Robust Analog Precoding Designs for Millimeter Wave MIMO Transceivers

Pengfei Xia*, Robert W. Heath Jr.†, and Nuria Gonzalez Prelcic‡
*Key Lab of Embedded System and Service Computing, Tongji University, Shanghai China
†The University of Texas at Austin, Austin, Texas, USA
‡University of Vigo, Vigo, Pontevedra, Spain

***Abstract*—Millimeter wave communication provides high data rates thanks to large arrays at the transmitter and receiver, coupled with large bandwidth channels. Exploiting the arrays is challenging due to the need to configure precoding at the transmitter based on the large frequency selective channel. In this paper we exploit the power iteration principle and propose a robust analog precoding training algorithm that can be applied in both frequency division duplex transmission (FDD) systems and time division duplex transmission (TDD) systems with or without RF calibration. We further analyze the convergence of the proposed algorithm and show how it converges to the singular value decomposition optimality exponentially. We propose null space projection on top of the power iteration to form multiple orthogonal beams at the transmitter and receiver. Strongest tap selection with proper energy pruning is used to collect as much precoding gain as possible from a frequency selective fading channel. The exponential effective SINR mapping performance is evaluated and demonstrates that the overall approach works effectively. Numerical simulation results demonstrate algorithm robustness and the algorithm works effectively not only for the simplified millimeter wave directional channels, but also for more general rich scattering channels.**

I. Introduction

Thanks to recent advances in RF technologies and the availability of wide bandwidth channels, millimeter wave (mmWave) wireless communications is being developed for low cost consumer applications. The new IEEE 802.11ad standard [12] provides multiple giga bit per second transmissions for indoor applications. More recently, the wireless industry is considering the mmWave frequency band for outdoor cellular systems. One of the major challenges for mmWave Gbps communications is the poor link budget due to the large bandwidth and the small antenna aperture size. Fortunately, because the carrier wavelength is on the order of several millimeters, it is possible to integrate a large number of antenna elements for both transmitter and receiver. This is the reason for the wide interest in multiple-input multiple-output (MIMO) transceiver techniques in mmWave systems [10].

We consider analog precoding in this work, where the transmitter precoding and receiver combining occur in the analog domain, i.e. after digital to analog converter at the transmitter side and before analog to digital converter at the receiver side. As a result, the number of RF chains needed depends only on the number of streams, rather than on the number of antenna elements. Notice that, although not included in the solution in this paper, digital precoding/combining may be applied on top of the analog precoding/combining operation, and is generally known as hybrid precoding/combining. Optimization of the transceiver precoding coefficients for a hybrid MIMO mmWave system is a topic of recent interest [2], [4], [5]; see also recent overviews of hybrid precoding for MIMO mmWave systems [3], [9], [10].

Joint design of analog and digital precoding leveraging sparsity in the channel was pursued in [5]. The joint design problem was formulated as a sparsity constrained matrix reconstruction problem; the orthogonal matching pursuit algorithm was used to solve the reformulated problem. The main limitations of [5] are that perfect channel state information is assumed known at the receiver side and that a narrowband directional channel model is assumed. An alternative approach was considered in [2] where a multi-resolution precoding codebook is proposed as a way to estimate the channel through a hierarchical search. This approach is still limited by the assumed directional channel model (see Eq. (2)). More recently, [4] studies hybrid precoding over frequency selective channels, where both baseband and RF precoders on the transmitter side are selected from quantized codebooks. Perfect channel knowledge is assumed at the receiver (mobile station) side, which may not be available in practice. In [6], the power iteration was proposed for blind MIMO antenna array training for TDD transmissions. Applications were shown for relatively a small number, say 4, of antennas. In such cases, however, the power iteration method does not demonstrate a clear advantage in terms of the training overhead, especially when the number of iterations is large. Furthermore, neither the constant modulus constraint nor the frequency selective fading channel is considered therein. Recently, the Arnoldi iteration method was proposed for mmWave MIMO subspace estimation [7]. Both algorithms work only for TDD transmissions, but not for FDD transmissions or TDD transmissions with uncalibrated antennas. Moreover, both [6] and [7] deal with flat fading channels only.

In this paper, we derive new algorithms for analog precoding for large scale MIMO systems with very large number of transmit and receive antennas and small number of spatial streams, with an emphasis on mmWave MIMO systems. In particular, we do not assume perfect estimate of the entire MIMO channel at the receiver side or the transmitter side. Instead, the precoding matrices are estimated by the transmitter and receiver in a distributed manner. As a result, the training

overhead is greatly reduced than the conventional approach of direct channel estimation. We develop a power iteration algorithm to iteratively estimate the transmit precoding and receive combining matrices, which is applicable in both FDD systems and TDD systems with/without uncalibrated antennas. We analyze the convergence of the proposed algorithm and show the exponential convergence to the optimal singular value decomposition solution. The proposed algorithms do not assume any particular structure of the underlying channel model. Numerical simulations demonstrate the proposed algorithms work effectively not only for the simplified structured directional channel in mmWave wireless, but also for rich scattering channels. Henceforth, the proposed algorithm is robust with regards to the channel propagation characteristics. We develop a strongest tap selection algorithm with energy pruning for frequency selective fading channels. Numerical results demonstrate effectiveness of the iterative algorithms in frequency selective fading channels. Numerical simulations also demonstrate that the power iteration algorithm, when modified to accommodate the constant modulus constraint in e.g. mmWave wireless communication systems, performs well in general and converges at a similar speed.

II. System and Channel Model

A. System Description

We consider a multi-stream mmWave MIMO transceiver with analog precoding, where a large number of transmit and receive antennas are used, as illustrated in Fig. 1. Without loss of generality, we select number of RF chains equal to the number of streams, although more RF chains per stream may be considered. Let $N_{\rm t}$ and $N_{\rm r}$ be the number of transmit and receive antennas respectively. The joint transceiver precoding techniques developed in the two stream system can be generalized to a higher order multi-stream precoding system. Naturally, a single stream system may be viewed as a special case of Fig. 1. For simplicity of exposition, we focus on the simple, yet general enough, two stream case.

