Performance Analysis of Cell-Free mmWave Massive MIMO with Low-Resolution DAC Quantization

Seung-Eun Hong
Mobile Communications Research Division
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, South Korea
iptvguru@etri.re.kr

***Abstract*—** This paper focuses on the downlink of a cell-free massive multiple-input multiple-output (MIMO) system in which each of the access nodes (ANs) is equipped with low-resolution digital-to-analog converters (DACs). In particular, the system is assumed to operate over mmWave frequency with practical considerations such as pilot contamination and hybrid precoding scheme. By mimicking the effect of few-bit DACs with the additive quantization noise model (AQNM), a tight approximate rate expression is derived and analyzed with some simulation results, which provide deep understanding of the impacts from the few-bit quantization as well as channel estimation error on the performance of cell-free mmWave massive MIMO.

Keywords—Cell-Free massive MIMO, mmWave, Hybrid precoding, Low-Resolution DAC, Additive quantization noise model

I. INTRODUCTION

Owing to the capabilities of providing uniformly higher data rates to everyone in anywhere, cell-free massive MIMO (CF-M^3) is considered as one of the key candidate technologies in future 5G+ and 6G wireless communication systems [1], [2]. More specifically, the CF-M^3 system has an architectural feature wherein, under the coordinating control of a central processing unit (CPU), a very large number of distributed ANs simultaneously and jointly serve a much smaller number of user equipments (UEs) across the cells at a geographical scale.

More recently, some synergies between mmWave and CF-M^3 were studied in [3], [4], where the combined technique so called Cell-Free MmWave Massive MIMO (CF-M^4) is shown that it can fully utilize three-dimensions, such as spectrum, spectral efficiency, and network densification in wireless capacity and it may become to be one of the future-proof wireless transmission techniques to accommodate explosively ever-increasing mobile traffic. The increase in the number of RF chains and ANs, however, significantly increases the network power consumption in the CF-M^4 where each of ANs is equipped with multiple RF chains. Therefore, by lowering the resolution of the DAC, the network power consumption in the CF-M^4 can be reduced, and the deployment cost can also be reduced due to the low cost of the low-resolution DAC. The authors in [5] derive a tight approximate downlink rate expression with the AQNM in closed-form to capture the quantization noise due to few-bit ADCs/DACs, under the condition that all ANs performs full-digitally simple conjugate precoding with their local channel state information (CSI). Only very recently, work in [6] considers a CF-M^4 system composed of ANs equipped with low-resolution DACs, and jointly analyzes the performance impact along with the compression quantization effect according to the fronthaul capacity constraint. However, the system of [6] is impractical because it assumes an ideal channel without pilot contamination and channel estimation error. In addition, although zero-forcing digital precoding is considered, the corresponding RF beamforming design method is not clearly presented as a hybrid precoding scheme for the cell-free mmWave operation.

Inspired by the related works, this paper presents the downlink performance on the achievable rate lower bounds of CF-M^4 system, where a mmWave hybrid precoding is performed at all APs equipped with low-resolution DAC quantization. To the aim, we adopt the combined zero-forcing baseband precoding and a simple RF beamforming as the hybrid precoding scheme, and leverage the AQNM model to cover some imperfect quantization error due to the few-bit DAC. Additionally, the negative effects of the pilot contamination and channel estimation error are also considered altogether.

II. SYSTEM MODEL FOR CF-M^4 WITH LOW-RESOLUTION DACS

This paper assumes the same CF-M^4 system model as that in [4], excepting that each AN is equipped with few-bit DACs. More specifically, M ANs are ultra-densely distributed in a geographical area and all of them jointly serve a K single-antenna UEs. Each AN is assumed to have L RF chains which are connected with an array of N antennas in a way of the fully-connected architecture where a pair of B-bit DACs are used to generate the in-phase and quadrature (I/Q) components of the downlink baseband signal in each RF chain at ANs. For simplicity, it is assumed that every AN uses the same B bits.

