CHANNEL ESTIMATION AND TRAINING DESIGN FOR HYBRID ANALOG-DIGITAL MULTI-CARRIER SINGLE-USER MASSIVE MIMO SYSTEMS

Jianshu Zhang[1], *Ivan Podkurkov*[1,2], *Martin Haardt*[1], *and Adel Nadeev*[2]
[1]Communications Research Laboratory, Ilmenau University of Technology, Ilmenau, Germany
[2]German-Russian Institute of Advanced Technologies, Tupolev Kazan National Research Technical University, Kazan, Russia

ABSTRACT

In this paper we study the channel estimation problem for a CP-OFDM based hybrid analog-digital massive MIMO system. In contrast to a conventional MIMO system, two additional constraints need to be fulfilled. First, the analog precoding is achieved using only a phase shift network, which imposes constant modulus constraints on the elements of the RF precoding and decoding matrices. Second, there is just one common equivalent RF precoding or decoding matrix for all subcarriers. These constraints lead to a challenging channel estimation task that includes the training design. To estimate the channel at the receiver, a least squares (LS) method and a compressed sensing (CS) method with a single-stage or a two-stage design are introduced. Compared to the single-stage designs, the two-stage designs have a lower computational complexity. Sufficient conditions for a unique channel estimation are derived for both methods. Simulation results show that the CS method provides more accurate channel estimates than the LS method under mild conditions.

Index Terms— mmWave Massive MIMO, hybrid precoding, OFDM, least squares, compressed sensing.

I. INTRODUCTION

Massive MIMO, which uses orders of magnitude more antennas (e.g., 100 or more), can provide significant MIMO gains [1]. When combined with millimeter wave (mmWave) technology, it will not only gain from large chunks of underutilized spectrum in the mmWave band [2] but will also benefit from a significantly reduced form factor of the massive MIMO array [3]. Hence, massive MIMO communication is a potential technique for future wireless networks [4]. However, if a large number of RF chains is implemented to steer the massive number of antenna elements, the involved power consumption and the hardware cost are too high and therefore are impractical. To exploit the MIMO multiplexing gain under a reasonable cost, one promising solution is to deploy hybrid analog-digital precoding schemes, realized using phase shifters or switches in the RF domain [5], and digital precoding schemes, implemented in the digital baseband domain as in conventional MIMO. If analog precoding is achieved using phase shifters only, the analog precoding matrix should have only constant modulus entries [6], [7], [8]. Furthermore, when a wideband multi-carrier system is considered, equivalently we get the same phase shifts for all subcarriers [9]. These two constraints are stringent such that they lead to significant challenges not only for the precoding of the transmitted data but also for the required channel estimation tasks [7], [10], [11], [12]. In [7] an adaptive compressed sensing (CS) based channel estimation algorithm is proposed to estimate the channel of a hybrid analog-digital massive MIMO system. This CS based channel estimation algorithm has been further extended in [10] by involving multiple measurement vectors (MMV) to improve the channel estimation accuracy. The CS based concept is also used in [11], where an adaptive multi-grid sparse recovery approach is applied instead. Finally, a multi-user hybrid analog-digital system is considered in [12] and a minimum mean squared error (MMSE) approach is developed to estimate the channel. Unfortunately, all the above papers deal with narrowband systems, or equivalently a flat fading channel. Their results cannot be directly used in a multicarrier system, or equivalently a frequency selective channel due to the fact that there is a common RF precoding and decoding matrix for all the subcarriers. Hence, this motivates us to design channel estimation algorithms as well as training sequences for single user multi-carrier hybrid massive MIMO systems.

In this paper we develop channel estimation algorithms for a single user multi-carrier hybrid massive MIMO system. A cyclic prefix OFDM (CP-OFDM) based multi-carrier modulation scheme is used and training using pilot tones is considered. To estimate the channel at the receiver side we study two different approaches, i.e., a least squares (LS) approach and a CS approach. The former one is a linear method while the latter one is realized using the orthogonal matching pursuit (OMP) algorithm, which is a non-linear method. We provide the sufficient condition for a unique channel recovery using the LS technique. Moreover, to reduce the computational complexity, an orthogonal training design via a two-stage channel estimation is proposed and an analytical expression is given for the achieved MSE of the LS channel estimates. The CS approach exploits the sparsity of the channel in the angular-delay domain. A single-stage design based on the OMP algorithm is proposed and a sufficient condition for unique channel estimate is derived. By consuming relatively more training resources compared to the single-stage design, a reduced complexity

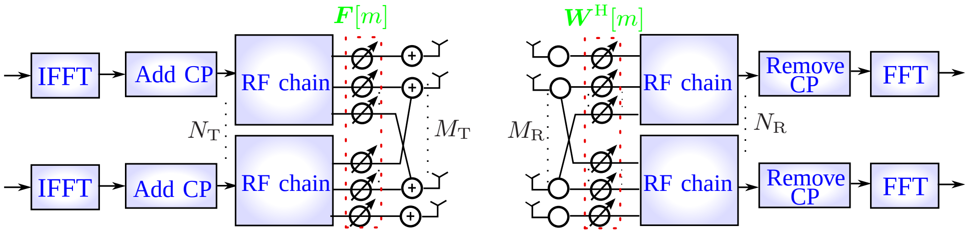

Fig. 1. A hybrid point-to-point massive MIMO-OFDM system with M_{T} transmit antennas and M_{R} receive antennas. There are $N_{\mathrm{T}} \ll M_{\mathrm{T}}$ transmit RF chains and $N_{\mathrm{R}} \ll M_{\mathrm{R}}$ receive RF chains. Matrices $\boldsymbol{F}[m]$ and $\boldsymbol{W}[m]$ denote the analog precoding and decoding matrices, which are implemented on the phase shift networks.

channel estimation method is developed via a two-stage OMP based sparse recovery. Simulation results show that the proposed CS approaches outperform the LS approaches and they require a smaller number of training symbols especially when the spatial frequencies of the estimated channel lie on uniform grids.

Notation: Upper-case and lower-case bold-faced letters denote matrices and vectors, respectively. The expectation, trace of a matrix, transpose, conjugate, Hermitian transpose, and Moore-Penrose pseudo inverse are denoted by $\mathbb{E}\{\cdot\}$, $\mathrm{Tr}\{\cdot\}$, $\{\cdot\}^{\mathrm{T}}$, $\{\cdot\}^{*}$, $\{\cdot\}^{\mathrm{H}}$, and $\{\cdot\}^{+}$, respectively. The Euclidean norm of a vector and the absolute value are denoted by $\|\cdot\|$ and $|\cdot|$, respectively. The Frobenius norm is denoted by $\|\cdot\|_{\mathrm{F}}$. The (c,d)-th element of a matrix is denoted by $(\cdot)_{c,d}$. The Kronecker product is $\otimes$. The $\mathrm{vec}\{\cdot\}$ operator stacks the columns of a matrix into a vector. The $\mathrm{unvec}_{M\times N}\{\cdot\}$ operator stands for the inverse function of $\mathrm{vec}\{\cdot\}$. The smallest integer that is greater than or equal to x is denoted by $\lceil x \rceil$. The largest integer that is smaller than or equal to x is denoted by $\lfloor x \rfloor$. The modulo operation is denoted by $\mathrm{mod}(\cdot)$. The p-th column of an identity matrix is denoted by $\boldsymbol{e}_p$. The operator $\angle(\cdot)$ computes the phase of a complex number. The uniform distribution within the interval $[a_1, a_2)$ is defined as $\mathcal{U}(a_1, a_2)$.

II. SYSTEM DESCRIPTION

II-A. System Model

We study a point-to-point massive MIMO system where a multi-antenna base station (BS) transmits data to a multi-antenna user equipment (UE) as depicted in Fig. 1. The BS has M_{T} transmit antennas and N_{T} RF chains. The UE has M_{R} receive antennas and N_{R} RF chains. The number of RF chains is assumed to be much smaller than the number of antenna elements, i.e., $M_{\mathrm{T}} \gg N_{\mathrm{T}}$ and $M_{\mathrm{R}} \gg N_{\mathrm{R}}$. A CP-OFDM based multi-carrier modulation scheme is applied to combat the multipath effect. The corresponding FFT size is N_{fft}. Let $\boldsymbol{s}_n[m] \in \mathbb{C}^{N_{\mathrm{T}}}$ represent the transmitted pilot vector on the n-th subcarrier in the m-th OFDM symbol over all available RF chains ($n \in \{k_1, \cdots, k_{N_{\mathrm{f}}}\} \subset \{1, \cdots, N_{\mathrm{fft}}\}$, $p \in \{1, \cdots, N_{\mathrm{f}}\}$, $m \in \{1, \cdots, N_{\mathrm{t}}\}$). Thereby, the training procedure consists of N_{t} OFDM symbols each with N_{f} pilot tones. The pilot tones and the data tones are interleaved on all the subcarriers and then passed through the IFFT filter. Furthermore, we assume that the pilot tones are assigned equally spaced and equally powered. A CP of length N_{CP} symbols is added, followed by an RF precoder $\boldsymbol{F}[m] \in \mathbb{C}^{M_{\mathrm{T}}\times N_{\mathrm{T}}}$ using analog circuitry. We assume that the RF precoder is implemented using analog phase shifters. Hence, constant modulus constraints should be fulfilled for each element of $\boldsymbol{F}[m] \in \mathbb{C}^{M_{\mathrm{T}}\times N_{\mathrm{T}}}$, i.e., $|(\boldsymbol{F}[m])_{a,b}| = 1$ for all $a \in \{1, \cdots, M_{\mathrm{T}}\}$ and $b \in \{1, \cdots, N_{\mathrm{T}}\}$. Finally, the total power of the pilot tones in one OFDM symbol is limited such that $\sum_{p=1}^{N_{\mathrm{f}}} \|\boldsymbol{F}[m]\boldsymbol{s}_{k_p}[m]\|^2 \leq P_{\mathrm{T}}$ for all m.

