta_{n_\Omega}\mathbf{c}_{n_\Omega}\mathbf{c}_{n_\Omega}^H\mathbf{q}_n$, $\beta_n = \sigma_1^2(1+\sigma_1^2\mathbf{q}_n^H\mathbf{c}_n)^{-1}$
      for $t = 1 : T-1$ **do**
        $\alpha_{n,t} = \alpha^{(d,j)} + \ln\frac{\beta_n}{\sigma_1^2} + \beta_n|\mathbf{c}_n^H\mathbf{z} + \frac{\mu_t}{\sigma_1^2}|^2 - \frac{|\mu_t|^2}{\sigma_1^2} + \ln\frac{\lambda_1}{\lambda_0}$
      end for
    end for
    $\hat{\mathbf{h}}^{(d,j)} = \sum_{k=1}^{j}\boldsymbol{\delta}_{[\mathbf{n}]_k}[\sigma_1^2\mathbf{c}_{[\mathbf{n}]_k}^H\mathbf{z} + \mu_{[\mathbf{t}]_k}]$
  end for
end for

D. Fast Metric Update

In the above search, metric α needs to be calculated for each possible SP. We adopted a fast metric metric update method [15] to reduce the computational complexity.

For the case that $[\mathbf{s}]_n = t$ and $[\mathbf{s}']_n = t'$, where $\mathbf{s}$ and $\mathbf{s}'$ are identical except for the nth coefficient. For brevity, we use $\mu_{t',t} \triangleq \mu_{t'} - \mu_t$, $\sigma_{t',t}^2 \triangleq \sigma_{t'}^2 - \sigma_t^2$ and $\Delta_{n,t'}(\mathbf{s},\mathbf{y}_v) \triangleq \alpha(\mathbf{s}',\mathbf{y}_v) - \alpha(\mathbf{s},\mathbf{y}_v)$ below. Note that the root node ($\mathbf{S}_\Omega^{(0)} = \mathbf{0}$) has the following metric

$$\alpha(\mathbf{0},\mathbf{y}_v) = -\frac{1}{\sigma^2}\|\mathbf{y}_v\|_2^2 - M\ln\sigma^2 - M\ln\pi + N\ln\lambda_0. \quad (15)$$

To derive the fast metric update, starting with property

$$\boldsymbol{\Phi}(\mathbf{s}') = \boldsymbol{\Phi}(\mathbf{s}) + \sigma_{t',t}^2\mathbf{q}_n\mathbf{q}_n^H, \quad (16)$$

where $\mathbf{q}_n$ is the nth column of $\mathbf{Q}$. The matrix inversion lemma implies

$$\boldsymbol{\Phi}(\mathbf{s}')^{-1} = \boldsymbol{\Phi}(\mathbf{s})^{-1} - \beta_{n,t'}\mathbf{c}_n\mathbf{c}_n^H \quad (17)$$

$$\mathbf{c}_n \triangleq \boldsymbol{\Phi}(\mathbf{s})^{-1}\mathbf{q}_n \quad (18)$$

$$\beta_{n,t'} \triangleq \sigma_{t',t}^2(1+\sigma_{t',t}^2\mathbf{q}_n^H\mathbf{c}_n)^{-1} \quad (19)$$

According to [15], we assume that $\sigma_{t'}^2 \neq \sigma_t^2$, (15)-(18) imply

$$\begin{aligned}\Delta_{n,t'}(\mathbf{s},\mathbf{y}_v) = {} & \beta_{n,t'}\left|\mathbf{c}_n^H\left(\mathbf{y}_v - \mathbf{Q}\boldsymbol{\mu}(\mathbf{s})\right) + \mu_{t',t}/\sigma_{t',t}^2\right| \\ & - |\mu_{t',t}|^2/\sigma_{t',t}^2 + \ln(\beta_{n,t'}/\sigma_{t',t}^2) \\ & + \ln(\lambda_{t'}/\lambda_t)\end{aligned} \quad (20)$$

where $\Delta_{n,t'}(\mathbf{s},\mathbf{y}_v)$ quantifies the change to $\alpha(\mathbf{s},\mathbf{y}_v)$ corresponding to the change of the nth index in $\mathbf{s}$ from t to t'. And then we can work out the metric for $\mathbf{s}'$ as $\alpha(\mathbf{s},\mathbf{y}_v) + \Delta_{n,t'}(\mathbf{s},\mathbf{y}_v)$. In this letter, $T = 2, t = 0, t' = 1$.

