Hybrid MIMO-OFDM Beamforming for Wideband mmWave Channels Without Instantaneous Feedback

Yuan-Pei Lin, *Senior Member, IEEE*

***Abstract*—In this paper, we consider the design of statistical multiple-input-multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) beamformers for millimeter wave (mmWave) channels. The transmitter designs the subcarrier beamformers based on the statistics of the channel, without instantaneous channel information. To overcome the radio frequency (RF) limitation in mmWave application, the subcarrier beamformers are implemented in a hybrid structure, which imposes constraints on the design of subcarrier beamformers. We analyze the unconstrained statistical subcarrier beamformers using spectral analysis of the subcarrier channels. The analysis shows that, for each subcarrier channel, the optimal statistical beamformer is approximately a linear combination of optimal statistical beamformers for some appropriately defined narrowband single-cluster subchannels. The result suggests a design of the subcarrier beamformers that can be readily implemented in a hybrid structure. Furthermore, a hybrid design for the receiver is proposed based on the concept of vector quantization. Simulations are given to show that the use of a hybrid beamforming structure incurs a minor degradation in transmission rate. With three RF chains, the performance is close to that of all digital statistical beamforming.**

***Index Terms*—Hybrid precoding, wideband mmWave, statistical precoding, hybrid multiple-input-multiple-output orthogonal frequency division multiplexing (MIMO-OFDM).**

I. Introduction

THE performance of a MIMO system is known to improve with the number of antennas at the transmitter and receiver. Recent advances show that it is feasible to pack a large number of antennas in a small area, particularly in mmWave communication systems that use small wavelengths [1]. However cost and power constraints often prohibit having one dedicated RF chain for each antenna [2]. A promising technique to overcome the RF limitation is the so called hybrid scheme [3], [4], in which analog processing of RF signals is combined with digital processing in the baseband. Analog RF processing, due to power and complexity consideration, is typically implemented using phase shifters and the elements of the analog precoder are of unit modulus [5], [6]. Recently, it is shown in [7]–[9] that if two phase shifters are used to implement each coefficient of the RF beamformer, the design of RF beamformers can be free from the unit modulus constraint.

Limited feedback for MIMO systems have been studied extensively [10]–[17]. Efficient feedback of channel information in massive MIMO systems, where a large number of antennas, is all the more important. To reduce the amount of feedback information in massive MIMO systems, antenna response vectors are used in [3], [4] as the column vectors of the RF precoders. Product codebook design is suggested in [18] for uniform planar arrays (UPA). A joint design of RF precoders and subcarrier baseband precoders is considered in [19] for MIMO-OFDM with limited feedback. In the context of feedback consideration, statistical beamforming and precoding [20]–[23], designed based on the channel statistics, have been proposed in the literature. A statistical design requires infrequent update of channel statistics but not instantaneous feedback. Optimal beamforming for maximizing the average capacity of Rayleigh fading channels is designed in [20]. Optimal statistical beamforming is considered in [21] for a useful class of channels. Solutions of statistical beamforming on the Grassmann manifold for two-user broadcast channel are given in [22]. Channel covariance matrices are exploited in [23]–[25] to reduce the dimension of channel state information in multiuser transmission. Hybrid MIMO-OFDM is designed in [26] by constructing the RF precoder based on statistics and the baseband precoder based on the instantaneous channel. It is generally assumed that the receiver has the channel information. Channel estimation for mmWave channels has been addressed in [27], [28]. Algorithms for the estimation of wideband mmWave channels are proposed in [29], [30].

In this paper we consider statistical narrowband and wideband MIMO-OFDM beamforming for point-to-point mmWave transmission with UPAs. The transmitter knows only the statistics of the channel but not instantaneous channel information. We first design statistical beamformers for narrowband transmission and then consider the extension to the wideband case. The wideband MIMO-OFDM differs from the narrowband case in that a beamformer is used for each subcarrier but the subcarrier beamformers are constrained due to limited RF chains. Each subcarrier beamformer consists of an RF precoder that is common to all the subcarriers and a baseband digital beamformer that is subcarrier dependent. We design the subcarrier beamformers based on spectral analysis of the covariance matrices of the subcarrier channels. A multi-cluster channel is decomposed into multiple single-cluster *subchannels*. We will see that the dominant eigenvector of each of the subcarrier covariance

Manuscript received February 26, 2018; revised July 21, 2018; accepted July 22, 2018. Date of publication August 20, 2018; date of current version September 4, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Subhrakanti Dey. This work was supported by Ministry of Science and Technology, Taiwan, under Grant MOST 106-2221-E-009-031-MY3.

The author is with the Department of Electrical Engineering, National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: ypl@mail.nctu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2018.2864610

matrix is approximately a linear combination of the dominant eigenvectors of the single-cluster subchannels. The observation lends itself to the design of subcarrier beamformers that can be readily implemented in a hybrid structure. Using these dominant eigenvectors to form the RF precoder allows us to form beams in directions that are more important statistically. As a result there is little degradation due to RF limitation. Furthermore, similarity is observed between the design of RF combiner and that of vector quantization, which leads to a hybrid design for the MIMO-OFDM receiver. With three RF chains, the constrained statistical beamforming system can achieve a rate close to that of a fully digital statistical beamforming system, in which each antenna is endowed with an RF chain. Some preliminary results on the narrowband case has been published in [31].

The main contributions are summarized as follows. For a hybrid MIMO-OFDM system, beamforming based only on statistics is proposed. Statistical beamforming has been proposed for narrowband systems [20]–[23], but not for hybrid MIMO-OFDM systems, for which the RF limitation poses new design challenges. The proposed RF precoder and baseband subcarrier beamformers are determined from statistics alone, thus less feedback overhead is needed. Spectral analysis of the covariance matrices of subcarrier channels reveals that the subcarrier statistical beamformers can be readily expressed in a hybrid manner. The finding gives rise to an algorithm for designing the RF precoder from the statistics in a straightforward manner; the design of the RF precoder does not involve the computation of all instantaneous subcarrier channels as in [19], [26]. We have also proposed the design of a hybrid MIMO-OFDM receiver, the design of which has not been addressed in earlier works.

Notation: The variance of a random variable x is denoted as σ_x^2 and the expectation of x by $E[x]$. The 2-norm of a vector $\mathbf{f}$ is denoted as $||\mathbf{f}||$. The notation $\mathbf{A}^\dagger$ represents the transpose and conjugate of a matrix $\mathbf{A}$.

II. System Model: Narrowband Transmission

Consider a MIMO channel with N_t transmit antennas and N_r receive antennas, represented by an $N_r \times N_t$ matrix $\mathbf{H}$. We adopt the clustered geometric channel representation that is useful for modeling mmWave propagation [32]. Suppose the channel consists of N_{cl} cluster and the ℓth cluster contains L_ℓ rays,

$$\mathbf{H} = \gamma \sum_{\ell=1}^{N_{cl}} \sum_{i=1}^{L_\ell} \alpha_{\ell,i} \mathbf{a}_r(\phi_{\ell,i}^r, \theta_{\ell,i}^r) \mathbf{a}_t^\dagger(\phi_{\ell,i}^t, \theta_{\ell,i}^t), \tag{1}$$

where $\gamma = \sqrt{N_t N_r / \sum_{\ell=1}^{N_{cl}} L_\ell}$ is a normalization factor, $\alpha_{\ell,i}$, denoting the complex gain of the ith ray in the ℓth cluster, are assumed to be independent random variables of zero mean. The azimuth (elevation) angle[1] of departures (AoD) are denoted by $\phi_{\ell,i}^t$ ($\theta_{\ell,i}^t$) and the azimuth (elevation) angle of arrival (AoA) by $\phi_{\ell,i}^r$ ($\theta_{\ell,i}^r$). The angles of departure $\phi_{\ell,i}^t$ ($\theta_{\ell,i}^t$) are independent, of mean $\overline{\phi}_\ell^t$ ($\overline{\theta}_\ell^t$) and standard deviations (also called angular spread) $\sigma_{\phi_\ell^t}$ ($\sigma_{\theta_\ell^t}$). The complex gain $\alpha_{\ell,i}$, AoD and AoA are assumed to be independent. The vectors $\mathbf{a}_t(\phi_{\ell,i}^t, \theta_{\ell,i}^t)$ and $\mathbf{a}_r(\phi_{\ell,i}^r, \phi_{\ell,i}^r)$ are, respectively, the transmit and receive antenna array response vectors. The array response vector for an UPA arranged on the yz-plane with size $N_z \times N_y$ (N_z in the z-direction and N_y in the y-direction) is given by $[\mathbf{a}(\phi,\theta)]_{m+nN_z} = \frac{1}{\sqrt{N_y N_z}} e^{j\xi(m\cos(\theta)+n\sin(\theta)\sin(\phi))}$, for $0 \le m < N_z$ and $0 \le n < N_y$, where $\xi = 2\pi d$ and d is the antenna spacing normalized by the wavelength. When $N_y = 1$, the antenna array becomes a uniform linear array (ULA) along the z-axis. We assume the transmit antenna array is $N_{t,z} \times N_{t,y}$.

