Abstract—In this article, we investigate channel estimation for wideband millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) under hybrid architecture with low-precision analog-to-digital converters (ADCs). To design channel estimation for the hybrid structure, both analog processing components and frequency-selective digital combiners need to be optimized. The proposed channel estimator follows the typical linear-minimum-mean-square-error (LMMSE) structure and applies for an arbitrary channel model. Moreover, for sparsity channels as in mmWave, the proposed estimator performs more efficiently by incorporating orthogonal matching pursuit (OMP) to mitigate quantization noise caused by low-precision ADCs. Consequently, the proposed estimator outperforms conventional ones as demonstrated by computer simulation results.

Index Terms—mmWave, channel estimation, hybrid, ADC.

I. INTRODUCTION

Millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) is an emerging technology for future wireless networks. Typical massive MIMO is equipped with a large number of radio frequency (RF) chains, which are cost- and power-hungry, especially for wideband mmWave systems. Hybrid architecture with limited RF chains recently attracts much attention for mmWave massive MIMO to reduce cost and complexity [1]. However, it imposes additional challenges for channel estimation because fully digital processing is no longer accessible. In [2], a subspace-based channel estimator has been presented for narrowband massive MIMO with the hybrid structure. In [3], the channel sparsity has been further utilized and an orthogonal matching pursuit (OMP)-based least-square (LS) estimator has been proposed. A distributed grid matching pursuit (DGMP) algorithm [4] has been proposed to solve the power leakage in uplink channel estimation for mmWave MIMO. Further in [5], a simultaneous weighted-OMP estimator has been developed.

For wideband mmWave MIMO, high-precision analog-to-digital converters (ADCs) are expensive and power-hungry [6]. In order to alleviate the burden, low-precision ADCs have been introduced. However, channel estimate is deteriorated due to the nonlinear quantization of low-precision ADCs. A linear-minimum-mean-square-error (LMMSE) estimator has been developed in [2] for massive MIMO with 1-bit ADCs.

In this article, we investigate channel estimation for mmWave massive MIMO with the hybrid architecture using low-precision ADCs. Main contributions are summarized as follows. In Section II, we introduce the system model. The channel estimation algorithm is presented in Section III. Simulation results and conclusions are presented in Sections IV and V, respectively.

Notations: $A^H$, $A^T$, and $A^*$ are the conjugate transpose, transpose, and conjugate of $A$, respectively. $\text{vec}(A)$ and $\text{diag}(A)$, respectively, return vectorization and the diagonal matrix containing diagonal elements of $A$. Operator $\otimes$ represents the Kronecker product. $E\{A\}$, $\text{Tr}(A)$, and $[A]_{ij}$ are the expectation, trace, and $(i,j)$th element of $A$, respectively. $\|a\|_p$ is the $l_p$-norm of vector $a$, and $\lceil a \rceil$ is the ceiling function of scalar $a$. $CN(0, 1)$ indicates circularly symmetric complex Gaussian distribution with mean 0 and variance 1.

II. SYSTEM MODEL

We consider a wideband mmWave massive MIMO-OFDM system with low-precision ADCs where hybrid precoder and combiner are used. The transmitter is equipped with $N_t$ antennas driven by a smaller number, $N_{RF}$, of RF chains. The receiver is equipped with $N_r$ antennas and $N_{RF}$ RF chains.

