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

Jianshu Zhang[1], *Ivan Podkurkov*[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 phase shift networks, which imposes constant modulus constraints on the RF precoding and decoding matrices. Second, there is just one common equivalent RF precoding or decoding matrix for all subcarriers. This also leads to a challenging channel estimation task and an optimal training design for the considered system. We derive both linear and non-linear channel estimation methods. More specifically, a linear estimation via lease squares (LS) together with a specific optimal pilot sequence design is proposed. Moreover, a non-linear channel estimation counterpart is developed via two-stage compressed sensing (CS). Initial simulation results show that the CS approach outperforms the LS approach.

Index Terms— Massive MIMO, hybrid precoding, mmWave, 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 under-utilized 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], [9]. Furthermore, when a wideband multi-carrier system is considered, equivalently we get the same phase shifts for all subcarriers [10]. 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 [8], [11], [12], [13]. In [8] 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 [11] by involving multiple measurement vectors (MMV) to improve the channel estimation accuracy. The CS based concept is also used in [12], where an adaptive multi-grid sparse recovery approach is applied instead. Finally, a multi-user hybrid analog-digital system is considered in [13] 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. The 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., the least squares (LS) approach, and the CS approach. The former one is a linear method while the latter one can be interpreted as a non-linear method. In both cases we have to transmit multiple OFDM training symbols. Nevertheless, it is still possible to build optimal pilot tones in the minimum squared error (MSE) sense of the LS estimate and one optimal pilot sequence design is provided in this paper. The CS approach exploits the sparsity of the channel model. To avoid a training design based on random numbers, which might result in practical implementation difficulties, we develop a two-stage CS based channel estimation method, where the pilot sequences and the RF matrices use a DFT-like training design. Our initial simulation results show that the CS approach outperforms the LS approach.

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 (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 following integer is denoted by $\lceil\cdot\rceil$.

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 equipments (UE). 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 multicarrier modulation technique is applied to combat the multipath effect. The corresponding FFT size is N_{fft}. Let $\boldsymbol{s}_n[m] \in \mathbb{C}^{N_{\mathrm{ss}}}$ represent the transmitted pilot sequence on the n-th subcarrier in the m-th OFDM symbol ($n \in \{k_1, \cdots, k_{N_{\mathrm{f}}}\} \subset \{1, \cdots, N_{\mathrm{fft}}\}$, $m \in \{1, \cdots, N_{\mathrm{t}}\}$). Therefore, the training procedure consists of N_{t} OFDM symbols each with N_{f} pilot tones. We have $N_{\mathrm{T}} \geq N_{\mathrm{ss}}$. The training pilots and the data are interleaved on all the subcarriers and then pass through the IFFT filter. Furthermore, we assume that the pilot tones are assigned equally spaced. 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., $|(F[m])_{a,b}| = 1$ 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 $\sum_{n=k_1}^{k_{N_{\mathrm{f}}}} \|\boldsymbol{F}[m]\boldsymbol{s}_n[m]\|^2 \leq P$ 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. 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. It is assumed that the channel remains unchanged during the training procedure. The received pilot signal on the n-th subcarrier in the m-th OFDM symbol is given by [10]

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

where $\boldsymbol{z}_n[m]$ is the 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. In this paper $\|\boldsymbol{W}[m]\|_{\mathrm{F}} = 1$ for all m. Note that 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 estimated at the receiver side.

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 angle of arrival (AoA) $\theta_{\mathrm{R},\ell} \in [0, 2\pi)$ and angle of departure (AoD) $\theta_{\mathrm{T},\ell} \in [0, 2\pi)$ for all $\ell \in \{0, \cdots, L-1\}$. The discrete CTF of the UE on the n-th subcarrier is modeled as [14], [15]

$$\boldsymbol{H}_n = \sum_{\ell=0}^{L-1} \underbrace{\frac{\alpha_\ell}{\sqrt{L}}\boldsymbol{a}(\theta_{\mathrm{R},\ell})\boldsymbol{a}^{\mathrm{H}}(\theta_{\mathrm{T},\ell})}_{\boldsymbol{H}_\ell \in \mathbb{C}^{M_{\mathrm{R}}\times M_{\mathrm{T}}}} e^{-j2\pi\frac{\ell \cdot n}{N_{\mathrm{FFT}}}}, \quad (2)$$