In Fig. 1, transmit precoding is represented by the two precoding vectors $\{v_{1,i}\}_{i=1}^{N_{\rm t}}$, $\{v_{2,i}\}_{i=1}^{N_{\rm t}}$, which are the precoding coefficients on the ith antenna for the 1st and 2nd stream; and receive combining is represented by the two precoding vectors $\{u_{1,j}\}_{j=1}^{N_{\rm r}}$, $\{u_{2,j}\}_{j=1}^{N_{\rm r}}$ which are the combining coefficients on the jth antenna for the 1st and 2nd stream. Typically, the number of antennas $N_{\rm t}, N_{\rm r}$ is much larger than the number of data streams $N_s = 2$. This is desired especially for mmWave MIMO, where the per-RF-chain cost is relatively high and the per-antenna cost is relatively low, or when the number of supportable streams is much less than the number of antennas.

For flat fading channels, the system equation may be represented by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where the size $N_{\rm r} \times N_{\rm t}$ matrix $\mathbf{H}$ represents the wireless channel, the size $N_{\rm r}\times 1$ vector $\mathbf{y}, \mathbf{n}$ represent the received signal and the additive Gaussian channel noise respectively,

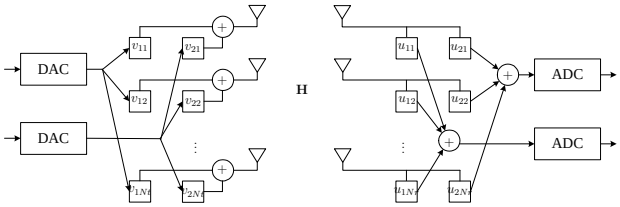

Fig. 1. Illustration of a two stream MIMO joint precoding system.

the size $N_{\rm t}\times 1$ vector $\mathbf{x}$ represents the transmitted signal. Frequency selective fading channels will be dealt with later.

The main problem we solve is how to compute the optimal precoding vectors at both transmitter and receiver sides. Toward this purpose, channel state information is normally used to design the transmit and receive beams. For MIMO systems operating at lower frequencies, full CSI is usually assumed at least at the receiver side. With full CSI at the receiver, different levels of CSI may be made available at the transmitter side.

B. Channel Model and RF Consideration

The mmWave wireless channel typically has limited scattering, and multipaths are mainly generated by LOS and 1st- and 2nd-order reflections. As a result, mmWave channels demonstrate clear directionality, where the physical angles of departure (AoDs) and angles of arrival (AoAs) in the azimuth and elevation domain play a critical role in determining the channel response. In such cases, the following simplified directional channel model may be used [5], [10]

$$\mathbf{H} = \sqrt{N_{\rm t}N_{\rm r}} \sum_{\ell=1}^{L}\sum_{q=1}^{Q} \lambda_{\ell,q}\ \mathbf{g}(\phi^r_{\ell,q}\theta^r_{\ell,q})\ \mathbf{g}(\phi^r_{\ell,q},\theta^r_{\ell,q})', \tag{2}$$

where L is the number of clusters, Q is the number of rays per cluster, $\lambda_{\ell,q}$ is the complex gain of the ℓ-th cluster and q-th ray, $\phi^t_{\ell,q}, \theta^t_{\ell,q}$ are the elevation and azimuth domain AoDs of the q-th ray within the ℓ-th cluster, and $\phi^r_{\ell,q}, \theta^r_{\ell,q}$ are the elevation and azimuth domain AoAs. In general, λ_ℓ may be modeled as complex Gaussian distributed, the cluster AoA/AoDs $\phi^r_\ell, \theta^r_\ell$ may be modeled as Laplacian distributed, and the ray AoA/AoDs $\phi^t_{\ell,q}, \theta^t_{\ell,q}, \phi^r_{\ell,q}, \theta^r_{\ell,q}$ may be modeled to be uniformly distributed around the cluster AoA/AoDs with a constant angular spread [10]. $\mathbf{g}(\cdot)$ is the steering vector depending on the antenna array geometry.

The above channel model assumes narrowband transmissions. The purported benefits of mmWave wireless communications are found though in channels with a large bandwidth. In Section IV, we will consider frequency selective fading channels. The simplified directional channel model in Eq. (2) may be extended to have multiple taps, with multiple non-overlapping taps in the delay domain and each tap containing multiple clusters/rays. See [4, Eq. (3)] for details on the extended channel model.

For analog precoding, the precoding operation can be represented by a phase shifter and an amplitude modulator per antenna branch. The phase shifter and the amplitude modulator on each antenna branch may be separately controlled. It is

well known that for a large number of antenna elements, equal gain combining/precoding is able to achieve most of the gain with the per-antenna peak power constraint (important for mmWave MIMO communications in the GHz frequency band). Because of this and also because of the extra cost to control the amplitudes independently in the analog domain, in many cases, one may place a constant modulus constraint for all antenna branches and only adjusts the phase shifters independently. This is the reason for the prevalence of phased arrays in mmWave systems. In this paper, we will develop the precoding algorithms with the constant modulus constraint in mind. As we will see in numerical results, most of the precoding gain can be maintained when the constant modulus constraint is applied. Once the analog precoding are formed between the transmitter and the receiver in the analog domain, extra digital precoding may be performed on top. This is also known as hybrid precoding and has been studied in the literature (see [3], [5], [10]).

Although the iterative analog precoding training algorithms in this paper are developed with mmWave wireless channels in mind, they may be applied in certain other cases (e.g. large scale MIMO systems in the centimeter wave frequency band), for which the above channel model (2) may not hold true or be sufficiently accurate. To verify the algorithm effectiveness in such cases, we will also test the developed algorithms in a more general MIMO fading channel model

$$\{\mathcal{H}\}_{i,j,\ell} \sim \text{CN}(0, \sigma_{h,\ell}^2) \tag{3}$$

where the channel coefficient $\{\mathcal{H}\}_{i,j,\ell}$ between the i-th receive and j-th transmit antenna on the ℓ-th tap is zero mean complex Gaussian with a tap-dependent channel variance $\sigma_{h,\ell}^2$. This channel model corresponds to the typical "rich scattering assumption" made in lower frequency MIMO systems. As we will see, the developed algorithm is robust in that it works almost equally well in those more challenging channel scenarios.

III. Analog Precoding in Flat Fading

In this section we focus on the flat fading channels, and our objective is to train the precoding vectors on the transmitter and receiver side jointly.

