Joint Channel Estimation and Beamforming for Millimeter Wave Cellular System

Huang Huang, Kunpeng Liu, Rong Wen, Yi Wang, Guangjian Wang
Communications Technology Laboratory, 2012 Laboratory
Huawei Technologies Co., Ltd., Shenzhen, P.R.China.
huanghuang, liukunpeng, wenrong, yi.wang, wangguangjian@huawei.com

Abstract—In this paper, we propose a joint channel estimation and beamforming (BF) scheme for the wide band millimeter wave (mmWave) cellular system. Specifically, low complexity compressive sensing (CS) based estimation algorithm is used to estimate the sparse mmWave channels. Based on the estimated channel, low complexity BF scheme is proposed to adapt to the channel for data communications. The algorithm is designed by considering the practical hardware and channel constraints. Furthermore, the complexity is very low without matrix inverse and singular value decomposition (SVD) compared with traditional algorithms, which is well suitable to practical realization. Finally, we show by simulations that the performance of proposed scheme is close to the optimal scheme with extremely less overhead.

Index Terms—mmWave, Beamforming, Large Scale Antennas, Compressive Sensing, Wide Band, Low Complexity

I. INTRODUCTION

Recently, cellular communications in mmWave band becomes a hot research topic in the literature. It is considered as a good complement for the traditional <6GHz frequency band, where it can provide several GHz bandwidth to meet the huge peak data rate requirement in 5G. However, the millimeter wave suffers much higher free space path loss compared with the current cellular frequency band. BF technique is a promising way to solve this problem, where large scale antenna array BF could provide sufficient gain to compensate the path loss.

In [1], the authors consider the BF design for single user wireless communications under the RF hardware constraints. In [2], the authors propose a joint spatial division and multiplexing scheme for multi-user communications under the mmWave channels. However, these BF designs are based on the fully knowledge of the physical channel at the transmitter and receivers without the channel estimation considerations. In fact, the physical channel is very difficult to acquire in practice, since large scale antenna array is used in mmWave communications. There are extensive researches for the channel estimation for low frequency massive MIMO systems [3]. However, these works just assume that each antenna is connected with one analog radio frequency (RF) chain. It is not applicable for mmWave system, where one RF chain is connecting multiple antennas. Due to the sparsity of the mmWave channel, compressive sensing (CS) technique is a promising way to the mmWave channel estimation. In [4], the authors propose an adaptive CS channel estimation scheme for mmWave system by exploring the sparsity of the mmWave channels, where angle of arrival (AoA) and angle of departure (AoD) of per path is estimated iteratively. However, it will incur several downlink and uplink switch between base station (BS) and mobile station (MS). In [5], the authors use the standard basis pursuit (BP) denoising algorithm to estimate the AoD only by assuming omi-directional antenna at receiver.

In this paper, we consider a point to point single RF chain mmWave system as illustrated in Fig. 1. Each RF chain is connected with multiple antennas, where the transmitted signal in each antenna is controlled by one analog phase shifter. We are interested in designing a low complexity joint channel estimation and BF scheme. However, there are various first order technical challenges that need to be addressed.

- **Wide band signal with multi-paths:** The key benefit for mmWave band is the Gbps available bandwidth. However, existing works [1], [2], [4], [5] assume narrow band signal to ignore the multi-paths issues. Actually, there are multiple clusters with different delays, and multiple paths within one cluster in practical scenario [6]. Gbps bandwidth should result in at least hundreds of coherent subband, which is impossible for CS based channel estimation in each subband. Furthermore, note that the phase shifter cannot support different BF control for different subband, even the channel is estimated in each subband. Therefore, the existing works cannot be extended directly to the wide band scenarios, and is quite challenging.