where α_ℓ is the random complex gain of the ℓ-th path, with zero mean and $\mathbb{E}\{|\alpha_\ell|^2\} = 1$. The vectors $\boldsymbol{a}(\theta_{\mathrm{T},\ell})$ and $\boldsymbol{a}(\theta_{\mathrm{R},\ell})$ are the array steering vectors of the BS and the UEs, respectively. As in [7], 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 consisting of M elements is then defined as

$$\boldsymbol{a}(\theta) = \begin{bmatrix} 1 & e^{-j2\pi\sin(\theta)} & \cdots & e^{-j2\pi(M-1)\sin(\theta)} \end{bmatrix}^{\mathrm{T}}. \quad (3)$$

Furthermore, for notational simplicity (2) implies that $\tau_\ell = \ell T_{\mathrm{s}}$, where T_{s} represents the sampling period. In general the developed channel estimation algorithms in this paper can be applied for other choices of τ_ℓ. Note that the proposed algorithms can be also applied when the rank of $\boldsymbol{H}_\ell$ is higher than one, e.g., considering L clusters each with N_{p} paths [7].

III. LEAST SQUARES APPROACH

In this section we study the LS based training design, which is a commonly used channel estimation scheme, e.g., [16]. By inserting (2) into (1) we obtain

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

where we have

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

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

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

where $\boldsymbol{C}[m] \in \mathbb{C}^{LM_\mathrm{T}\times N_\mathrm{f}}$ is computed by

$$\boldsymbol{C}[m] = \left[\boldsymbol{w}_{k_1} \otimes (\boldsymbol{F}[m]\boldsymbol{s}_{k_1}[m]) \quad \cdots \quad \boldsymbol{w}_{k_{N_\mathrm{f}}} \otimes (\boldsymbol{F}[m]\boldsymbol{s}_{k_{N_\mathrm{f}}}[m])\right].$$

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 $\boldsymbol{y}_\mathrm{s} = \left[\boldsymbol{y}^\mathrm{T}[1] \quad \cdots \quad \boldsymbol{y}^\mathrm{T}[N_\mathrm{t}]\right]^\mathrm{T}$, $\boldsymbol{z}_\mathrm{s} = \left[\boldsymbol{z}^\mathrm{T}[1] \quad \cdots \quad \boldsymbol{z}^\mathrm{T}[N_\mathrm{t}]\right]^\mathrm{T}$, and

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

Conventionally, to recover $\boldsymbol{h}_u$ from (7) over one OFDM symbol, i.e., $N_\mathrm{t} = 1$, the LS design requires that the $(N_\mathrm{R}N_\mathrm{f})$-by-$(LM_\mathrm{T}M_\mathrm{R})$ matrix product $(\boldsymbol{C}^\mathrm{T}[m] \otimes \boldsymbol{W}^\mathrm{H}[m])$ is of full column rank $LM_\mathrm{T}M_\mathrm{R}$, i.e., $N_\mathrm{R}N_\mathrm{f} \geq LM_\mathrm{T}M_\mathrm{R}$. However, it can be proven that this is not possible due to the fact that there is only one RF precoding matrix for all the subcarriers. Therefore, to recover $\boldsymbol{h}_u$ we have to use multiple OFDM symbols as in (7). Similarly, this requires that $\boldsymbol{P}_1$ has full column rank, i.e., $N_\mathrm{t}N_\mathrm{f}N_\mathrm{R} \geq LM_\mathrm{T}M_\mathrm{R}$. To this end, there should be no more than $N_\mathrm{f} = N_\mathrm{T}$ pilot tones in one OFDM symbol. As we have $M_\mathrm{T} \gg N_\mathrm{T}$ and $M_\mathrm{R} \gg N_\mathrm{R}$, this means that many training time slots (at least $\left\lceil \frac{LM_\mathrm{T}M_\mathrm{R}}{N_\mathrm{T}N_\mathrm{R}} \right\rceil$) are needed. Finally, the LS channel estimate is computed by