A. Power Iteration Introduction

We review here the method of the power iteration, which is used in numerical matrix analysis to compute matrix eigenvalue decompositions and singular value decompositions [8]. Write the SVD decomposition of $\mathbf{H}$ as

$$\mathbf{H} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1' + \sigma_2 \mathbf{u}_2 \mathbf{v}_2' + \cdots + \sigma_P \mathbf{u}_P \mathbf{v}_P' \tag{4}$$

where $\mathbf{u}_1, \cdots, \mathbf{u}_P$ are left singular vectors of size $N_{\text{r}} \times 1$, $\mathbf{v}_1, \cdots, \mathbf{v}_P$ are right singular vectors of size $N_{\text{t}} \times 1$, and $\sigma_1 \geq \cdots \geq \sigma_P$ are singular values in a non-increasing order. For a positive integer m, define

$$\mathbf{H}^{2m} := (\mathbf{H}'\mathbf{H})^m; \quad \mathbf{H}^{2m-1} = \mathbf{H}\mathbf{H}^{2m-2} \tag{5}$$

Thanks to the orthogonality between singular vectors, we have that

$$\begin{aligned} \mathbf{H}^{2m} &= \sigma_1^{2m} \quad \mathbf{v}_1\mathbf{v}_1' + \sigma_2^{2m} \quad \mathbf{v}_2\mathbf{v}_2' \cdots + \sigma_P^{2m} \quad \mathbf{v}_P\mathbf{v}_P' \\ \mathbf{H}^{2m-1} &= \sigma_1^{2m-1}\mathbf{u}_1\mathbf{v}_1' + \sigma_2^{2m-1}\mathbf{u}_2\mathbf{v}_2' \cdots + \sigma_P^{2m-1}\mathbf{u}_P\mathbf{v}_P'. \end{aligned} \tag{6}$$

As the positive integer m increases, $\sigma_i^{2m}/\sigma_1^{2m}$ decreases exponentially, $\forall i = 2, \cdots, P$. Note that $\sigma_i \leq \sigma_1, \forall i > 1$. Therefore,

$$\begin{aligned} \lim_{m\to\infty} \mathbf{H}^{2m} &= \sigma_1^{2m}\mathbf{v}_1\mathbf{v}_1', \\ \lim_{m\to\infty} \mathbf{H}^{2m-1} &= \sigma_1^{2m-1}\mathbf{u}_1\mathbf{v}_1'. \end{aligned} \tag{7}$$

In other words, only the strongest eigenmode $\mathbf{u}_1, \mathbf{v}_1$ remains in the extreme case.

B. Iterative Analog Precoding Algorithm

The optimal joint transceiver precoding can be obtained via the singular value decomposition (SVD) of the channel. As a result, the MIMO channel $\mathbf{H}$ can be simplified to a diagonal channel to support multiple non-interfering data streams.

Starting from (7), an arbitrary vector $\mathbf{q}$ in the range space of $\mathbf{H}$ can be written as $\mathbf{q} = \sum_{i=1}^{P} c_i\mathbf{v}_i$ where $c_i\mathbf{v}_i$ is the contribution of $\mathbf{q}$ along $\mathbf{v}_i$. Thanks to (7) and the orthogonality between singular vectors, we have

$$\lim_{m\to\infty} \mathbf{H}^{2m}\mathbf{q} = \sigma_1^{2m}\mathbf{v}_1\mathbf{v}_1' \left(\sum_{i=1}^{P} c_i\mathbf{v}_i\right) \tag{8}$$

where the right hand side is equivalent to $\mathbf{v}_1$ after normalization. In other words, the leading transmit precoding vector $\mathbf{v}_1$ may be obtained by multiplying a random vector $\mathbf{q}$ by $\mathbf{H}^{2m}$. Similarly, we have

$$\lim_{m\to\infty} \mathbf{H}^{2m-1}\mathbf{q} = \sigma_1^{2m-1}\mathbf{u}_1\mathbf{v}_1' \left(\sum_{i=1}^{P} c_i\mathbf{v}_i\right) \tag{9}$$

where the right hand side is equivalent to $\mathbf{u}_1$ after normalization. In other words, the leading receive combining vector $\mathbf{u}_1$ may be obtained by multiplying a random vector $\mathbf{q}$ by $\mathbf{H}^{2m-1}$. It is noted that the random vector $\mathbf{q}$ is in the range space of $\mathbf{H}$ with probability one. Henceforth, Equations (8) and (9) hold true with probability one.

Our algorithm is presented in Algorithm 1 and also illustrated in Fig. 2. The algorithm is an iterative process involving two steps, i.e., optimizing the receive combining vector by fixing the transmit precoding vector, and optimizing the transmit precoding vector by fixing the receive combining vector.

Particularly, in step 1, destination device (receiver) training is performed by sweeping across a number of orthogonal beams at the receiver side while fixing the transmit precoding to a random initial vector $\mathbf{q}$. N_{r} time slots are needed because N_{r} combining coefficients are to be estimated. Note that vector normalization is applied always. On the other hand, if constant modulus constraint is needed e.g. in the mmWave frequency band, amplitude information of each combining efficient is stripped and only phase information is retained.

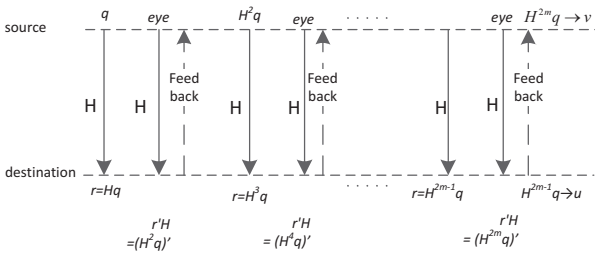

Fig. 2. Power iteration illustration (FDD).

In step 2, source device (transmitter) training is performed by sweeping across a number of orthogonal beams at the transmitter side while fixing the receive combining vector to be the newly computed result from the previous step. N_{t} time slots are needed because N_{t} precoding coefficients are to be estimated. Note that vector normalization is applied always.