We consider a frequency selective quasi-static block fading channel. Assume that N_{CP} has the same length as the maximum excess delay of the channel such that the inter-symbol interference is avoided. After passing through the channel, first, an RF decoder $\boldsymbol{W}^{\mathrm{H}}[m] \in \mathbb{C}^{N_{\mathrm{R}}\times M_{\mathrm{R}}}$ is used at the UE. The RF decoder is also implemented using phase shifters and therefore all the elements of $\boldsymbol{W}^{\mathrm{H}}[m]$ have unit magnitude, i.e., $|(\boldsymbol{W}[m])_{c,d}| = 1$ for all $c \in \{1, \cdots, M_{\mathrm{R}}\}$ and $d \in \{1, \cdots, N_{\mathrm{R}}\}$. Afterwards, the CP is removed from the received signal and by using the FFT filter the time domain signal is transformed into the frequency domain. Let $\boldsymbol{H}_n \in \mathbb{C}^{M_{\mathrm{R}}\times M_{\mathrm{T}}}$ denote the discrete channel transfer function (CTF) on n-th subcarrier of the UE. The received training signal on the n-th subcarrier in the m-th OFDM symbol is given by [9]

$$\boldsymbol{y}_n[m] = \boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{H}_n\boldsymbol{F}[m]\boldsymbol{s}_n[m] + \boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{z}_n[m], \quad (1)$$

where $\boldsymbol{z}_n[m]$ represents zero mean circularly symmetric complex Gaussian (ZMCSCG) noise with covariance matrix $\mathbb{E}\{\boldsymbol{z}_n[m]\boldsymbol{z}_n^{\mathrm{H}}[m]\} = \sigma_{\mathrm{n}}^2\boldsymbol{I}_{M_{\mathrm{R}}}$ for all n and m. Note that equation (1) implies that the channel remains unchanged during the training procedure. In our design the data symbols are not used for channel estimation.

Our goal is to design $\boldsymbol{W}[m]$, $\boldsymbol{F}[m]$, and $\boldsymbol{s}_n[m]$, $\forall n, m$, such that the channel can be accurately estimated at the receiver.

II-B. Channel Model

In our paper we consider an analytical channel model consisting of a finite number of scatterers, i.e., L scatterers. Each scatterer contributes to a single propagation path between the BS and the UE, which accounts for one time delay τ_ℓ and one pair of spatial frequencies $(\mu_{\mathrm{T},\ell}, \mu_{\mathrm{R},\ell})$ for $\ell \in \{0, \cdots, L-1\}$. The frequency domain representation of the channel is given by [13]

$$\boldsymbol{H}(f) = \sum_{\ell=0}^{L-1} \underbrace{\alpha_\ell \boldsymbol{a}(\mu_{\mathrm{R},\ell}) \boldsymbol{a}^{\mathrm{H}}(\mu_{\mathrm{T},\ell})}_{\boldsymbol{H}_\ell \in \mathbb{C}^{M_\mathrm{R} \times M_\mathrm{T}}} e^{-j2\pi\tau_\ell f}, \tag{2}$$

where α_ℓ is the random complex gain of the ℓ-th path, with zero mean and $\mathbb{E}\{|\alpha_\ell|^2\} = 1/L$, $\forall \ell$. The vectors $\boldsymbol{a}(\mu_{\mathrm{T},\ell})$ and $\boldsymbol{a}(\mu_{\mathrm{R},\ell})$ are the array steering vectors of the BS and the UE, respectively. Note that the developed algorithms in this paper do not depend on whether a one-dimension (1-D) or two-dimension (2-D) array geometry is used. The defined array steering vectors here are just illustrative examples. Furthermore, for notational simplicity we assume that $\tau_\ell = \ell T_\mathrm{s}$, where $T_\mathrm{s} = 1/(N_\mathrm{fft} \cdot \Delta f)$ represents the sampling period and Δf denotes the subcarrier spacing. Then the sampled CTF on the n-th subcarrier is modeled as [14]

$$\boldsymbol{H}_n = \sum_{\ell=0}^{L-1} \underbrace{\alpha_\ell \boldsymbol{a}(\mu_{\mathrm{R},\ell}) \boldsymbol{a}^{\mathrm{H}}(\mu_{\mathrm{T},\ell})}_{\boldsymbol{H}_\ell \in \mathbb{C}^{M_\mathrm{R} \times M_\mathrm{T}}} e^{-j2\pi \frac{\ell \cdot n}{N_\mathrm{fft}}}, \tag{3}$$

III. LEAST SQUARES APPROACH

In this section we study the LS based training design, which is a commonly used channel estimation scheme, e.g., [15]. More specifically, a general LS solution via a single-stage design is proposed in Section III-A while a reduced complexity LS method via a two-stage orthogonal design is introduced in Section III-B.

III-A. LS estimation via a single-stage design

By inserting (3) into (1) we obtain

$$\begin{aligned} \boldsymbol{y}_n[m] &= \boldsymbol{W}^{\mathrm{H}}[m] \sum_{\ell=0}^{L-1} \boldsymbol{H}_\ell e^{-j2\pi\frac{\ell n}{N_\mathrm{fft}}} \boldsymbol{F}[m]\boldsymbol{s}_n[m] \\ &\quad + \boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{z}_n[m] \\ &= \boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{H}_u(\boldsymbol{w}_n \otimes (\boldsymbol{F}[m]\boldsymbol{s}_n[m])) \\ &\quad + \boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{z}_n[m], \end{aligned} \tag{4}$$

where we have

$$\boldsymbol{H}_u = [\boldsymbol{H}_0 \quad \cdots \quad \boldsymbol{H}_{L-1}] \in \mathbb{C}^{M_\mathrm{R} \times L M_\mathrm{T}},$$

and

$$\boldsymbol{w}_n = \begin{bmatrix} 1 & \cdots & e^{-j2\pi\frac{(L-1)n}{N_\mathrm{fft}}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{L}.$$

By stacking $\boldsymbol{y}_n[m]$ next to each other along the frequency domain (the n-dimension) we obtain a matrix $\boldsymbol{Y}[m] = \begin{bmatrix}\boldsymbol{y}_{k_1}[m] & \cdots & \boldsymbol{y}_{k_{N_\mathrm{f}}}[m]\end{bmatrix} \in \mathbb{C}^{N_\mathrm{R} \times N_\mathrm{f}}$, which is expressed as

$$\boldsymbol{Y}[m] = \boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{H}_u\boldsymbol{C}[m] + \boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{Z}[m], \tag{5}$$

where $\boldsymbol{Z}[m] = \begin{bmatrix}\boldsymbol{z}_{k_1}[m] & \cdots & \boldsymbol{z}_{k_{N_\mathrm{f}}}[m]\end{bmatrix} \in \mathbb{C}^{M_\mathrm{R} \times N_\mathrm{f}}$ and $\boldsymbol{C}[m] \in \mathbb{C}^{L M_\mathrm{T} \times N_\mathrm{f}}$ is computed by

$$\boldsymbol{C}[m] = \begin{bmatrix} \boldsymbol{w}_{k_1}^{\mathrm{T}} \otimes (\boldsymbol{F}[m]\boldsymbol{s}_{k_1}[m])^{\mathrm{T}} \\ \vdots \\ \boldsymbol{w}_{k_{N_\mathrm{f}}}^{\mathrm{T}} \otimes (\boldsymbol{F}[m]\boldsymbol{s}_{k_{N_\mathrm{f}}}[m])^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}.$$

Let $\boldsymbol{y}[m] = \mathrm{vec}\{\boldsymbol{Y}[m]\}$, where $\boldsymbol{h}_u = \mathrm{vec}\{\boldsymbol{H}_u\}$ and $\boldsymbol{z}[m] = \mathrm{vec}\{\boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{Z}[m]\}$. Then the vectorized version of (5) is expressed as

$$\boldsymbol{y}[m] = (\boldsymbol{C}^{\mathrm{T}}[m] \otimes \boldsymbol{W}^{\mathrm{H}}[m])\boldsymbol{h}_u + \boldsymbol{z}[m] \tag{6}$$