A. Uplink Training

According to time-division duplexing (TDD) mode, all UEs need to simultaneously send their pilot sequences to all ANs, each of which can receive them through its multiple RF chains estimate and estimate the uplink effective channels $\boldsymbol{g}_{mk} = (\boldsymbol{W}_m^{RF})^T \boldsymbol{h}_{mk}$ where $\boldsymbol{h}_{mk} \in \mathbb{C}^{N\times 1}$ is the channel vector from user k to AN m and $\boldsymbol{W}_m^{RF} \in \mathbb{C}^{N\times L}$ is the RF beamforming matrix of AN m. In this paper, the SLNRmax-SA method is used for the design of $\mathbf{W}_m^{RF}$ with the guidelines in [4, Section 2.3]. More specifically, it is assumed that every AN is able to acquire its

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-01659,5G Open Intelligence-Defined RAN (ID-RAN) Technique based on 5G New Radio).

spatial channel covariance matrix $\boldsymbol{R}_{mk} = \mathbb{E}\{\boldsymbol{h}_{mk}\boldsymbol{h}_{mk}^{\mathrm{H}}\}, \forall k, m$ through some large-scale channel estimation methods [3] and it calculates the simple average given by $\overline{\boldsymbol{R}}_m = \sum_{k=1}^{K} \boldsymbol{R}_{mk} / K$ in each the pre-defined multiple-times of the coherence interval. Then, $\boldsymbol{W}_m^{RF}$ is designed as follows:

$$\boldsymbol{W}_m^{RF} = \left[\exp\left(j\measuredangle \boldsymbol{w}_{m,1}^{RF}\right) \cdots \exp\left(j\measuredangle \boldsymbol{w}_{m,L}^{RF}\right)\right] \quad (1)$$

where $\boldsymbol{w}_{m,l}^{RF}$ is the eigenvector associated with l-th largest eigenvalue of $\overline{\boldsymbol{R}}_m$ and $\measuredangle \boldsymbol{w}$ represents the phase angles (in unit of radian) for each element of the eigenvector $\boldsymbol{w}$. In the sequel, AN $m, m = 1, \dots, M$ calculates the estimates $\widehat{\boldsymbol{g}}_{mk}, \forall k$ by using the minimum mean square error (MMSE) channel estimation method [7].

Let $\boldsymbol{\varphi}_k \in \mathbb{C}^{\tau_p \times 1}$ represent the pilot sequence assigned to UE k, where τ_p depicts the length of the pilot sequence. Then, the received pilot sequence at the AN m can be shown as

$$\boldsymbol{y}_m^{\tau_p} = \sum_{k=1}^{K} \sqrt{\rho_{\tau_p}} (\boldsymbol{W}_m^{RF})^T \boldsymbol{h}_{mk} \, \boldsymbol{\varphi}_k^T + \boldsymbol{n}_m^{\tau_p} \quad (2)$$

where ρ_{τ_p} is the total transmit power of the pilot sequence and $\boldsymbol{n}_m^{\tau_p} \in \mathbb{C}^{L \times \tau_p}$ models the AWGN matrix. All entries in $\boldsymbol{n}_m^{\tau_p}$ are independent and identically distributed (i.i.d) elements that follow the distribution $\mathcal{CN}(0, \sigma^2)$. It is generally assumed to be $\tau_p < K$ so that the allocated pilot sequences are not fully orthogonalized among UEs, which can cause some pilot contamination between co-pilot UEs. Based on the MMSE principle, the MMSE channel estimate $\widehat{\boldsymbol{g}}_{mk}$ is given by

$$\widehat{\boldsymbol{g}}_{mk} = \sqrt{\rho_{\tau_p}} \boldsymbol{R}_{\boldsymbol{g}_{mk}} \boldsymbol{\Phi}_{mk}^{-1} \boldsymbol{y}_m^{\tau_p} \boldsymbol{\varphi}_k^* \quad (3)$$