The estimated transmit precoding vector is then complex conjugated and fed back to the transmitter. Different feedback technologies maybe used. For example, when constant modulus constraint is not forced (e.g. for large scale MIMO systems in centimeter wave frequency band), limited feedback transmit precoding may be used with the codebook designed based on isotropic packing of the Grassmannian manifold; when constant modulus is forced (e.g. for large scale MIMO systems in the mmWave frequency band), amplitude information of each combining efficient is stripped and only phase information is retained. Also note that the feedback (dashed curve in Fig. 2) may be done in a different frequency band from which the antenna array training is performed (solid curve in Fig. 2), or via a different radio.

Algorithm 1 Iterative Analog Precoding Training

0. Start with a random initial vector $\mathbf{q}$ at the source device.
1. Transmit at the source with $\mathbf{q}$ as the transmit precoding vector, and receive at the destination with the i-th column of $\mathbf{I}_{N_{\mathrm{r}}}$ as the receive combining vector. Repeat this with $i = 1, \cdots, N_{\mathrm{r}}$ over N_{r} time slots. Collect the N_{r} received samples over the N_{r} time slots, and obtain at the destination $\mathbf{r} = \mathbf{H}\mathbf{t}$. Apply constant modulus constraint if needed. Apply vector normalization.

2. Transmit at the source with the j-th column of N_{t}, and receive at the destination with the recently computed $\mathbf{r}$ as the receive combining vector. Repeat this with $j = 1, \cdots, N_{\mathrm{t}}$ over N_{t} time slots. Collect the N_{t} received samples over the N_{t} time slots, and obtain at the destination $\mathbf{r}'\mathbf{H}$. Apply constant modulus constraint if needed. Apply vector normalization. Apply complex conjugate operation to obtain $\mathbf{t} = \mathbf{H}'\mathbf{r}$ and feed it back to the source device.

3. Repeat steps 1 and 2 until convergence.

Power iteration has also been studied in [6]. Several important differences exist though. Firstly, perfect downlink/uplink channel reciprocity is assumed in [6], which requires TDD transmissions with perfectly calibrated RF frontend. This however is not required in Algorithm 1, which may be used in FDD transmissions, or TDD transmissions with/without RF calibration. Secondly, in [6], a complete transceiver on the frequency band in consideration is required at both the destination device and the source device. For example when mmWave wireless communications is considered, the source device need to have a transmitter in the mmWave frequency band and a receiver in the mmWave frequency band. So is the destination device. This however is not required in Algorithm 1. It can be seen clearly from Fig. 2 that it is only required that the source device have a transmitter in the mmWave frequency band, and that the destination device have a receiver in the mmWave frequency band. Note that the feedback link may be implemented on a different frequency band via a separate, hopefully cheaper, radio.

C. Convergence Analysis

In [6], the power iteration convergence is studied with the performance metric selected as the mean square error between the right singular vector and the ideal transmit precoding vector, or the mean square error (MSE) between the left singular vector and the ideal receive combining vector, or the average of the two. Numerical simulations demonstrate that the mean square error converges after around $m = 10$ iterations for a 3×3 system in the noiseless case. We argue that the convergence behavior may be better evaluated by looking at the effective channel. From Fig. 2, Vector $\mathbf{r}$ converges to the desired receive combining vector $\mathbf{u}_1$: $\mathbf{r} = \mathbf{H}^{2m-1}\mathbf{q} \to \mathbf{u}_1$, and vector $\mathbf{t}$ converges to the desired transmit precoding vector $\mathbf{v}_1$: $\mathbf{t} = \mathbf{H}^{2m}\mathbf{q} \to \mathbf{v}_1$, where m is the iteration index. The optimal achievable precoding gain can be used as a reference and written as

$$g_0 = |\mathbf{s}_1| = |\mathbf{u}_1'\mathbf{H}\mathbf{v}_1|. \tag{10}$$

At the m-th iteration, the convergence of the transmit precoding vector may be individually evaluated as

$$g_{\mathrm{t}}(m) = |\mathbf{u}_1'\mathbf{H}\mathbf{t}| = |\mathbf{u}_1'\mathbf{H}\mathbf{H}^{2m}\mathbf{q}| = |\mathbf{u}_1'\mathbf{H}^{2m+1}\mathbf{q}|, \tag{11}$$

where the current transit precoding vector $\mathbf{t}$ is paired with the ideal receive combining vector $\mathbf{u}_1$, and can be viewed as the angular distance between $\mathbf{t}$ and the ideal transmit precoding vector $\mathbf{v}_1$ (corresponding to the distance on the Grassmannian manifold $\mathcal{G}(N_{\mathrm{t}}, 1)$). Note that Euclidean distance between $\mathbf{t}$ and $\mathbf{v}_1$ is used [6]. Similarly, convergence of the receive combining vector may be individually evaluated as

$$g_{\mathrm{r}}(m) = |\mathbf{r}'\mathbf{H}\mathbf{v}_1| = |\mathbf{q}'(\mathbf{H}^{2m-1})'\mathbf{H}\mathbf{v}_1| = |\mathbf{v}_1'\mathbf{H}^{2m}\mathbf{q}| \tag{12}$$

where the receive combining vector $\mathbf{r}$ is paired with the ideal transmit precoding vector $\mathbf{v}_1$, and can be viewed as the angular distance between $\mathbf{r}$ and the ideal receive combining vector $\mathbf{u}_1$. From (11) and (12), the converging speed of the transmit precoding/receive combining vector individually is on the order of $(\sigma_2/\sigma_1)^{2m}$ where the second largest singular value σ_2 would dictate the convergence in general.

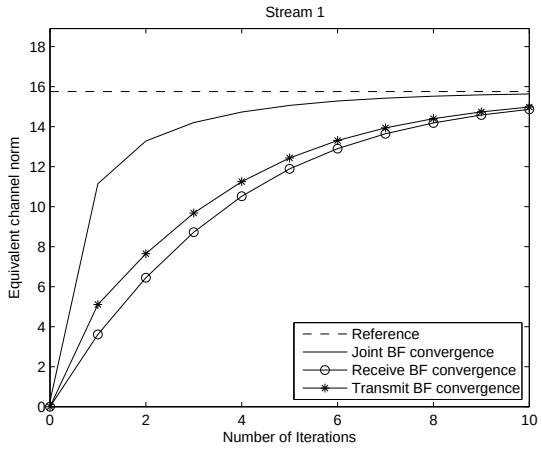

Fig. 3. Convergence behavior.