- **Hardware constraints:** For cost and power considerations, analog BF is often implemented using analog phase shifters. In other words, only finite phase can be controlled, which places constant modulus constraints on the elements of analog BF. Even if perfect channel is acquired, the traditional optimal BF by SVD of the channel is not applicable, where the amplitude tuning is required. Furthermore, unlike the digital BF, the variation of analog BF needs switch time for the analog phase shifter. For example, in 802.11ad [7], a short BF inter frame space interval is 1us, whereas one OFDM symbol time is only 0.2us. Therefore, one analog BF pattern should serve multiple OFDM symbols, otherwise it will introduce significant overhead. The BF design in [1] based on instantaneous channel is not applicable in practice. The analog BF should be adaptive to the second order statistics of the channel, for example, AoD information. It further complicates...
Sensing beam design: Sensing matrix design is very important in CS frameworks, which is sensing beam design in our channel estimation. The sensing beams with entries randomly generated from \( \{\pm 1, \pm j\} \) has been proposed for mmWave wave channel estimation in [8]. In [9], the authors propose to use the random rows of orthogonal matrix. However, all these works do not consider the real scenarios where AoAs and AoDs are from some given region. A randomly generated beam always spread its power in the full spatial areas. If we could concentrate the sensing beams in the interested region, then we could highly improve the estimation performance. As a result, new, random sensing beam design should be exploited for the local area case.

In this paper, we propose a joint channel estimation and BF scheme for wide band mmWave cellular system with large scale antennas, where practical hardware and channel constraints are considered. Specifically, low complexity CS based estimation algorithm is used to estimate the sparse mmWave channels. Based on the estimated channel, low complexity BF scheme is proposed to adapt to the channel for data communications. For hardware constraints, we only consider the finite phase control with limited quantization bits for analog BF. For channel constraints, we consider wide band channel with multiple clusters and paths having different delays. The complexity is very low without matrix inverse and SVD compared with the traditional algorithms, which is well suitable to practical realization. Finally, we show by simulations that the performance of proposed scheme is close to the optimal scheme with extremely less overhead.

Notation: \( \mathbb{C} \) denotes the set of complex numbers; \( | \cdot | \) denotes the element-wise absolute value function or the cardinality function for a set. \( \| \cdot \|_p \) denotes the \( p \)-norm of a vector. \( \cdot^T \) and \( \cdot^\dagger \) denote the transpose and Hermitian transpose, respectively; \( \otimes \) denotes the Kronecker product; \( \{x, x, X\} \) represents the scaler, vector and matrix respectively. \( x_{[i]} \) denotes the \( i \)-th element of vector \( x \). \( \{X_{[i,j]}, X_{[i,j]}, X_{[i,j]}\} \) denotes the \((i, j)\)-th element, whole \( i \)-th row elements and whole \( j \)-th column elements of matrix \( X \), respectively. \( \mathrm{vec}(\cdot) \) is the vectorizing function.

II. SYSTEM MODEL

A. System Topology

In this paper, we consider a point to point single RF chain mmWave communication system with uniform planar array (UPA) antennas as illustrated in Fig. 1. Specifically, one RF chain is connected with the UPA. The transmitted and received signal in each antenna is controlled by one analog phase shifter, which is quantized by \( Q \) bits. The transmitter and receiver UPA are equipped with \( M = M_\theta M_\phi \) and \( N = N_\theta N_\phi \) antenna elements respectively. The transmitter tries to transmit one data streams to the receiver due to the RF chain limit.

B. Channel Model

We adopt the clustered channel model, which is widely used in the academic and industry literature [1], [2], [4], [6]. Specifically, the channel is assumed to be a sum of the contributions of \( N_{cl} \) clusters, each of which contributes \( N_{path} \) propagation paths. The paths in the same cluster have the same propagation delay, and different clusters have different delays. The channel of the \( l \)-th cluster at time \( t \) is given by [6]:

\[
\mathbf{H}(l, t) = \sum_{n=1}^{N_{path}} \kappa_{l,n} \mathbf{a}_R(\theta_{R,l,n}, \phi_{R,l,n}) \cdot \mathbf{a}_T^\dagger(\theta_{T,l,n}, \phi_{T,l,n}) e^{j\eta t ||v_{l,n}||^2} \tag{1}
\]

where \( \kappa_{l,n} \) is the complex gain of the \( n \)-th path in cluster \( l \). It includes the transmit and receive antenna element gain. \( (\theta_{R,l,n}, \phi_{R,l,n}, \theta_{T,l,n}, \phi_{T,l,n}) \) are the physical azimuth and elevation angles of arrival and departure respectively. \( \{\mathbf{a}_R(\cdot), \mathbf{a}_T(\cdot)\} \) are the normalized receive and transmit array response vectors respectively. For a UPA, the array response vector is given by [2]:

