Off-Grid Sparse Bayesian Learning Based Channel Estimation for MmWave Massive MIMO Uplink

Haoyue Tang, Jintao Wang, *Senior Member, IEEE*, Longzhuang He, *Student Member, IEEE*

***Abstract*—In this letter, an angle domain off-grid channel estimation algorithm for the uplink (UL) millimeter wave (mmWave) massive multiple-input and multiple-output (MIMO) systems is proposed. By exploiting spatial sparse structure in mmWave channels, the proposed method is capable of identifying the angles and gains of the scatter paths. Comparing with the conventional channel estimation methods for mmWave system, the proposed method achieves better performance in terms of mean square error (MSE). Numerical simulation results are provided to verify the superiority of the proposed algorithm.**

***Index Terms*—mmWave massive MIMO, direction of arrival (DoA), spatial basis expansion model (SBEM).**

I. Introduction

MASSIVE multiple-input and multiple-output (MIMO) systems operating in the millimeter wave (mmWave) frequency band [1] is an enabling technology for current and future wireless communication systems. Equipped with massive antennas at the base station (BS), massive MIMO is able to serve multiple users simultaneously with higher energy efficiency, broader coverage and higher spectral efficiency [2].

Since all the potential gains require beamforming [3] and precoding, the acquisition of precise channel state information (CSI) at the receiver [4] is extremely important. As a result of the fading attenuation characteristics in mmWave frequency band [5], only a limited number of scatter paths are observed [1] and spatial basis expansion model (SBEM) [6] can be established for massive MIMO channel modeling.

By exploiting the sparsity of mmWave and MIMO channels, compressed sensing (CS) has been widely used in channel estimation [7]–[11]. A two-stage adaptive CS algorithm is proposed in [12], theoretic analysis as well as simulation results are provided to demonstrate its robustness in low SNR regions. Recent research adopted the sparse Bayesian learning algorithm for CSI recovery [13]. The estimation accuracy of these algorithms is limited by the coarsely pre-divided angle grids. Although off-grid can precisely solve CS problem with unknown basis, its computation complexity is extremely high [14].

Motivated by the above issues, we propose a channel estimation method based on the off-grid sparse Bayesian learning algorithm for the multi-user mmWave massive MIMO system. The proposed method exploits channel characteristic with higher estimation accuracy, which is measured by the mean square error (MSE) of the CSI matrix.

The remainder of the letter is organized as follows. In Section II, we introduce the uplink mmWave massive MIMO channel model and discuss channel estimation from the angle domain perspective. Section III demonstrates the proposed improved sparse Bayesian learning (ISBL) based channel estimation strategy. Simulation results are provided in Section IV. Section V draws the conclusion and discusses future work.

Notations: Matrices and vectors are written in boldface letters. The determinant, trace, transpose and Hermitian of matrix $\mathbf{A}$ are denoted as $|\mathbf{A}|$, $\mathrm{Tr}(\mathbf{A})$, $\mathbf{A}^T$ and $\mathbf{A}^H$ respectively. The distribution of the random variable X conditioned on Y with θ as a parameter is denoted as $P(X|Y;\theta)$. Notations $\Re(\cdot)$ and $\Im(\cdot)$ represent the real and imagine part of a complex, respectively.

II. Problem Formulation

A. System Model

In this letter, we consider an uplink multiuser mmWave massive MIMO system [15], where the base station (BS) is equipped with $M \gg 1$ antennas in the form of uniform linear arrays (ULA). The BS serves U single-antenna[1] users simultaneously in the coverage area. The propagation from user u to BS is assumed to compose P scatter paths, and the UL channel vector can be expressed as:

$$\mathbf{h}_u = \frac{1}{\sqrt{P}}\sum_{p=1}^{P}\alpha_{up}\mathbf{a}(\theta_{up}), \quad u = 1, 2, ..., K, \tag{1}$$

where α_{up} and $\theta_{up} \in [-\frac{\pi}{2}, \frac{\pi}{2})$ denote the attenuation factor and the direction of arrival (DOA) for the p^{th} scatter path of user u, respectively. The vector $\mathbf{a}(\theta)$ is:

$$\mathbf{a}(\theta) = \left[1, e^{\left(j\frac{2\pi d}{\lambda}\sin\theta\right)}, ..., e^{\left(j(M-1)\frac{2\pi d}{\lambda}\sin\theta\right)}\right]^T, \tag{2}$$

where λ and d denote the wave length and distance of antenna elements, respectively.