Channel estimation in OFDM is usually performed in the frequency domain. Let $N_c$ and $K$ be the maximal delay tap and the number of subcarriers, respectively. The channel response matrix of the $l$th path can be expressed as [8]

$$H_d = \sum_{d=0}^{N_p} \alpha_l \delta(d - \tau_l) a_R(\theta_{RL}) a_T^H(\theta_{TT}),$$

for $d = 0, 1, \ldots, \frac{N_c}{2} - 1$, where $\alpha_l$ is the channel gain of the $l$th path, $\tau_l$ is the normalized path delay, and $a_R(\theta_{RL}) \in \mathbb{C}^{N_c \times 1}$ and $a_T(\theta_{TT}) \in \mathbb{C}^{N_t \times 1}$ are antenna array response vectors under uniform linear array (ULA) setup at the receiver and the transmitter, respectively.

$$a_R(\theta_{RL}) = \frac{1}{\sqrt{N_t}} [1, e^{-j2\pi \theta_{RL}}, \ldots, e^{-j2\pi (N_t - 1) \theta_{RL}}]^T,$$

where $\theta_{RL} = \frac{\lambda}{2} \cos(\phi_{RL})$ is the directional cosine with carrier wavelength, $\lambda$, antenna spacing, $s$ ($s \geq \frac{\lambda}{2}$), and angle of arrival (AoA), $\phi_{RL}$. $a_T(\theta_{TT})$ can be similarly expressed.
the frequency-domain channel response at the $k$th subcarrier can be represented as
\[
H[k] = \sum_{d=0}^{N_{t}-1} H_{d} e^{-j2\pi \frac{kd}{N_{t}}} = \sum_{l=1}^{N_{r}} \alpha_{l} \left( \sum_{d=0}^{N_{t}-1} \delta(d - \tau_{l}) e^{-j2\pi \frac{kd}{N_{t}}} \right) a_{R}(\theta_{Rl}) a_{T}^{H}(\theta_{Tl}).
\]

In hybrid massive MIMO illustrated in Fig. 1, the main task is to recover $N_{r}N_{t}$ channel coefficients. Let each training be a transmission of $N_{RF}$ orthogonal pilots formed by signals at $N_{RF}$ RF chains. $N_{RF}, N_{RF}$ coefficients can be estimated at the receiver using $N_{RF}$ RF chains with each training. Thus, the multiple of $L \triangleq \left\lceil \frac{N_{r}N_{t}}{N_{RF}} \right\rceil$, trainings are needed.

Let $M$ be the times of channel use within a coherence time. Then $MN_{RF}$ observations are utilized to estimate the $N_{r}N_{t}$-dimensional channel vector. For uncorrelated channels, at least $M = \left\lceil \frac{N_{r}N_{t}}{N_{RF}} \right\rceil$ channel uses are needed. Let $s_{m1}[k], s_{m2}[k] \in \mathbb{C}^{N_{RF} \times 1}$ be pilots at the $k$th subcarrier during the $m_{1}$th and $m_{2}$th $(m_{1}, m_{2} \in \{1, \ldots, M\})$ channel use,
\[
s_{m1}[k]s_{m2}[k] = \begin{cases} \mathbb{C}, & m_{1} \neq m_{2}, \\ 0, & m_{1} = m_{2}. \end{cases}
\]