\[
\mathbf{a}_T(\theta, \phi) = \frac{1}{\sqrt{M}} \left[ 1, \ldots, e^{j\eta d(m \sin(\theta) \sin(\phi) + n \cos(\phi))}, \ldots, e^{j\eta d((M_\theta - 1) \sin(\theta) \sin(\phi) + (M_\phi - 1) \cos(\phi))} \right]^T \tag{2}
\]

where \( \eta = \frac{2\pi}{\lambda} \), \( \lambda \) is the wavelength and \( d \) is the inter-element space. \( v_{l,n} \) is the relative speed of the transmitter and receiver for the \( n \)-th path in the \( l \)-th cluster.

Note that the AoA and AoD \((\theta_{R,l,n}, \phi_{R,l,n}, \theta_{T,l,n}, \phi_{T,l,n})\) do not change during the considered communication stage.

III. CS BASED CHANNEL ESTIMATION

In this section, we shall propose a CS based algorithm to do the channel estimation. We assume that channel in (1) does not change during the estimation stage, since the time duration is very small. It is a minor assumption widely adopted in the academic literature [10] and the industry standards 802.11ad [7] and LTE. Furthermore, it can be justified in the simulations. Therefore, by merging the time index \( t \) into the complex gain \( \kappa_{l,n} \) in (1), we have

\[
\mathbf{H}(l) = \sum_{n=1}^{N_{path}} \kappa_{l,n} \mathbf{a}_R(\theta_{R,l,n}, \phi_{R,l,n}) \cdot \mathbf{a}_T^\dagger(\theta_{T,l,n}, \phi_{T,l,n}) \tag{3}
\]

A. Sparse Formulation of the mmWave Channel

Note that the paths of AoA and AoD in (3) come from the continuous value, which is very difficult to do estimation. Therefore, in [4] the authors assume that the AoAs and AoDs
are taken from a uniform grid of \(N_L\) points, where \(N_L\) is much larger than the number of paths, which is off-grid based approximation. In [5], the authors explicitly propose the DFT basis approximation of the grid for the uniform linear array (ULA). Actually, it follows the virtual channel representation (VCR) of the channels [11], where a unitary DFT matrix is applied at the receive and transmit side respectively to transform the channel matrix from spatial domain to virtual angle domain. VCR is widely used in the literature for the sparse channel estimation [12]. However, they consider the simple ULA case. In this work, we shall extend to the UPA case. Specifically, the VCR of the channel \(\mathbf{H}(l)\) in (3) for UPA is also given by:

\[
\mathbf{H}_V(l) = \mathbf{U}_R^\dagger \mathbf{H}(l) \mathbf{U}_T \tag{4}
\]

where \(\mathbf{U}_T\) is a unitary matrix given by: \(\mathbf{U}_T = \mathbf{U}_{T,\theta} \otimes \mathbf{U}_{T,\phi}\), where \(\{\mathbf{U}_{T,\theta}, \mathbf{U}_{T,\phi}\}\) are the traditional DFT matrices in ULA case given the antennas \(M_0\) and \(M_0\) respectively. \(\mathbf{U}_R\) follows the same way at the receiver side.

The sparsity of the virtual channel \(\mathbf{H}_V(l)\) is guaranteed by the following Lemma.

**Lemma 1 (VCR of UPA Channel):** Given the approximation as in the VCR of ULA channel [11]

\[
f_N(\theta) \approx \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} e^{-j2\pi \theta i} = \begin{cases} \sqrt{N} & \text{if } -\frac{1}{2N} \leq \theta < \frac{1}{2N} \\ 0 & \text{otherwise} \end{cases}
\]

There are at most \(N_{\text{path}}\) non-zero elements in the VCR channel \(\mathbf{H}_V(l)\) in (4).