Suppose the user u sends a training sequence $\mathbf{x}_u$ of length L. To realize near-optimal training of all the users, we use orthogonal training sequences here. The received training signal $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_L]$ is:

$$\mathbf{Y} = \sum_{u=1}^{U}\mathbf{h}_u\mathbf{x}_u^H + \mathbf{N} = \mathbf{H}\mathbf{X}^H + \mathbf{N}, \tag{3}$$

[1] Single uplink antenna model widely used in LTE systems. Future work will address the case of more antennas at the UE end.

The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: tanghaoyue13@tsinghua.org.cn; wangjintao@tsinghua.edu.cn).

This work was supported in part by the National Key R&D Program of China (Grant No.2017YFE011230) and the National Natural Science Foundation of China (Grant No.61471221).

Manuscript received March 21, 2018.

where $\mathbf{N}=[\mathbf{n}_1, \mathbf{n}_2, ..., \mathbf{n}_U]$, $\mathbf{n}_u \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$ is additive white Gaussian noise on the receiver, and $\mathbf{H}=[\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_U]$ is the CSI matrix. The pilot signal training power $\sigma_p^2 = \frac{1}{L}\mathbf{x}_i^H \mathbf{x}_i$ and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_L]$.

B. Channel Estimation from the Angle Domain

Due to the orthogonality of the training sequences, we can obtain a coarse estimated channel information:

$$\tilde{\mathbf{h}}_u = \frac{1}{L\sigma_p^2}\mathbf{Y}\mathbf{x}_u. \tag{4}$$

In this way, the precision of estimated CSI matrix is severely contaminated by the noise on the receiver. This can be mitigated by taking the structure of CSI formulation in Eqn. (1) into account. Thus, we need to estimate the DoA and attenuation information of the scatter paths, which can be done by solving the following problem:

$$\min P', \quad \text{s. t } \|\tilde{\mathbf{h}}_u - \sum_{i=1}^{P'} \hat{\alpha}_{ui}\mathbf{a}(\hat{\theta}_{ui})\| < \epsilon. \tag{5}$$

III. Improved off-grid Sparse Bayesian Learning (ISBL) Based Channel Estimator

To solve the above non-convex problem, we propose a two-stage ISBL algorithm based channel estimation algorithm. In the first stage, on-grid channel estimation using sparse Bayesian learning (SBL) is carried out to determine the coarse location of the angle domain information with the help of predivided grids. The second stage is precise off-grid angle domain search using expectation maximization (EM) algorithm.

A. On-Grid Sparse Bayesian Learning

By dividing $[-\frac{\pi}{2}, \frac{\pi}{2})$ into grids of number N and denoting ω_i as the angle of the i^{th} grid, we construct a matrix $\mathbf{A} = [\mathbf{a}(\omega_1), \mathbf{a}(\omega_2), ..., \mathbf{a}(\omega_n)]$.