To utilize the training resource along the time domain (the m-dimension), we stack $\boldsymbol{y}[m]$ on top of each other as

$$\boldsymbol{y}_\mathrm{s} = \boldsymbol{P}_1\boldsymbol{h}_u + \boldsymbol{z}_\mathrm{s} \in \mathbb{C}^{N_\mathrm{R} N_\mathrm{f} N_\mathrm{t}}, \tag{7}$$

where $\boldsymbol{y}_\mathrm{s} = \begin{bmatrix}\boldsymbol{y}^{\mathrm{T}}[1] & \cdots & \boldsymbol{y}^{\mathrm{T}}[N_\mathrm{t}]\end{bmatrix}^{\mathrm{T}}$, $\boldsymbol{z}_\mathrm{s} = \begin{bmatrix}\boldsymbol{z}^{\mathrm{T}}[1] & \cdots & \boldsymbol{z}^{\mathrm{T}}[N_\mathrm{t}]\end{bmatrix}^{\mathrm{T}}$, and

$$\boldsymbol{P}_1 = \begin{bmatrix} \boldsymbol{C}^{\mathrm{T}}[1] \otimes \boldsymbol{W}^{\mathrm{H}}[1] \\ \vdots \\ \boldsymbol{C}^{\mathrm{T}}[N_\mathrm{t}] \otimes \boldsymbol{W}^{\mathrm{H}}[N_\mathrm{t}] \end{bmatrix} \in \mathbb{C}^{N_\mathrm{t} N_\mathrm{f} N_\mathrm{R} \times L M_\mathrm{T} M_\mathrm{R}}.$$

Conventionally, the LS estimate of $\boldsymbol{h}_u$ from (7) is computed by

$$\hat{\boldsymbol{h}}_u = \boldsymbol{P}_1^{+}\boldsymbol{y}_\mathrm{s}. \tag{8}$$

This requires that $\boldsymbol{P}_1$ has full column rank, i.e., $\mathrm{rank}(\boldsymbol{P}_1) = L M_\mathrm{T} M_\mathrm{R} \leq N_\mathrm{t} N_\mathrm{f} N_\mathrm{R}$. This condition also implies that the required total number of time-frequency resources is

$$N_\mathrm{t} N_\mathrm{f} \geq \frac{L M_\mathrm{T} M_\mathrm{R}}{N_\mathrm{R}}. \tag{9}$$

The LS problem can be solved using the Cholesky decomposition, which yields a computational complexity of order $\mathcal{O}((L M_\mathrm{T} M_\mathrm{R})^3/6)$ [16].

III-B. A two-stage orthogonal design

In the following we propose an orthogonal training design such that only matrix multiplications instead of the matrix pseudoinverse (8) are required during the channel estimation process. Thereby, the involved computational complexity is reduced.

Let us divide N_t time slots into $N_{\mathrm{t,R}}$ frames, where each frame consists of $N_{\mathrm{t,T}}$ OFDM symbols, i.e., $N_\mathrm{t} = N_{\mathrm{t,T}} \cdot N_{\mathrm{t,R}}$. The RF decoding matrix stays constant during each frame while the RF precoding matrices and the training symbols used in different frames are the same. That is, we have

$$\begin{aligned} \boldsymbol{F}[m] &= \boldsymbol{F}_i, \quad i = \mathrm{mod}(m-1, N_{\mathrm{t,T}}) + 1 \\ \boldsymbol{s}_n[m] &= \boldsymbol{s}_{n,i}, \quad i = \mathrm{mod}(m-1, N_{\mathrm{t,T}}) + 1 \\ \boldsymbol{W}[m] &= \boldsymbol{W}_j, \quad j = \lfloor m - 1/N_{\mathrm{t,T}} \rfloor + 1, \end{aligned} \tag{10}$$

where $i \in \{1, \cdots N_{\mathrm{t,T}}\}$ and $j \in \{1, \cdots, N_{\mathrm{t,R}}\}$.

Our methodology is to first estimate the matrix product $\boldsymbol{H}_{u,j} = \boldsymbol{W}_j^{\mathrm{H}}\boldsymbol{H}_u$ during each frame, where $\boldsymbol{W}_j^{\mathrm{H}}$ is the analog decoding matrix used during the j-th frame, and then to estimate $\boldsymbol{H}_u$ by using the combined $\boldsymbol{W}_j\boldsymbol{H}_u$ over all frames.

In the first stage, the LS estimate of $\boldsymbol{H}_{u,j}$ is computed by

$$\hat{\boldsymbol{H}}_{u,j} = [\boldsymbol{Y}[(j-1)N_{\mathrm{t,T}} + 1] \quad \cdots \quad \boldsymbol{Y}[jN_{\mathrm{t,T}}]]\, \boldsymbol{C}_\mathrm{T}^{+}, \tag{11}$$

where $\boldsymbol{C}_{\mathrm{T}} = [\boldsymbol{C}[1] \quad \cdots \quad \boldsymbol{C}[N_{\mathrm{t,T}}]] \in \mathbb{C}^{LM_{\mathrm{T}} \times N_{\mathrm{f}} N_{\mathrm{t,T}}}$ should have a full row rank, i.e.,

$$LM_{\mathrm{T}} \leq N_{\mathrm{f}} N_{\mathrm{t,T}}. \tag{12}$$

An orthogonal design of $\boldsymbol{C}_{\mathrm{T}}$ requires that

$$\begin{aligned}\boldsymbol{C}_{\mathrm{T}}\boldsymbol{C}_{\mathrm{T}}^{\mathrm{H}} &= \sum_{m=1}^{N_{\mathrm{t,T}}} \boldsymbol{C}[m]\boldsymbol{C}^{\mathrm{H}}[m] \qquad (13)\\ &= \sum_{i=1}^{N_{\mathrm{t,T}}}\sum_{p=1}^{N_{\mathrm{f}}} (\boldsymbol{w}_{k_p} \otimes \boldsymbol{f}_{k_p,i})(\boldsymbol{w}_{k_p} \otimes \boldsymbol{f}_{k_p,i})^{\mathrm{H}} = \beta \boldsymbol{I}_{LM_{\mathrm{T}}},\end{aligned}$$

where $\boldsymbol{f}_{k_p,i} = \boldsymbol{F}_i \boldsymbol{s}_{k_p,i}$ and $\beta > 0$. When (13) holds, the computational complexity of (11) is dominated by one matrix multiplication of order $\mathcal{O}(N_{\mathrm{R}} N_{\mathrm{f}} N_{\mathrm{t,T}} LM_{\mathrm{T}})$. Let $\log_2(N_{\mathrm{f}})$ and $\log_2(L)$ be integer values and $\log_2(N_{\mathrm{f}}) \geq \log_2(L)$. Inspired by [15], one design of $\boldsymbol{f}_{k_p}[m]$, which satisfies (13) but does not take into account the constant modulus constraint of $\boldsymbol{F}_i$, is given by

$$(\boldsymbol{f}_{k_p,i})_a = \sqrt{\frac{P_{\mathrm{T}}}{M_{\mathrm{T}} N_{\mathrm{f}}}} e^{\frac{-j2\pi(i-1+(p-1)N_{\mathrm{t,T}})L(a-1)}{N_{\mathrm{f}} \cdot N_{\mathrm{t,T}}}}, \tag{14}$$

where $a \in \{1, \cdots, M_{\mathrm{T}}\}$ and the scaling factor comes from the power limitation of the transmitted training signal. This results in a $\beta = P_{\mathrm{T}} \cdot N_{\mathrm{t,T}}/M_{\mathrm{T}}$. With the constant modulus constraint and based on (14), we propose the design of $\boldsymbol{F}_i$ and $\boldsymbol{s}_{k_p,i}$ only for two cases, i.e., $N_{\mathrm{f}} = L$ and $N_{\mathrm{T}} \geq N_{\mathrm{f}} > L$. The proposed designs of the analog precoding matrix $\boldsymbol{F}_i$ are given by

$$(\boldsymbol{F}_i)_{a,b} = \begin{cases} e^{\frac{-j2\pi(i-1)L(a-1)}{N_{\mathrm{f}} \cdot N_{\mathrm{t,T}}}}, & N_{\mathrm{f}} = L \\ e^{\frac{-j2\pi(i-1+(p-1)N_{\mathrm{t,T}})L(a-1)}{N_{\mathrm{f}} \cdot N_{\mathrm{t,T}}}}, & N_{\mathrm{T}} \geq N_{\mathrm{f}} > L, \end{cases} \tag{15}$$

and the proposed corresponding designs of the pilot vectors $\boldsymbol{s}_{k_p,i}$ are

$$\boldsymbol{s}_{k_p,i} = \begin{cases} \sqrt{\frac{P_{\mathrm{T}}}{M_{\mathrm{T}} N_{\mathrm{f}} N_{\mathrm{T}}^2}} \boldsymbol{1}_{N_{\mathrm{T}}}, & N_{\mathrm{f}} = L \\ \sqrt{\frac{P_{\mathrm{T}}}{M_{\mathrm{T}} N_{\mathrm{f}}}} \boldsymbol{e}_p, & N_{\mathrm{T}} \geq N_{\mathrm{f}} > L. \end{cases} \tag{16}$$