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

The next question is whether it is possible to develop optimal pilot tones which minimize the MSE of the LS channel estimate. Moreover, we should have $\boldsymbol{P}_1^\mathrm{H}\boldsymbol{P}_1 = \alpha\boldsymbol{I}_{LM_\mathrm{T}M_\mathrm{R}}$ with $\alpha > 0$ [17]. Inspired by [16], we describe such a design in the following. Let us divide the 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 used in different frames are the same. Our methodology is to first estimate the matrix product $\boldsymbol{W}_j^\mathrm{H}\boldsymbol{H}_u$ during each frame ($j \in \{1, \cdots, N_\mathrm{t,R}\}$) and then estimate $\boldsymbol{H}_u$ by using the combined $\boldsymbol{W}_j\boldsymbol{H}_u$ over all frames. To this end, the concatenated matrix $\left[\boldsymbol{W}_1^* \quad \cdots \quad \boldsymbol{W}_{N_\mathrm{t,R}}^*\right]^\mathrm{T} \in \mathbb{C}^{N_\mathrm{R}N_\mathrm{t,R}\times M_\mathrm{R}}$ should have orthogonal columns. One possible choice is to select from the first M_R columns of a $N_\mathrm{R}N_\mathrm{t,R}$-by-$N_\mathrm{R}N_\mathrm{t,R}$ DFT matrix, which is further scaled by $1/\sqrt{M_\mathrm{R}N_\mathrm{R}}$. Next, we describe a specific design of $\boldsymbol{f}_n[i] = \boldsymbol{F}[i]\boldsymbol{s}_n[i] \in \mathbb{C}^{M_\mathrm{T}}$ such that

$$\begin{aligned}\sum_{i=1}^{N_\mathrm{t,T}} \boldsymbol{C}[i]\boldsymbol{C}^\mathrm{H}[i] &= \sum_{i=1}^{N_\mathrm{t,T}}\sum_{p=1}^{N_\mathrm{T}} (\boldsymbol{w}_{k_p} \otimes \boldsymbol{f}_{k_p}[i])(\boldsymbol{w}_{k_p} \otimes \boldsymbol{f}_{k_p}[i])^\mathrm{H} \\ &= \beta\boldsymbol{I}_{LM_\mathrm{T}},\end{aligned} \tag{9}$$

where $\beta > 0$. Note that to make the best use of pilot tones in the frequency domain we set $N_\mathrm{f} = N_\mathrm{T}$ in (9). One choice of $\boldsymbol{f}_{k_p}[m]$ which satisfies (9) is computed as

$$(\boldsymbol{f}_{k_p}[i])_a = \sqrt{\frac{P}{M_\mathrm{T}N_\mathrm{T}}} e^{\frac{-j2\pi(i-1+(p-1)N_\mathrm{t,T})L(a-1)}{N_\mathrm{T}\cdot N_\mathrm{t,T}}}, \tag{10}$$

$\forall a \in \{1, \cdots, M_\mathrm{T}\}$. *All the derivations will be elaborated in the full paper.*

IV. COMPRESSED SENSING APPROACH

In contrast to the LS approach, the CS approach exploits the sparsity in the delay domain. To apply the CS approach, we assume that the AoAs and the AoDs lie on uniform grids $G_\mathrm{R} \geq M_\mathrm{R}$ and $G_\mathrm{T} \geq M_\mathrm{T}$. This is a common assumption in the literature, e.g., [8], [11], [12], [13], [18]. Under this assumption, the discrete CTF in (2) can be rewritten as

$$\boldsymbol{H}_n = \sum_{\ell=0}^{L-1} \boldsymbol{A}_\mathrm{R}\boldsymbol{H}_{\nu,\ell}\boldsymbol{A}_\mathrm{T}^\mathrm{H} e^{-j2\pi\frac{\ell\cdot n}{N_\mathrm{FFT}}}, \tag{11}$$