In summary, the proposed Bayesian Matching Pursuit based method is a non-exhaustive tree-search using the SP selection metric (14) with fast metric update. According to the characteristics of mmWave channel, we choose to apply $T = 2, (\mu_0, \sigma_0^2) = (0,0), (\mu_1, \sigma_1^2) = (0, 100P), D = 5, L' = 5, \lambda_1 = L^{'}/N, J = \lceil \text{erfc}^{-1}(2J_0)\sqrt{2N\lambda_1(1-\lambda_1)} + N\lambda_1 \rceil$, $J_0 = 0.005$. The algorithm is shown in Algorithm 1, where $\boldsymbol{\delta}$ represents approximate posterior probability of $\mathbf{s}$ using the renormalized estimate

$$p(\mathbf{s}|\mathbf{y}_v) = \frac{\exp\{\alpha(\mathbf{s},\mathbf{y}_v)\}}{\sum_{\mathbf{s}'\in\mathbf{S}}\exp\{\alpha(\mathbf{s}',\mathbf{y}_v)\}} \approx \frac{\exp\{\alpha(\mathbf{s},\mathbf{y}_v)\}}{\sum_{\mathbf{s}'\in\mathbf{S}_\Omega}\exp\{\alpha(\mathbf{s}',\mathbf{y}_v)\}}. \quad (21)$$

When the search ends, the algorithm would return the MMSE estimation of $\mathbf{h}$ using (12).

V. SIMULATION RESULTS

The performance of the proposed method is examined via computer simulation. ULAs are assumed at both BS and MS with $N_T = N_R = 32$. We use $N_T^{Beam} = 32$ training beams at BS and $N_R^{Beam} = 32$ combining beams at MS. All simulation results are averaged over 500 channel realizations with a carrier frequency of 60GHz. At each channel realization, the number of scatterers is $L = 7$. We sample $[0,\pi)$ uniformly with $G = 64$ samples. The design of hybrid precoding and combining matrices have been extensively investigated, so we just adopt the precoder and combiner presented in [17]. $\mathbf{F} = (\boldsymbol{\Lambda}_F^{-1/2}\mathbf{U}_F^H)^T$ where $\mathbf{U}_F$ and $\boldsymbol{\Lambda}_F$ are the matrices of the eigenvectors and eigenvalues of $\mathbf{A}_T^*(\mathbf{A}_T^*)^H$. $\mathbf{W} = (\boldsymbol{\Lambda}_W^{-1/2}\mathbf{U}_W^H)^H$ where $\mathbf{U}_W\boldsymbol{\Lambda}_W\mathbf{U}_W^H = \mathbf{A}_R(\mathbf{A}_R)^H$. $G = 64$ is used to satisfy RIP for applying CS algorithms. For BCS and SBL, true noise power are provided based on SNR. For the proposed method, noise power are measured as $1/100$ of the variance of the received signal. Parameters are selected as explained in section IV-D. Note that, we also use a large virtual sparsity 10 to compare with the algorithm which uses small virtual sparsity 5. Proposed algorithms are named as Proposed S and Proposed L for small virtual sparsity and large virtual sparsity respectively.

In Fig. 2, we compare methods OMP, SBL, BCS, the Proposed S and the Proposed L. The performance of channel estimation precision is measured by the normalized

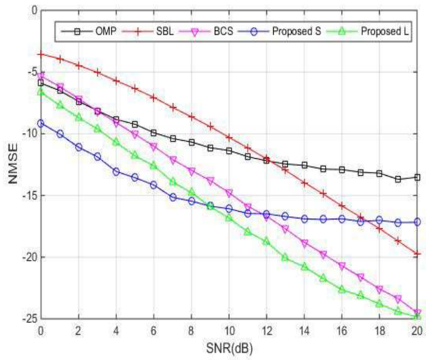

Fig. 2: NMSE at different SNRs (dB).