We can use these coarse divided angles to approximately reconstruct the uplink CSI for each user $u = 1, 2, ..., U$, by finding a set of coefficient $\mathbf{w}_u = [w_{u1}, w_{u2}, ..., w_{un}]^T$ to represent the coarsely estimated channel $\tilde{\mathbf{h}}_u$, i.e,

$$\tilde{\mathbf{h}}_u = \mathbf{A}\mathbf{w}_u + \tilde{\mathbf{n}}. \tag{6}$$

If the DoA of a real scatter path θ_{up} lies around a coarsely divided grid angle ω_n, then the corresponding coefficient w_{un} is similar to the gain of the scatter path α_{up}. Otherwise, the corresponding coefficient is nearly zero. From the above analysis, it can be seen that the coefficient vector $\mathbf{w}_u$ will possess a sparse characteristic if we select the division number N much larger than the number of scatter paths. Hence, we need to find a vector $\mathbf{w}_u$, satisfying

$$\min \|\mathbf{w}_u\|_0, \text{s.t } \|\tilde{\mathbf{h}}_u - \mathbf{A}\mathbf{w}_u\| < \delta. \tag{7}$$

The SBL algorithm can solve the above compressive sensing problem with no prior information about the sparsity (number of scatter paths). For simplicity, denote $\tilde{\mathbf{h}}$ as $\tilde{\mathbf{h}}_u$ and $\mathbf{w}_u$ as $\mathbf{w}$, each item w_i in vector $\mathbf{w}$ corresponds to a normal distribution $w_i \sim \mathcal{N}(\mu_i, \lambda_i^{-1})$, and $\tilde{\mathbf{n}}$ is a noise vector, each item corresponds to $\mathcal{N}(0, \sigma^2)$. We need to find out a set of parameters $\{\mu, \sigma^2, \lambda\}$ such that $p(\mathbf{w}, \lambda, \sigma^2|\tilde{\mathbf{h}})$ is maximized. According to the probability characteristic:

$$p(\mathbf{w}, \lambda, \sigma^2|\tilde{\mathbf{h}}) = p(\mathbf{w}|\lambda, \sigma^2, \tilde{\mathbf{h}})p(\lambda, \sigma^2|\tilde{\mathbf{h}}). \tag{8}$$

On-grid angle domain estimation adopts the EM algorithm.

1) E-step (Inference): Based on the parameters λ, σ^2, the most likely $\mathbf{w}$ is computed. Denote the $(\cdot)^{(k)}$ as the parameters we obtained in the k^{th} iteration. Optimizing $p(\mathbf{w}^{(k+1)}|\lambda^{(k)}, \sigma^{2(k)}, \mathbf{h})$ is equivalent as optimizing its corresponding $\ln$ function, which is the following problem:

$$\mathbf{w}^{(k+1)} = \arg\min_{\mathbf{w}} \frac{1}{\sigma^{2(k)}}\|\tilde{\mathbf{h}} - \mathbf{A}\mathbf{w}\|_2^2 + \sum_{i=1}^{N} \lambda_i^{(k)}\|w_i\|^2, \tag{9}$$

of which the solution can be obtained directly due to its convexity,

$$\mathbf{w}^{(k+1)} = \left(\mathbf{A}^H\mathbf{A} + \sigma^{2(k)}\mathbf{\Lambda}^{(k)}\right)^{-1}\mathbf{A}^H\tilde{\mathbf{h}}, \tag{10}$$

where the matrix $\mathbf{\Lambda}^{(k)} = \text{diag}\left(\lambda_1^{(k)}, \lambda_2^{(k)}, ..., \lambda_N^{(k)}\right)$.

2) M-step (Hyperparameter Prediction): Based on the coefficient $\mathbf{w}^{(k+1)}$ obtained from the E-step, the iteratively re-estimated parameters α and μ has the following closed-form expression.

$$\mu^{(k+1)} = \mathbf{w}^{(k+1)} = \sigma^{(k+1)^{-2}}\mathbf{\Sigma}^{(k+1)}\mathbf{A}^H\tilde{\mathbf{h}}, \tag{11}$$

where $\mathbf{\Sigma}^{(k+1)} = \left(\sigma^{2(k)}\mathbf{A}^H\mathbf{A} + \mathbf{\Lambda}^{(k)}\right)^{-1}$.