**Proof:** Please refer to Appendix A.

Finally, if \(\mathbf{H}_V(l)\) is known, the original channel \(\mathbf{H}(l)\) is given by:

\[
\mathbf{H}(l) = \mathbf{U}_R \mathbf{H}_V(l) \mathbf{U}_T^\dagger \tag{6}
\]

Note that the general VCR expression are defined on the global areas, where there are no constraints on the AoA and AoD. Actually, we care about the local areas in practice, where the AoA and AoD direction have physical constraints. For example, a BS cell may cover only 60° horizontal area, and there is only 140° angle spread for UE [6]. In the following, we shall show that we could reduce the dimension of the VCR channel \(\mathbf{H}_V(l)\) from the prior knowledge of the AoA and AoD directions, which could largely reduces the estimation overhead. From the proof of Lemma 1, we know that for a given AoA and AoD region, we could find the sets of positive integers \(\{\Omega_R, \Omega_T\}\), such that

\[
\mathbf{H}_V(l)|_{i,j} = 0, \text{ if } i \not\in \Omega_R \text{ or } j \not\in \Omega_T \tag{7}
\]

Let \(\mathbf{U}_R = \mathbf{U}_{R,\Omega_R}\), \(\mathbf{U}_T = \mathbf{U}_{T,\Omega_T}\), and \(\mathbf{H}_V(l) = \mathbf{H}_V(l)|_{\Omega_R,\Omega_T}\). Then \(\mathbf{H}(l)\) in (6) is given by:

\[
\mathbf{H}(l) = \mathbf{U}_R \mathbf{H}_V(l) \mathbf{U}_T^\dagger \tag{8}
\]

**B. VCR Channel Estimation**

In this section, we shall propose a low complexity algorithm to estimate the VCR channel \(\mathbf{H}(l)\) in (8). Specifically, the frame structure of the channel estimation constitutes of multiple pseudo-random sequence transmission slot as illustrated in Fig. 2. The length of the PN sequence is \(T\). Different transmit and/or receive sensing beams are applied in each slot. BF gap is inserted at the end of each slot, which is used for sensing beam switch either at the transmitter and/or receiver side. There are totally \(J\) slots.

![Frame structure of the channel estimation.](image)

At the \(j\)-th slot, where the transmit BF is \(\mathbf{w}_{T,j}\) and receive BF is \(\mathbf{w}_{R,j}\), the received signal is given by:

\[
\mathbf{y}_j(n) = \sum_l \mathbf{w}_{R,j}^\dagger \mathbf{H}(l) \mathbf{w}_{T,j} s(n - t_l) + \mathbf{w}_{R,j}^\dagger \mathbf{z}_j(n) \tag{9}
\]

where \(s(\cdot)\) is the PN sequence, \(t_l\) is the path delay of the \(l\)-th cluster. The cross correlation output is given by:

\[
\gamma \begin{cases} 
\frac{\mathbf{w}_{R,j}^\dagger \mathbf{H}(l) \mathbf{w}_{T,j} + \mathbf{z}_{a,j}}{\mathbf{z}_{a,j}} & \text{if } a = t_l \\
\mathbf{z}_{a,j} & \text{otherwise}
\end{cases} \tag{10}
\]

where \(\gamma = \sum_n s(n)s^*(n)\) is the correlation gain. \(\mathbf{z}_{a,j} = \sum_n \mathbf{w}_{R,j}^\dagger \mathbf{z}_j(a + n)s^*(n)\) is the noise.

Note that \(\tilde{y}_j(l)\) only consists the paths that come from the \(l\)-th cluster. Furthermore, if there exists a cluster at the path delay \(t_l\), there should be a correlation peak at corresponding time instant. Therefore, \(\tilde{y}_j(l)\) can be applied for the \(l\)-th cluster paths estimation. For simplicity, we shall ignore the index \(l\) hereafter. Replace \(\mathbf{H}(l)\) in (8), we have

\[
\tilde{y}_j = \gamma \mathbf{w}_{R,j}^\dagger \mathbf{H}_V \mathbf{U}_R \mathbf{H}_{T,j} + \mathbf{z}_{a,j} = \gamma (\mathbf{g}_{T,j}^\dagger \otimes \mathbf{g}_{R,j}^\dagger) \mathbf{vec}(\mathbf{H}_V) + \mathbf{z}_{a,j} \tag{11}
\]