Let the transmitted symbol be s and the $N_t \times 1$ transmit beamformer be $\mathbf{f}$, then the transmission power is $P_t = ||\mathbf{f}||^2 \sigma_s^2$. The output of the receiver is $r = \mathbf{g}^\dagger \mathbf{H}\mathbf{s} + \mathbf{g}^\dagger \mathbf{n}$, where $\mathbf{g}$ is the $N_r \times 1$ receive combiner, $\mathbf{n}$ is the $N_r \times 1$ channel noise vector. Assume the elements of $\mathbf{n}$ are independent, of variance N_0 and zero mean, and the beamformer and the combiner are normalized such that $||\mathbf{f}|| = 1$ and $||\mathbf{g}|| = 1$, then the output SNR is $\text{SNR} = \frac{P_t}{N_0} |\mathbf{g}^\dagger \mathbf{H} \mathbf{f}|^2$. The transmission rate is given by $\log_2(1 + \text{SNR})$.

III. Statistical Beamforming

The transmitter knows only the statistics of the channel, but has no instantaneous feedback of the channel information. The beamformer $\mathbf{f}$ is designed based on the statistics. The receiver has the channel information and knows the statistical beamformer used at the transmitter. In this case, the optimal combiner is $\mathbf{g} = \mathbf{H}\mathbf{f}$. Averaging the SNR over the random channel, we get

$$E[\text{SNR}] = \frac{P_t}{N_0} E[||\mathbf{H}\mathbf{f}||^2]. \tag{2}$$

To maximize the average SNR, we design the statistical beamformer to solve the following problem: $\max_{\mathbf{f}\text{ s.t. }||\mathbf{f}||=1} \mathbf{f}^\dagger \mathbf{\Sigma}_t \mathbf{f}$, where $\mathbf{\Sigma}_t = E[\mathbf{H}^\dagger \mathbf{H}]$. The optimal beamformer is the unit eigenvector of the transmit covariance matrix $\mathbf{\Sigma}_t$ that corresponds to the largest eigenvalue. In what follows we show that for the channel model given in (1), the matrix $\mathbf{\Sigma}_t$ can be approximated in a closed form when the angular spreads are small.

Notice that the channel in (1) can be written in a matrix form as

$$\mathbf{H} = \gamma \sum_{\ell=1}^{N_{cl}} \mathbf{A}_{r,\ell} \mathbf{D}_\ell \mathbf{A}_{t,\ell}^\dagger, \tag{3}$$

where $\mathbf{A}_{t,\ell}$ is the $N_t \times L_\ell$ matrix whose column vectors are the transmit antenna array response vectors corresponding to the ℓth cluster: $\mathbf{a}_t(\phi_{\ell,1}^t, \theta_{\ell,1}^t), \ldots, \mathbf{a}_t(\phi_{\ell,L_\ell}^t, \theta_{\ell,L_\ell}^t)$, whereas $\mathbf{A}_{r,\ell}$ is the $N_r \times L_\ell$ matrix whose column vectors are the L_ℓ receive antenna array response vectors. The matrix $\mathbf{D}_\ell$, of size $L_\ell \times L_\ell$, is diagonal with diagonal elements $\alpha_{\ell,1}, \alpha_{\ell,2}, \ldots, \alpha_{\ell,L_\ell}$.

Lemma 1: Consider the clustered channel $\mathbf{H}$ in (1). The transmit covariance matrix $\mathbf{\Sigma}_t = E[\mathbf{H}^\dagger \mathbf{H}]$ is given by

$$\mathbf{\Sigma}_t = \gamma^2 \sum_{\ell=1}^{N_{cl}} \sigma_{\alpha_\ell^2} \mathbf{C}_\ell, \quad \text{where } \mathbf{C}_\ell = E\left[\mathbf{A}_{t,\ell} \mathbf{A}_{t,\ell}^\dagger\right]. \tag{4}$$

[1]The elevation angle of a ray is the angle between the ray and the z-axis whereas the azimuth angle is the angle between the x-axis and the orthogonal projection of the ray on the xy-plane.

In particular

$$[\mathbf{\Sigma}_t]_{m+nN_{t,z},q+kN_{t,z}} = \frac{\gamma^2}{N_t}\sum_{\ell=1}^{N_{cl}}\sigma_{\alpha_\ell}^2 L_\ell \mu_\ell(m-q,n-k), \quad (5)$$

for $0 \le m,q < N_{t,z}$ and $0 \le n,k < N_{t,y}$, where $\mu_\ell(m,n) = E[e^{j\xi(m\cos\theta_{\ell,i}^t + n\sin\theta_{\ell,i}^t \sin\phi_{\ell,i}^t)}]$.

See Appendix A for a proof. The function $\mu_\ell(m,n)$ depends on the statistics of the AoD.

Theorem 1: When the angular spreads $\sigma_{\theta_\ell^t}$ and $\sigma_{\phi_\ell^t}$ are small, the transmit covariance matrix can be approximated by

$$\begin{aligned}[\mathbf{\Sigma}_t]&_{m+nN_{t,z},q+kN_{t,z}} \\ &\approx \frac{\gamma^2}{N_t}\sum_{\ell=1}^{N_{cl}} L_\ell \sigma_{\alpha_\ell}^2 e^{j\xi((m-q)\cos\overline{\theta}_\ell^t + (n-k)\sin\overline{\theta}_\ell^t \sin\overline{\phi}_\ell^t))} \\ &\quad\times \nu_{\Delta\theta_{\ell,1}^t}\left(-(m-q)\xi\sin\overline{\theta}_\ell^t + (n-k)\xi\cos\overline{\theta}_\ell^t\sin\overline{\phi}_\ell^t\right) \\ &\quad\times \nu_{\Delta\phi_{\ell,i}^t}\left((n-k)\xi\sin\overline{\theta}_\ell^t\cos\overline{\phi}_\ell^t\right), \end{aligned} \quad (6)$$

for $0 \le m,q < N_{t,z}$ and $0 \le n,k < N_{t,y}$, where $\Delta\theta_{\ell,1}^t = \theta_{\ell,1}^t - \overline{\theta}_\ell^t$, $\Delta\phi_{\ell,1}^t = \phi_{\ell,1}^t - \overline{\phi}_\ell^t$, and $\nu_x(\beta) = E[e^{j\beta x}]$.

A proof is given in Appendix B. Note that both $\Delta\theta_{\ell,i}^t$ and $\Delta\phi_{\ell,i}^t$ are zero-mean random variables, for which truncated Laplacian and Gaussian distributions have been suggested [33][34]. In the case of truncated Laplacian distribution, we can verify that

$$\nu_x(\beta) = \frac{1}{c_0\left(1+\frac{\beta^2\sigma^2}{2}\right)}\left[1 + e^{-\frac{\pi\sqrt{2}}{\sigma}}\left(\frac{\beta\sigma}{\sqrt{2}}\sin(\beta\pi) - \cos(\beta\pi)\right)\right], \quad (7)$$

where $c_0 = 1 - e^{-\pi\sqrt{2}/\sigma}$ is a constant, whereas for Gaussian distribution, we have

$$\nu_x(\beta) = 1 + \frac{(\beta\sigma)^2}{2}\left(\frac{\sqrt{2\pi}}{c_1\sigma}e^{-\pi^2/(2\sigma^2)} - 1\right), \quad (8)$$

where $c_1 = \frac{1}{\sqrt{2\pi}}(F_g(\pi/\sigma) - F_g(-\pi/\sigma))$ and $F_g(x)$ is the CDF of a zero-mean Gaussian random variable with unit variance.

With the above closed-form approximation of $\mathbf{\Sigma}_t$, the beamforming vector can be computed accordingly. The derivations of (6) are based on the first-order Taylor series approximations of $\sin(x)$ and $\cos(x)$ for small x, where x is $\Delta\theta_{\ell,i}^t$ or $\Delta\phi_{\ell,i}^t$. The approximations are accurate with up to % 2 error for $|x| < \pi/9$. The probability that $|x|$ is larger than 4 times its standard deviation is less than 0.7% for the Laplacian case and less than 0.0063% for the Gaussian case. Therefore the approximation is a good one when the standard deviation is less than $\pi/36$, i.e, 5°. Although the derivation of $\mathbf{\Sigma}_t$ in Theorem 1 requires the assumption that the angular spreads are small, simulation examples will be given to demonstrate that the approximation is accurate and the resulting beamformer useful even for larger angular spread. When $N_y = 1$, the antennas form a uniform linear array (ULA) along the z-axis and the results derived for UPA are valid for ULA as well by setting $N_y = 1$. The autocorrelation of the vectorized channel for a single-cluster case has been derived in [38] using a similar approach for ULA and uniform circular array.

To analyze the theoretical performance of the statistical beamformer, let us consider a channel that has only one cluster of a single ray and the antenna array is ULA. In this case $\mathbf{H} = \alpha\mathbf{a}_r(\theta^r)\mathbf{a}_t^\dagger(\theta^t)$. When the transmitter has full channel information and employs optimal beamforming, the average SNR in (2) becomes $E[\text{SNR}_{opt}] = \frac{P_t\sigma_\alpha^2}{N_0}$. For the statistical beamformer, the average SNR in (2) is given by $E[\text{SNR}_{stat}] = \frac{P_t\sigma_\alpha^2}{N_0}E[|\mathbf{a}_t^\dagger(\theta^t)\mathbf{f}|^2]$. The ratio of the two $\rho = E[|\mathbf{a}_t^\dagger(\theta^t)\mathbf{f}|^2]$ reflects the SNR loss of statistical beamforming compared to optimal beamforming.