with two definitions of $\boldsymbol{R}_{\boldsymbol{g}_{mk}} = (\mathbf{W}_m^{RF})^T \boldsymbol{R}_{mk} (\mathbf{W}_m^{RF})^*$ and $\boldsymbol{\Phi}_{mk} = \rho_{\tau_p} \sum_{k=1}^{K} \boldsymbol{R}_{\boldsymbol{g}_{mk}} |\boldsymbol{\varphi}_k^T \boldsymbol{\varphi}_k^*|^2 + \sigma^2 \mathbf{I}_L$. The statistics of the MMSE channel estimate are given by

$$\widehat{\boldsymbol{g}}_{mk} \sim \mathcal{CN}\left(0, \rho_{\tau_p} \boldsymbol{R}_{\boldsymbol{g}_{mk}} \boldsymbol{\Phi}_{mk}^{-1} \boldsymbol{R}_{\boldsymbol{g}_{mk}}^H\right) \quad (4)$$

Due to the well-known uncorrelated relationships between $\boldsymbol{g}_{mk}$ and $\widehat{\boldsymbol{g}}_{mk}$, the channel estimation error can be quantified as $\widetilde{\boldsymbol{g}}_{mk} = \boldsymbol{g}_{mk} - \widehat{\boldsymbol{g}}_{mk}$.

B. Downlink Data Transmission

In this subsection, the signal model that sheds light on the downlink data transmission with few-bit DACs in CF-M[4] system is addressed. Let $\boldsymbol{s} = [s_1, s_2, \dots, s_K]^T \in \mathbb{C}^{K \times 1}$ depict the signal vector for the UEs that follows $\mathbb{E}\{\boldsymbol{s}\boldsymbol{s}^H\} = \boldsymbol{I}_K$. Let us also define the effective MIMO channel matrix between the UEs and ANs as $\boldsymbol{G} = [\boldsymbol{G}_1^T \cdots \boldsymbol{G}_M^T]^T$, with $\boldsymbol{G}_m = [\boldsymbol{g}_{m1}, \cdots, \boldsymbol{g}_{mK}] \in \mathbb{C}^{L \times K}$. By exploiting the instantaneously estimated $\widehat{\boldsymbol{g}}_{mk}$ (short-term CSI) and their long-term statistics in the uplink training phase, every AN digitally precodes the weighted (with power control diagonal matrix $\boldsymbol{P} = \mathrm{diag}([p_1, \cdots, p_K]) \in \mathbb{R}^{K \times K}$, $p_k \geq 0$) data signals intended for all UEs in its base-band module, of which the output depicted by $\boldsymbol{x}_m^{uq}$ in AN m is inputted into low-resolution DACs. For avoiding inter-user interferences in cell-free operation, zero-forcing (ZF) digital precoding [8] is assumed to be used in this paper.

TABLE I. THE VALUES OF α_B FOR B-BIT QUANTIZATION

B	1	2	3	4	5	6-15
α_B	0.6366	0.8825	0.96546	0.990503	0.997501	$1-\frac{\pi\sqrt{3}}{2}2^{-2B}$

Then, using these notations and assumptions, the unquantized ZF-precoded signal $\boldsymbol{x}_m^{uq}$ is represented by

$$\boldsymbol{x}_m^{uq} = \boldsymbol{W}_m^{BB} \boldsymbol{P}^{\frac{1}{2}} \boldsymbol{s} = \widehat{\boldsymbol{G}}_m^* \left(\widehat{\boldsymbol{G}}^T \widehat{\boldsymbol{G}}^*\right)^{-1} \boldsymbol{P}^{\frac{1}{2}} \boldsymbol{s} \quad (5)$$

where $\boldsymbol{W}_m^{BB} = \widehat{\boldsymbol{G}}_m^* \left(\widehat{\boldsymbol{G}}^T \widehat{\boldsymbol{G}}^*\right)^{-1}$ is the ZF linear precoder for mth AN, $\widehat{\boldsymbol{G}}_m = \boldsymbol{G}_m - \widetilde{\boldsymbol{G}}_m$, and $\widehat{\boldsymbol{G}} = \boldsymbol{G} - \widetilde{\boldsymbol{G}}$.