In the second stage, the LS estimate of $\boldsymbol{H}_u$ is obtained as

$$\hat{\boldsymbol{H}}_u = \boldsymbol{W}_{\mathrm{R}}^{+} \left[\hat{\boldsymbol{H}}_{u,1}^{\mathrm{T}} \quad \cdots \quad \hat{\boldsymbol{H}}_{u,N_{\mathrm{t,R}}}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{17}$$

where $\boldsymbol{W}_{\mathrm{R}} = \left[\boldsymbol{W}_1^* \quad \cdots \quad \boldsymbol{W}_{N_{\mathrm{t,R}}}^*\right]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{R}} N_{\mathrm{t,R}} \times M_{\mathrm{R}}}$ should have a full column rank, i.e.,

$$N_{\mathrm{R}} N_{\mathrm{t,R}} \geq M_{\mathrm{R}}. \tag{18}$$

An orthogonal training design requires that $\boldsymbol{W}_{\mathrm{R}}$ has orthogonal columns, i.e., $\boldsymbol{W}_{\mathrm{R}}^{\mathrm{H}} \boldsymbol{W}_{\mathrm{R}} = \alpha \boldsymbol{I}_{M_{\mathrm{R}}}$. One option is to select the first M_{R} columns of a $N_{\mathrm{R}} N_{\mathrm{t,R}}$-by-$N_{\mathrm{R}} N_{\mathrm{t,R}}$ DFT matrix. This results in an $\alpha = N_{\mathrm{R}} N_{\mathrm{t,R}}$. With an orthogonal design of $\boldsymbol{W}_{\mathrm{R}}$ the computational complexity of (17) is dominated by $\mathcal{O}(M_{\mathrm{R}} N_{\mathrm{R}} N_{\mathrm{t,R}} LM_{\mathrm{T}})$. The overall computational complexity via this two-stage orthogonal design is $\mathcal{O}(N_{\mathrm{R}} N_{\mathrm{f}} N_{\mathrm{t}} LM_{\mathrm{T}} + M_{\mathrm{R}} N_{\mathrm{R}} N_{\mathrm{t,R}} LM_{\mathrm{T}})$, which is much lower compared to the single-stage LS design.

***Example*:** Assume that (12) and (18) are satisfied with equality, which implies that $N_{\mathrm{t,T}} = LM_{\mathrm{T}}/N_{\mathrm{f}}$, $N_{\mathrm{t,R}} = M_{\mathrm{R}}/N_{\mathrm{R}}$, and (9) is also satisfied with equality. The overall computational complexity is then $\mathcal{O}(LM_{\mathrm{T}} M_{\mathrm{R}}(LM_{\mathrm{T}} + M_{\mathrm{R}}))$. Let us define

$$\begin{aligned}
\boldsymbol{Z}_j &= [\boldsymbol{Z}[(j-1)N_{\mathrm{t,T}}+1] \quad \cdots \quad \boldsymbol{Z}[jN_{\mathrm{t,T}}]] \in \mathbb{C}^{M_{\mathrm{R}} \times N_{\mathrm{f}} N_{\mathrm{t,T}}}\\
\boldsymbol{Z}_{\mathrm{A}} &= \left[\boldsymbol{Z}_1^{\mathrm{T}} \quad \cdots \quad \boldsymbol{Z}_{N_{\mathrm{t,R}}}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{t,R}} M_{\mathrm{R}} \times N_{\mathrm{f}} N_{\mathrm{t,T}}}\\
\boldsymbol{Q}_j &= \boldsymbol{W}_j \boldsymbol{W}_j^{\mathrm{H}} \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{R}}}\\
\boldsymbol{Q}_{\mathrm{A}} &= \left[\boldsymbol{Q}_1 \quad \cdots \quad \boldsymbol{Q}_{N_{\mathrm{t,R}}}\right] \in \mathbb{C}^{M_{\mathrm{R}} \times N_{\mathrm{t,R}} M_{\mathrm{R}}},
\end{aligned}$$

and set $N_{\mathrm{f}} = \max(L, N_{\mathrm{T}})$. Then the achieved MSE via the proposed two-stage orthogonal design is computed as

$$\begin{aligned}
\mathbb{E}\{\|\hat{\boldsymbol{H}}_u - \boldsymbol{H}_u\|_{\mathrm{F}}^2\} &= \mathbb{E}\left\{\left\|\boldsymbol{W}_{\mathrm{R}}^{+}\begin{bmatrix}\boldsymbol{W}_1^{\mathrm{H}}\boldsymbol{Z}_1\\ \vdots \\ \boldsymbol{W}_{\mathrm{t,R}}^{\mathrm{H}}\boldsymbol{Z}_{\mathrm{t,R}}\end{bmatrix}\boldsymbol{C}_{\mathrm{T}}^{+}\right\|_{\mathrm{F}}^2\right\}\\
&= \frac{1}{\alpha^2\beta^2}\mathrm{Tr}\{\boldsymbol{Q}_{\mathrm{A}}\mathbb{E}\{\boldsymbol{Z}_{\mathrm{A}}\boldsymbol{C}_{\mathrm{T}}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{T}}\boldsymbol{Z}_{\mathrm{A}}^{\mathrm{H}}\}\boldsymbol{Q}_{\mathrm{A}}^{\mathrm{H}}\}\\
&= \frac{1}{\alpha^2\beta}\mathrm{Tr}\{\boldsymbol{Q}_{\mathrm{A}}\mathbb{E}\{\boldsymbol{Z}_{\mathrm{A}}\boldsymbol{Z}_{\mathrm{A}}^{\mathrm{H}}\}\boldsymbol{Q}_{\mathrm{A}}^{\mathrm{H}}\} = \frac{N_{\mathrm{f}} N_{\mathrm{t,T}}\sigma_{\mathrm{n}}^2}{\alpha^2\beta}\mathrm{Tr}\{\boldsymbol{Q}_{\mathrm{A}}\boldsymbol{Q}_{\mathrm{A}}^{\mathrm{H}}\}\\
&= \frac{N_{\mathrm{f}} N_{\mathrm{t,T}}\sigma_{\mathrm{n}}^2}{\alpha^2\beta}\sum_{j=1}^{N_{\mathrm{t,R}}}\mathrm{Tr}\{\boldsymbol{W}_j\boldsymbol{W}_j^{\mathrm{H}}\boldsymbol{W}_j\boldsymbol{W}_j^{\mathrm{H}}\}\\
&= \frac{M_{\mathrm{R}} N_{\mathrm{f}} N_{\mathrm{t,T}}\sigma_{\mathrm{n}}^2}{\alpha^2\beta}\sum_{j=1}^{N_{\mathrm{t,R}}}\mathrm{Tr}\{\boldsymbol{W}_j\boldsymbol{W}_j^{\mathrm{H}}\} = \frac{M_{\mathrm{R}}^2 N_{\mathrm{f}} N_{\mathrm{t,T}}\sigma_{\mathrm{n}}^2}{\alpha\beta}\\
&= \frac{M_{\mathrm{R}} M_{\mathrm{T}} N_{\mathrm{f}}\sigma_{\mathrm{n}}^2}{P_{\mathrm{T}}} = \frac{\max(L, N_{\mathrm{T}}) M_{\mathrm{R}} M_{\mathrm{T}}\sigma_{\mathrm{n}}^2}{P_{\mathrm{T}}},
\end{aligned} \tag{19}$$

where the facts $\boldsymbol{C}_{\mathrm{T}}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{T}} = \beta \boldsymbol{I}_{N_{\mathrm{f}} N_{\mathrm{t,T}}}$ and $\boldsymbol{W}_j^{\mathrm{H}}\boldsymbol{W}_j = M_{\mathrm{R}}\boldsymbol{I}_{N_{\mathrm{R}}}$ are used during the derivation. Note that the LS approach does not depend on the structure of $\boldsymbol{H}_\ell$.