Theorem 2: Consider the channel in (1) with a single ray and ULA antenna arrays. Suppose the angular spread is small. When the mean AoD $\overline{\theta}^t$ is such that $\sin(\overline{\theta}^t) \approx 0$, we have $\rho \approx 1$. Otherwise, the SNR loss has the approximate lower bound $\rho \gtrsim \rho_l$ when θ^t is of the truncated Laplacian distribution and $\rho \gtrsim \rho_g$ when θ^t is of the truncated Gaussian distribution. The bounds ρ_l and ρ_g are, respectively, given by,

$$\begin{aligned}\rho_l &= 1 - \frac{1}{12}\zeta^2 + \frac{1}{12}e^{-\sqrt{2}\pi/\zeta}\left(\pi^2 + \sqrt{2}\pi\zeta + \zeta^2 - 12\right), \\ \rho_g &= erf\left(\frac{\pi}{\sqrt{2}\zeta}\right)\left(1 - \zeta^2/12\right) + \frac{\sqrt{\pi}}{6\sqrt{2}}\zeta e^{-(\pi/\zeta)^2/2}, \end{aligned} \quad (9)$$

where $\zeta = \xi N_t\sigma_{\theta^t}\sin(\overline{\theta}^t)$ and the error function $erf(x)$ is defined as $\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$.

In either Laplacian and Gaussian case, the lower bound depends on ζ i.e, depends on the antenna size N_t, the mean AoD $\overline{\theta}^t$ and angular spread σ_{θ^t}. Simulations show that the lower bound gives a good estimate of the actual loss.

Magnitude response of the beamformer: To better understand the derived statistical beamformer, we look into the beamformer design from the frequency domain point of view. Let us re-index the beamforming coefficients as an array $f_{m,n}$, for $0 \le m < N_{t,z}$ and $0 \le n < N_{t,y}$. Let $F(\omega_1,\omega_2) = \sum_{m=0}^{N_{t,z}-1}\sum_{n=0}^{N_{t,y}-1} f_{m,n}e^{-j(m\omega_1+n\omega_2)}$ be the Fourier transform for the two-dimensional sequence $f_{m,n}$. The expression of $\mathbf{\Sigma}_t$ given in (4) means that the average SNR is of the form $E[\text{SNR}] = \gamma^2\frac{P_t}{N_0}\sum_{\ell=1}^{N_{cl}}\sigma_{\alpha_\ell^2}E[||\mathbf{A}_{t,\ell}^\dagger\mathbf{f}||]$. Observe that the i-th element of the vector $\mathbf{A}_{t,\ell}^\dagger\mathbf{f}$ is equal to $\frac{1}{\sqrt{N_t}}G(\phi_{\ell,i}^t,\theta_{\ell,i}^t)$, where $G(\phi,\theta) = F(\xi\cos(\theta),\xi\sin(\theta)\sin(\phi))$ is the array factor [35]. As $\theta_{\ell,i}^t$ and $\phi_{\ell,i}^t$ are identically distributed for $i = 1,2,\ldots,L_\ell$, we have

$$E[\text{SNR}] = \frac{\gamma^2 P_t}{N_0 N_t}E\left[\sum_{\ell=1}^{N_{cl}} L_\ell\sigma_{\alpha_\ell}^2|G(\phi_{\ell,1}^t,\theta_{\ell,1}^t)|^2\right]. \quad (10)$$

This implies that to maximize the SNR, we are to design $\mathbf{f}$ so that $|G(\phi,\theta)|$ is large in a more likely range of (θ,ϕ). We will see in simulations that $|F(\omega_1,\omega_2)|$ has peaks at frequencies corresponding to the means of the AoD.

Spectral analysis: Notice that from (6) the average SNR can be expressed as $E[\text{SNR}] = \frac{P_t}{N_0}\mathbf{f}^\dagger\sum_{\ell=1}^{N_{cl}}\sigma_{\alpha_\ell}^2\mathbf{C}_\ell\mathbf{f}$. We can expect that the eigenvectors of $\mathbf{C}_\ell$ that correspond to the first few most significant eigenvalues are filters with

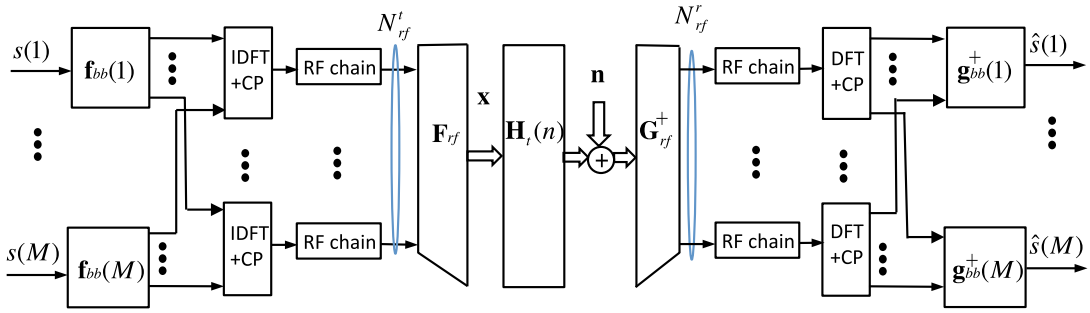

Fig. 1. A hybrid MIMO-OFDM beamforming system.

passband around $(\xi\cos(\bar{\theta}_\ell^t), \xi\sin(\bar{\theta}_\ell^t)\sin(\bar{\phi}_\ell^t))$. Approximate $\mathbf{C}_\ell$ as $\mathbf{C}_\ell \approx \mathbf{V}_\ell\mathbf{\Lambda}_\ell\mathbf{V}_\ell^\dagger$, where $\mathbf{\Lambda}_\ell$ is diagonal matrix whose diagonal elements are the first few dominant eigenvalues of $\mathbf{C}_\ell$ and $\mathbf{V}_\ell$ is a semi-unitary matrix that consists of the eigenvectors corresponding to the eigenvalues contained in $\mathbf{\Lambda}_\ell$. With the above approximation of $\mathbf{C}_\ell$, we have

$$\mathbf{\Sigma}_t \approx \gamma^2 \sum_{\ell=1}^{N_{cl}} \sigma_{\alpha_\ell}^2 \mathbf{V}_\ell\mathbf{\Lambda}_\ell\mathbf{V}_\ell^\dagger. \tag{11}$$

Let $\mathbf{V} = \begin{pmatrix}\mathbf{V}_1 & \mathbf{V}_2 & \cdots & \mathbf{V}_{N_{cl}}\end{pmatrix}$. We can rewrite (11) as $\mathbf{\Sigma}_t \approx \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\dagger$, where $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are those of $\sigma_{\alpha_\ell}^2\mathbf{\Lambda}_\ell$. Thus the optimal statistical beamformer is a linear combination of the vectors in $\mathbf{V}$ and its Fourier transform $F(\omega_1, \omega_2)$ may have multiple passbands, depending on the linear combination coefficients. This property can be exploited in Sec. V to design hybrid beamformers for wideband MIMO-OFDM systems. It has also been observed in [36] that the optimal beamformer for a given instantaneous channel is a linear combination of eigenvectors corresponding to individual clusters. Steering based on the dominant clusters has been considered in [37].

Remark: When there are multiple clusters and the clusters are well separated, i.e., the AoD cluster means spatially separated, the beams of the dominant eigenvectors of different clusters do not overlap and they are almost orthogonal, particularly so when N_t is large. That is, $\mathbf{V}_\ell^\dagger\mathbf{V}_m \approx \mathbf{0}$ for $\ell \neq m$. In this case $\mathbf{V}$ is approximately semi-unitary and $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^\dagger$ is an approximate eigen decomposition of $\mathbf{\Sigma}_t$. The optimal statistical beamformer is a column vector of $\mathbf{V}$ and the optimal solution in the multi-cluster case is the similar to the solution corresponding to a certain cluster.