where $P$ is the pilot power, with $m_{1} = iN_{RF} + j_{1}$ and $m_{2} = iN_{RF} + j_{2}$ with $i \in \{0, 1, \ldots, M_{k} - 1\}$ where $M_{k} = MN_{RF}$. The corresponding received signal at the $k$th subcarrier is
\[
r_{m}[k] = H[k]F_{Am} s_{m}[k] + v_{m}[k],
\]
where $F_{Am} \in \mathbb{C}^{N_{r} \times N_{RF}}$ is the analog precoder and $v_{m}[k] \sim \mathcal{CN}(0, \sigma_{v}^{2}I_{N_{r}})$ is the noise vector. At the receiver, an analog estimator, $W_{Am} \in \mathbb{C}^{N_{t} \times N_{RF}}$, is first implemented. Both $F_{Am}$ and $W_{Am}$ represent the operations of a phase-shifter network connecting the large antenna array to limited RF chains. The phase shifter adjusts only the phase of input signal without changing its amplitude. Thus, each element in $F_{Am}$ and $W_{Am}$ is restricted as a unity-magnitude value. Then after the ADC quantization, the channel coefficients can be estimated via a linear digital estimator, $W_{Dm}[k] \in \mathbb{C}^{N_{r} \times N_{t}}$, incorporating with the former $W_{Am}$. Note that $F_{Am}$ and $W_{Am}$ are performed on the wideband signals in the time domain. Consequently, the same analog processing components apply for all subcarriers, which are thus described as frequency-independent. Digital processing is performed in the frequency domain and thus can be different across subcarriers, which is termed as frequency-selective. The channel estimating problem is
\[
\hat{h}[k] = \arg \min_{F_{Am}, W_{Am}, W_{Dm}[k]} \mathbb{E}\left\{||\hat{h}[k] - \text{vec}(H[k])||_{2}^{2}\right\},
\]
\[
\begin{align*}
\hat{h}[k] &= W_{Dm}[k]y[k], \\
W_{Dm}[k] \triangleq [W_{D1}[k], \ldots, W_{DM}[k]]^{T} \\
y_{m}[k] = Q\left(\frac{H[k]F_{Am} s_{m}[k]}{W_{Am} y_{m}[k]} + W_{Am} v_{m}[k]\right)
\end{align*}
\text{s.t.}
\]

where $Q(\cdot)$ represents the ADC quantization operation.

III. WIDEBAND HYBRID CHANNEL ESTIMATION

The wideband channel estimation problem in (6) is challenging. This section focuses on designing the linear hybrid estimators, including $W_{Am}$, $W_{Dm}[k]$, and $F_{Am}$.

With limited RF chains, we only have access to a much smaller number of observations per estimation than the number of channel coefficients to be estimated. The design of $W_{Dm}[k]$ is frequency-selective while the design of analog ones, $F_{Am}$ and $W_{Am}$, are frequency-independent. Moreover, we have to take into account the ADC quantization $Q$.

A. Channel Estimation Formulation

From (3), it is convenient to project channel coefficients onto the angular domain. We use dictionary matrices consisting of ULA response vectors, whose sizes are chosen as $N_{t}$ and $N_{r}$ which denote the numbers of resolvable angles at transmitter and receiver, respectively. Specifically, let
\[
A_{t} = [a_{T}(\theta_{T,1}), a_{T}(\theta_{T,2}), \ldots, a_{T}(\theta_{T,N_{t}})] \in \mathbb{C}^{N_{t} \times N_{r}}.
\]

be the dictionary matrix consisting of columns $a_{T}(\theta_{T,p}) (p \in \{1, 2, \ldots, N_{t}\})$ with $\theta_{T,p} \in (-0.5, 0.5)$ drawn from a fixed equal interval as $\theta_{T,p} \approx \frac{1}{N_{t}} (p - \frac{N_{t} + 1}{2})$, and let
\[
A_{r} = [a_{R}(\theta_{R,1}), a_{R}(\theta_{R,2}), \ldots, a_{R}(\theta_{R,N_{r}})] \in \mathbb{C}^{N_{r} \times N_{t}},
\]

be the dictionary matrix composed of $a_{R}(\theta_{R,q}) (q \in \{1, 2, \ldots, N_{r}\})$ with $\theta_{R,q} \approx \frac{1}{N_{r}} (q - \frac{N_{r} + 1}{2})$. The equivalent channel matrix after a whole-space projection is defined as
\[
H[k] \triangleq A_{r}^{H} H[k] A_{t},
\]

Note that the channel model for uniform planar array (UPA) [9] shares a similar structure as (9). The only difference lies in the dictionary matrices for UPA contain an extra quantized angle grid on vertical. Thus the following proposal can also apply for UPA with corresponding subtle changes.

