NOMA Systems With Content-Centric Multicast Transmission for C-RAN

Sang-Rim Lee, *Member, IEEE*, Seok-Hwan Park, *Member, IEEE*, and Inkyu Lee, *Fellow, IEEE*

***Abstract*—This letter considers a multicast transmission for the downlink of a cloud radio access network. In this system, each user equipment (UE) requests a file from a library and the requested files are served by a set of remote radio heads (RRHs) that are connected to a baseband processing unit (BBU) through finite-capacity fronthaul links. A conventional approach to deliver the requested files to the UEs is to perform coordinated linear precoding at the BBU and then send quantized versions of the precoded signals to the RRHs. In this letter, we investigate a non-orthogonal multiple access (NOMA) transmission scheme under the assumption that the receiving UEs perform successive interference cancellation decoding. The problem of maximizing the minimum delivery rate of the requested files subject to per-RRH fronthaul capacity and transmit power constraints is tackled. From numerical results, we validate the advantages of the proposed NOMA multicast scheme.**

***Index Terms*—C-RAN, NOMA, content-centric multicasting.**

I. Introduction

IT IS envisioned that in the next generation wireless communication systems, current connection-centric communications will be replaced with content-centric communications owing to the increasing demand for multimedia services which include video streaming and mobile TV [1]. Content delivery from the network to requesting user equipments (UEs) can be interpreted as a generalized multicast scenario which was examined under various network settings such as multi-user multiple-input multiple-output (MIMO) systems [2], [3] and cloud radio access network (C-RAN) [1], [4]. Specifically, in [1] and [4], the optimization of signal processing strategies, such as coordinated multicast beamforming and fronthaul transmission, was studied with criteria of minimizing a compound cost function and maximizing the delivery rate, respectively.

In this letter, we investigate non-orthogonal multiple access (NOMA) transmission for content delivery in the downlink of a C-RAN system. The NOMA technique has been conceived as an efficient means to provide high-connectivity service in 5G wireless networks (e.g., [5]–[9]). The main idea of NOMA is to leverage power-domain multiplexing of independent signals by adopting superposition coding at the transmitter and successive interference cancellation (SIC) decoding at the receiver. Since an initial report in [5] on the potential advantages of NOMA with single antenna nodes, several works have tackled the optimization of the NOMA scheme under different MIMO scenarios such as multi-user [7] and multi-cell downlink systems [6], [8], [9]. However, the effect of adopting the NOMA technique in content-centric C-RAN system has not been well studied in the existing literature.

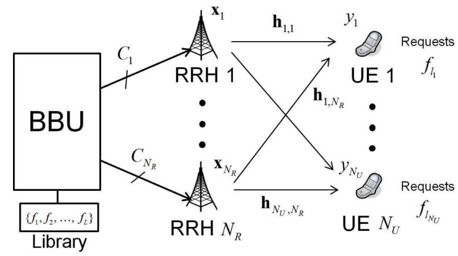

Fig. 1. Illustration of the downlink of C-RAN system with content-centric multicasting.

In this letter, we address the joint optimization of NOMA transmission and fronthaul compression strategies for the content-centric multicast downlink of a C-RAN system. Here we assume finite-capacity fronthaul links from the baseband processing unit (BBU) to remote radio heads (RRHs). After formulating the optimization problem of maximizing the minimum delivery rate of the requested files, we propose an iterative algorithm based on the concave convex procedure (CCCP) approach [1], [4], [10]. From numerical results, we evaluate the proposed NOMA scheme compared to the conventional linear beamforming scheme in terms of the average minimum delivery rate as well as the connectivity support level.

II. System Model

As illustrated in Fig. 1, we consider the downlink of a C-RAN where a BBU serves N_U single-antenna UEs through N_R RRHs. As in [11] and [12], we assume that the ith RRH ($i = 1, \ldots, N_R$) is connected to the BBU via digital fronthaul link of capacity C_i bit/symbol and uses $n_{R,i}$ antennas. The sets of RRHs and UEs are defined as $\mathcal{N}_R \triangleq \{1, 2, \ldots, N_R\}$ and $\mathcal{N}_U \triangleq \{1, 2, \ldots, N_U\}$, respectively, and we denote $n_R \triangleq \sum_{i\in\mathcal{N}_R} n_{R,i}$ as the number of total RRH antennas.