IV. System Model: Wideband Transmission

In the narrowband case the time delays of different clusters are small compared to the sampling period T_s and the time delays are ignored. For wideband transmission, the delays of different clusters need to be taken into consideration. With sampling period T_s, a wideband geometric channel model of N_{cl} clusters is of the form [32]

$$\mathbf{H}_t(n) = \gamma \sum_{\ell=1}^{N_{cl}} \sum_{i=1}^{L_\ell} \alpha_{\ell,i} p(nT_s - \tau_{\ell,i}) \mathbf{a}_r(\phi_{\ell,i}^r, \theta_{\ell,i}^r)\mathbf{a}_t^\dagger(\phi_{\ell,i}^t, \theta_{\ell,i}^t), \tag{12}$$

where $\alpha_{\ell,i}$, γ, $\mathbf{a}_t(\phi_{\ell,i}^t, \theta_{\ell,i}^t)$ and $\mathbf{a}_r(\phi_{\ell,i}^r, \theta_{\ell,i}^r)$ are as in (1). The function $p(t)$ represents the lumped pulse shaping functions of the transmitter and receiver, and $\tau_{\ell,i}$ is the delay associated with the ith ray in the ℓth cluster. The inter-symbol interference due to multipath can be removed through the insertion of proper cyclic prefix and applying DFT and IDFT in MIMO-OFDM. Let the DFT size be M and the length of cyclic prefix be N_{cp}. When N_{cp} is longer than the largest time delay, the equivalent channel of the kth subcarrier is

$$\mathbf{H}(k) = \frac{1}{\sqrt{M}} \sum_{n=0}^{N_{cp}} \mathbf{H}_t(n) e^{-j2\pi kn/M}, \quad k = 0, 1, \ldots, M-1. \tag{13}$$

Achieving the optimal beamforming gain requires as many RF chains as antennas. In practice the number of RF chain is limited and a hybrid structure, as shown in Fig. 1, can be used to reduce the degradation due to the RF chain limitation [19]. The number of RF chains at the transmitter N_{rf}^t is usually much smaller than the number of antennas N_t. The equivalent subcarrier beamformers are of the form [3][19]

$$\mathbf{f}(k) = \mathbf{F}_{rf}\mathbf{f}_{bb}(k), \tag{14}$$

where $\mathbf{F}_{rf}$ is an $N_t \times N_{rf}^t$ matrix that represents the RF analog processing, as shown in Fig. 1, and $\mathbf{f}_{bb}(k)$ is the $N_{rf}^t \times 1$ baseband beamformer of the kth subcarrier before the IDFT. Similarly, at the receiver we have an RF combiner $\mathbf{G}_{rf}$ and baseband subcarrier combiners $\mathbf{g}_{bb}(k)$. When the transmission power is P_t and all the transmitted symbols are of the same variance, the SNR of the kth subcarrier is

$$\text{SNR}(k) = \frac{P_t}{N_0} \frac{|\mathbf{g}_{bb}^\dagger(k)\mathbf{G}_{rf}^\dagger\mathbf{H}(k)\mathbf{F}_{rf}\mathbf{f}_{bb}(k)|^2}{||\mathbf{g}_{bb}^\dagger(k)\mathbf{G}_{rf}^\dagger||^2||\mathbf{F}_{rf}\mathbf{f}_{bb}(k)||^2}. \tag{15}$$

The achievable sum rate is $\sum_{k=0}^{M-1}\log_2(1+\text{SNR}(k))$.

V. Beamforming for Wideband MIMO-OFDM Systems

Consider the wideband channel model in (12) and the MIMO-OFDM system in Fig. 1. In what follows we will first examine the subcarrier channels and approximate them as narrowband clustered channels like the one given in (1). Based on this result, a statistical design of hybrid subcarrier beamformers is given (Sec. 5.1). The design of hybrid combiners is addressed in Sec. 5.2.

We can write $\tau_{\ell,i} = \overline{\tau}_\ell + \Delta\tau_{\ell,i}$, where $\overline{\tau}_\ell$ is the mean delay of the ℓth cluster and $\Delta\tau_{\ell,i}$ is the delay relative to $\overline{\tau}_\ell$. Assume, reasonably, that the rays in the same cluster have small time delay variations. In particular, when $\Delta\tau_{\ell,i} << T_s$, we have $p(nT_s - \tau_{\ell,i}) \approx p(nT_s - \overline{\tau}_\ell)$. Defining $p_{\ell,k}$ as $p_{\ell,k} = \sum_{n=0}^{N_{cp}} p(nT_s - \overline{\tau}_\ell)e^{-j2\pi kn/M}$, we get

$$\mathbf{H}(k) \approx \gamma \sum_{\ell=1}^{N_{cl}} \sum_{i=1}^{L_\ell} \beta_{\ell,i,k} \mathbf{a}_r(\phi_{\ell,i}^r, \theta_{\ell,i}^r)\mathbf{a}_t^\dagger(\phi_{\ell,i}^t, \theta_{\ell,i}^t), \quad (16)$$

where $\beta_{\ell,i,k} = \alpha_{\ell,i}p_{\ell,k}$. We observe that the right hand side of the above equation is of the form in (1); each $\mathbf{H}(k)$ is a narrow-band geometric clustered channel as in (1). For the kth subchannel, the complex gain of the ith ray in the ℓth cluster is $\beta_{\ell,i,k}$, which depends on $\alpha_{\ell,i}$ and also on the M-point DFT of $p(nT_s - \overline{\tau}_\ell)$. It follows that $\sigma^2_{\beta_{\ell,i,k}} = \sigma^2_{\alpha_\ell}|p_{\ell,k}|^2$. Due to the DFT and IDFT in OFDM, the clusters of different time delays are spread to all frequency bins to form narrowband subcarrier channels. The spreading is affected by $|p_{\ell,k}|^2$ and differs from subcarrier to subcarrier.

A. Design of Hybrid Beamformers

Given the statistics of the channel, we can design statistical beamformers for each subchannel $\mathbf{H}(k)$ as in Sec. 3. However the resulting statistical beamformers can not be implemented in a hybrid manner as in Fig. 1; they require a fully digital implementation in general. In the hybrid structure, the beamformers are of the form in (14); only the baseband processing $\mathbf{f}_{bb}(k)$ varies from subcarrier to subcarrier and $\mathbf{F}_{rf}$ is common to all the subcarriers. Therefore each subcarrier beamformer $\mathbf{f}(k)$ is constrained to be a linear combination of the column vectors of $\mathbf{F}_{rf}$. In what follows, a statistical approach to the design of the RF precoder $\mathbf{F}_{rf}$ and baseband beamformers $\mathbf{f}_{bb}(k)$ is given below.

Assume the receiver knows the channel, which can be estimated using algorithms developed for mmWave channels [29], [30]. Suppose the receiver combining vector $\mathbf{g}(k)$ is not constrained by the hybrid structure and $\mathbf{g}(k) = \mathbf{H}(k)\mathbf{f}(k)$, for a given $\mathbf{f}(k)$. The SNR maximization problem for the k-th subcarrier is

$$\max_{\mathbf{f}(k) \text{ s.t. } ||\mathbf{f}(k)||=1} \mathbf{f}(k)^\dagger \mathbf{\Sigma}_t(k)\mathbf{f}(k),$$
$$\text{where } \mathbf{\Sigma}_t(k) = E[\mathbf{H}^\dagger(k)\mathbf{H}(k)]. \quad (17)$$

Notice that $\mathbf{\Sigma}_t(k)$ can be expressed as

$$\mathbf{\Sigma}_t(k) = \gamma^2 \sum_{\ell=1}^{N_{cl}} \sigma^2_{\alpha_\ell}|p_{\ell,k}|^2 \mathbf{C}_\ell,$$

where $\mathbf{C}_\ell = E[\mathbf{A}_{t,\ell}\mathbf{A}_{t,\ell}^\dagger]$ is as defined in (4). As in Sec. 3, we approximate $\mathbf{C}_\ell$ as $\mathbf{C}_\ell \approx \mathbf{V}_\ell \mathbf{\Lambda}_\ell \mathbf{V}_\ell^\dagger$, where the diagonal elements of $\mathbf{\Lambda}_\ell$ are in non increasing order, $\lambda_{\ell,1} \geq \lambda_{\ell,2} \geq \cdots$, and the column vectors of $\mathbf{V}_\ell$ are $\{\mathbf{v}_{\ell,i}\}$. Thus $\mathbf{\Sigma}_t(k) \approx \gamma^2 \mathbf{V}\mathbf{\Lambda}(k)\mathbf{V}^\dagger$, where $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are those of $\sigma^2_{\alpha_\ell}|p_{\ell,k}|^2\mathbf{\Lambda}_\ell$ and $\mathbf{V} = \begin{pmatrix}\mathbf{V}_1 & \mathbf{V}_2 & \cdots & \mathbf{V}_{N_{cl}}\end{pmatrix}$. Notice that $\mathbf{V}$ does not depend on k, only $\mathbf{\Lambda}(k)$ does. The dominant eigenvector of $\mathbf{\Sigma}_t(k)$ is approximately a linear combination of the column vectors of $\mathbf{V}$. The linear combination is different for different subcarriers due to the factor $|p_{\ell,k}|^2$. This observation suggests the following: we can use the column vectors in $\mathbf{V}$ to construct $\mathbf{F}_{rf}$. Having chosen $\mathbf{F}_{rf}$, we can then design the linear combination coefficients for the kth subcarrier, i.e., $\mathbf{f}_{bb}(k)$. Thus the beamformer of the kth subcarrier $\mathbf{f}(k)$ can be expressed as $\mathbf{f}(k) = \mathbf{F}_{rf}\mathbf{f}_{bb}(k)$, a form that can be implemented using a hybrid structure. Algorithm 1 gives a procedure to choose column vectors of $\mathbf{V}$ to form $\mathbf{F}_{rf}$. In the case the number of RF chains is the same as that of clusters, we can collect the first vector of $\mathbf{V}_\ell$ to form $\mathbf{F}_{rf}$. Otherwise priority is given to those $\mathbf{v}_{\ell,1}$ that are associated with a larger $\sigma^2_{\alpha_\ell}\lambda_{\ell,1}$, which is the value that determines the average SNR when the channel consists of a single cluster.

Algorithm 1: Design of $\mathbf{F}_{rf}$.

The algorithm finds a set S. The vectors in S are used to form the matrix $\mathbf{F}_{rf}$.