For modeling the output signal $\boldsymbol{x}_m^q$ of few-bits DACs with the input $\boldsymbol{x}_m^{uq}$, the DACs installed in all ANs are assumed to have the same resolution (B-bit), without loss of generality, and the linear AQNM is introduced. Note that the AQNM can well approximate the nonlinear distortion of low-resolution DACs as a linear gain with the additive quantization noise [9], which has been extensively used in the quantized MIMO systems [10]. Consequently, the quantized signal $\boldsymbol{x}_m^q$ can be expressed as

$$\boldsymbol{x}_m^q = \mathbb{Q}_B\left(\boldsymbol{x}_m^{uq}\right) \approx \alpha_B \boldsymbol{x}_m^{uq} + \boldsymbol{q}_B \quad (6)$$

where $\mathbb{Q}_B(\cdot)$ denotes the scalar quantization function mapping the input $\boldsymbol{x}_m^{uq}$ into the output $\boldsymbol{x}_m^q$ according to AQNM and $\alpha_B (= 1 - \beta_B)$ is a linear gain with β_B denoting the distortion parameter of the B-bit DACs. Specifically, β_B is defined to be a ratio of the quantizer error variance to quantizer input variance [9] and the exactly approximated values of α_B are listed in Table I. In addition, $\boldsymbol{q}_B$ denotes the additive Gaussian quantization noise vector following $\mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{R}_{\boldsymbol{q}_B}\right)$, where the covariance matrix $\boldsymbol{R}_{\boldsymbol{q}_B}$ is represented by

$$\boldsymbol{R}_{\boldsymbol{q}_B} = \alpha_B (1 - \alpha_B) \mathrm{diag}\left(\sum_{k=1}^{K} \mathbf{w}_{mk}^{BB} (\mathbf{w}_{mk}^{BB})^H p_k\right) \quad (7)$$

Letting $\boldsymbol{x}_m$ depict the transmitted signal at mth AN, $\boldsymbol{x}_m = \boldsymbol{W}_m^{RF} \boldsymbol{x}_m^q$ and the signal vector received at UEs $\boldsymbol{y} \in \mathbb{C}^{K \times 1}$ can be represented as

$$\begin{aligned} \boldsymbol{y} &= \sum_{m=1}^{M} \boldsymbol{H}_m^T \boldsymbol{x}_m + \boldsymbol{n} \\ &= \sum_{m=1}^{M} \left\{\alpha_B \boldsymbol{H}_m^T \boldsymbol{W}_m^{RF} \boldsymbol{W}_m^{BB} \boldsymbol{P}^{\frac{1}{2}} \boldsymbol{s} + \boldsymbol{H}_m^T \boldsymbol{W}_m^{RF} \boldsymbol{q}_B\right\} + \boldsymbol{n} \\ &= \sum_{m=1}^{M} \left\{\alpha_B \boldsymbol{G}_m^T \widehat{\boldsymbol{G}}_m^* \left(\widehat{\boldsymbol{G}}^T \widehat{\boldsymbol{G}}^*\right)^{-1} \boldsymbol{P}^{\frac{1}{2}} \boldsymbol{s} + \boldsymbol{G}_m^T \boldsymbol{q}_B\right\} + \boldsymbol{n} \quad (8) \end{aligned}$$

where $\boldsymbol{n}$ is AWGN vector. From (8), the signal received at UE k can be re-casted as

$$\begin{aligned} y_k = & \underbrace{\alpha_B \sqrt{p_k} s_k}_{\text{desired signal (DS}_k)} + \underbrace{\alpha_B \widetilde{\boldsymbol{g}}_k^T \widehat{\mathbf{G}}^* \left(\widehat{\mathbf{G}}^T \widehat{\mathbf{G}}^*\right)^{-1} \mathbf{P}^{\frac{1}{2}} \mathbf{s}}_{\text{channel estimation error (CEE}_k)} \\ & + \underbrace{\sum_{m=1}^{M} \boldsymbol{g}_{mk}^T \boldsymbol{q}_{\mathrm{B},dm}}_{\text{quantization noise (QN}_k)} + n_k \quad (9) \end{aligned}$$

where $\widetilde{\boldsymbol{g}}_k \in \mathbb{C}^{ML \times 1}$ is kth column vector in the corresponding matrix $\widetilde{\boldsymbol{G}} = \boldsymbol{G} - \widehat{\boldsymbol{G}}$.