where $\boldsymbol{A}_\mathrm{R} = [\boldsymbol{a}(\theta_{\mathrm{R},1}) \quad \cdots \quad \boldsymbol{a}(\theta_{\mathrm{R},G_\mathrm{R}})] \in \mathbb{C}^{M_\mathrm{R}\times G_\mathrm{R}}$, $\boldsymbol{A}_\mathrm{T} = [\boldsymbol{a}(\theta_{\mathrm{T},1}) \quad \cdots \quad \boldsymbol{a}(\theta_{\mathrm{T},G_\mathrm{T}})] \in \mathbb{C}^{M_\mathrm{T}\times G_\mathrm{T}}$, and $\boldsymbol{H}_{\nu,\ell} \in \mathbb{C}^{G_\mathrm{R}\times G_\mathrm{T}}$ contains just one non-zero element $\alpha_\ell/\sqrt{L}$. In other words, $\boldsymbol{H}_{\nu,\ell}$ is sparse. Since $\boldsymbol{A}_\mathrm{R}$ and $\boldsymbol{A}_\mathrm{T}$ are fixed, we only need to estimate $\boldsymbol{H}_{\nu,\ell}$.

By inserting (11) into (1) and applying some algebraic manipulations we obtain

$$\begin{aligned}\boldsymbol{y}_n[m] = &\boldsymbol{W}^\mathrm{H}[m]\boldsymbol{A}_\mathrm{R}\boldsymbol{H}_\nu(\boldsymbol{w}_n \otimes \boldsymbol{A}_\mathrm{T}^\mathrm{H})\boldsymbol{F}[m]\boldsymbol{s}_n[m] \\ &+ \boldsymbol{W}^\mathrm{H}[m]\boldsymbol{z}_n[m],\end{aligned} \tag{12}$$

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 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

$$\boldsymbol{y}[m] = \left(\boldsymbol{B}^\mathrm{T}[m] \otimes (\boldsymbol{W}^\mathrm{H}[m]\boldsymbol{A}_\mathrm{R})\right)\boldsymbol{h}_\nu + \boldsymbol{z}[m], \tag{13}$$

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 and obtain

$$\boldsymbol{y}_\mathrm{s} = \boldsymbol{P}_2\boldsymbol{h}_u + \boldsymbol{z}_\mathrm{s} \in \mathbb{C}^{N_\mathrm{R}N_\mathrm{f}N_\mathrm{t}}, \tag{14}$$

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 (14) fulfills a sparse recovery problem and thus any sparse approximation algorithm in [19] can be applied. 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 or constructing $\boldsymbol{P}_2$ to possess randomly selected columns of a DFT matrix [14]. These two constructions are not suitable for our application. The former one can be achieved by determining the entries of $\boldsymbol{W}[m]$, $\boldsymbol{F}[m]$, and $\boldsymbol{s}_n[m]$ to be drawn from Gaussian distributions. Obviously, this relies on a random codebook, which is difficult to implement in practice. Nevertheless, this feasible choice is chosen as a benchmark in our numerical experiments. The latter one requires a matrix decomposition which decomposes a matrix into a product of an arbitrary matrix and a matrix with constant modulus entries. Such a matrix decomposition cannot be exact for an arbitrary matrix, even if it exists.

Instead, we propose a two-stage sparse recovery algorithm with a more structured design procedure. We use a similar training procedure as described in Section III, i.e., dividing the total training OFDM symbols into $N_{\mathrm{t,R}}$ frames each with $N_{\mathrm{t,T}}$ OFDM symbols. 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 following facts. First, each row of $\boldsymbol{H}_{\mathrm{R}}$ involves at most $L \ll LG_{\mathrm{T}}$ non-zero elements. Second, let us stack $\boldsymbol{y}_n[m]$ next to each other and obtain the following equation

$$\bar{\boldsymbol{Y}}_j = \boldsymbol{P}_3 \cdot \boldsymbol{H}_{\mathrm{R},j}^{\mathrm{T}} + \bar{\boldsymbol{Z}}_j, \tag{15}$$