1. *Initialization:*
   Let S be an empty set.
   $N_d := \lfloor N_{rf}^t/N_{cl} \rfloor$
   $T := \{1, 2, \ldots, N_{cl}\}$
2. *Construction of S:*
   Add the first N_d column vectors of $\mathbf{V}_\ell$ to S for $\ell = 1, 2, \ldots, N_{cl}$.
   for $k = 1 : N_{rf}^t - N_d N_{cl}$ **do**
      $j := \text{argmax}_{m \in T} \sigma^2_{\alpha_m} \lambda_{m,N_d+1}$
      Add $\mathbf{v}_{j,N_d+1}$ to S
      Remove j from T
   end for
3. Use the vectors in S as the column vectors of $\mathbf{F}_{rf}$.

For a given $\mathbf{F}_{rf}$, the effective channel becomes $\mathbf{H}(k)\mathbf{F}_{rf}$. The SNR maximization in this case is $\max_{\mathbf{f}_{bb}(k)} \mathbf{f}_{bb}^\dagger(k)\mathbf{F}_{rf}^\dagger \mathbf{\Sigma}_t(k)\mathbf{F}_{rf}\mathbf{f}_{bb}(k)$, subject to $||\mathbf{F}_{rf}\mathbf{f}_{bb}(k)|| = 1$. The optimal $\mathbf{f}_{bb}(k)$ is given by (a proof given in Appendix D)

$$\mathbf{f}_{bb}(k) = \mathbf{Q}^{-1/2}\mathbf{b}(k)/||\mathbf{F}_{rf}\mathbf{Q}^{-1/2}\mathbf{b}(k)||, \quad (18)$$

where $\mathbf{Q} = \mathbf{F}_{rf}^\dagger \mathbf{F}_{rf}$ and $\mathbf{b}(k)$ is the eigenvector of $(\mathbf{F}_{rf}\mathbf{Q}^{-1/2})^\dagger \mathbf{\Sigma}_t(k)\mathbf{F}_{rf}\mathbf{Q}^{-1/2}$ that corresponds to the largest eigenvalue.

Remarks:

1) When the clusters are separated, $\mathbf{V}_\ell^\dagger \mathbf{V}_m \approx \mathbf{0}$ for $\ell \neq m$ as in the narrowband case. Thus, we have $\mathbf{F}_{rf}^\dagger \mathbf{F}_{rf} \approx c\mathbf{I}_{N_{rf}}$, which generally holds when N_t is large. Then the solution of $\mathbf{f}_{bb}(k)$ in (18) becomes the dominant eigenvector of $\mathbf{F}_{rf}^\dagger \mathbf{\Sigma}_t(k)\mathbf{F}_{rf}$ with norm normalization. In this case, we can extend single-stream transmission to multi-stream case. For N_s-substream transmission a suboptimal solution is to replace $\mathbf{f}_{bb}(k)$ with an $N_{rf} \times N_s$ matrix $\mathbf{F}_{bb}(k)$ by choosing the N_s eigenvectors of $\mathbf{F}_{rf}^\dagger \mathbf{\Sigma}_t(k)\mathbf{F}_{rf}$ that correspond to the largest N_s eigenvalues.

2) In the above design neither the RF precoder $\mathbf{F}_{rf}$ nor the subcarrier beamformer depends on the instantaneous channels. They are determined only by the statistics of the channel, $\sigma^2_{\alpha_\ell}$,

mean AoD $\overline{\theta}_\ell^t$ and $\overline{\phi}_\ell^t$, angular spread $\sigma_{\theta_\ell^t}$ and $\sigma_{\phi_\ell^t}$, and mean delay $\overline{\tau}_\ell$. These statistics can be quantized and fed back to the transmitter infrequently. To reduce feedback, we can send back only the statistics of the clusters whose eigenvectors are chosen to form $\mathbf{F}_{rf}$. Simulations show that the feedback of the quantized statistics incurs a minor degradation. In addition, the RF precoder can be obtained directly from the statistics using Algorithm 1, not involving subcarrier channels as in [19], [26].

3) With the hybrid structure, the subcarrier beamformers are constrained to be linear combinations of the column vectors of $\mathbf{F}_{rf}$. Using the eigen vectors from individual clusters to from $\mathbf{F}_{rf}$ allow us to transmit signals in directions that are more important statistically. Thus there is little degradation due to RF limitation, particularly when there are three or more RF chains.

4) The RF precoder $\mathbf{F}_{rf}$ designed using Algorithm 1 does not satisfy the unit modulus constraint that is typically imposed on the RF precoder for phase-shifter implementation [2], [4]. It can be implemented using the two-phase-shifter-per-coefficient (THIC) method [7], [9] by expressing each coefficient as the sum of two phase shifters. The phase shifters can then be further quantized for finite resolution implementation. For single-phase-shifter implementation, the solution of $\widetilde{\mathbf{F}}_{rf}$ that minimizes $||\mathbf{F}_{rf} - \widetilde{\mathbf{F}}_{rf}||$ has been found in [5] to be given by $[\widetilde{\mathbf{F}}_{rf}]_{i,j} = e^{j\angle[\mathbf{F}_{rf}]_{i,j}}$.

B. Design of Hybrid Combiners

In the discussion in Sec. 5.1, we assume that the combiner for the k-th subcarrier is not constrained and the ideal combiner is $\mathbf{g}_{uc}(k) = \mathbf{H}(k)\mathbf{f}(k)$, where we have added the subscript 'uc' to indicate that they are the ideal combiners, not constrained by a hybrid structure. When a hybrid implementation used, the subcarrier combiners are of the form $\mathbf{g}(k) = \mathbf{G}_{rf}\mathbf{g}_{bb}(k)$, where $\mathbf{G}_{rf}$ is an N_r by N_{rf}^r matrix that is independent of subcarriers and the baseband combiner $\mathbf{g}_{bb}(k)$ can be different for each subcarrier. In the hybrid structure, each subcarrier combiner $\mathbf{g}(k)$ is a linear combination of the N_{rf}^r column vectors of $\mathbf{G}_{rf}$. In other words, M combiners are represented using only N_{rf}^r vectors. This is similar to the problem considered in the LBG (Linde-Buzo-Gray) algorithm, which finds a set of vectors to represent a larger set. We treat the column vectors of the RF combiner, denoted as $\mathbf{g}_{rf,i}$ for $i = 1, 2, \cdots N_{rf}^r$, as the N_{rf}^r vectors that are used to represent the ideal unconstrained combiners $\{\mathbf{g}_{uc}(k)\}$. In the LBG algorithm, the ideal combiners are clustered into N_{rf}^r groups and $\mathbf{g}_{rf,i}$ is the representative of the the ideal combiners in the ith group. For the measure of distance, we use the projection F-norm distance [40], as the minimization of which also maximizes the SNR in (15). The projection F-norm distance, defined as the Frobenius norm of $\frac{1}{\sqrt{2}}(\mathbf{A}\mathbf{A}^\dagger - \mathbf{B}\mathbf{B}^\dagger)$ for two $n \times p$ semiunitary matrices $\mathbf{A}$ and $\mathbf{B}$, reduces to $1 - |\mathbf{a}^\dagger\mathbf{b}|^2$ for two column vectors of unit norm. The LBG algorithm iterates the following nearest neighbour and centroid conditions [39].

- *Nearest neighbor condition:* The ideal combiners are grouped according to the nearest neighbour rule. The ith region is given by $\mathcal{R}_i = \{k : |\mathbf{g}_{rf,i}^\dagger \mathbf{g}_{uc}(k)| \geq |\mathbf{g}_{rf,m}^\dagger \mathbf{g}_{uc}(k)|, m = 1, 2, \ldots, N_{rf}^r\}$.
- *Centroid condition:* Within the ith group, we choose $\mathbf{g}_{rf,i}$ as the unit vector that minimizes the distortion $\sum_{k\in\mathcal{R}_i} 1 - |\mathbf{g}_{rf,i}^\dagger \mathbf{g}_{uc}(k)|^2$. This means the optimal solution of $\mathbf{g}_{rf,i}$ is the unit eigenvector of $\sum_{k\in\mathcal{R}_i} \mathbf{g}_{uc}(k)\mathbf{g}_{uc}^\dagger(k)$ that corresponds to its largest eigenvalue.

For a given $\mathbf{G}_{rf}$, we can design $\mathbf{g}_{bb}(k)$ to maximize the kth subcarrier SNR. Using a technique similar to that given in the proof of Lemma 3, we can show that the optimal $\mathbf{g}_{bb}(k)$ can be given in a closed-form by $\mathbf{g}_{bb}(k) = (\mathbf{G}_{rf}\mathbf{G}_{rf}^\dagger)^{-1/2}\mathbf{G}_{rf}^\dagger\mathbf{H}(k)\mathbf{f}(k)$. The above discussion is for the single-stream case. When there are N_s streams with $N_s \geq N_{rf}^t$ and $N_s \geq N_{rf}^r$, one possible approach is to design the RF combiner as in the single-stream case. The baseband combiner $\mathbf{G}_{bb}(k)$, now an $N_{rf}^r \times N_s$ matrix, can be chosen so that $\mathbf{G}_{bb}^\dagger(k)\mathbf{G}_{rf}^\dagger\mathbf{H}(k)\mathbf{F}_{rf}\mathbf{F}_{bb}(k)$ is an identity matrix and we can have a zero forcing receiver.