From Lemma 1, we know that there are approximately at most \(N_{\text{path}}\) non-zero elements in \(\mathbf{H}_V\). Therefore, we desired to estimate sparse \(\mathbf{H}_V\) based on the CS theory. Let \(\tilde{y}_j\) be the \(j\)-th element of \(\tilde{y}\), we have

\[
\mathbf{y} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_j \\ \vdots \\ \tilde{y}_J \end{bmatrix} = \gamma \begin{bmatrix} \mathbf{g}_{T,1}^\dagger \otimes \mathbf{g}_{R,1}^\dagger \\ \vdots \\ \mathbf{g}_{T,J}^\dagger \otimes \mathbf{g}_{R,J}^\dagger \end{bmatrix} \mathbf{vec}(\mathbf{H}_V) + \mathbf{z} \tag{12}
\]

where \(\mathbf{z} \in \mathbb{C}^{J \times 1}\) is the noise vector. Then the problem of estimating \(\mathbf{H}_V\) is given by following standard CS optimization
problem:
\[
\begin{align*}
\arg\min_{\mathbf{H}_v} & \quad ||\text{vec}(\mathbf{H}_v)||_0 \\
\text{s.t.} & \quad ||\tilde{y} - \mathbf{A}\text{vec}(\mathbf{H}_v)||_2 \leq \delta
\end{align*}
\] (13)

where \(\delta > 0\) bounds the noise effect in the received signal.

C. Sensing Beam Design

The first issue to solve the problem (13) is to design good sensing matrix \(w_{R,j}\) and \(w_{T,j}\). We shall propose a low complexity sensing matrix design for local areas. Specifically, the sensing beams are given by:
\[
\begin{align*}
w_{R,j} = \text{Quant}(\tilde{U}_{R,j}^\dagger ; \Omega_R) & \quad \text{for} \quad j = 1, \ldots, Q \\
w_{T,j} = \text{Quant}(\tilde{U}_{T,j}^\dagger ; \Omega_R) & \quad \text{for} \quad j = 1, \ldots, Q
\end{align*}
\] (14)

where \(r_{R,j} \in \mathbb{C}^{Q_R \times 1}, r_{T,j} \in \mathbb{C}^{Q_T \times 1}\) are the random matrices where each element follows a same random distribution, for example the [0,1] uniform distribution. \(\text{Quant}()\) is the quantization function that map the sensing beams to the achievable beams under practical hardware conditions. It is an element-wise function such that the angle of each element is mapped to the nearest available quantized angle.

D. Channel Estimation Algorithm

There are many algorithms to solve the problem (13) based on the CS frame structure, such as basis pursuit (BP), orthogonal matching pursuit (OMP), stage-wise orthogonal matching pursuit (st-OMP) and subspace pursuit (SP) [3]. In addition to these standard estimation algorithms, we propose a low complexity algorithm, where there is no matrix inverse addition to these standard estimation algorithms, we propose matching pursuit (st-OMP) and subspace pursuit (SP) [3]. In this section, we shall propose a low complexity BF algorithm given below

Algorithm 3 (BF Algorithm with Full Channel Estimation): Let \(\mathbf{H}_v(l) = \sum_j |\mathbf{H}_v(l)|^2\), the index of the maximal value of \(\mathbf{H}_v(l)\) is given by \((i^*, j^*)\). Then the BF vector of the transmitter and receiver is given by:
\[
\begin{align*}
w_T = \text{Quant}(\mathbf{U}_{T,[j^*, \cdot]}), & \quad \text{and} \quad w_R = \text{Quant}(\mathbf{U}_{R,[i^*, \cdot]})
\end{align*}
\] (17)

Remark 2 (Interpretation of Algorithm 2): The transmit and receive BF vectors are chosen to one AoD and AoA direction, which has the largest power gain from all the clusters and paths. Note that the BF vectors naturally meet the constant modulus constraint with only phase quantizations. Furthermore, the complexity and overhead is extremely low. The receiver only needs to feedback the index of largest AoD to the transmitter rather than the total BF vector as the traditional closed loop MIMO.

Note that Algorithm 2 needs the full estimation of the channel by standard CS estimation algorithms. Based on the proposed low complexity Algorithm 1, we propose a simple BF algorithm as follows:

Algorithm 1 (BF Algorithm with Full Channel Estimation): Let \(q = \mathbf{A}^\dagger \tilde{y}_2\) the estimated best AoD and AoA direction \(\mathbf{q}^*\) in \(\text{vec}(\mathbf{H}_v)\) is given by:
\[
q^* = \arg\max |q|
\] (15)

Remark 1 (Complexity of Algorithm 1): The complexity of Algorithm 1 is very low, where there is only multiply operations without any matrix inverse. On the other hand, the standard CS based algorithms to obtain the whole channel \(\mathbf{H}_v\) needs up to \(N_{path} \times N_{path}\) matrix inverse, for example the OMP algorithm [13].