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

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

If $N_{\mathrm{f}} N_{\mathrm{t,T}} < LG_{\mathrm{T}}$ or even $N_{\mathrm{f}} N_{\mathrm{t,T}} \ll LG_{\mathrm{T}}$, then (15) is a sparse recovery problem and more precisely, a MMV problem [20]. To solve (15), in this abstract we consider only the orthogonal matching pursuit (OMP) algorithm for a single measurement vector case. Thereby, we estimate $\boldsymbol{H}_{\mathrm{R},j}^{\mathrm{T}}$ column-wise. A direct solution to (15) via the M-OMP algorithm in [20] is also possible and the performance will be demonstrated in the full paper. 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

$$\hat{\boldsymbol{H}}_{\mathrm{R},j} = \boldsymbol{W}_j^{\mathrm{H}} \boldsymbol{A}_{\mathrm{R}} \boldsymbol{H}_\nu + \boldsymbol{\Delta}_j, \tag{16}$$

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

$$\hat{\boldsymbol{H}}_{\mathrm{cum}} = \boldsymbol{W}_{\mathrm{cs}} \boldsymbol{A}_{\mathrm{R}} \boldsymbol{H}_\nu + \boldsymbol{\Delta}_{\mathrm{cum}} \in \mathbb{C}^{N_{\mathrm{t,R}} N_{\mathrm{R}} \times LG_{\mathrm{T}}} \tag{17}$$

where $\hat{\boldsymbol{H}}_{\mathrm{cum}} = \left[\hat{\boldsymbol{H}}_{\mathrm{R},1}^{\mathrm{T}} \quad \cdots \quad \hat{\boldsymbol{H}}_{\mathrm{R},N_{\mathrm{t,R}}}^{\mathrm{T}}\right]^{\mathrm{T}}$, $\boldsymbol{\Delta}_{\mathrm{cum}} = \left[\boldsymbol{\Delta}_1^{\mathrm{T}} \quad \cdots \quad \boldsymbol{\Delta}_{N_{\mathrm{t,R}}}^{\mathrm{T}}\right]^{\mathrm{T}}$, and $\boldsymbol{W}_{\mathrm{cs}} = \left[\boldsymbol{W}_1^* \quad \cdots \quad \boldsymbol{W}_{N_{\mathrm{t,R}}}^*\right]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{t,R}} N_{\mathrm{R}} \times M_{\mathrm{R}}}$. If $N_{\mathrm{t,R}} N_{\mathrm{R}} \ll G_{\mathrm{R}}$, recovering $\boldsymbol{H}_\nu$ from (17) is still a MMV sparse recovery problem. However, the sparse profile on each column of $\boldsymbol{H}_\nu$ is not the same. We apply the single measurement vector solution. Furthermore, to reduce the computational complexity, we split equation (17) into L sub-equations and each sub-equation is calculated as

$$\hat{\boldsymbol{H}}_{\mathrm{cum},\ell} = \boldsymbol{W}_{\mathrm{cs}} \boldsymbol{A}_{\mathrm{R}} \boldsymbol{H}_{\nu,\ell} + \boldsymbol{\Delta}_{\mathrm{cum},\ell} \in \mathbb{C}^{N_{\mathrm{t,R}} N_{\mathrm{R}} \times G_{\mathrm{T}}}, \tag{18}$$

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 $\boldsymbol{\Delta}_{\mathrm{cum}}$, respectively. To solve $\boldsymbol{H}_{\nu,\ell}$, in general we can vectorize both sides of (18) with respect to $\boldsymbol{H}_{\nu,\ell}$ and the following relation yields

$$\hat{\boldsymbol{h}}_{\mathrm{cum},\ell} = \boldsymbol{P}_4 \boldsymbol{h}_{\nu,\ell} + \boldsymbol{\delta}_{\mathrm{cum},\ell} \in \mathbb{C}^{N_{\mathrm{t,R}} N_{\mathrm{R}} G_{\mathrm{T}}}, \tag{19}$$