We can evaluate how fast the transmit precoding vector converges to the ideal by comparing $g_t(m)$ with g_0, and how fast the receive combining vector converges to the ideal by comparing $g_r(m)$ with g_0. Yet, they can be jointly evaluated and the achieved gain at the end of the m-th iteration is

$$g_{\mathrm{t,r}}(m) = |\mathbf{r}'\mathbf{H}\mathbf{t}| = |\mathbf{q}'(\mathbf{H}^{2m-1})'\mathbf{H}\mathbf{H}^{2m}\mathbf{q}| = |\mathbf{q}'\mathbf{H}^{4m}\mathbf{q}| \tag{13}$$

In fact, although we are interested in the convergence behavior of the transmit precoding/receive combining vector individually (11) (12), what matters more is the overall achieved channel gain (13). The convergence speed of the transmit precoding and receive combining vector jointly is on the order of $(\sigma_2/\sigma_1)^{4m}$, doubling the individual speed. This can be witnessed from Fig. 3 where $N_{\mathrm{t}} = N_{\mathrm{r}} = 64$ is used, and the joint precoding/combining performance corresponds to (13) and the individual precoding/combining performance corresponds to (11) and (12) respectively. Note that the joint precoding/combining performance converges much faster than the individual precoding/combining performance.

In many cases in practice, we are interested in the effective channel SNR (signal to noise ratio). It is straightforward that the channel SNR is proportional to the channel gain squared

$$g^2_{\mathrm{t,r}}(m) = |\mathbf{q}'\mathbf{H}^{4m}\mathbf{q}|^2 = |\mathbf{q}'\mathbf{H}^{8m}\mathbf{q}|, \tag{14}$$

Thus, the effective SNR converges with a speed on the order of $(\sigma_2/\sigma_1)^{8m}$, further doubling the convergence speed.

In Algorithm 1, estimation is done for $\mathbf{Hq}$ as a whole, and $\mathbf{H}^2\mathbf{q}$ as a whole, and so on and so forth. The underlying channel $\mathbf{H}$ is never estimated explicitly. The number of estimates is simply $m \times (N_{\mathrm{t}} + N_{\mathrm{r}})$. In comparison, the number of channel estimates required to estimate the original $\mathbf{H}$ directly is $N_{\mathrm{t}} \times N_{\mathrm{r}}$. Clearly, the proposed algorithm saves computation complexity and training air time especially when number of antennas $N_{\mathrm{t}}, N_{\mathrm{r}}$ are large and number of spatial streams is small. This is exactly the case for mmWave wireless communications.

D. Extension to the Second Stream

In this subsection, we continue the antenna array training for the second stream, where the second leading transmit and receive combining vectors $\mathbf{v}_2, \mathbf{u}_2$ are to be acquired. Let $\mathbf{t}_\varnothing$ be a random vector orthogonal to $\mathbf{v}_1$, which can be obtained by projecting a random vector $\mathbf{t}$ onto the null space of $\mathbf{v}_1$, where

$$\mathbf{t}_\varnothing = \mathbf{t} - (\mathbf{v}_1'\mathbf{t})\mathbf{v}_1 = \sum_{i=2}^{P} c_i\mathbf{v}_i. \tag{15}$$

is orthogonal to $\mathbf{v}_1$.

Similar to (8), it can be shown that

$$\begin{aligned} \mathbf{H}^{2m}\mathbf{t}_\varnothing &= \sum_{i=2}^{P} c_i\sigma_i^{2m}\mathbf{v}_i, \\ \mathbf{H}^{2m-1}\mathbf{t}_\varnothing &= \sum_{i=2}^{P} c_i\sigma_i^{2m-1}\mathbf{u}_i. \end{aligned} \tag{16}$$

and furthermore

$$\begin{aligned} \lim_{m\to\infty} \mathbf{H}^{2m}\mathbf{t}_\varnothing &= c_2\sigma_2^{2m}\mathbf{v}_2 \\ &\propto \mathbf{v}_2. \end{aligned} \tag{17}$$

In other words, when m is large enough, we may use $\mathbf{H}^{2m}\mathbf{t}_\varnothing$ to obtain the second transmit precoding vector $\mathbf{v}_2$, where we need to make sure that all the contribution from $\mathbf{v}_1$ is excluded. Similarly,

$$\begin{aligned} \lim_{m\to\infty} \mathbf{H}^{2m-1}\mathbf{t}_\varnothing &= c_2\sigma_2^{2m-1}\mathbf{u}_2 \\ &\propto \mathbf{u}_2 \end{aligned} \tag{18}$$

and we may use $\mathbf{H}^{2m-1}\mathbf{t}_\varnothing$ to obtain the second receive combining vector $\mathbf{u}_2$, where we need to make sure that all the contribution from $\mathbf{u}_1$ is excluded.

Algorithm 1 from Section III can be directly generalized to the second stream. The only difference is that for each step a null space projection is needed. In particular, at the end of step 1 before applying the constant modulus constraint and vector normalization, the obtained receive combining vector need be projected onto the null space of $\mathbf{u}_1$, which has been obtained for stream 1. Constant modulus constraint and vector normalization can then be applied. Similarly, at the end of step 2 before applying the constant modulus constraint and vector normalization, the obtained transmit precoding vector need be projected onto the null space of $\mathbf{v}_1$, which has been obtained for stream 1. Constant modulus constraint and vector normalization can then be applied.

E. Numerical Results

We carry out numerical simulations over the simplified directional channel model in Eq. (2), which may be used for certain mmWave applications. A system setup with $N_{\mathrm{t}} = 32$ and $N_{\mathrm{r}} = 32$ antennas is considered. $L = 10$ multipaths are generated, with the path gain complex normal distributed, and the AoAs and AoDs uniformly distributed within $[0, \pi]$. Fig. 4 illustrates the convergence behavior, while the constant modulus constraint is applied. The dashed lines represent the

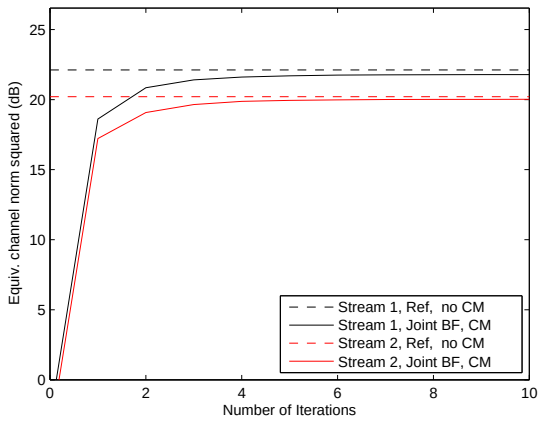

Fig. 4. Convergence in simplified mmWave directional channels.