IV. Compressed Sensing Approach

In contrast to the LS approach, the CS approach exploits the sparsity in the angular-delay domain. To reduce the mutual coherence of the array manifold of the ULA, we assume that the spatial frequencies $\mu_{\mathrm{R},\ell}$ and $\mu_{\mathrm{T},\ell}$ lie on uniform grids of dimensions $G_{\mathrm{R}} \geq M_{\mathrm{R}}$ and $G_{\mathrm{T}} \geq M_{\mathrm{T}}$, respectively. Thereby, we have $\mu_{\mathrm{R},\ell} = (\bar{i}-1)\Delta_{\mathrm{R}}$ with $\Delta_{\mathrm{R}} = 2\pi/G_{\mathrm{R}}$ ($\bar{i} \in \{1, \cdots, G_{\mathrm{R}}\}$) and $\mu_{\mathrm{T},\ell} = (\bar{j}-1)\Delta_{\mathrm{T}}$ with $\Delta_{\mathrm{T}} = 2\pi/G_{\mathrm{T}}$ ($\bar{j} \in \{1, \cdots, G_{\mathrm{T}}\}$), respectively. This is a common assumption in the literature, e.g., [7], [10], [11], [12], [17]. Under this assumption, the array manifold $\boldsymbol{A}_{\mathrm{R}} = [\boldsymbol{a}(\mu_{\mathrm{R},1}) \quad \cdots \quad \boldsymbol{a}(\mu_{\mathrm{R},G_{\mathrm{R}}})] \in \mathbb{C}^{M_{\mathrm{R}} \times G_{\mathrm{R}}}$ and $\boldsymbol{A}_{\mathrm{T}} = [\boldsymbol{a}(\mu_{\mathrm{T},1}) \quad \cdots \quad \boldsymbol{a}(\mu_{\mathrm{T},G_{\mathrm{T}}})] \in \mathbb{C}^{M_{\mathrm{T}} \times G_{\mathrm{T}}}$ are known matrices. The discrete CTF in (3) can be rewritten as

$$\boldsymbol{H}_n = \sum_{\ell=0}^{L-1} \boldsymbol{A}_{\mathrm{R}} \boldsymbol{H}_{\nu,\ell} \boldsymbol{A}_{\mathrm{T}}^{\mathrm{H}} e^{-j2\pi\frac{\ell \cdot n}{N_{\mathrm{fft}}}}, \tag{20}$$

where $\boldsymbol{H}_{\nu,\ell} \in \mathbb{C}^{G_{\mathrm{R}} \times G_{\mathrm{T}}}$ contains just one non-zero element α_ℓ. In other words, $\boldsymbol{H}_{\nu,\ell}$ is sparse. Equation (20) is now a linear function of $\boldsymbol{H}_{\nu,\ell}$. The rest of this section is also split into two parts, where a general CS based channel estimation method via a single-stage design is developed in Section IV-A and its reduced-complexity version is proposed in Section IV-B.

IV-A. A single-stage CS approach

By inserting (20) into (1) and applying some algebraic manipulations we obtain

$$\begin{aligned}\boldsymbol{y}_n[m] = \boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{A}_{\mathrm{R}}\boldsymbol{H}_{\nu}(\boldsymbol{w}_n \otimes \boldsymbol{A}_{\mathrm{T}}^{\mathrm{H}})\boldsymbol{F}[m]\boldsymbol{s}_n[m] \\ + \boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{z}_n[m],\end{aligned} \tag{21}$$

where

$$\boldsymbol{H}_{\nu} = [\boldsymbol{H}_{\nu,0} \quad \cdots \quad \boldsymbol{H}_{\nu,L-1}] \in \mathbb{C}^{G_{\mathrm{R}} \times LG_{\mathrm{T}}}.$$

Similarly as in equation (6) of Section III, to fully exploit the time-frequency resources, we can first stack $\boldsymbol{y}_n[m]$ on top of each other along the frequency domain, which yields

$$\boldsymbol{y}[m] = \left(\boldsymbol{B}^{\mathrm{T}}[m] \otimes (\boldsymbol{W}^{\mathrm{H}}[m]\boldsymbol{A}_{\mathrm{R}})\right)\boldsymbol{h}_{\nu} + \boldsymbol{z}[m], \tag{22}$$

where $\boldsymbol{h}_{\nu} = \mathrm{vec}\{\boldsymbol{H}_{\nu}\} \in \mathbb{C}^{LG_{\mathrm{R}}G_{\mathrm{T}}}$ is a sparse vector containing $L \ll LG_{\mathrm{R}}G_{\mathrm{T}}$ non-zero elements. The p-th column of $\boldsymbol{B}[m] \in \mathbb{C}^{LG_{\mathrm{T}} \times N_{\mathrm{f}}}$ is given by $(\boldsymbol{w}_{k_p} \otimes \boldsymbol{A}_{\mathrm{T}}^{\mathrm{H}})\boldsymbol{F}[m]\boldsymbol{s}_{k_p}[m]$ for $p \in \{1, \cdots, N_{\mathrm{f}}\}$. Then we can stack $\boldsymbol{y}[m]$ along the time domain on top of each other as in (7) and obtain

$$\boldsymbol{y}_{\mathrm{s}} = \boldsymbol{P}_2\boldsymbol{h}_{\nu} + \boldsymbol{z}_{\mathrm{s}} \in \mathbb{C}^{N_{\mathrm{R}}N_{\mathrm{f}}N_{\mathrm{t}}}, \tag{23}$$

where

$$\boldsymbol{P}_2 = \begin{bmatrix} \boldsymbol{B}^{\mathrm{T}}[1] \otimes (\boldsymbol{W}^{\mathrm{H}}[1]\boldsymbol{A}_{\mathrm{R}}) \\ \vdots \\ \boldsymbol{B}^{\mathrm{T}}[N_{\mathrm{t}}] \otimes (\boldsymbol{W}^{\mathrm{H}}[N_{\mathrm{t}}]\boldsymbol{A}_{\mathrm{R}}) \end{bmatrix} \in \mathbb{C}^{N_{\mathrm{t}}N_{\mathrm{f}}N_{\mathrm{R}} \times LG_{\mathrm{T}}G_{\mathrm{R}}}.$$

The formulation (23) fulfills a sparse recovery problem and thus any sparse approximation algorithm in [18] can be applied. In this paper we apply the OMP algorithm. To ensure that $\boldsymbol{h}_u$ can be uniquely and stably determined, the matrix $\boldsymbol{P}_2$ should be constructed such that the restricted isometry property (RIP) is satisfied. In practice there are no algorithms which could check the RIP for a given matrix in polynomial time. But there are certain probabilistic constructions of matrices satisfying the RIP with high probability, i.e., constructing $\boldsymbol{P}_2$ with randomly distributed elements, e.g., Gaussian or sub-Gaussian distributions, or constructing $\boldsymbol{P}_2$ to possess randomly selected rows of a DFT matrix [14]. The latter one is difficult to achieve because $\boldsymbol{P}_2$ has a complicated structure and the analog precoding and decoding matrix must consist of constant modulus entries. The former one is our choice in this paper. According to [19], when $\boldsymbol{P}_2$ is sufficiently random, an asymptotic reliable recovery of $\boldsymbol{h}_{\nu}$ with a sparsity order of L can be achieved if

$$N_{\mathrm{t}}N_{\mathrm{f}}N_{\mathrm{R}} \geq 2L\log(LG_{\mathrm{T}}G_{\mathrm{R}}) \tag{24}$$

as $L \to +\infty$. Numerically we find that this condition is approximately true when the phases of the elements of $\boldsymbol{F}[m]$ and $\boldsymbol{W}[m]$ are uniformly distributed over the interval $[0, 2\pi)$. As the pilot vectors $\boldsymbol{s}_{k_p,i}$ have almost no influence on the performance, they are identical over all the OFDM symbols. We propose

$$\boldsymbol{s}_{k_p,i} = \begin{cases} \sqrt{\frac{P_{\mathrm{T}}}{M_{\mathrm{T}}N_{\mathrm{T}}}}\boldsymbol{e}_p, & N_{\mathrm{f}} \leq N_{\mathrm{T}} \\ \sqrt{\frac{P_{\mathrm{T}}}{M_{\mathrm{T}}N_{\mathrm{T}}N_{\mathrm{f}}}}\boldsymbol{d}_p, & \mathrm{Otherwise}, \end{cases} \tag{25}$$

where $\boldsymbol{d}_p = \left[1 \quad \ldots \quad e^{-j2\pi\frac{(N_{\mathrm{T}}-1)(p-1)}{N_{\mathrm{f}}}}\right]^{\mathrm{T}}$. The computational complexity of the OMP algorithm is dominated by $\mathcal{O}(N_{\mathrm{t}}N_{\mathrm{f}}N_{\mathrm{R}}L^2G_{\mathrm{T}}G_{\mathrm{R}})$. Note that in the numerical experiments we use a uniform linear array.

IV-B. A reduced complexity two-stage CS approach

Now we introduce a two-stage sparse recovery algorithm. It requires relatively more training symbols compared to the single-stage method discussed above. But it has a more tractable training design, which might be useful for theoretical analysis, and has a reduced computational complexity. The same training procedure as described in Section III-B is used, i.e., the training process is split into $N_{\mathrm{t,R}}$ frames each with $N_{\mathrm{t,T}}$ OFDM symbols. Again, it is assumed that the analog decoding matrices are reused within each frame and the analog precoding matrices as well as the training symbols are reused in different frames. Mathematically, equation (10) holds.