IV. BEAM PATTERN DESIGN

In this section, we shall propose a low complexity BF scheme based on the estimated channel in Section III.

A. BF Algorithm

Note that, the BF vectors at the receiver and transmit that maximize the average capacity under the multi-path channel \(\mathbf{H}(l, t)\) in (1) is given by:
\[
\begin{align*}
\max_{w_R} E_t \left[ \sum_k \log_2(1 + \rho |w_R^\dagger \mathbf{H}^k(t) w_T|^2) \right] \\
\text{s.t.} \quad w_R \in \mathcal{W}_R, w_T \in \mathcal{W}_T
\end{align*}
\] (16)

where \(\mathbf{H}^k(t)\) is the channel of the \(k\)-th subcarrier out of the total \(K\) coherent subcarriers from the channel \(\mathbf{H}(l, t)\). \(\rho\) is the SNR. \(\mathcal{W}_R, \mathcal{W}_T\) are the feasible sets under hardware constraints. However, the solution to problem (16) needs the knowledge of the statistics of the channel \(\mathbf{H}(l, t)\), whereas only the knowledge of instant channel \(\mathbf{H}(l, t)\) are estimated in Section III. Furthermore, even if the statistics is perfectly known, the solution is also very complex, since it is a non-convex problem.

Note that the AoA and AoD directions do not change, and if we have the estimated virtual channel \(\mathbf{H}_v(l, t)\) from the standard CS algorithms, we propose a low complexity BF algorithm given below

Algorithm 2 (BF Algorithm with Full Channel Estimation): Let \(\mathbf{H}_v(l) = \sum_j |\mathbf{H}_v(l)|^2\), the index of the maximal value of \(\mathbf{H}_v(l)\) is given by \((i^*, j^*)\). Then the BF vector of the transmitter and receiver is given by:
\[
\begin{align*}
w_T = \text{Quant}(\mathbf{U}_{T,[j^*, \cdot]}), & \quad \text{and} \quad w_R = \text{Quant}(\mathbf{U}_{R,[i^*, \cdot]})
\end{align*}
\] (17)

Remark 2 (Interpretation of Algorithm 2): The transmit and receive BF vectors are chosen to one AoD and AoA direction, which has the largest power gain from all the clusters and paths. Note that the BF vectors naturally meet the constant modulus constraint with only phase quantizations. Furthermore, the complexity and overhead is extremely low. The receiver only needs to feedback the index of largest AoD to the transmitter rather than the total BF vector as the traditional closed loop MIMO.

Note that Algorithm 2 needs the full estimation of the channel by standard CS estimation algorithms. Based on the proposed low complexity Algorithm 1, we propose a simple BF algorithm as follows:

Algorithm 3 (BF Algorithm with Full Channel Estimation): Let \(i^*\) be the best cluster index of the maximal value of \(|q_{i^*, \cdot}^*(l)|, \forall l\}, the BF vector of the transmitter and receiver is given by:
\[
\begin{align*}
w_T = \text{Quant}(\mathbf{U}_{T,[j^*, \cdot]}), & \quad \text{and} \quad w_R = \text{Quant}(\mathbf{U}_{R,[i^*, \cdot]})
\end{align*}
\] (18)

where \((i^*, j^*)\) is the AoA and AoD index of \(\mathbf{H}_v(l^*)\) corresponding to the \(a^*(l^*)\) index of \(\text{vec}(\mathbf{H}_v(l^*))\).
it is a common case in practice, such as in 802.11ad beam track stage, where only transmit or receive BF vector is trained and the other side BF vector remains unchange [7]. Since there is only one receive BF vector \( w_R \), the received signal in (12) can be revised as

\[
\tilde{y} = \gamma w_R^\dagger \tilde{U}_R \tilde{H} \tilde{U}_T^\dagger \tilde{W}_T + \tilde{z}
\]

where \( h_{AoD} \) is the AoD information we want to estimate.

\[
W_T = [w_{T,1}, \cdots, w_{T,J}] \]

Algorithm 2 and 3 still apply by just fixing the receive BF vector.