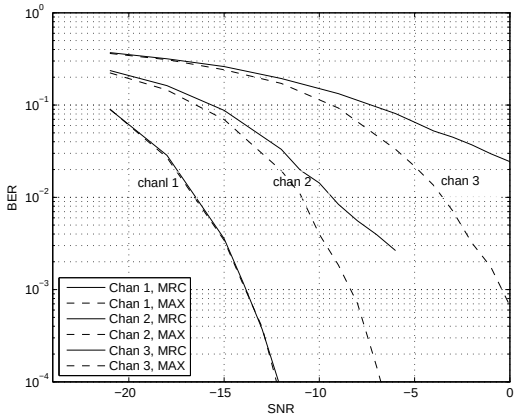

Fig. 5. MRC-type combining vs MAX-type selection in the delay domain.

upper-bounds in the first and second streams promised by the ideal singular value decomposition, solid lines represent the achieved channel norm squared with the constant modulus constraint. For both streams, the power iteration is able to converge to the optimal gain quickly, after 3 to 4 iterations.

IV. Analog Precoding in Frequency Selective Fading

In this work, we employ the so-called exponential effective SINR mapping (EESM) SNR [13] as the performance metric, where

$$\text{SINR}_{\text{eff}} = -\beta \log\left(\frac{1}{N_c}\frac{1}{N_s}\sum_{k=1}^{N_c}\sum_{n_s=1}^{N_s}\exp\left(-\frac{\text{SINR}_{k,n_s}}{\beta}\right)\right) \tag{19}$$

where SINR_{k,n_s} is the post-equalization SINR on the k-th OFDM subcarrier and the n_sth stream, and β is a MCS-dependent constant. A linear Zero Forcing (ZF) equalizer is assumed per subcarrier, and the post-equalization SINR per stream is readily available. Essentially, an OFDM system with per subcarrier SINRs $\{\text{SINR}_k\}_{k=1}^{N_c}$ would achieve a similar block error rate performance as that of a single tap AWGN channel with SNR of SINR_{eff}.

A. Tap Selection for the First Stream

Following Section III-B, let $\mathbf{u}_1$, and $\mathbf{v}_1$ be the receive and transmit precoding vector to be acquired. It is clear that the effective channel may be written as $\mathcal{H}_{\text{e}} = \{\mathbf{u}_1'\mathbf{H}_1\mathbf{v}_1, \cdots, \mathbf{u}_1'\mathbf{H}_L\mathbf{v}_1\}$ in the time domain. In the frequency domain, the overall channel may be written as $\mathcal{H}_{f,e} = \{\mathbf{u}_1'\mathbf{H}_{f,1}\mathbf{v}_1, \cdots, \mathbf{u}_1'\mathbf{H}_{f,N_c}\mathbf{v}_1\}$ where $\{\mathbf{H}_{f,1}, \cdots, \mathbf{H}_{f,N_c}\}$ is the frequency domain channel with N_c being the number of OFDM subcarriers. The per subcarrier effective SNR is thus $\text{SINR}_k = |\mathbf{u}_1'\mathbf{H}_{f,k}\mathbf{v}_1|^2\rho$ where ρ is the per-subcarrier SINR without precoding/combining. The overall EESM mapping can then be written as

$$\text{SINR}_{\text{eff}} = -\beta \log\left(\frac{1}{N_c}\sum_{n=1}^{N_c}\exp\left(\frac{|\mathbf{u}_1'\mathbf{H}_{f,k}\mathbf{v}_1|^2\rho}{\beta}\right)\right). \tag{20}$$

Direct optimization of the effective SINR based on EESM mapping is difficult, due to the exponential-sum-log operation. Therefore we resort to some simpler heuristic alternatives. The difficulty is that we have only one pair of transmit and receive combining vectors. Yet, we need to deal with multiple channel taps (or frequency subcarriers). Conceptually, one need to decide to focus the beam onto one particular tap or one particular frequency subcarrier, or to the focus the beam onto a combination of multiple taps or a combination of multiple subcarriers. One heuristic method is to select only the tap with the largest tap energy, and let the transmitter and receiver form the beams based on the channel matrix on top of this particular tap. let ℓ_1^* be the tap with highest tap energy (or largest average channel Frobenius norm square)

$$\ell_1^* = \arg\max_{\ell} F_\ell^2 \tag{21}$$

where F_ℓ^2 is the average Frobenius norm square of the channel on the ℓ-th tap, and is porportional to $\sigma_{h,\ell}^2$ in Eq. (3). Similarly, let ℓ_2^* be the tap with second highest tap energy to be used later. Essentially, in stream 1, the receiver would always form the beam based only on the ℓ_1^*-th column and ignores contribution from all other taps.

We carry out numerical simulations in the following. The channel model in Eq. (3) is adopted. Three different channel power profiles are used. Channel power profile 1 is simply single tap flat fading channel; channel profile 2 is 16-tap frequency selective channel with an exponentially decaying power profile with power decaying factor 0.7; and channel profile 3 is 16-tap frequency selective channel with same energy across all 16 taps. OFDM modulation is used, with perfect synchronization assumed and cyclic prefix large enough to enable ISI-free transmissions. QPSK modulation is used and we compare the two methods in terms of uncoded BERs in Fig. 5. In Fig. 5, the MRC method refers to the scheme where the beam is formed based on the MRC-combined channel across all taps, while the MAX method refers to the strongest tap selection method.