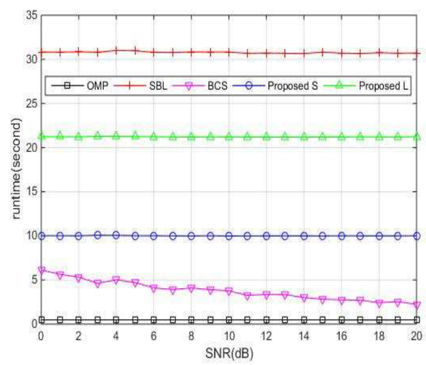

Fig. 3: Runtime at different SNRs (dB).

mean square error (NMSE) defined as $10\log_{10}\left(\mathrm{E}(\|\mathbf{H}-\mathbf{H}^{\text{estimate}}\|_F^2/\|\mathbf{H}\|_F^2)\right)$. As shown, our proposed methods perform better than any other CS algorithms at low SNRs. The proposed S achieves the best performance with 3-4 dB improvement compared with BCS when $SNR < 9dB$. For higher SNRs, the proposed L can achieve 2dB improvement over BCS. We found that smaller virtual sparsity works better for low SNRs, but bigger virtual sparsity is required for higher SNRs. This is because we did not consider off-grid error mitigation in this letter. The accuracy of channel estimation is affected by noise and off-grid errors. For higher SNRs, where the errors caused by off-grid error dominates, the additional active elements can help mitigate off-grid error impact and improve the estimation performance. On the contrary, noise dominates at lower SNRs. In such case, adding extra active elements which are redundancy for MMSE estimation will lead to worse performance. Apparently, larger virtual sparsity requires higher complexity, so we have to consider the trade-off between complexity and estimation accuracy. For mmWave channel estimation, using small virtual sparsity provides sufficient accuracy even for high SNRs, with a much lower complexity as shown in Fig. 3.

Fig. 3 displays the average runtime of all CS based methods. Our proposed method is significantly faster than SBL, on the same order of BCS, significantly slower than OMP. The result shows that our proposed method can greatly improve channel estimation performance with affordable computation.

VI. Conclusion

In this letter, we propose a novel method based on Bayesian matching pursuit algorithm for channel estimation in mmWave MIMO communication. Through selecting appropriate parameters according to the characteristics of mmWave channel, we utilize Bayesian model to implement MMSE channel estimation using a set of candidate SPs. The simulation results demonstrated that our algorithm can outperform all existing methods while requiring an affordable computational complexity.

References

[1] 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.

[2] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” *IEEE communications magazine*, vol. 49, no. 6, pp. 101–107, 2011.

[3] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially sparse precoding in millimeter wave mimo systems,” *IEEE transactions on wireless communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[4] P. F. Smulders and L. Correia, “Characterisation of propagation in 60 ghz radio channels,” *Electronics & communication engineering journal*, vol. 9, no. 2, pp. 73–80, 1997.

[5] E. Candes and T. Tao, “Decoding by linear programming,” *arXiv preprint math/0502327*, 2005.

[6] T. Kim and D. J. Love, “Virtual aoa and aod estimation for sparse millimeter wave mimo channels,” in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2015, pp. 146–150.

[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] S. Sun and T. S. Rappaport, “Millimeter wave mimo channel estimation based on adaptive compressed sensing,” in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2017, pp. 47–53.

[9] J. Lee, G.-T. Gil, and Y. H. Lee, “Exploiting spatial sparsity for estimating channels of hybrid mimo systems in millimeter wave communications,” in *2014 IEEE Global Communications Conference*. IEEE, 2014, pp. 3326–3331.

[10] A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, “Sparse bayesian learning-based channel estimation in millimeter wave hybrid mimo systems,” in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2017, pp. 1–5.

[11] M. U. Aminu, M. Codreanu, and M. Juntti, “Bayesian learning based millimeter-wave sparse channel estimation with hybrid antenna array,” in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2018, pp. 1–5.

[12] Y. You and L. Zhang, “Ip aided omp based channel estimation for millimeter wave massive mimo communication,” in *Proceedings of 2019 IEEE WCNC*. IEEE, 2019.

[13] B. Qi, W. Wang, and B. Wang, “Off-grid compressive channel estimation for mm-wave massive mimo with hybrid precoding,” *IEEE Communications Letters*, vol. 23, no. 1, pp. 108–111, 2018.

[14] J. Dai, X. Bao, W. Xu, and C. Chang, “Root sparse bayesian learning for off-grid doa estimation,” *IEEE Signal Processing Letters*, vol. 24, no. 1, pp. 46–50, 2016.

[15] P. Schniter, L. C. Potter, and J. Ziniel, “Fast bayesian matching pursuit,” in *2008 Information Theory and Applications Workshop*. IEEE, 2008, pp. 326–333.

[16] H. V. Poor, *An introduction to signal detection and estimation*. Springer Science & Business Media, 2013.

[17] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, “Sensing matrix optimization for block-sparse decoding,” *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4300–4312, 2011.