In the first stage, i.e., in each frame, we estimate the matrix product $\boldsymbol{H}_{\mathrm{R},j} = \boldsymbol{W}_j^{\mathrm{H}}\boldsymbol{A}_{\mathrm{R}}\boldsymbol{H}_{\nu} \in \mathbb{C}^{N_{\mathrm{R}} \times LG_{\mathrm{T}}}$ using the CS approximation algorithm. This is motivated by the fact that matrix $\boldsymbol{H}_{\mathrm{R},j}$ contains at most $L \ll LG_{\mathrm{T}}$ columns with non-zero elements. Then, by stacking $\boldsymbol{y}_n[m]$ next to each other we obtain the following equation

$$\bar{\boldsymbol{Y}}_j = \boldsymbol{P}_3 \cdot \boldsymbol{H}_{\mathrm{R},j}^{\mathrm{T}} + \bar{\boldsymbol{Z}}_j, \tag{26}$$

where $\bar{\boldsymbol{Z}}_j = \left[\boldsymbol{W}_j^{\mathrm{H}}\boldsymbol{Z}[(j-1)N_{\mathrm{t,T}}+1] \quad \cdots \quad \boldsymbol{W}_j^{\mathrm{H}}\boldsymbol{Z}[jN_{\mathrm{t,T}}]\right]^{\mathrm{T}}$, $\bar{\boldsymbol{Y}}_j = [\boldsymbol{Y}[(j-1)N_{\mathrm{t,T}}+1] \quad \cdots \quad \boldsymbol{Y}[jN_{\mathrm{t,T}}]]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{f}}N_{\mathrm{t,T}} \times N_{\mathrm{R}}}$, and

$$\boldsymbol{P}_3 = [\boldsymbol{B}[1] \quad \cdots \quad \boldsymbol{B}[N_{\mathrm{t,T}}]]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{f}}N_{\mathrm{t,T}} \times LG_{\mathrm{T}}}.$$

Since each column of $\boldsymbol{H}_{\mathrm{R},j}^{\mathrm{T}}$ has at most $L \ll LG_{\mathrm{T}}$ non-zero elements, (26) is a sparse recovery problem and more precisely, a MMV problem [20]. Nevertheless, to solve (26), we apply the OMP algorithm column wise, i.e., considering only a single measurement vector. Thereby, the matrix $\boldsymbol{H}_{\mathrm{R},j}^{\mathrm{T}}$ is estimated column by column. Within this stage, the involved computational complexity is $\mathcal{O}(N_{\mathrm{f}}N_{\mathrm{t,T}}N_{\mathrm{R}}L^2G_{\mathrm{T}})$. The asymptotic recovery condition for OMP requires that

$$N_{\mathrm{f}}N_{\mathrm{t,T}} \geq 2L\log(LG_{\mathrm{T}}). \tag{27}$$

In the second stage, we recover $\boldsymbol{H}_{\nu}$ using the estimated $\hat{\boldsymbol{H}}_{\mathrm{R},j}$. Without loss of generality, $\hat{\boldsymbol{H}}_{\mathrm{R},j}$ can be rewritten as

$$\hat{\boldsymbol{H}}_{\mathrm{R},j} = \boldsymbol{W}_j^{\mathrm{H}}\boldsymbol{A}_{\mathrm{R}}\boldsymbol{H}_{\nu} + \boldsymbol{\Delta}_j, \tag{28}$$

where $\boldsymbol{\Delta}_j$ represents the estimation error. We stack $\hat{\boldsymbol{H}}_{\mathrm{R},j}$ on top of each other and obtain

$$\hat{\boldsymbol{H}}_{\mathrm{cum}} = \boldsymbol{P}_4\boldsymbol{H}_{\nu} + \boldsymbol{\Delta}_{\mathrm{cum}} \in \mathbb{C}^{N_{\mathrm{t,R}}N_{\mathrm{R}} \times LG_{\mathrm{T}}} \tag{29}$$

Table I. A comparison of the required training resource and the computational complexity under ideal conditions

Algorithm	Training Procedure	Minimum required training resource ($N_\mathrm{f} \cdot N_{\mathrm{t,T}} \cdot N_{\mathrm{t,R}}$)	Computational complexity
LS	Single-stage	$LM_\mathrm{T}M_\mathrm{R}/N_\mathrm{R}$	$\mathcal{O}((LM_\mathrm{T}M_\mathrm{R})^3/6)$
	Two-stage	$LM_\mathrm{T}M_\mathrm{R}/N_\mathrm{R}$	$\mathcal{O}(LM_\mathrm{T}M_\mathrm{R}(LM_\mathrm{T}+M_\mathrm{R}))$
CS	Single-stage	$\lceil 2L\log(LG_\mathrm{T}G_\mathrm{R})/N_\mathrm{R}\rceil$	$\mathcal{O}(2L^3G_\mathrm{T}G_\mathrm{R}\log(LG_\mathrm{T}G_\mathrm{R}))$
	Two-stage	$\lceil 4L\log(LG_\mathrm{T})\log(G_\mathrm{R})/N_\mathrm{R}\rceil$	$\mathcal{O}(2L^3N_\mathrm{R}G_\mathrm{T}\log(LG_\mathrm{T})+2LG_\mathrm{T}G_\mathrm{R}\log(G_\mathrm{R}))$

Table II. An overview of the proposed training design

Algorithm	$\boldsymbol{s}_{k_p}[m]$ [a]	$\boldsymbol{F}[m]$	$\boldsymbol{W}[m]$
Orthogonal LS design	(16)	(15)	$\boldsymbol{W}_j^\mathrm{H}$ is the j-th sub-matrix containing consecutive N_R rows of a DFT matrix
Single-stage CS	(25)	$\angle((\boldsymbol{F}[m])_{a,b}) \sim \mathcal{U}(0,2\pi)$	$\angle((\boldsymbol{W}[m])_{c,d}) \sim \mathcal{U}(0,2\pi)$
Two-stage CS	(25)	$\angle((\boldsymbol{F}_i)_{a,b}) \sim \mathcal{U}(0,2\pi)$	$\angle((\boldsymbol{W}_j)_{c,d}) \sim \mathcal{U}(0,2\pi)$

[a] Equal-powered and equal-spaced pilot tones are used.

where $\hat{\boldsymbol{H}}_\mathrm{cum} = \begin{bmatrix}\hat{\boldsymbol{H}}_{\mathrm{R},1}^\mathrm{T} & \cdots & \hat{\boldsymbol{H}}_{\mathrm{R},N_{\mathrm{t,R}}}^\mathrm{T}\end{bmatrix}^\mathrm{T}$, $\boldsymbol{P}_4 = \boldsymbol{W}_\mathrm{cs}\boldsymbol{A}_\mathrm{R} \in \mathbb{C}^{N_{\mathrm{t,R}}N_\mathrm{R}\times G_\mathrm{R}}$, $\boldsymbol{W}_\mathrm{cs} = \begin{bmatrix}\boldsymbol{W}_1^* & \cdots & \boldsymbol{W}_{N_{\mathrm{t,R}}}^*\end{bmatrix}^\mathrm{T} \in \mathbb{C}^{N_{\mathrm{t,R}}N_\mathrm{R}\times M_\mathrm{R}}$, and $\boldsymbol{\Delta}_\mathrm{cum} = \begin{bmatrix}\boldsymbol{\Delta}_1^\mathrm{T} & \cdots & \boldsymbol{\Delta}_{N_{\mathrm{t,R}}}^\mathrm{T}\end{bmatrix}^\mathrm{T}$. Recovering $\boldsymbol{H}_\nu$ from (29) is again a MMV sparse recovery problem. But the sparsity profile on each column of $\boldsymbol{H}_\nu$ is not the same. A single measurement vector based sparse recovery method is used. We split equation (29) into L sub-equations and each sub-equation is given by

$$\hat{\boldsymbol{H}}_{\mathrm{cum},\ell} = \boldsymbol{P}_4\boldsymbol{H}_{\nu,\ell} + \boldsymbol{\Delta}_{\mathrm{cum},\ell} \in \mathbb{C}^{N_{\mathrm{t,R}}N_\mathrm{R}\times G_\mathrm{T}}, \tag{30}$$

where $\hat{\boldsymbol{H}}_{\mathrm{cum},\ell}$ and $\boldsymbol{\Delta}_{\mathrm{cum},\ell}$ denote the ℓ-th sub-matrix of $\hat{\boldsymbol{H}}_\mathrm{cum}$ and $\boldsymbol{\Delta}_\mathrm{cum}$ consisting of G_T columns, respectively. To estimate $\boldsymbol{H}_{\nu,\ell}$, the vectorized representation of (30) is used and yields

$$\hat{\boldsymbol{h}}_{\mathrm{cum},\ell} = \boldsymbol{P}_5\boldsymbol{h}_{\nu,\ell} + \boldsymbol{\delta}_{\mathrm{cum},\ell} \in \mathbb{C}^{N_{\mathrm{t,R}}N_\mathrm{R}G_\mathrm{T}}, \tag{31}$$