As we can see, in the flat fading channel, both methods yield the same performance as expected. Yet, as the channel becomes more and more frequency selective (with the channel power profile becoming less and less like a delta function), the strongest tap selection method usually achieves a better performance demonstrating a higher diversity order. We thus stick to the strongest tap selection method in the following work in stream 2. In this paper, we assume that the channel power profile ($\{\sigma_{h,\ell}^2\}$ in Eq. (3)) has been estimated. We emphasize that accurate estimation of the channel power profile (and in particular the tap energies on several strongest taps) is important and may be obtained with a small overhead. Detailed treatment of the simplified channel power profile estimation is beyond the scope of this paper.

B. Extension to the Second Stream

In stream 1, we resort to the heuristic of selecting the strongest tap ℓ_1^*, and let the transmitter/receiver form the beams based on $\mathbf{H}_{\ell_1^*}$. Let the transmitter beam be represented by $\mathbf{v}_1$, the receiver beam be represented by $\mathbf{u}_1$, and the effective channel formed on the ℓ_1^*-th tap be $s_1 = \mathbf{u}_1'\mathbf{H}_{\ell_1^*}\mathbf{v}_1$. Now we face the problem of training the second transmit and receive beams. The major difference unique to the multi-tap frequency selective channel is that one transmit/receive beam has been formed on the ℓ_1^*-th tap. One may form the second transmit/receive beam on the second strongest tap ℓ_2^*. Yet, this may not be the best tap selection strategy for the second stream, because it excludes ℓ_1^*-th tap from the selection completely. Instead we propose the following new tap selection for stream 2:

$$\ell_2^p = \arg\max_{\ell} \left(F_\ell^2 - G_\ell^2\right) \tag{22}$$

where $G_{\ell_1^*} = |s_1|$ and 0 otherwise. Here $s_1 = \mathbf{u}_1'\mathbf{H}_{\ell_1^*}\mathbf{v}_1$ and can be estimated at the end of stream 1 iteration. Essentially, we propose to first remove the contribution of stream 1 precoding/combining on the ℓ_1^*-th tap, and then selects the beam with the largest energy afterwards. A detailed description of the algorithm is shown in Algorithm 2.

In the spatial domain, we perform null space projection such that the new beam obtained has no contribution from the first stream. in the delay/tap domain, we perform energy pruning on the selected tap to make sure that contribution from stream 1 has been properly removed as well. Thus, both null space projection and tap selection with energy pruning serve the same design principle.

We perform numerical simulations in the following for two different channel profiles, under channel model Eq. (3). Channel profile 4 has 16 taps with the channel amplitude on each tap [0.11, 0.9, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11,0.11, 0.11, 0.11], where we see one strong tap overall. Channel profile 5 has 16 taps also with the channel amplitude on each tap [0.1, 0.6, 0.7, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1] where we see two strong taps. For both channel profiles, we perform the iterative algorithms for 1000 channel realizations, and plot the EESM channel abstraction (19) (averaged across 1000 channel realizations) for each iteration (iteration 1 to 6 for stream 1 and iteration 6 to 11 for stream 2). For tap selection in stream 2, ℓ_2^p-th tap is used if pruning is used, and ℓ_2^*-th tap is used if pruning is not used.

Algorithm 2 Stream Two Training over Frequency Selective Fading Channels

0. Start at the source device with a random initial vector $\mathbf{q}$ that is orthogonal to $\mathbf{v}_1$. Prune the energy $G_{\ell_1^*}^2$ from the ℓ_1^*-th tap and then pick the strongest tap with index ℓ_2^p.

1. Transmit at the source with $\mathbf{t}$ as the transmit precoding vector, and receive at the destination with the i-th column of $\mathbf{I}_{N_\mathrm{r}}$ as the receive combining vector. Repeat this with $i = 1, \cdots, N_\mathrm{r}$ over N_r time slots. Collect the N_r received samples over the N_r time slots on the ℓ_2^p-th tap, and obtain at the destination $\mathbf{r} = \mathbf{H}_{\ell_2^p}\mathbf{t}$. Project $\mathbf{r}$ onto the null space of $\mathbf{u}_1$. Apply constant modulus constraint and vector normalization.

2. Transmit at the source with the j-th column of N_t, and receive at the destination with the recently computed $\mathbf{r}$ as the receive combining vector. Repeat this with $j = 1, \cdots, N_\mathrm{t}$ over N_t time slots. Collect the N_t received samples over the N_t time slots on the ℓ_2^p-th tap, and obtain at the destination $\mathbf{r}'\mathbf{H}_{\ell_2^p}$. Apply complex conjugate operation and obtain $\mathbf{t} = \mathbf{H}_{\ell_2^p}'\mathbf{r}$. Feed $\mathbf{t}$ back to the source device.

3. Project $\mathbf{t}$ onto the null space of $\mathbf{v}_1$. Apply constant modulus constraint and vector normalization.

3. Repeat steps 1, 2, and 3 until convergence.

For channel profile 4, tap selection with pruning is able to improve the effective SINR by roughly 23 dB on average, while the tap selection without pruning is able to improve the effective SINR only by roughly 9 dB. Note that for the latter method, most of the gain is actually gleaned in stream 1 (common to both methods), which demonstrates that the tap selection without pruning essentially fails to work for such a channel profile. Clearly, this is because channel profile 4 has only one strong tap. If the strongest tap is selected in stream 1 and excluded in stream 2 (by ℓ_2^*), the performance improvement is expected to be minimum. For channel profile 5, tap selection with pruning is able to improve the effective SINR by roughly 20 dB on average, while the tap selection without pruning is able to improve the effective SINR roughly 19 dB. Note that in this case, two strong taps exist and hence, selecting the second tap without pruning, although not optimal, still provides large SINR improvement over the sub-optimal tap selection. Note that in both channel profiles, the strongest tap selection with energy pruning is always providing better performance gain than the tap selection without energy pruning.