After defining $\Psi \triangleq A_{r} \ast A_{t}$ and using matrix vectorization for notational simplicity, we denote $h_{c}[k] \triangleq \text{vec}(H[k])$ as the equivalent channel coefficient vector to be estimated. From (9) and the unitary properties of $A_{t}$ and $A_{r}$, we write
\[
y_{m}[k] = Q\left(\left(\frac{s_{m}[k]F_{Am} \ast W_{Am}^{H}}{W_{Am} y_{m}[k]} + W_{Am} v_{m}[k]\right)\right)
\]

\[
= Q\left(\Phi_{m}[k] \Psi h_{c}[k] + e_{m}[k]\right),
\]

where $\Phi_{m}[k] \triangleq \left(\frac{s_{m}[k]F_{Am}^{T}}{W_{Am}^{H}} + e_{m}[k] \triangleq \mathbb{W}_{Am} v_{m}[k]\right), W_{Am} y_{m}[k]$.

We have to use at least $M$ trainings within a coherence time to estimate a complete $h_{c}[k]$. Stacking the received signal vectors corresponding to the $M$ trainings, we have
\[
y[k] = Q\left(\Phi_{1}[k] \Psi h[k] + e_{1}[k]\right), \ldots, y_{M}[k] = Q\left(\Phi_{M}[k] \Psi h_{c}[k] + e_{M}[k]\right),
\]

where $\Phi_{m}[k]$ and $e_{m}[k]$ are the corresponding stacked vectors of $\Phi_{m}[k]$ and $e_{m}[k]$ $(m \in \{1, 2, \ldots, M\})$, respectively.

The quantization of ADCs is in general non-linear. Thanks to studies [10][11] which applied the Bussgang theorem [12] on modelling non-linear quantization, it showed that the output of the non-linear quantizer with Gaussian input can be expressed in closed form by decomposing it into a desired signal component and an uncorrelated quantization distortion, e.g.

\[
y[k] = (1 - \eta_{b})\Phi_{m}[k] \Psi h_{c}[k] + \hat{e}[k],
\]

\[
\hat{e}[k] = \mathbb{Q}\left(\frac{\Phi_{m}[k] \Psi h_{c}[k] + e_{m}[k]}{\sigma_{e}^{2}I_{N_{t}}}ight).
\]
where \( \eta_b \) is the distortion factor in terms of the number of quantization bits of ADCs, i.e., \( b \), and

\[
\hat{e}[k] = (1 - \eta_b)e[k] + e_q
\]

(13)
is the equivalent noise including both the ADC quantization error and the AWGN. The value of \( \eta_b \) is determined by the quantization precision. In the condition of high-precision optimal non-uniform quantizations, \( \eta_b \) can be approximately determined by a closed-form expression \( [10] \). For a general ADC precision, there is no explicit expression for determining \( \eta_b \). While in \( [13] \), typical values of \( \eta_b \) corresponding to various precisions are exemplified in Table I. Through the digital combiners at the receiver, the estimate in (6) is represented in the virtual-angular domain as

\[
\hat{h}_v[k] = W_H^H[k]\left((1 - \eta_b)\Phi[k]\Psi h_v[k] + \hat{e}[k]\right),
\]

(14)

where \( W_D[k] = [W_{D1}[k], \cdots, W_{DM}[k]]^T \). In (14), the channel estimation problem is converted into finding proper \( W_D[k] \) and \( \Phi[k] \), which requires careful evaluation of the randomness of terms in (14) and will be discussed subsequently.

Without priori channel statistics or exploiting channel sparsity, we have to solve the estimation problem in (14) with \( N_vN_c \) coefficients in \( h_v[k] \), or equivalently \( h_v[k] \). When a priori channel sparsity, \( N_c \) \( (N_v \ll N_vN_c) \), is exploited, problem (14) is rewritten to incorporate compressed sensing (CS) techniques, e.g., OMP, for complexity and pilot overhead reduction. With OMP, we use a uniform selective matrix to pick the dominant coefficients for estimation. It yields

\[
h_v^{NZ}[k] = P_v^T[k]h_v[k] \triangleq [e_{\pi(1)}, \cdots, e_{\pi(N_v)}] h_v[k],
\]

(15)

where \( e_{\pi(i)} (\pi(i) \in \{1, 2, \cdots, N_vN_c\}) \) is a selecting vector with the \( \pi(i) \)th element being 1. According to (15), we only need to estimate \( N_vN_c \) non-zero channel coefficients. Potentially, \( M \) can be at most reduced to \[ N_vN_c/N_{RF} \].