where $\boldsymbol{P}_5 = \boldsymbol{I}_{G_\mathrm{T}} \otimes \boldsymbol{P}_4 \in \mathbb{C}^{N_{\mathrm{t,R}}N_\mathrm{R}G_\mathrm{T}\times G_\mathrm{R}G_\mathrm{T}}$, $\hat{\boldsymbol{h}}_{\mathrm{cum},\ell} = \mathrm{vec}\{\hat{\boldsymbol{H}}_{\mathrm{cum},\ell}\}$, $\boldsymbol{h}_{\nu,\ell} = \mathrm{vec}\{\boldsymbol{H}_{\nu,\ell}\}$, and $\boldsymbol{\delta}_{\mathrm{cum},\ell} = \mathrm{vec}\{\boldsymbol{\Delta}_{\mathrm{cum},\ell}\}$. Now $\boldsymbol{h}_{\nu,\ell}$ can be recovered by using the OMP algorithm. Nevertheless, recalling that $\boldsymbol{H}_{\nu,\ell}$ contains only one non-zero element, the OMP algorithm can be modified so that a more computationally efficient algorithm is obtained. The proposed modified OMP algorithm for solving (30) contains two steps, first, the p-th column of $\boldsymbol{P}_4$, which is most correlated with the q-th column of $\hat{\boldsymbol{H}}_{\mathrm{cum},\ell}$ is computed by $(p,q) = \arg\max_{\bar{j}}\max_{\bar{i}} |\boldsymbol{p}_{4,\bar{i}}^\mathrm{H}\cdot\hat{\boldsymbol{h}}_{\mathrm{cum},\ell,\bar{j}}|$, where $\boldsymbol{p}_{4,\bar{i}}$ and $\hat{\boldsymbol{h}}_{\mathrm{cum},\ell,\bar{j}}$ denote the $\bar{i}$-th and the $\bar{j}$-th column of $\boldsymbol{P}_4$ and $\hat{\boldsymbol{H}}_{\mathrm{cum},\ell}$, respectively. Second, the LS solution of the corresponding non-zero element is given by $b_\ell = \boldsymbol{p}_{4,p}^\mathrm{H}\cdot\hat{\boldsymbol{h}}_{\mathrm{cum},\ell,q}/\|\boldsymbol{p}_{4,p}\|^2$. In this stage the computational complexity is dominated by $\mathcal{O}(N_{\mathrm{t,R}}N_\mathrm{R}LG_\mathrm{T}G_\mathrm{R})$. The asymptotic recovery condition for OMP requires that

$$N_\mathrm{R}N_{\mathrm{t,R}} \geq 2\log(G_\mathrm{R}). \tag{32}$$

Note that a similarly training scheme as in the single-stage approach will be used for constructing $\boldsymbol{W}_j$, $\boldsymbol{F}_i$, and $\boldsymbol{s}_{k_p,i}$, $\forall n,m$.

Summarizing, let us assume that the asymptotic recovery conditions (24), (27), and (32) for the OMP algorithms are tight, and sufficient conditions (9), (12), and (18) for the LS algorithms are satisfied with equality. A comparison between the LS approach and the CS approach with a single-stage or a two-stage estimation procedure in terms of the total number of required time-frequency resources and the involved computational complexity is summarized in Table I. The proposed training design is specified in Table II. It is can be seen that in general the CS approach requires less training and has a lower computational complexity. However, the CS approach is sensitive to the sparsity profile of $\boldsymbol{H}_{\nu,\ell}$.

V. SIMULATION RESULTS

The proposed algorithms are evaluated using Monte-Carlo simulations. The maximum allowable power P_T is set to unity. The SNR for channel estimation is defined as $\mathrm{SNR}_\mathrm{pilot} = 1/(N_\mathrm{fft}\sigma_n^2)$ while the SNR for data transmission is defined as $\mathrm{SNR}_\mathrm{sig} = P_\mathrm{sig}/(N_\mathrm{data}\sigma_n^2)$, where N_data denotes the number of active data subcarriers. The maximum power used for data transmission P_sig is also set to unity. We set $N_\mathrm{fft} = 128$ and $L = 8$. As in [6], a uniform linear array (ULA) geometry is used at both ends. The inter-element spacing of the ULA is equal to half of the wavelength. The array steering vector at both ends (BS and UE) consisting of M elements is then defined as

$$\boldsymbol{a}(\mu) = \begin{bmatrix}1 & e^{-j\mu} & \cdots & e^{-j(M-1)\mu}\end{bmatrix}^\mathrm{T}, \tag{33}$$

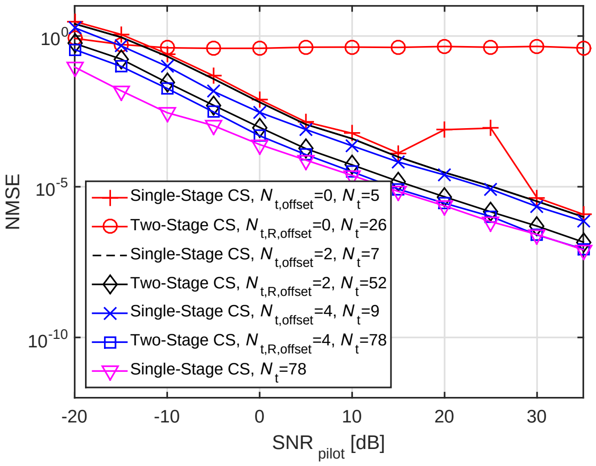

Fig. 2. A reliability test of the asymptotic recovery conditions (24), (27), and (32) for the OMP algorithm for $M_{\mathrm{T}} = 64$, $M_{\mathrm{R}} = 32$, $N_{\mathrm{T}} = 8$, and $N_{\mathrm{R}} = 4$.

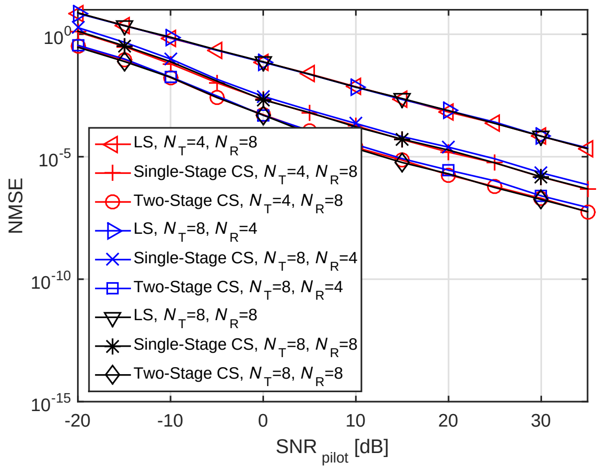

Fig. 3. Achievable NMSE vs. $\mathrm{SNR}_{\mathrm{pilot}}$ for different number of RF chains with $M_{\mathrm{T}} = 64$ and $M_{\mathrm{R}} = 32$.

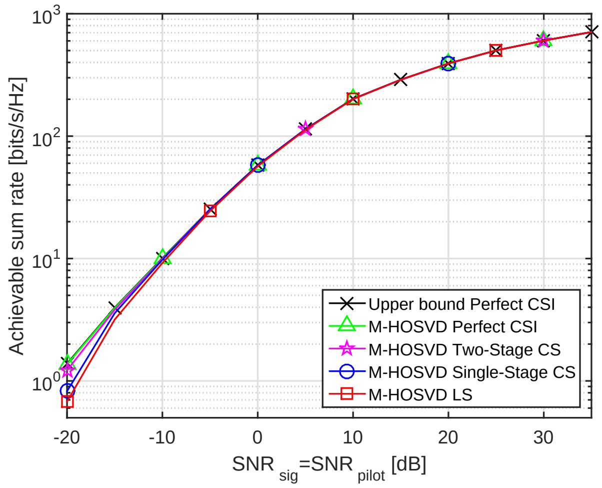

Fig. 4. Achievable sum rate vs. $\mathrm{SNR}_{\mathrm{sig}}$ for $M_{\mathrm{T}} = 64$, $M_{\mathrm{R}} = 32$, $N_{\mathrm{T}} = 8$, $N_{\mathrm{R}} = 4$, $N_{\mathrm{data}} = 16$, and $N_{\mathrm{ss}} = 4$.

where μ denotes the spatial frequency. Moreover, we set $G_{\mathrm{T}} = M_{\mathrm{T}}$ and $G_{\mathrm{R}} = M_{\mathrm{R}}$. We set $N_{\mathrm{f}} = \max(N_{\mathrm{T}}, L)$ for all the proposed algorithms, which accounts for a pilot overhead of $\frac{\max(N_{\mathrm{T}},L)}{N_{\mathrm{fft}}}\%$ per OFDM symbol. The LS method via the orthogonal design, the single-stage sparse recovery method, and the two-stage sparse recovery method are denoted as "LS", "Single-Stage CS", and "Two-Stage CS", respectively. The simulation results are obtained by averaging over 100 channel realizations.