**Theorem 1 (Optimality of the Algorithm 2 and 3):**
Suppose there is only one cluster, i.e., \( N_{cl} = 1 \). Each element of equivalent channel \( h_{AoD} \) has fixed power gain with random phase distributions. The optimal solution to problem (13) without phase quantization is given by:

\[
w_T = U_T[j^*, \cdot] \quad (20)
\]

where \( j^* \) is the index of the maximal value of \( |h_{AoD}| \).

**Proof:** please refer to Appendix B.

**Remark 4 (Interpretation of Theorem 1):** Theorem 1 indicates that under the given case, the optimal transmit BF vector is transmitting towards the largest AoD. It is exactly the Algorithm 2 or 3, where the two algorithms are the same under one cluster condition. One cluster assumption is reasonable in some cases. For example, in ITU channel model [6], the first cluster may cover 90% of the total power in some scenarios. In addition, the \( h_{AoD} \) assumption in Theorem 1 may hold, since most of the channel models in the literature assume that the path gain does not change, where the phase changes w.r.t. the time [6], [7].

Note that the same results hold when the transmit BF vector does not change by considering the uplink and downlink duality.

**V. SIMULATION AND DISCUSSION**

In this section, we shall compare the performance of the proposed algorithms with the various baselines under practical channel model. Specifically, the channel is given by the ITU UMi model [6]. The mobility of the UE is 10Km/h. We assume V-band 802.11ad [7] single carrier physical parameters for channel estimation. The number of antennas of transmitter and receiver is \( \{M = 8 \times 8 = 64, N = 4 \times 4 = 16\} \) with \( d = \lambda/2 \). The quantization bits is \( Q = 4 \). We assume that the transmitter covers a 60° azimuth and elevation angle area and the receiver covers the whole angle area. Therefore, according to the proof of Lemma 1, we have that \( |\Omega_T| = 16 \) and \( |\Omega_R| = 16 \). That is we need to estimate a \( 256 \times 1 \) vector \( \text{vec}(\tilde{H}_v) \) in (13). We set \( J = 64 \) in the simulation, and hence it saves 75%=1-64/256 overhead.

Fig. 3 compares the performance of the proposed sensing beam design versus the baselines. Specifically, baseline 1 is that the elements of sensing beams are chosen randomly from \( \{\pm 1, \pm j\} \) [8]. Baseline 2 is given by the random Zadoff-Chu (ZC) sequence [9]. It illustrates the successful probability of detecting the best AoD and AoA direction under the channel estimation algorithm 1 for different SNR.

Fig. 4 compares the performance of the proposed scheme versus SNR. Specifically, the comparing schemes are given below.

- **Baseline 3, Near-Optimal BF Scheme [2]:** The algorithm is the near-optimal BF scheme to solve the problem (16) such that the BF vector is adaptive to the statistics of the channel given by: \( H = \sum_k E_t ([H_k(t)]^\dagger H_k(t)) \), the BF vector is the eigenvector of the largest eigenvalue of \( H \). Note that this algorithm requires the knowledge of the statistics of the channel which needs a long time observation, and hence it is not applicable in practice. Furthermore, it does not satisfy the hardware constraints.
- Baseline 4, BF Algorithm 2 with Perfect Channel Information: This algorithm is the upper bound of the proposed joint estimation and BF algorithms, since there is perfect channel information without estimation errors.
- Proposed Scheme 1, BF Algorithm 2 with Full Channel Estimation: The channel estimation is the traditional OMP algorithm [13] to estimate the full channels. Then BF Algorithm 2 is utilized from the fully estimated channel.
- Proposed Scheme 2, BF Algorithm 3 with Channel Estimation Algorithm 1: This algorithm has the lowest complexity, since it does not need any matrix inverse.

It illustrates the capacity performance versus SNR. Note that the proposed scheme 2 has the similar performance as proposed scheme 1, where proposed scheme 2 has extremely low complexity. It comes from the fact that the first cluster dominates the channel power in given ITU channel model. Furthermore, the proposed schemes have only 2~3dB loss compared with the baseline 2 where it has perfect channel information. It means that the performance loss due to the channel estimation error is limited and ignorable. Finally, the proposed schemes have about 8dB loss compared with the proposed schemes has about 8dB loss compared with the proposed schemes has about 8dB loss compared with the proposed schemes has about 8dB loss compared with the proposed schemes has about 8dB loss compared with the proposed schemes has about 8dB loss compared with the proposed schemes has about 8dB loss compared with the proposed schemes.