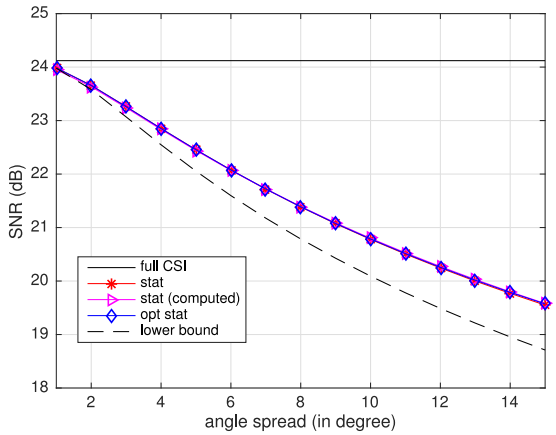

Fig. 2. Average SNR for a single-ray ULA channel with $(N_t, N_r) = (16, 16)$.

VI. Simulation Examples

Consider the geometric channel model in (1). The complex gains $\{\alpha_\ell\}$ are assumed to be Gaussian random variables of zero mean and unit variance, unless mentioned otherwise. UPAs or ULAs with omnidirectional antennas are assumed. The antenna spacing is half wavelength, thus $d = 1/2$ and truncated Laplacian distribution is used for AoD and AoA. Narrowband transmission is considered in Examples 1–3, and wideband MIMO-OFDM in Example 4.

Example 1: Consider the channel model in (1) with only one cluster of one ray and $P_t/N_0 = 0$ dB. For ULA with $(N_t, N_r) = (16, 16)$ and $\overline{\theta}^t = \pi/4$, the average SNR is shown as a function of the angular spread σ_{θ^t} in Fig. 2 . The antennas are arranged in a line along the z-axis. The statistical beamformer (labeled as 'stat') is obtained through the approximation in Theorem 1. The optimal statistical solution (labeled as 'optimal stat') is obtained by first evaluating the average of $\mathbf{H}^\dagger\mathbf{H}$ using 10^4 realizations and then computing the dominant eigenvector. Also shown is 'stat (computed)', which corresponds to the maximum eigen-

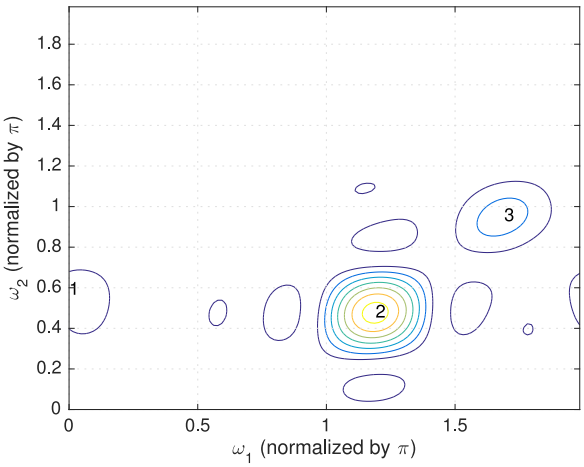

Fig. 3. Magnitude response of the statistical beamformer $|F(\omega_1, \omega_2)|$ for a channel with three clusters of fixed means.

value of the transmit covariance matrix given in (6). We see that the three curves overlap in this case and 'stat (computed)' gives a good estimate of the actual SNR. We have also shown the average SNR computed using the lower bound given in (9). We see the bound is more accurate for small angular spread. The gap with 'stat' is less than 0.9 dB when the spread is 15°.

Example 2: Consider the UPA channel model in (1) with three clusters, each of ten rays and $(N_t, N_r) = (64, 16)$. Suppose the cluster means are fixed and the means of the AoD in elevation and azimuth $(\bar{\theta}_\ell^t, \bar{\phi}_\ell^t)$ for the three clusters are respectively $(0.5\pi, 0.2\pi)$, $(0.8\pi, 0.3\pi)$ and $(0.6\pi, 0.5\pi)$. The angular spreads of AoD and AoA in azimuth and elevation are 8° for each cluster. Fig. 3 shows the magnitude response $|F(\omega_1, \omega_2)|$ as a function of ω_1 and ω_2, where we have marked the cluster number at frequencies corresponding to the mean of AoD, i.e., $(\xi \cos(\bar{\theta}_\ell^t), \xi \sin(\bar{\theta}_\ell^t) \sin(\bar{\phi}_\ell^t))$. We see that the peaks of the magnitude response $|F(\omega_1, \omega_2)|$ agree with the marked frequencies. But the peaks of $|F(\omega_1, \omega_2)|$ are not of the same height in these frequencies, i.e., power allocation done implicitly. The largest peak is at the second mean $(0.8\pi, 0.3\pi)$, a smaller peak at the third mean and a even smaller one at the first.

Example 3: In this example, we evaluate the performance of the statistical beamformer for the UPA channel in (1) with 3 clusters, each of 10 rays and $(N_t, N_r) = (64, 16)$. The means of the AoD in elevation and azimuth are uniformly distributed over $[0, 2\pi]$. Fig. 4 (a) shows the average output SNR as a function of the angular spreads ($\sigma_{\theta_\ell^t} = \sigma_{\phi_\ell^t}$) for $P_t/N_0 = 0$ dB. The curves of the statistical beamformer ('stat') and the optimal statistical solution ('opt stat') are close, especially for angular spread smaller than 5°. The formula in Theorem 1, though obtained under small angular spread assumption, is useful even for moderate spread. As a benchmark, we also show in the figure the performance of the optimal beamformer when the transmitter has full CSI. The difference is around 2.3 dB for a small angular spread and increases to around 3.5 dB when the angular spread is 10°. Also shown in Fig. 4(a) is the average SNR of the statistical beamformer when the coefficients are quantized due to the use of finite resolution phase shifters. Each beamforming

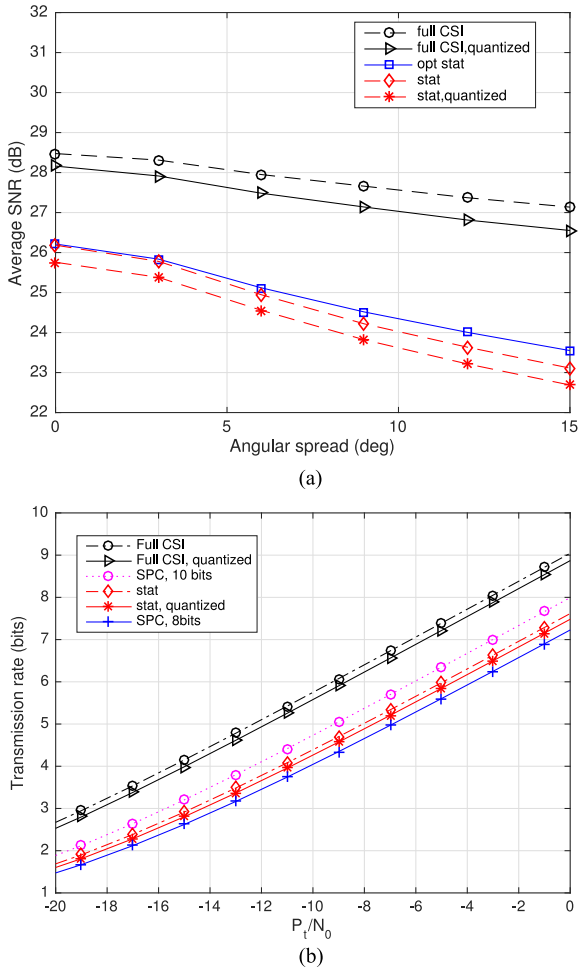

Fig. 4. Statistical beamforming for UPA with $(N_t, N_r) = (64, 16)$, (a) average SNR vs. angular spread, and (b) transmission rate performance.

weight is implemented as the sum of two phase shifters that are of three-bit resolution [7], [9]. We see that the degradation due to quantization is around 0.4 dB. Fig. 4(b) shows the transmission rate as a function of P_t/N_0 when the angular spread is 5°. For comparison, we have shown the rate of sparse precoding and combining (SPC) [4] for two cases of feedback bits, 8 and 10 bits. The beamforming vector therein is in the form of an array response vector and only the elevation and azimuth angles need to be fed back to the transmitter. The phase shifters in SPC are also quantized using three bits. We see that the performance of the quantized beamformer is comparable to SPC with 8 feedback bits.