We next carry out numerical simulations over the simplified directional channel model in Eq. (2). To capture the long delay spread behavior, the $L = 12$ multipaths are grouped into two

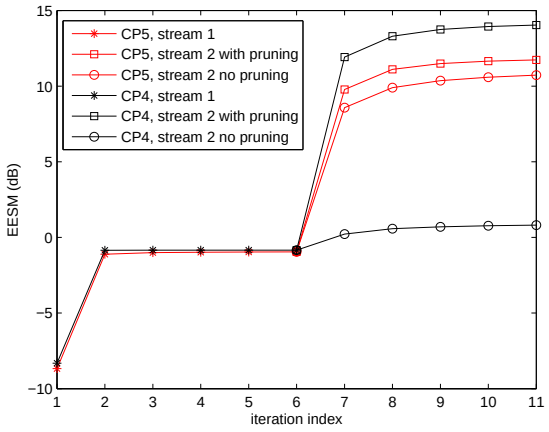

Fig. 6. EESM SINR comparisons.

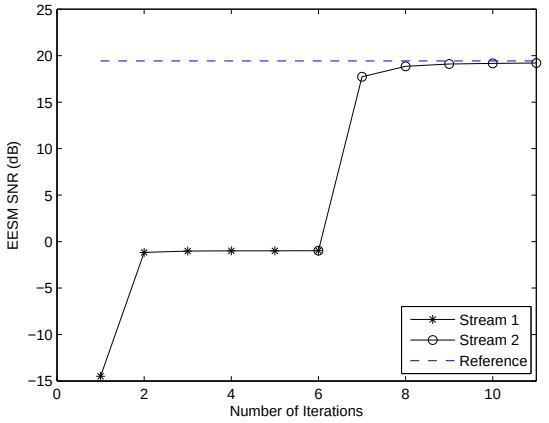

Fig. 7. Simplified directional channel: EESM SINR comparisons.

clusters, with each cluster occupying a different tap. Within each cluster, 6 multipaths/rays are randomly generated in the same way as in Eq. (2), i.e, with the path gain complex normal distributed, and the AoAs and AoDs uniformly distributed within $[0, \pi]$. The two clusters are assumed to have the same power per multipath/ray, and hence the same total power per cluster/tap. Fig. 7 illustrates the convergence behavior as a result of the iterative algorithms, with the constant modulus constraint applied. The reference curve shows the ideal EESM SNR performance, where the two transceiver beams are formed onto the two strongest paths (perfectly acquired), and a perfect zero forcing equalizer is performed at the receiver side. As we can see, the iterative algorithm is also able to achieve the ideal performance quickly.

V. Summary

In this work, we study multi-stream transceiver precoding for mmWave MIMO systems over general frequency selective fading channels. The proposed algorithm enjoys low complexity and training overhead on the order of $m \times (N_{\rm t} + N_{\rm r})$, where m is the number of iterations. Compared with the conventional approach of direct channel estimation with a training overhead on the order of $N_{\rm t} \times N_{\rm r}$, the saving is especially significant when both $N_{\rm t}, N_{\rm r}$ are large and number of spatial streams is small, which is exactly the case for mmWave MIMO communications. The proposed algorithm is robust in that it works in FDD transmissions as well as TDD transmissions with/without RF calibration. Analysis shows that the algorithm converges to the optimal solution exponentially at a speed of $(\sigma_2/\sigma_1)^{8m}$ in terms of the equivalent channel SNR. Numerical results show that the convergence is achieved usually after 3 to 4 iterations. When constant modulus constraint is applied in mmWave wireless communications, most of the gain can be maintained, and the convergence speed is almost the same. The proposed algorithm is also robust in that it works effectively not only for the structured directional mmWave wireless channel, but also for rich scattering channels. Henceforth, it may be used even when the structured directional model is not perfectly accurate, e.g. for non-mmWave wireless communications.

References

[1] S. Alalusi and S. Brodersen, "A 60Ghz phased array in CMOS," *IEEE Conference on Custom Integrated Circuits*, Sep. 2006, pp. 393 – 396.

[2] A. Alkhateeb, O. E. Ayach, G. Leus and R. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal on Sel. Topics in Sig. Proc.*, vol. 8, no. 5, Oct. 2014.

[3] A. Alkhateeb and J. Mo and N. G. Prelcic and R. W. Heath Jr, "MIMO Precoding and Combining Solutions for Millimeter-Wave Systems," *IEEE Comm. Magazine*, vol.52, no.12, pp.122 – 131, December 2014.

[4] A. Alkhateeb and R. W. Heath Jr., "Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems," *submitted to IEEE Transactions on Communications*, arXiv 1510.00609, Oct. 2015.

[5] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems,", *IEEE Trans. on Wireless Comm.*, vol. 13, no. 3, pp. 1499 – 1513, March 2014.

[6] T. Dahl and N. Christophersen and D. Gesbert, "Blind MIMO eigenmode transmission based on the algebraic power method", *IEEE Transactions on Signal Processing*, vol. 52, no. 9, pp. 2424 – 2431, Sept. 2004.

[7] H. Ghauch and T. Kim and M. Bengtsson and M. Skoglund, "Subspace Estimation and Decomposition for Large Millimeter-Wave MIMO systems," *arXiv preprint 1507.00287*, July 2015.

[8] G. H. Golub and C. F. Van Loan, "Matrix Computations," *The Johns Hopkins University Press*, 1990.

[9] S. Han, C. I., Z. Xu and C. Rowell. "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G", *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, 2015.

[10] R. W. Heath Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," to appear in the *IEEE Journal on Special Topics in Signal Processing* in 2016.

[11] D. Huang and K. B. Letaief, "Symbol based space diversity for coded OFDM systems," *IEEE Trans. on Wireless Comm*, Jan. 2004.

[12] IEEE 802.11ad Standard Draft D0.1, [Online]. Available: www.ieee802.org/11/Reports/tgad update.htm

[13] R. Srinivasan et al., "IEEE 802.16m Evaluation Methodology Document". *IEEE 802.16m-08/004r2*.

[14] T. A. Thomas, H. C. Nguyen, G. R. MacCartney Jr., and T. S. Rappaport, "3D mmWave Channel Model Proposal", *IEEE Vehicular Technology Conference*, pp. 1 – 6, 2014.

[15] WirelessHD Consortium, *WirelessHD specification 1.0*, January 2008.

[16] P. Xia, S. K. Yong, J. Oh and C. Ngo, "Multi-Stage Antenna Training Millimeter Wave Communication Systems," *IEEE Globecom Conference 2008*.

[17] S. K. Yong, "TG3c Channel Modeling Sub-committee Final Report," *IEEE 15–07–0584–01–003c*, Mar. 2007.