The determination of \( P_v[k] \) depends on whether we have priori channel sparsity information. If there is no priori sparsity information, we simply have \( P_v[k] = I_{N_vN_c} \). For sparse channels without knowing \( N_vN_c \), we apply OMP to obtain locations of \( N_vN_c \) dominant channel coefficients. In this case, the selecting matrix, \( P_v[k] \), can be elaborated in Appendix A.

\[
\text{TABLE I}
\]

| \( b \) | 1    | 2    | 3    | 4    | 5    |
|--------|-----|-----|-----|-----|-----|
| \( \eta_b \) | 0.3634 | 0.1175 | 0.03454 | 0.009497 | 0.002399 |

B. Linear Channel Estimator Optimization

The channel estimation problem in (15) and (14) is to design \( W_D[k] \) and \( \Phi[k] \), equivalently \( F_{Am} \) and \( W_{Am} \), to estimate \( h_v^{NZ}[k] \). By substituting (15) into (12), we have

\[
y[k] = (1 - \eta_b)(\Phi[k]\Psi h_v[k] + \hat{e}[k]) = (1 - \eta_b)\Omega[k]h_v^{NZ}[k] + \hat{e}[k].
\]

(16)

Without any priori channel directivity information, it is reasonable to apply isotropic pilot directions, which corresponds to i.i.d Gaussian \( F_{Am} \) and corresponding \( W_{Am} \). Under the hybrid architecture, however, generating i.i.d Gaussian matrix is infeasible due to analog hardware limitation. Alternatively, we choose that \( F_{Am} \) and \( W_{Am} \) have phases drawn uniformly from \([0, 2\pi)\) via phase shifters with unimodular constraints. In practice, we can generate fixed analog processing matrices corresponding to the uniform distribution. These fixed matrices form a codebook, in which each matrix is a codeword. The codebook can be predetermined and shared by both sides.

Our goal remains to optimize \( W_D[k] \in \mathbb{C}^{M_{RF} \times N_c} \) \((N_v \leq N_c)\) to estimate \( h_v^{NZ}[k] \) from \( y[k] \) in (16), i.e.,

\[
\hat{h}_v^{NZ}[k] = (1 - \eta_b)W_D^H[k]\Omega[k]h_v^{NZ}[k] + W_D^H[k]\hat{e}[k].
\]

(17)

From (17) and Appendix B, the optimal digital estimator in terms of MMSE is derived as

\[
W_D^* = \frac{\Omega[k]}{1 - \eta_b}
\]

\[
[\Omega[k]^H[\Omega[k] + \sigma^2 h^{-1} N_c\right]^{-1},
\]

(18)

where \( \sigma^2_h \) is the variance of each element of the effective noise vector, \( \hat{e}[k] \), and \( \sigma^2_h \) is the large-scale fading factor of \( h_v[k] \).

To this end, we have obtained the channel estimate as

\[
\hat{h}[k] = (A_1 \otimes A_r)(P_v^T[k])^{-1}(W_D^*[k])^HY[k],
\]

(19)

where \( (P_v^T[k])^{-1} \) is the inverse operation of \( P_v^T[k] \).

For completeness, if correlations across the frequency domain are considered, pilots can be inserted every few subcarriers. The minimum required length for channel training is \( \lfloor K\Delta f/B_c \rfloor \) in the frequency domain, where \( B_c \) and \( \Delta f \) are the coherence bandwidth and subcarrier spacing, respectively.