Denote

$$\gamma_i^{(k+1)} = 1 - \lambda_i^{(k)}\Sigma_{ii}^{(k+1)}, \tag{12}$$

the parameter λ is obtained via:

$$\lambda_i^{(k+1)} = \frac{\gamma_i^{(k+1)}}{\left(\mu_i^{(k+1)}\right)^2}. \tag{13}$$

The estimated noise variance also needs to be update:

$$\sigma^{2(k)} = \frac{\|\tilde{\mathbf{h}} - \mathbf{A}\mathbf{w}^{(k+1)}\|_2}{M - \sum_i \left(\lambda_i^{(k)}\right)^{-1}}. \tag{14}$$

According to the theory of SBL, if the corresponding $\mu_i = w_i$ is near 0, then the corresponding λ_i will be quite large. This will prevent w_i from increasing too much in the following iterations, hence guarantee the sparse structure in $\mathbf{w}$.

3) Parameter Pruning: To reduce computational complexity, we force the items w_i and μ_i permanently to zeros if the corresponding α_i is larger than a threshold. The relationship between $|\alpha|$ and $\sin(\omega)$ after three training phases are plotted in Fig. 1, which implies SBL algorithm can locate the angle grid in the neighborhood of the real DoA angles within a few iterations.

After the iterations end, those w_i not equal to 0 indicate the corresponding angle ω_i is near the real scatter paths' DoA. We establish $\mathbf{\Omega}^{(0)} = [\omega_{q1}, \omega_{q2}, ...\omega_{qN}]$, the corresponding w_{qi} of each ω_{qi} satisfies $w_{qi} \neq 0$.

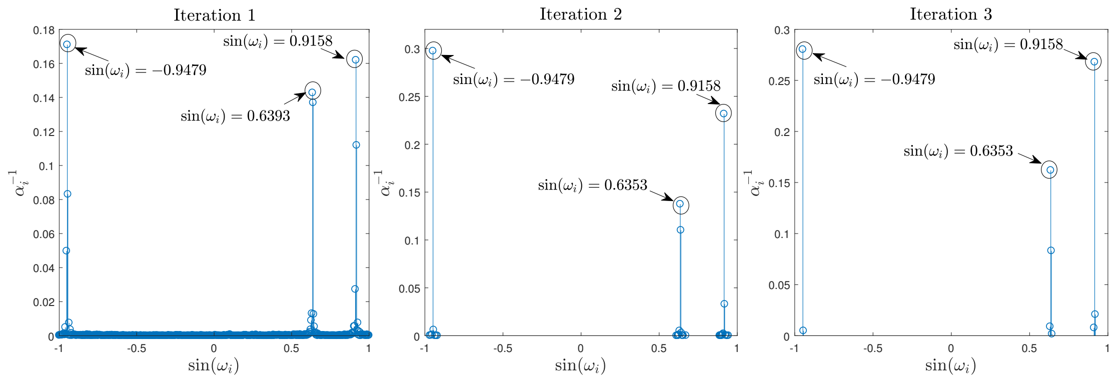

Fig. 1. Example of λ^{-1} with a certain user whose DoA of scatter paths $\sin(\theta)$ equals $[-0.9474, 0.6371, 0.9172]$ after three on-grid training phases.

B. Off-Grid CS based Angle-Domain Estimator

The estimation accuracy is limited by the number of grids in the sensing matrix. To avoid this, off-grid optimization is needed. We need to find a set of angles $\mathbf{\Omega} = [\hat{\omega}_1, \hat{\omega}_2, ..., \hat{\omega}_N]^T$, such that the basis $\mathbf{A}(\mathbf{\Omega}) = [\phi(\hat{\omega}_1), \phi(\hat{\omega}_2), ..., \phi(\hat{\omega}_N)]$ can be used to reconstruct the CSI. To simplify further analysis, denote $\mathbf{s} = [s_1, s_2, ..., s_N]^T$, where $s_n = \sin(\hat{\omega}_n)$. The task is to find $\mathbf{\Omega} = \arcsin(\mathbf{s}^*)$ such that

$$[\mathbf{s}^*, \mathbf{w}^*] = \arg\left(\min_{\mathbf{s},\mathbf{w}} \|\mathbf{h} - \mathbf{A}\left(\sin^{-1}(\mathbf{s})\right)\mathbf{w}\|_2^2\right). \tag{15}$$