In Fig. 2 a reliability test of the asymptotic recovery conditions (24), (27), and (32) has been performed for the proposed single-stage and two-stage CS approaches. The normalized MSE (NMSE) is defined as $\mathrm{NMSE} = \mathbb{E}\{\|\boldsymbol{H}_u - \boldsymbol{H}_{u,\mathrm{est}}\|_{\mathrm{F}}^2 / \|\boldsymbol{H}_u\|_{\mathrm{F}}^2\}$. For the single-stage method we increase the number of measurements by adding $N_{\mathrm{t,offset}}$ to the lower bound of $N_{\mathrm{t,low}} = \lceil 2\log(LG_{\mathrm{T}}G_{\mathrm{R}})/N_{\mathrm{R}} \rceil$. For the two-stage method $N_{\mathrm{t,T}}$ is fixed as $\lceil 2\log(LG_{\mathrm{T}}) \rceil$ and the total number of measurements is increased by adding $N_{\mathrm{t,R,offset}}$ to the lower bound of $N_{\mathrm{t,R,low}} = \lceil 2\log(G_{\mathrm{R}})/N_{\mathrm{R}} \rceil$. Numerical results show that this bound is not tight for our problem. When the two approaches use the same number of OFDM training symbols, the single-stage method is slightly better than the two-stage method. To preserve a stable performance, in the following simulations we will set $N_{\mathrm{t,offset}} = N_{\mathrm{t,R,offset}} = 4$. Note that even with this increased number of training symbols the CS approaches still require much less OFDM training symbols compared to the LS method. For example, with the same number of antennas and RF chains the LS method requires at least $N_{\mathrm{t}} = 512$.

The performance of different channel estimation algorithms with varying numbers of RF chains is compared in Fig. 3. In general the CS based algorithms outperforms the LS based algorithm. The LS algorithm provides almost the same results under different settings. This coincides with our analytic result (19) since for all settings $N_{\mathrm{f}} = 8$. The performance of the CS based algorithms is insensitive to the change of N_{T} but gets slightly better when N_{R} increases.

To compare the achievable sum rate of the system with estimated channel state information (CSI), we apply a modified truncated higher-order SVD (M-HOSVD) based suboptimal solution together with the waterfilling power allocation scheme. The difference between the M-HOSVD algorithm and the original HOSVD algorithm in [21] lies on the design of the digital precoder and decoder per subcarrier. Let $\boldsymbol{F}_{\mathrm{RF}}$ and $\boldsymbol{W}_{\mathrm{RF}}^{\mathrm{H}}$ denote the obtained RF precoder and decoder using the HOSVD algorithm. Define the economic SVD of $\boldsymbol{F}_{\mathrm{RF}} = \boldsymbol{U}_{T,s}\boldsymbol{\Sigma}_{T,s}\boldsymbol{V}_{T,s}^{\mathrm{H}}$ and $\boldsymbol{W}_{\mathrm{RF}}^{\mathrm{H}} = \boldsymbol{U}_{R,s}\boldsymbol{\Sigma}_{R,s}\boldsymbol{V}_{R,s}^{\mathrm{H}}$. Define the rank-$N_{\mathrm{ss}}$ truncated SVD of $\boldsymbol{\Sigma}_{R,s}^{-1}\boldsymbol{U}_{R,s}^{\mathrm{H}}\boldsymbol{H}_n\boldsymbol{V}_{T,s}\boldsymbol{\Sigma}_{T,s}^{-1} \approx \boldsymbol{U}_{\mathrm{trun}}\boldsymbol{\Sigma}_{\mathrm{trun}}\boldsymbol{V}_{\mathrm{trun}}^{\mathrm{H}}$. Then the digital baseband precoder and decoder for the n-th data subcarrier are calculated as $\boldsymbol{F}_{\mathrm{BB},n} = \boldsymbol{V}_{T,s}\boldsymbol{\Sigma}_{T,s}^{-1}\boldsymbol{V}_{\mathrm{trun}} \in \mathbb{C}^{N_{\mathrm{T}}\times N_{\mathrm{ss}}}$ and $\boldsymbol{W}_{\mathrm{BB},n}^{\mathrm{H}} = \boldsymbol{U}_{\mathrm{trun}}^{\mathrm{H}}\boldsymbol{\Sigma}_{R,s}^{-1}\boldsymbol{U}_{R,s}^{\mathrm{H}} \in \mathbb{C}^{N_{\mathrm{ss}}\times N_{\mathrm{R}}}$, respectively. Moreover, the digital upper bound in [21] using the perfect CSI and the waterfilling power allocation scheme is also plotted as a benchmark. Since our focus is on the effects of imperfect CSI rather than the overall system spectral efficiency, the demonstrated sum rate comparison in Fig. 4 accounts only for the data transmission phase. During the data transmission phase, the analog as well as digital precoding and decoding schemes are modified according the M-HOSVD algorithm and the estimated channel. In other words, the proposed analog training matrices will be replaced. Moreover, to reduce the computational complexity we use $N_{\text{ss}} = 4$ spatial streams [21]. Fig. 4 shows that the achievable sum rate is only slightly affected when the estimated CSI is used. There is almost no difference when different channel estimation schemes are used and the SNR is not too small (e.g., ≥ -10 dB).

VI. CONCLUSION

In this paper we have developed pilot based channel estimation methods for estimating the CIR for a multi-carrier single-user massive MIMO system. Two different estimation methods are used, i.e., the LS method and the CS method. For both methods sufficient conditions for a unique channel estimate are derived. Reduced-complexity versions of both methods are developed based on a two-stage channel estimation procedure. Simulation results show that the CS based methods require a reduced number of training symbols and provide a better performance compared to the LS method. Moreover, the channel estimation error only slightly affects the achievable system sum rate when a HOSVD based hybrid precoding and decoding scheme is used. This conclusion is obtained assuming that the spatial frequencies of the MIMO channel lie on uniform grids.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support by the Carl-Zeiss-Foundation (http://carl-zeiss-stiftung.de/).

VII. REFERENCES

[1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: opportunities and challenges with very large arrays,” *Signal Processing Magazine*, vol. 30, no. 1, 2013.

[2] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, “Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results,” *IEEE Communications Magazine*, pp. 106–113, Feb. 2014.

[3] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, “Millimeter-wave massive MIMO: The next wireless revolution?,” *IEEE Communications Magazine*, pp. 56–62, Sept. 2014.

[4] S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter-wave cellular wireless networks: Potentials and challenges,” *Proceedings of the IEEE*, vol. 102, pp. 366–385, Mar. 2014.

[5] R. Mendez-Rial, C. Rusu, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, Jr., “Channel estimation and hybrid combining for mmWave: phase shifters or switches,” in *Proc. of Information Theory and Applications Workshop (ITA 2015)*, UCSD, USA, Feb. 2015.

[6] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 13, pp. 1499–1513, Mar. 2014.

[7] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., “Channel estimation and hybrid precoding for millimeter wave cellular systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, pp. 831–846, Oct. 2014.

[8] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., “Achievable rates of multi-user millimeter wave systems with hybrid precoding,” in *Proc. IEEE International Conference on Communications (ICC 2015)*, London, UK, June 2015.

[9] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, “On the number of RF chains and phase shifters, and scheduling design with hybrid analog-digital beamforming,” *submitted to IEEE Transactions on Wireless Communications*, pp. 1–30, 2015.

[10] R. M.-Rial, C. Rusu, A. Alkhateeb, N. G.-Prelcic, and R. W. Heath, Jr., “Channel estimation and hybrid combining for mmwave: phase shifters or switches,” in *Proc. IEEE Information Theory and Applications Workshop (ITA)*, 2015.

[11] J. Lee, G.-T. Gil, and Y. H. Lee, “Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications,” in *Proc. IEEE Global Communications Conference (GLOBECOM 2014)*, Austin, TX USA, Dec. 2014.

[12] T. E. Bogale, L. B. Le, and X. Wang, “Hybrid analog-digital channel estimation and beamforming: Training-throughput tradeoff,” *submitted to IEEE Transactions on Communications, available at http://http://arxiv.org/abs/1509.0509*, pp. 1–15, 2015.

[13] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösti, D. Laurenson, G. Matz, A. F. Molisch, C. Oestges, and H. Özcelik, “Survey of channel and radio propagation models for wireless MIMO systems,” *EURASIP Journal on Wireless Communications and Networking*, Jan. 2007.

[14] 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, pp. 1058–1076, June 2010.

[15] I. Barhumi, G. Leus, and M. Moonen, “Optimal training design for MIMO OFDM systems in mobile wireless channels,” *IEEE Transactions on Signal Processing*, vol. 51, pp. 1615–1624, June 2003.

[16] A. Björck, *Numerical methods for least squares problems*, Society for Industrial and Applied Mathematics, 1996.

[17] P. Schniter and A. Sayeed, “Channel estimation and precoder design for millimeter-wave communications: the sparse way,” in *Proc. of the 48th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2014.

[18] J. A. Tropp and S. J. Wright, “Computational methods for sparse solution of linear inverse problems,” *Proceedings of the IEEE*, vol. 98, pp. 948–958, June 2010.

[19] A. Fletcher and S. Rangan, “Orthogonal matching pursuit from noisy random measurements: a new analysis,” in *Proc. 234d Ann. Conf. Neural Information Processing Systems*, Vancouver, Canada, Dec. 2009.

[20] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 53, pp. 2477–2488, July 2005.

[21] J. Zhang, A. Wiesel, and M. Haardt, “Low rank approximation based hybrid precoding schemes for multi-carrier single-user massive MIMO systems,” in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016.