Example 4: Let us consider a MIMO-OFDM system over the frequency selective channel model in (12), in which each cluster has 10 rays, angular spread 8° and mean delay uniformly distributed over $[0, N_{cp} T_s]$. The pulse shaping function $p(t)$ is the raised-cosine filter with roll off factor 0.5. The MIMO-OFDM system has DFT size $M = 128$ and CP length $= 16$. UPA antenna arrays with $(N_t, N_r) = (64, 16)$ are used. Fig. 5 (a) shows the transmission rate for different number of clusters for $P_t/N_0 = 0$ dB. The receiver is not constrained to be of a hybrid structure, i.e., fully digital. As the number of clusters

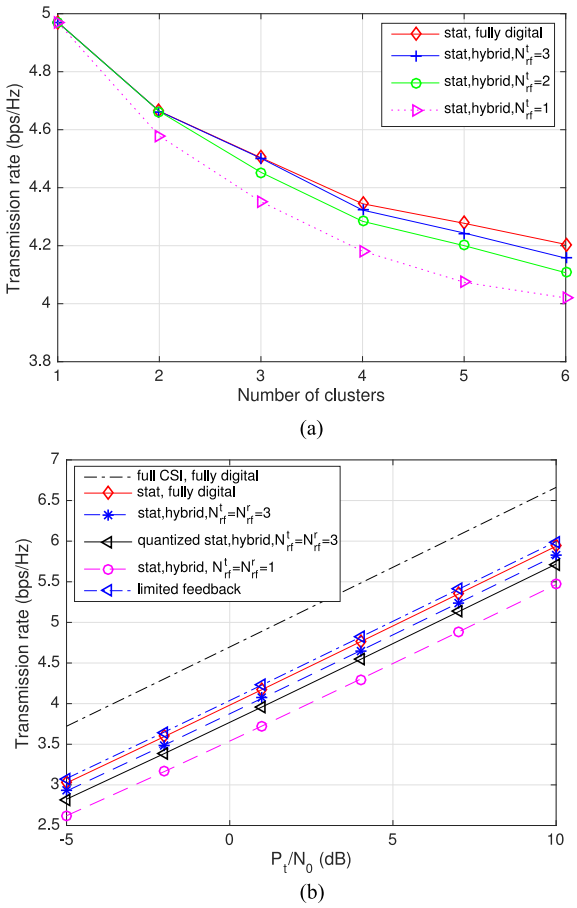

Fig. 5. Performance of statistical MIMO-OFDM beamforming. (a) Transmission rate vs. number of clusters for $P_t/N_0 = 0$ dB. (b) Transmission rate vs. SNR for $N_{cl} = 3$ and uniform distribution of $\sigma^2_{\alpha_\ell}$ over $[0, 1]$.

increases, there is more spatial diversity. But due to the normalization factor γ in (12), which reduces the SNR by a factor of N_{cl}, the gain due to more clusters is offset and the rate shows a slight decrease with N_{cl}. We see that when the number of RF chains at the transmitter is the same as the number of clusters in the channel, there is no rate loss in using hybrid beamforming. When the number of clusters is more than three, the use of three RF chains incurs little loss. This can be understood as follows. When there is a large number of clusters, it is likely that one or more of them are more important than the others. Choosing the dominant eigenvectors corresponding these clusters to form the RF precoder yields a good representation of the subcarrier beamformers and thus a minor degradation.

Fig. 5(b) shows the transmission rate when the channel has three clusters and $\sigma^2_{\alpha_\ell}$ is uniformly distributed over $[0, 1]$. The combiners at the receiver are also of a hybrid structure. With three RF chains at both the transmitter and receiver, we have a small gap between the hybrid system and the one that uses fully digital statistical beamformers at the transmitter and fully digital combiners at the receiver. For comparison, we have shown the rate of the hybrid MIMO-OFDM system [19] with limited feedback, 10 feedback bits for the RF precoder and 6 feedback bits for each baseband subcarrier beamformer. Also shown in Fig. 5(b) is the optimal fully digital MIMO-OFDM system that has full CSI at the transmitter. When there is only one RF chain at the transmitter and receiver, the hybrid statistical beamformer is around 1.5 bits away from the optimal unconstrained system. The difference narrows to around 1 bit when there are three RF chains at both the transmitter and receiver. For the three-RF-chain case, we have shown the rate when the RF precoder and baseband subcarrier beamformers are obtained from quantized statistics. The mean AoDs are each quantized using 5 bits while angular spread, $\sigma^2_{\alpha_\ell}$ and mean delays are each quantized using 4 bits. We can see that the quantization incurs a minor degradation.

VII. Conclusion

In this paper we consider statistical beamforming for mmWave systems. With the assumption of small angular spread, we show that the transmit covariance matrix can be approximated in a closed-form for narrowband mmWave channels. The results obtained for the narrowband case also shed light on the the design of hybrid beamformers for wideband MIMO-OFDM. Spectral analysis suggests that each subcarrier beamformer can be approximated as a linear combination of the statistical beamformers corresponding to dominant clusters. As a result subcarrier beamformers can be readily implemented in a hybrid structure. For either narrowband and wideband systems, the degradation in transmission rate due to statistical designs is a small one, as demonstrated through simulations. Theoretical characterization of the degradation would be an interesting topic worth further investigation.

Appendix A
Proof of Lemma 1

Using the property that the gain $\{\alpha_{\ell,i}\}$ are independent, we can express $\mathbf{\Sigma}_t$ as

$$\mathbf{\Sigma}_t = \sum_{\ell=1}^{N_{cl}} E_{\mathbf{H}} \left[\mathbf{A}_{t,\ell} \mathbf{D}_\ell^\dagger \mathbf{A}_{r,\ell}^\dagger \mathbf{A}_{r,\ell} \mathbf{D}_\ell \mathbf{A}_{t,\ell}^\dagger \right],$$

where the subscript of the expectation means the average is performed over $\mathbf{H}$. As the AoD, AoA and the complex gains $\{\alpha_{\ell,i}\}$ are independent we get $\mathbf{\Sigma}_t = \sum_{\ell=1}^{N_{cl}} E_{\{\phi^t_{\ell,i},\theta^t_{\ell,i}\}}[\mathbf{A}_{t,\ell}\mathbf{E}_\ell\mathbf{A}_{t,\ell}^\dagger]$, where the $N_\ell \times N_\ell$ matrix $\mathbf{E}_\ell$ is given by $E_{\{\alpha_\ell\}}[\mathbf{D}_\ell^\dagger E_{\{\phi^r_{\ell,i},\theta^r_{\ell,i}\}}[\mathbf{A}_{r,\ell}^\dagger\mathbf{A}_{r,\ell}]\mathbf{D}_\ell]$. Notice that $\mathbf{E}_\ell$ is a diagonal matrix as $\{\alpha_{\ell,i}\}$ are independent and of zero mean. The diagonal elements of the matrix $E_{\{\phi^r_{\ell,i},\theta^r_{\ell,i}\}}[\mathbf{A}_{r,\ell}^\dagger\mathbf{A}_{r,\ell}]$ are all equal to one because $\mathbf{a}_r(\phi^r_{\ell,i}, \theta^r_{\ell,i})$ is of unit norm. Thus the diagonal elements of $\mathbf{E}_\ell$ is equal to the variance of $\alpha_{\ell,i}$, i.e, $\sigma^2_{\alpha_\ell}$. Using the definition of the antenna response vector, we can obtain $\mathbf{\Sigma}_t$ as in Lemma 1.

Appendix B
Proof of Theorem 1

With the definition of $\Delta\theta^t_{\ell,i} = \theta^t_{\ell,i} - \overline{\theta}^t_\ell$, we can write the term $\cos\theta^t_{\ell,i}$ as $\cos(\overline{\theta}^t_\ell)\cos(\Delta\theta^t_{\ell,i}) - \sin(\overline{\theta}^t_\ell)\sin(\Delta\theta^t_{\ell,i})$. When the angular spread is small, $\Delta\theta^t_{\ell,i}$ is small and we have

$\cos(\Delta\theta^t_{\ell,i}) \approx 1$ and $\sin(\Delta\theta^t_{\ell,i}) \approx \Delta\theta^t_{\ell,i}$. Using a similar approximation for $\sin\theta^t_{\ell,i}$, we have $\sin\theta^t_{\ell,i}\sin\phi^t_{\ell,i} \approx \sin(\overline{\theta}^t_\ell)\sin(\overline{\phi}^t_\ell) + \cos(\overline{\theta}^t_\ell)\sin(\overline{\phi}^t_\ell)\Delta\theta^t_{\ell,i} + \sin(\overline{\theta}^t_\ell)\cos(\overline{\phi}^t_\ell)\Delta\phi^t_{\ell,i}$, where we have ignored the second order term of $\Delta\theta^t_{\ell,i}\Delta\phi^t_{\ell,i}$. Then $\mu_\ell(m,n)$ can be further expressed as

$$\begin{aligned}\mu_\ell(m,n) \approx\ & e^{j\xi(m\cos\overline{\theta}^t_\ell + n\sin\overline{\theta}^t_\ell\sin\overline{\phi}^t_\ell)} \\ & \times \nu_{\Delta\theta^t_{\ell,1}}\left(-m\xi\sin\overline{\theta}^t_\ell + n\xi\cos\overline{\theta}^t_\ell\sin\overline{\phi}^t_\ell\right) \\ & \times \nu_{\Delta\phi^t_{\ell,1}}\left(n\xi\sin\overline{\theta}^t_\ell\cos\overline{\phi}^t_\ell\right) \qquad (19)\end{aligned}$$

The result in Theorem 1 follows.