Denote $f(\mathbf{w}, \mathbf{s}) = \|\mathbf{h} - \mathbf{A}\left(\sin^{-1}(\mathbf{s})\right)\mathbf{w}\|_2^2$, the function can be approximated using Taylor expansion

$$\begin{aligned} f(\mathbf{w}, \mathbf{s}) \approx\, & f(\mathbf{w}, \mathbf{s}_0) + \nabla f(\mathbf{w}, \mathbf{s}_0)^H(\mathbf{s} - \mathbf{s}_0) \\ & + \frac{1}{2}(\mathbf{s} - \mathbf{s}_0)^H \nabla_{\mathbf{s}}^2 f(\mathbf{w}, \mathbf{s})(\mathbf{s} - \mathbf{s}_0). \end{aligned} \tag{16}$$

To find the optimum solutions for $\mathbf{w}$ and $\mathbf{\Omega}$ of the above convex function, we use the EM algorithm again.

1) Channel Reconstruction Coefficient Estimation (E-step): Based on the angle domain information $\mathbf{\Omega}^{(k)}$ obtained from the k^{th} iteration, establish the new basis $\mathbf{A}^{(k)}$ the optimum coefficient is obtained:

$$\mathbf{w}^{(k+1)} = \left(\mathbf{A}^{(k)H}\mathbf{A}^{(k)}\right)^{-1}\mathbf{A}^{(k)H}\mathbf{h}. \tag{17}$$

2) Angle Domain Information Estimation (M-step): Based on the coefficient $\mathbf{w}^{(k+1)}$, we adopt a backtracking line search algorithm to find the optimum $\sin(\mathbf{\Omega})$ and refresh the basis. The gradient of f with regard to $\mathbf{s}$ in the $(k+1)^{th}$ iteration is:

$$\nabla_{\mathbf{s}} f^{(k+1)} = \Re\left[\left(\mathbf{y} - \mathbf{A}(\mathbf{\Omega}^{(k)})\mathbf{w}^{(k+1)}\right)^H \mathbf{\Psi}^{(k)}\mathbf{w}^{(k+1)}\right], \tag{18}$$

where

$$\mathbf{\Psi}^{(k)} = j\frac{2\pi d}{\lambda}\text{diag}(0, 1, ..., M-1)\mathbf{A}\left(\mathbf{\Omega}^{(k)}\right). \tag{19}$$

Start from the initial point $\mathbf{s}^{(k)}$, the novel $\mathbf{s}^{(k+1)}$ is searched in the opposite direction of the gradient:

$$\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} - t\nabla_{\mathbf{s}} f^{(k+1)}. \tag{20}$$

The termination condition is

$$f\left(\mathbf{w}^{(k+1)}, \mathbf{s}^{(k)}\right) - \beta\|t\nabla_{\mathbf{s}} f^{(k+1)}\| \geq f\left(\mathbf{w}^{(k+1)}, \mathbf{s}^{(k+1)}\right). \tag{21}$$

If the termination condition cannot be satisfied, we have to lower the step size along the search direction, $t = \alpha t$. The variable $\alpha \in [0.1, 0.9]$ and $\beta \in [0.1, 1)$ are two parameters for backtracking line search algorithm.