A. Channel Model

Under a flat-fading channel model, the received signal y_k at the kth UE is given by

$$y_k = \sum_{i\in N_R} \mathbf{h}_{k,i}^{\dagger}\mathbf{x}_i + z_k = \mathbf{h}_k^{\dagger}\mathbf{x} + z_k, \tag{1}$$

Manuscript received March 21, 2018; revised April 16, 2018; accepted April 16, 2018. Date of publication April 23, 2018; date of current version October 11, 2018. This work was supported by the National Research Foundation through Ministry of Science, ICT and Future Planning of Korea Government under Grant 2017R1A2B3012316. The associate editor coordinating the review of this paper and approving it for publication was X. Chu.
(Corresponding author: Inkyu Lee.)

S.-R. Lee is with the Advanced Standard Research and Development Laboratory, LG Electronics, Seoul 06772, South Korea (e-mail: sangrim.lee@lge.com).

S.-H. Park is with the Division of Electronic Engineering, Chonbuk National University, Jeonju 54896, South Korea (e-mail: seokhwan@jbnu.ac.kr).

I. Lee is with the School of Electrical Engineering, Korea University, Seoul 136-713, South Korea (e-mail: inkyu@korea.ac.kr).
Digital Object Identifier 10.1109/LWC.2018.2829514

where $\mathbf{h}_{k,i} \in \mathbb{C}^{n_{R,i}\times 1}$ is the channel vector from RRH i to UE k, $\mathbf{x}_i \in \mathbb{C}^{n_{R,i}\times 1}$ indicates the signal vector transmitted by RRH i, $z_k \sim \mathcal{CN}(0,1)$ represents the additive noise at UE k, $(\cdot)^\dagger$ denotes the Hermitian transpose operation, and we define the vectors $\mathbf{h}_k \triangleq [\mathbf{h}_{k,1}^\dagger \dots \mathbf{h}_{k,N_R}^\dagger]^\dagger$ and $\mathbf{x} \triangleq [\mathbf{x}_1^\dagger \dots \mathbf{x}_{N_R}^\dagger]^\dagger$. We impose per-RRH power constraint as $\mathbb{E}||\mathbf{x}_i||^2 \le P_i$ for all $i \in \mathcal{N}_R$. We assume that perfect time synchronization is available among the RRHs and the UEs, and leave a robust design under imperfect synchronization as a future work [13].

B. Content-Centric Multicasting

UE k requests a file f_{l_k} from a library $\mathcal{F} \triangleq \{f_1, f_2, \dots, f_L\}$ of size L at the BBU. We consider the Zipf's distribution [1], [4] for the popularity such that the probability $\Pr[f_{l_k} = f_l]$ of UE k requesting the file f_l is given as $\Pr\left[f_{l_k} = f_l\right] = c\, l^{-\gamma}$, where c is chosen such that $\sum_{l\in\mathcal{L}} \Pr[f_{l_k} = f_l] = 1$ with $\mathcal{L} \triangleq \{1, 2, \dots, L\}$, and $\gamma \ge 0$ is a given constant.

Once the demand vector $\mathbf{f}_{\text{req}} \triangleq [f_{l_1}, \dots, f_{l_{N_U}}]$ of the UEs is realized, we can obtain the set $\mathcal{L}_{\text{req}} \triangleq \cup_{k\in\mathcal{N}_U}\{l_k\}$ of the requested file indices, where $L_{\text{req}} \triangleq |\mathcal{L}_{\text{req}}|$ equals the number of requested files. The L_{req} distinct indices in the set $\mathcal{L}_{\text{req}}$ are denoted as $\tilde{l}_1, \dots, \tilde{l}_{L_{\text{req}}}$. Note that as in [3], the scenario at hand has L_{req} multicasting groups, where the gth group $\mathcal{N}_{U,g} \triangleq \{k|\, l_k = \tilde{l}_g\}$ $(g = 1, \dots, L_{\text{req}})$ is the set of UEs that request the file $f_{\tilde{l}_g}$. Without loss of generality, we assume that the indices $\tilde{l}_1, \dots, \tilde{l}_{L_{\text{req}}}$ are sorted in the ascending order in terms of the worst-case channel power of the requesting UEs, i.e.,