**APPENDIX A: PROOF OF LEMMA 1**

Define $\alpha_{R,l,n} = \frac{\eta d}{N_{o}} \sin(\theta_{R,l,n}) \sin(\phi_{R,l,n})$, $\beta_{R,l,n} = \frac{\eta d}{N_{o}} \cos(\theta_{R,l,n})$, and $\{\alpha_{T,l,n}, \beta_{T,l,n}\}$ are defined on the AoD $\{\theta_{T,l,n}, \phi_{T,l,n}\}$ accordingly. From (3), the $(s,k)$-th element of $\mathbf{H}(l)$ is given by:

$$
\mathbf{H}(l)_{s,k} = \sum_{n} \kappa_{l,n} e^{j(\phi_{R,l,n} + \phi_{R,l,n})} e^{j(k_{o} \alpha_{T,l,n} + k_{o} \beta_{T,l,n})}
$$

where $\{s_{o} \leq N_{o} - 1, s_{o} \leq N_{o} - 1, k_{o} \leq M_{o} - 1, k_{o} \leq M_{o} - 1\}$ are non negative integers that satisfy the equations $s = s_{o} N_{o} + s_{o}$ and $k = k_{o} M_{o} + k_{o}$. Therefore, $\mathbf{H}(l)$ can be rewritten in Kronecker form as:

$$
\mathbf{H}(l) = \sum_{n} \kappa_{l,n} \mathbf{b}(\alpha_{R,l,n}) \otimes \mathbf{b}(\beta_{R,l,n})
$$

where $\mathbf{b}(\alpha) = \frac{1}{\sqrt{N_{o}}} [1, e^{j\alpha}, \ldots, e^{j(N_{o}-1)\alpha}]^{T}$ is the array response vector of ULA.

Apply the Kronecker product equation $(a \otimes b)(c \otimes d) = (a \otimes b) \mathbf{H}_{v}(l)$ where $\mathbf{H}_{v}(l)$ is given by:

$$
\mathbf{H}_{v}(l)_{p,q} = U_{R}^{\dagger} \cdot p \cdot H(l)_{T} \cdot U_{T} \cdot q
$$

where $\mathbf{b}(\alpha) = \frac{1}{\sqrt{N_{o}}} [1, e^{j\alpha}, \ldots, e^{j(N_{o}-1)\alpha}]^{T}$ is the array response vector of ULA.

**APPENDIX B: PROOF OF THEOREM 1**

Let $\mathbf{w}_{T}$ represented by the basis of $\mathbf{U}_{T}$ as follows:

$$
\mathbf{w}_{T} = \sum_{m} \sqrt{P_{m}} \mathbf{U}_{T} \cdot [m]_{T}
$$

where $P_{m}$ is the power basis we want to solve, and satisfies $\sum_{m=1}^{M} P_{m} = 1$ since $\mathbf{w}_{T} \cdot \mathbf{w}_{T}^{\dagger} = 1$. Under the given conditions, the objective in (13) is given by:

$$
\max_{\{P_{m}\}} \mathbb{E} \left[ \log_{2} \left( 1 + \rho \mathbf{H}_{w}^{\dagger} \mathbf{w}_{T} \mathbf{w}_{T}^{\dagger} \mathbf{H}_{w} \right) \right]
$$

$$
= \max_{\{P_{m}\}} \mathbb{E} \left[ \log_{2} \left( 1 + \rho \mathbf{H}_{\text{AoD}}^{\dagger} \mathbf{U}_{T} \mathbf{w}_{T} \mathbf{U}_{T} \mathbf{h}_{\text{AoD}} \right) \right]
$$

$$
\leq \log_{2} \left( 1 + \rho \mathbf{h}_{\text{AoD}}^{\dagger} \mathbf{J}_{T} \mathbf{h}_{\text{AoD}} \right)
$$

where the equation holds when $P_{j} = 1$ and $P_{i} = 0, \forall i \neq j$.

It finishes the proof.

**REFERENCES**