After K iterations, the CSI of the k^{th} user can be reconstructed through

$$\hat{\mathbf{h}}_u = \mathbf{A}^{(K)}\mathbf{w}^{(K)}. \tag{22}$$

Algorithm 1 Improved sparse Bayesian learning based mmWave channel estimator

1: **initialization**: select user u, $\mathbf{h} = \tilde{\mathbf{h}}_u$, establish initial dictionary $\mathbf{A}(\mathbf{\Omega})$, initial values for α and γ.
2: **repeat**
3: $\quad k \leftarrow k + 1$
4: $\quad \mathbf{w}^{(k+1)} \leftarrow \left(\mathbf{A}^H\mathbf{A} + \left(\sigma^{(k)}\right)^2 \mathbf{\Lambda}^{(k)}\right)^{-1} \mathbf{A}^H\mathbf{h}$
5: $\quad$ Refresh $\lambda^{(k+1)}, \gamma^{(k+1)}$ using (11)(12)
6: **until** $\|\mathbf{w}^{(k-1)} - \mathbf{w}^{(k)}\| < \delta_1$ $\quad \triangleright$ On-grid SBL search
7: $k \leftarrow 0$, establish $\mathbf{\Omega}^{(0)}$, $\mathbf{A}^{(0)}$
8: **repeat**
9: $\quad k \leftarrow k + 1$
10: $\quad \mathbf{w}^{(k+1)} \leftarrow \left(\mathbf{A}^{(k)H}\mathbf{A}^{(k)}\right)^{-1}\mathbf{A}^{(k)H}\mathbf{h}$
11: $\quad$ Compute $\nabla_{\mathbf{s}} f^{(k+1)}$ using (15), $t \leftarrow 1$
12: $\quad$ **while** $f\left(\mathbf{w}^{(k+1)}, \mathbf{s}^{(k)}\right) - \beta\|t\nabla_{\mathbf{s}} f^{(k+1)}\| < f\left(\mathbf{w}^{(k+1)}, \mathbf{s}^{(k+1)}\right)$ **do**
13: $\quad\quad t \leftarrow \alpha t$, $\mathbf{s}^{(k+1)} \leftarrow \mathbf{s}^{(k)} - t\nabla_{\mathbf{s}} f^{(k+1)}$
14: $\quad$ **end while**
15: **until** $\|\mathbf{w}^{(k-1)} - \mathbf{w}^{(k)}\| < \delta_2$ $\quad \triangleright$ Off-grid refinement
16: $\hat{\mathbf{h}}_u = \mathbf{A}^{(K)}\mathbf{w}^{(K)}$ $\quad \triangleright$ CSI reconstruction

IV. Simulation Results

A. Channel Estimation Precision

In this part, we compare the performance of the proposed ISBL method with other channel estimation algorithms, the performance of channel estimation precision is measured by the channel mean square error (MSE)

$$\text{MSE} = \frac{1}{U}\sum_{u=1}^{U} \frac{\|\hat{\mathbf{h}}_u - \mathbf{h}_u\|_2^2}{\|\mathbf{h}_u\|_2^2}.$$

Fig. 2 illustrates the MSE performance of uplink channel estimation, as a function of the signal to noise (SNR) ratio with different channel estimation algorithms. Simulations are carried out with users $U = 20$, the uplink scatter paths of each is $P = 3$, the number of antennas in the base station $M = 256$. The SNR is defined as $SNR = \sigma_p^2/\sigma_n^2$. We take the grid number $N = 800$ for the orthogonal matching pursuit (OMP)[2] and SBL based algorithm. For low SNR regions, the proposed algorithm receives significant MSE decrease. For high SNR regions, it receives 20 dB MSE gain than on grid SBL algorithm in the high-SNR regions. The OMP and SBL algorithms encounter error floor in medium to high SNR regions. This is caused by the limited number of grid numbers.