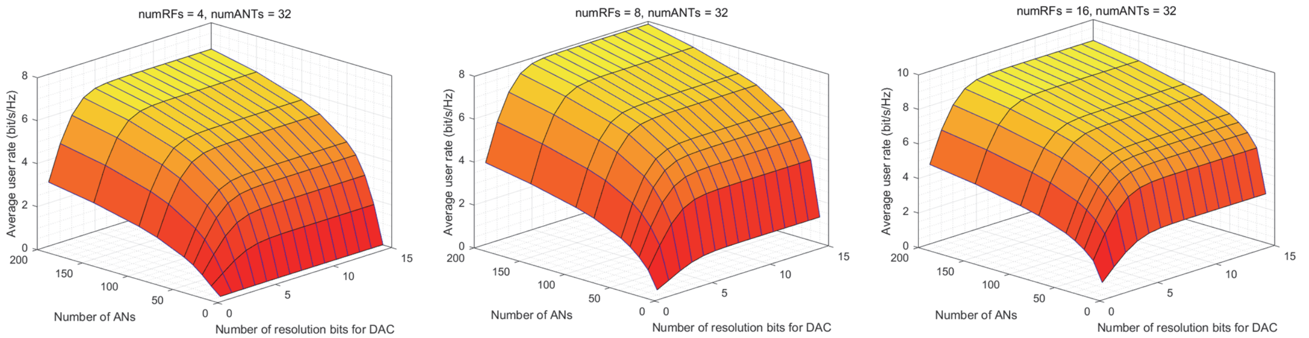

Fig. 1. Average user rate of CF-M[4] versus number of ANs and resolution bits for DAC under a fixed number of antennas N=32; (left) number of RF chains L=4, (middle) number of RF chains L= 8, (right) number of RF chains L=16.

III. Performance Analysis for Downlink Data Rate

In this section, a tight approximate data rate expression for the low-resolution DAC quantized CF-M[4] system in downlink is derived, taking into consideration both channel estimation error due to pilot contamination and quantization noise captured with AQNM. In (9), the sum of the CEE_k, QN_k, and n_k on the right-hand side (RHS) is so called *interference-plus-quantization-plus-noise* (IQN), which is uncorrelated with DS_k. Therefore, the use-and-then-forget (UtaF) capacity lower bounding techniques in [1], [3-6] can be applied to derive an achievable rate of UE k using the hybrid precoders of $\boldsymbol{W}_m^{RF}$ and $\boldsymbol{W}_m^{BB}$, for all $m \in \{1, \dots, M\}$, as $R_k = \log_2(1 + \mathrm{SIQNR}_k)$, where SIQNR_k is the signal-to-IQN ratio (SIQNR) of UE k and is given by

$$\begin{aligned}\mathrm{SIQNR}_k &= \frac{\mathbb{E}\{|\mathrm{DS}_k|^2\}}{\mathbb{E}\{|\mathrm{CEE}_k|^2\} + \mathbb{E}\{|\mathrm{QN}_k|^2\} + \sigma^2} \\ &= \frac{\alpha_B^2 p_k}{\alpha_B^2 \sum_{i=1}^K \gamma_i p_i + \alpha_B(1-\alpha_B)\Gamma + \sigma^2}\end{aligned} \quad (10)$$

Here, γ_i is the ith element of the vector consisting of the diagonal components in the matrix $\mathbb{E}\{(\mathbf{W}^{BB})^H \mathbb{E}\{\tilde{\boldsymbol{g}}_k^* \tilde{\boldsymbol{g}}_k^T\}\mathbf{W}^{BB}\}$ and $\Gamma = \sum_{m=1}^M \mathbb{E}\{\boldsymbol{g}_{mk}^T \mathrm{diag}(\sum_{k=1}^K \mathbf{w}_{mk}^{BB}(\mathbf{w}_{mk}^{BB})^H p_k)\boldsymbol{g}_{mk}^*\}$. In addition, this paper assumes the per-UE equal power control coefficients given by [4, equations (25) and (26)].