$$\min_{k\in\mathcal{N}_{U,g_1}} \|\mathbf{h}_k\|^2 \le \min_{k\in\mathcal{N}_{U,g_2}} \|\mathbf{h}_k\|^2, \tag{2}$$

for all $g_1 < g_2 \in \mathcal{G} \triangleq \{1, 2, \dots, L_{\text{req}}\}$.

III. Proposed Content-Centric NOMA Scheme

In this section, we propose a content-centric NOMA multicast scheme for the C-RAN system outlined in Section II. The BBU first encodes the files intended for the UEs and obtains the baseband signals $\{s_g\}_{g\in\mathcal{G}}$, where s_g encodes the file $f_{\tilde{l}_g}$ for the gth multicasting group and is distributed as $s_g \sim \mathcal{CN}(0,1)$ under Gaussian encoding. Then, the BBU linearly precodes the encoded signals and performs superposition coding for downlink multicasting as

$$\tilde{\mathbf{x}} = \left[\tilde{\mathbf{x}}_1^\dagger \cdots \tilde{\mathbf{x}}_{N_R}^\dagger\right]^\dagger = \sum\nolimits_{g\in\mathcal{G}} \mathbf{v}_g s_g, \tag{3}$$

where $\mathbf{v}_g \in \mathbb{C}^{n_R\times 1}$ represents the beamforming vector for the signal s_g, and $\tilde{\mathbf{x}}_i \in \mathbb{C}^{n_{R,i}\times 1}$ is the ith subvector of $\tilde{\mathbf{x}}$ that needs be transferred to the ith RRH. Since the fronthaul links to the RRHs have finite capacity, the subvectors $\tilde{\mathbf{x}}_i$ should be quantized and compressed before being transferred on the fronthaul links.

For the fronthaul transmission with finite capacity, the BBU sends a quantized version $\mathbf{x}_i$ of $\tilde{\mathbf{x}}_i$ to the ith RRH. Here the quantized signal $\mathbf{x}_i$ can be modeled as $\mathbf{x}_i = \tilde{\mathbf{x}}_i + \mathbf{q}_i$, where $\mathbf{q}_i$ represents the quantization noise. Assuming a Gaussian test channel for the compression as in [11] and [12], the quantization noise $\mathbf{q}_i$ is independent of the source signal $\tilde{\mathbf{x}}_i$ and distributed as $\mathbf{q}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Omega}_i)$.

Under the point-to-point compression strategy [14], [15] where the signals $\{\tilde{\mathbf{x}}_i\}_{i\in\mathcal{N}_R}$ are quantized separately, in order for the signal $\mathbf{x}_i$ to be reliably recovered by RRH i, the beamforming vectors $\mathbf{v} \triangleq \{\mathbf{v}_k\}_{k\in\mathcal{N}_U}$ and the quantization noise covariance $\boldsymbol{\Omega}_i$ should satisfy the constraint [16, Ch. 3]

$$\begin{aligned} g_i(\mathbf{v}, \boldsymbol{\Omega}_i) \triangleq I\left(\tilde{\mathbf{x}}_i; \mathbf{x}_i\right) = \log_2 \det\left(\sum\nolimits_{g\in\mathcal{G}} \mathbf{E}_i^\dagger \mathbf{v}_g \mathbf{v}_g^\dagger \mathbf{E}_i + \boldsymbol{\Omega}_i\right) \\ -\ \log_2 \det(\boldsymbol{\Omega}_i) \le C_i, \end{aligned} \tag{4}$$

where we define the shaping matrix $\mathbf{E}_i \in \mathbb{C}^{n_R\times n_{R,i}}$ that has all zero elements except for the rows from $\sum_{j=1}^{i-1} n_{R,j} + 1$ and $\sum_{j=1}^{i} n_{R,j}$ filled with an identity matrix of size $n_{R,i}$.