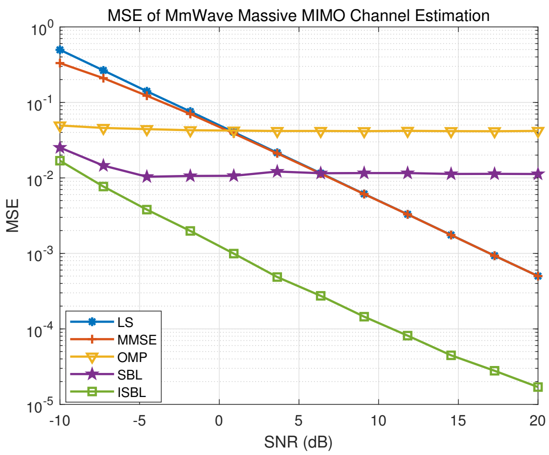

Fig. 2. MSE comparisons of various mmWave massive MIMO CSI estimation algorithm

B. Complexity Analysis

Here we analyze the computational complexity of the proposed algorithm and compare it with various methods. Denote the number of coarse division grids as N, the matrix $\mathbf{\Sigma}$ in Eqn. (8) is a Hermitian matrix, hence its inverse can be computed using Cholesky factorization with the multiplication times a square scale to its dimension. The variable K_1 and K_2 denote the number of loops in the on-grid SBL search and off-grid dictionary refinement respectively. In simulations, it takes $K_1 = 5 \sim 20$ loops for on-grid SBL search to terminate. To have a concrete comparison of the computational complexity, average multiplication times under 10dB case using these algorithms is provided. From the table, the proposed ISBL algorithm can achieve better effectiveness in sparse mmWave channel recovery with low additional computational cost compared with SBL algorithm, and receives compatible MSE performance with computational complexity reduction compared with Atom norm based algorithm.

V. Conclusions

In this letter, we exploited the physical structure of CSI of uplink mmWave massive MIMO system, based on this, we propose an ISBL algorithm that can achieve higher estimation accuracy. Simulation results were provided to demonstrate the effectiveness and the superiority of the proposed method. Combining the proposed method with user grouping [10] to mitigate the effect of pilot contamination will be our future work.

[2]The OMP algorithm is implemented with prior knowledge about the precise number of scatter paths, the performance of the OMP algorithm detoriates without this information, hence the improvement compared with OMP algorithm is larger in reality.

TABLE I. Computational Complexity and Runtime Comparisons

Method	Computational Complexity	Runtime
OMP	$\mathcal{O}(NMP)$	0.0884s
SBL [13]	$\mathcal{O}\left(K_1(N^2+NM)\right)$	0.4999s
ISBL	$\mathcal{O}(K_1(N^2+NM)+K_2(P^2+MP))$	1.005s

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] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, Oct 2014.

[3] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath and O. Koymen, ”Beamforming Tradeoffs for Initial UE Discovery in Millimeter-Wave MIMO Systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 543-559, April 2016.

[4] A. M. Sayeed and V. Raghavan, “Maximizing MIMO Capacity in Sparse Multipath With Reconfigurable Antenna Arrays,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 1, pp. 156-166, June 2007.

[5] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An overview of signal processing techniques for millimeter wave MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[6] H. Xie, F. Gao, S. Zhang, and S. Jin, “A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3170–3184, April 2017.

[7] W. U. Bajwa, J. Haupt, A. M. Sayeed and R. Nowak, “Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058-1076, June 2010.

[8] D. E. Berraki, S. M. Armour, and A. R. Nix, “Application of compressive sensing in sparse spatial channel recovery for beamforming in mmWave outdoor systems,” in *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*. IEEE, 2014, pp. 887–892.

[9] A. Alkhateeb, G. Leus, and R. W. Heath, “Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?” in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 2909–2913.

[10] Z. Gao, L. Dai, Z. Wang, and S. Chen, “Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO,” *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6169–6183, Dec 2015

[11] V. Raghavan and A. M. Sayeed, “Sublinear Capacity Scaling Laws for Sparse MIMO Channels,” *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 345-364, Jan. 2011.

[12] Y. Han and J. Lee, “Two-stage compressed sensing for millimeter wave channel estimation,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 860–864.

[13] Y. Wu, Y. Gu, and Z. Wang, “Channel estimation for mmWave MIMO with transmitter hardware impairments,” *IEEE Communications Letters*, vol. 22, no. 2, pp. 320–323, Feb 2018.

[14] G. Tang, B. N. Bhaskar, P. Shah and B. Recht, “Compressed Sensing Off the Grid,“ *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465-7490, Nov. 2013.

[15] O. E. 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, March 2014.