IV. Simulation Results

The performance is shown in Fig. 2 and Fig. 3 using the model in (12), where the normalized mean square error (NMSE) is defined as \( \text{NMSE} = \mathbb{E}\left\{\|h_v^{NZ}[k] - h_v^{NZ}[k]\|^2_2/\|h_v^{NZ}[k]\|^2_2\right\} \).
MSE = \mathbb{E} \left\{ \left\| \mathbf{h}^{NZ}[k] - \mathbf{h}_\nu^{NZ}[k] \right\|_2^2 \right\} = \mathbb{E} \left\{ \left\| (1 - \eta_b) \mathbf{W}_D[k] \mathbf{H}[k] - I_{N_t} \right\| \mathbf{h}_\nu^{NZ}[k] + \mathbf{W}_D[H][k] \mathbf{e}[k] \right\|_2^2 \right\}
= (1 - \eta_b)^2 \sigma_h^2 \left( 1 - \eta_b \right) \text{Tr} \left( \mathbf{W}_D[k] \mathbf{H}[k] \mathbf{W}_D[k] \right) - \text{Tr} \left( \mathbf{W}_D[k] \mathbf{H}[k] + \mathbf{H}[k] \mathbf{W}_D[k] \right) + \sigma_y^2 N_r + \sigma_e^2 \text{Tr} \left( \mathbf{W}_D[k] \mathbf{W}_D[k] \right) \tag{27}

\frac{\partial \text{MSE}}{\partial \mathbf{W}_D[k]} = (1 - \eta_b)^2 \sigma_h^2 \mathbf{W}_D[k] \mathbf{H}[k] \mathbf{W}_D[k] - (1 - \eta_b) \sigma_y^2 \mathbf{W}_D[k] + \sigma_e^2 \mathbf{W}_D[k] \tag{28}

V. CONCLUSION

We have proposed a general estimator under arbitrary channel statistics for wideband mmWave MIMO with hybrid architecture and low-precision ADCs.

REFERENCES

[1] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," IEEE Wireless Commun. Lett., vol. 3, no. 6, pp. 653–656, Oct. 2014.
[2] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 528–542, Apr. 2015.
[3] K. Venugopal et al., "Channel estimation for hybrid architecture-based wideband millimeter wave systems," IEEE J. Sel. Areas Commun., vol. 35, no. 9, pp. 1996–2009, Sept. 2017.
[4] Z. Gao et al., "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," IEEE Commun. Lett., vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
[5] J. P. G.-Coma et al., "Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems," IEEE J. Sel. Topics Signal Process., vol. 12, no. 2, pp. 353–367, May 2018.
[6] R. Walden, "Analytic-to-digital converter survey and analysis," IEEE J. Sel. Areas Commun., vol. 17, no. 4, pp. 539–550, Apr. 1999.
[7] Y. Li et al., "Channel estimation and performance analysis of one-bit massive MIMO systems," IEEE Trans. Signal Process., vol. 65, no. 15, pp. 4075–4087, Aug. 2017.
[8] A. Alkhateeb et al., "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.
[9] P. Zhao et al., "Hybrid precoding with two-layer phase shifter feeding network for mmWave FD-MIMO systems," in Proc. IEEE GLOBECOM, Singapore, Singapore, Dec. 2017, pp. 1–6.
[10] A. Mezghani and J. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in Proc. IEEE ISIT, Cambridge, MA, USA, July 2012.
[11] J. Xu, W. Xu, and F. Gong, "On performance of quantized transceiver in multiuser massive MIMO downlinks," IEEE Wireless Commun. Lett., vol. 6, no. 5, pp. 562–565, Oct. 2017.
[12] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Res. Lab. Electron., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 216, Mar. 1952.
[13] J. Max, "Quantizing for minimum distortion," IEEE Trans. Inf. Theory, vol. 6, no. 1, pp. 7–12, Mar. 1960.
[14] J. Mo, P. Schniter, and R. W. Heath, “Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs,” *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.