To provide simulation results on the achievable user rate in (10), all ANs and UEs are randomly distributed within a square of size 200×200 m^2, where the number of UEs is fixed to 30 and τ_p is set to 16 for the pilot contamination. As pilot allocation strategies selecting co-pilot users, a balanced random pilot assignment (BRPA) scheme in [3] is used, where a pilot sequence for UEs is selected with cyclic sequential patterns from the set of total pilot sequences. For mmWave channel model, the outdoor channel model in [4] is assumed and a decorrelation distance of 50 m is set in this paper. Unless otherwise specified, the same system parameters in [4] are used.

Fig. 1 plots the average user rate of CF-M4 versus number of ANs and resolution bits for DAC under the condition of AN with the number of RF chains $L = \{4, 8, 16\}$ and the fixed number of antennas $N = 32$. As you can see from the three graph results, the user rate is improved as either the number of RF chains or the number of ANs increases. The interesting point is that, irrespective of the numbers of ANs and RF chains, the user rate of 6-bits DAC resolution approaches that of the full-resolution under a tolerable loss, which implies that low-resolution DAC is a promising solution to the energy-efficient CF-M[4] system.

IV. Conclusion

In this paper, the performance analysis of cell-free mmWave massive MIMO with low-resolution DAC quantization is evaluated in terms of the downlink user rate, with the focus on the influences of various system parameters such as the number of ANs, number of RF chains, and number of resolution bits for DAC. For this goal, a tight approximate rate expression has been derived and analyzed with the Monte-Carlo simulations, which provide deep understanding of the impacts from the few-bit quantization as well as channel estimation error on the mmWave hybrid precoding scheme in cell-free operation. Simulation results show that the user rate increases according to the numbers of ANs and RF chains while it converges at only 6-bits of the DAC quantization resolutions.

References

[1] Özlem Tugfe Demir, Emil Björnson, Luca Sanguinetti, Foundations of user-centric cell-free massive MIMO, now, 2021.

[2] I. F. Akyildiz, A. Kak, S. Nie, "6G and Beyond: The Future of Wireless Communications Systems," IEEE Access, Vol. 8, 2020.

[3] G. Femenias, F. Riera-Palou, "Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity," IEEE Access, Vol. 7, Apr. 2019.

[4] S.-E. Hong, "On the effect of shadowing correlation and pilot assignment on hybrid precoding performance for cell-free mmWave massive MIMO UDN system," ICT Express, Vol. 7, Mar 2021.

[5] Y. Zhang, H. Cao, M. Zhou, X. Qiao, S. Wu, L. Yang, "Cell-Free Massive MIMO with Few-bit ADCs/DACs: AQNM versus Bussgang," In Proc. IEEE VTC2020Spring, May 2020.

[6] I. Kim, J. Choi, "Performance of Cell-Free MmWave Massive MIMO Systems with Fronthaul Compression and DAC Quantization," In Proc. IEEE WCNC2021, Mar. 2021.

[7] P. Pasangi, M. Atashbar, M.M. Feghhi, "Blind Downlink Channel Estimation for TDD-based Multiuser Massive MIMO In the Presence of Nonlinear HPA," ETRI Journal, Mar. 2019.

[8] M. Hu, Y. Chang, T. Zeng, B. Wang, "SINR Loss and User Selection in Massive MU-MISO Systems with ZFBF," ETRI Journal, Aprl. 2019.

[9] O.T. Demir and E. Bjornson, "The Bussgang Decomposition of Nonlinear Systems – Basic theory and MIMO extensions," IEEE Signal Processing Magazine, Jan. 2021.

[10] J. Zhang et la., "Performance Analysis of Mixed-ADC Massive MIMO Systems Over Rician Fading Channels," IEEE JSAC, Jun. 2017