Appendix C
Proof of Theorem 2

The SNR loss ρ can be expressed as $\rho = \frac{1}{N_t}E[|F(e^{j\xi\cos(\theta^t)}|^2]$, where $F(e^{j\omega})$ is the Fourier transform of the beamformer $\mathbf{f}$. Let $\Delta = \theta^t - \overline{\theta}^t$. When $\sigma^2_{\theta^t}$ is small, we can approximate $\cos(\theta^t)$ as $\cos(\overline{\theta}^t) - \sin(\overline{\theta}^t)\Delta$, as in the proof of Theorem 1. It follows that $\rho \approx \frac{1}{N_t}E[|F(e^{j\xi(\cos(\overline{\theta}^t)-\sin(\overline{\theta}^t)\Delta)})|^2]$. Thus $\rho \approx \frac{1}{N_t}\int_{-\pi}^{\pi}|F(e^{j\xi(\cos(\overline{\theta}^t)-\sin(\overline{\theta}^t)\Delta)})|^2 P(\Delta)d\Delta$, where $P(\Delta)$ is the probability density function of Δ. For a lower bound of ρ, let us consider the case that the beamformer is the antenna response vector $[\mathbf{f}]_n = \frac{1}{\sqrt{N_t}}e^{j\xi\cos(\overline{\theta}^t)n}$, for $n = 0, 1, \ldots, N_t - 1$. When $\sin\overline{\theta}^t \approx 0$, we see $\rho \approx 1$. Otherwise, we have $|F(e^{j\xi(\cos(\overline{\theta}^t)-\sin(\overline{\theta}^t)\Delta)})|^2 = \frac{1}{N_t^2}|\sin(\pi N_t \sin(\overline{\theta}^t)\Delta/2)/\sin(\pi\sin(\overline{\theta}^t)\Delta/2)|^2$, which can be further approximated as the ratio $|\sin(\pi N_t\sin(\overline{\theta}^t)\Delta/2)/(\pi N_t\sin(\overline{\theta}^t)\Delta/2)|^2$ for small Δ. Notice that the ratio is in the form of the squared sinc function; it is small when $\Delta > \pi/\beta$, where $\beta = \pi N_t\sin(\overline{\theta}^t)$. It follows that $\rho \gtrsim 2\int_{-\pi/\beta}^{\pi/\beta}(1-\cos(\beta\Delta))/(\beta\Delta)^2 P(\Delta)d\Delta$. Approximating $1-\cos(\beta\Delta)$ as $(\beta\Delta)^2/2 - (\beta\Delta)^4/4!$, and plugging in $P(\Delta)$ for Laplacian and Gaussian distributions, we get the bounds in (9).

Appendix D
Proof of (18)

The constraint can be written as $\mathbf{f}_{bb}(k)^\dagger\mathbf{Q}\mathbf{f}_{bb}(k) = 1$, where $\mathbf{Q} = \mathbf{F}^\dagger_{rf}\mathbf{F}_{rf}$. Let $\mathbf{b}(k) = \mathbf{Q}^{1/2}\mathbf{f}_{bb}(k)$, then the constraint is simply $||\mathbf{b}_k||^2 = 1$. As $\mathbf{f}_{bb}(k) = \mathbf{Q}^{-1/2}\mathbf{b}_k$, the objective can be rewritten as $E||\mathbf{H}(k)\mathbf{F}_{rf}\mathbf{Q}^{-1/2}\mathbf{b}(k)||$. The optimal solution of $\mathbf{b}(k)$ is the eigenvector of $(\mathbf{F}_{rf}\mathbf{Q}^{-1/2})^\dagger\mathbf{\Sigma}_t(k)\mathbf{F}_{rf}\mathbf{Q}^{-1/2}$. Using $\mathbf{f}_{bb}(k) = \mathbf{Q}^{-1/2}\mathbf{b}(k)$ and applying normalization such that $||\mathbf{f}(k)|| = 1$, we get the results in (18).

References

[1] C. H. Doan, S. Emami, D. A. Sobel, A. M. Niknejad, and R. W. Brodersen, "Design considerations for 60 GHz CMOS radios," *IEEE Commun. Mag.*, vol. 42, no. 12, pp. 132–140, Dec. 2004.

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[3] C. Kim, T. Kim, and J. Y. Seol, "Multibeam transmission diversity with hybrid beamforming for MIMO-OFDM systems," in *Proc. IEEE Globecom Workshops*, 2013, pp. 61–65.

[4] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath Jr, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[5] X. Zhang, A. F. Molisch, and S. Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

[6] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[7] E. Zhang and C. Huang, "On achieving optimal rate of digital precoder by RF-baseband codesign for MIMO systems," in *Proc. IEEE Veh. Tech. Conf.*, 2014.

[8] 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," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.

[9] Y.-P. Lin, "On the quantization of phase shifters for hybrid precoding systems," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2237–2246, May 2017.

[10] N. Jindal, " MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[11] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3417–3428, Jul. 2007.

[12] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[13] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.

[14] D. Hammarwall, M. Bengtsson, and B. Ottersten, "Acquiring partial CSI for spatially selective transmission by instantaneous channel norm feedback," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1188–1204, Mar. 2008.

[15] D. Hammarwall, M. Bengtsson, and B. Ottersten, "Utilizing the spatial information provided by channel norm feedback in SDMA systems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3278–3293, Jul. 2008.

[16] V. Raghavan and V. V. Veeravalli, "Ensemble properties of RVQ-based limited-feedback beamforming codebooks," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8224–8249, Dec. 2013.

[17] J. Choi, V. Raghavan, and D. J. Love, "Limited feedback design for the spatially correlated multiantenna broadcast channel," in *Proc. IEEE Global Commun. Conf.*, 2013, pp. 3481–3486.

[18] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3-D channel model," in *Proc. IEEE Int. Conf. Commun.*, 2014, pp. 5865–5870.

[19] A. Alkhateeb and R. W. Heath Jr, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[20] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[21] V. Raghavan, A. S. Y. Poon, and V. V. Veeravalli, "MIMO systems with arbitrary antenna array architectures: Channel modeling, capacity and low-complexity signaling," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2007, pp. 1219–1223.

[22] V. Raghavan, S. V. Hanly, and V. V. Veeravalli, "Statistical beamforming on the grassmann manifold for the two-user broadcast channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6464–6489, Oct. 2013.

[23] J. Park, S. Park, A. Y. Panah, and R. W. Heath Jr, "Exploiting spatial channel covariance for hybrid precoding in massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3818–3831, Jul. 2017.

[24] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[25] A. Adhikary *et al.*, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, Jun. 2014.

[26] D. Zhu, B. Li, and P. Liang, "A novel hybrid beamforming algorithm with unified analog beamforming by subspace construction based on partial CSI for massive MIMO-OFDM systems," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 594–607, Feb. 2017.
[27] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
[28] S. Payami, M. Shariat, M. Ghoraishi, and M. Dianati, "Effective RF codebook design and channel estimation for millimeter wave communication systems," in *Proc. IEEE Int. Conf. Commun. Workshop*, 2015, pp. 1226–1231.
[29] G. Destino, J. Saloranta, M. Juntti, and S. Nagaraj, "Robust 3-D MIMO-OFDM channel estimation with hybrid analog-digital architecture," in *Proc. Eur. Signal Process. Conf.*, 2016, pp. 1990–1994.
[30] K. Venugopal, A. Alkhateeb, N. Gonzlez Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
[31] Y.-P. Lin and S.-H. Tsai, "Beamforming with no instantaneous feedback for mmWave transmission," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, 2017, pp. 1–5.
[32] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. SAC-5, no. 2, pp. 128–137, Feb. 1987.
[33] V. Erceg *et al.*, *TGn Channel Models*, IEEE Standard 802.11-03/940r4, May 2004.
[34] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on 3D Channel Model for LTE (Release 12), Jun. 2015.
[35] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2002.
[36] V. Raghavan, S. Subramanian, J. Cezanne, and A. Sampath, "Directional beamforming for millimeter-wave MIMO systems," arXiv:1601.02380, 2016.
[37] V. Raghavan, S. Subramanian, J. Cezanne, A. Sampath, O. H. Koymen, and J. Li, "Single-user versus multiuser precoding for millimeter wave MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1387–1401, Jun. 2017.
[38] A. Forenza, D. J. Love, and R. W. Heath, "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924–1934, Jul. 2007.
[39] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA, USA: Kluwer, 1991.
[40] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, pp. 303–353, 1999.

Yuan-Pei Lin (S'93–M'97–SM'03) was born in Taipei, Taiwan, 1970. She received the B.S. degree in control engineering from the National Chiao-Tung University, Hsinchu, Taiwan, in 1992, and the M.S. degree and the Ph.D. degree, both in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 1993 and 1997, respectively.

She joined the Department of Electrical and Control Engineering, National Chiao-Tung University, in 1997. She has coauthored two books, *Signal Processing and Optimization for Transceiver Systems* (Cambridge Univ. Press, 2010), and *Filter Bank Transceivers for OFDM and DMT Systems* (Cambridge Univ. Press, 2010). Her research interests include digital signal processing, multirate filter banks, and signal processing for digital communications. She was a distinguished Lecturer of the IEEE Circuits and Systems Society for 2006–2007. She was an Associate Editor for the IEEE TRANSACTION ON SIGNAL PROCESSING, the IEEE TRANSACTION ON CIRCUITS AND SYSTEMS II, the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTION ON CIRCUITS AND SYSTEMS I, *EURASIP Journal on Applied Signal Processing*, and *Multidimensional Systems and Signal Processing* (Academic). She was a recipient of Ta-You Wu Memorial Award in 2004.