where $\boldsymbol{P}_4 = \boldsymbol{I}_{G_{\mathrm{T}}} \otimes (\boldsymbol{W}_{\mathrm{cs}} \boldsymbol{A}_{\mathrm{R}}) \in \mathbb{C}^{N_{\mathrm{t,R}} N_{\mathrm{R}} G_{\mathrm{T}} \times G_{\mathrm{R}} G_{\mathrm{T}}}$, $\hat{\boldsymbol{h}}_{\mathrm{cum},\ell} = \mathrm{vec}\{\hat{\boldsymbol{H}}_{\mathrm{cum},\ell}\}$, $\hat{\boldsymbol{h}}_{\nu,\ell} = \mathrm{vec}\{\hat{\boldsymbol{H}}_{\nu,\ell}\}$, and $\boldsymbol{\delta}_{\mathrm{cum},\ell} = \mathrm{vec}\{\boldsymbol{\Delta}_{\mathrm{cum},\ell}\}$. Then $\boldsymbol{h}_{\nu,\ell}$ can be estimated by using an arbitrary sparse approximation algorithm, e.g., the OMP algorithm [5].

The remaining question is how to create $\boldsymbol{P}_3$ and $\boldsymbol{P}_4$, which fulfill the RIP. Again, a general design rule does not exist. We resort to heuristic solutions. More specifically, to create $\boldsymbol{P}_3$ we use the same design as in Section III, i.e., $N_{\mathrm{f}} = N_{\mathrm{T}}$ and equation (10) is applied. For $\boldsymbol{P}_4$ the entries of $\boldsymbol{W}_{\mathrm{cs}}$ are drawn randomly from a Gaussian distribution and then normalized such that $\|\boldsymbol{W}_j\|_{\mathrm{F}} = 1$ for all j. *Further details regarding the CS approach will be provided in the full paper.*

V. SIMULATION RESULTS

The proposed algorithms are evaluated using Monte-Carlo simulations. The maximum allowable power P is fixed to unity. The SNR is defined as $\mathrm{SNR} = 1/(N_{\mathrm{FFT}} \sigma_n^2)$. For computational simplicity, we set $N_{\mathrm{FFT}} = 64$. Moreover, we set $N_f = N_{\mathrm{T}}$, $N_{t,T} = LM_{\mathrm{T}}/N_{\mathrm{T}}$ and $N_{t,R} = M_{\mathrm{R}}/N_{\mathrm{R}}$. The LS method, the CS method using (14), and the two-stage recovery method are denoted as "LS", "Direct CS", and "Two-Stage CS", respectively. The simulation results are obtained by fixing the channel realization but averaging over 10 noise realizations.

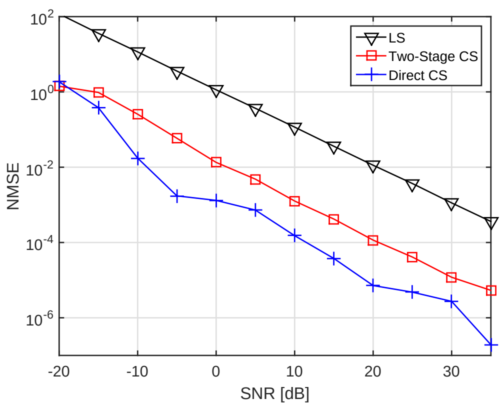

Fig. 1. Achievable NMSE vs. SNR for $M_{\mathrm{T}} = 64$, $M_{\mathrm{R}} = 16$, $N_{\mathrm{T}} = 16$, $N_{\mathrm{R}} = 8$, and $L = 4$.

Our initial results in Fig. 1 show that the CS methods perform better than the proposed LS approach.

More simulation results and further insights will be provided in the full paper, e.g., the performance of a MMSE estimator using the proposed training design, the impact of imperfect channel knowledge on the achievable sum rate, and the impact of imperfect AoA and AoD positions on the performance of channel estimation.

VI. 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] V. Venkateswaran and A. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4131–4143, Aug. 2010.

[7] 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.

[8] 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.

[9] 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.

[10] 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, available at http://http://arxiv.org/abs/1410.2609*, pp. 1–30, 2015.

[11] 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.

[12] 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.

[13] 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.

[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] 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.

[16] 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.

[17] T. L. Tung, K. Yao, and R. E. Hudson, "Channel estimation and adaptive power allocation for performance and capacity improvement of multiple-antenna OFDM systems," in *Proc. IEEE Signal Processing Workshop Signal Process. Adv. Wireless Commun.*, 2001.

[18] 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.

[19] 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.

[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.