Following the principle of NOMA schemes (e.g., [7] and [8]), we assume that UE $k \in \mathcal{N}_{U,g}$ that belongs to the gth multicasting group decodes the file $f_{\tilde{l}_g}$ by performing SIC decoding of the signals $\{s_{g'}\}_{g'=1}^{g}$ with the order[1] of $s_1, s_2, \dots, s_g$. In order to guarantee that the signal s_g is successfully decoded at the UEs $\mathcal{N}_{U,g}$ in the target multicasting group g as well as at the UEs $\cup_{g'=g}^{L_{\text{req}}} \mathcal{N}_{U,g'}$ in the next multicasting groups, the rate R_g of the signal s_g can be achieved if the following condition is satisfied for all $g' \ge g$ and $k \in \mathcal{N}_{U,g'}$ as

$$R_g \le \log_2\left(1 + \gamma_{k,g}(\mathbf{v}, \boldsymbol{\Omega})\right), \tag{5}$$

where the signal to interference-plus-noise ratio (SINR) $\gamma_{k,g}(\mathbf{v}, \boldsymbol{\Omega})$ of the signal s_g at UE k is expressed as

$$\gamma_{k,g}(\mathbf{v}, \boldsymbol{\Omega}) = \frac{\left|\mathbf{h}_k^\dagger \mathbf{v}_g\right|^2}{1 + \mathbf{h}_k^\dagger \boldsymbol{\Omega} \mathbf{h}_k + \sum_{g'=g+1}^{L_{\text{req}}} \left|\mathbf{h}_k^\dagger \mathbf{v}_{g'}\right|^2}, \tag{6}$$

and $\boldsymbol{\Omega}$ is denoted as $\boldsymbol{\Omega} \triangleq \text{diag}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \dots, \boldsymbol{\Omega}_{N_R})$.

IV. Optimization of the Proposed NOMA Scheme

We tackle the problem of optimizing the beamforming vectors $\mathbf{v}$ and the quantization noise covariance matrix $\boldsymbol{\Omega}$ with the goal of maximizing the minimum delivery rate $R_{\min} \triangleq \min_{g\in\mathcal{G}} R_g$ of the requested files while satisfying the per-RRH fronthaul capacity and power constraints. To this end, we formulate the problem as

$$\begin{aligned} \max_{\mathbf{v}, \boldsymbol{\Omega}, R_{\min}} \quad & R_{\min} \\ \text{s.t.} \quad & R_{\min} \le \log_2\left(1 + \gamma_{k,g}(\mathbf{v}, \boldsymbol{\Omega})\right) \\ & \text{for } k \in \cup_{g'\ge g} \mathcal{N}_{U,g'},\ g \in \mathcal{G}, \\ & g_i(\mathbf{v}, \boldsymbol{\Omega}_i) \le C_i \ \text{ for } i \in \mathcal{N}_R, \\ & \sum\nolimits_{g\in\mathcal{G}} \text{tr}\left(\mathbf{E}_i^\dagger \mathbf{v}_g \mathbf{v}_g^\dagger \mathbf{E}_i\right) + \text{tr}(\boldsymbol{\Omega}_i) \le P_i \ \text{ for } i \in \mathcal{N}_R. \end{aligned} \tag{7}$$

To solve problem (7), let us define the variable $\mathbf{V}_g \triangleq \mathbf{v}_g \mathbf{v}_g^\dagger$ for $g \in \mathcal{G}$. Recasting problem (7) with respect to the variables $\mathbf{V} \triangleq \{\mathbf{V}_g\}_{g\in\mathcal{G}}$, $\boldsymbol{\Omega}$ and $R_{\min}$ by removing the constraint $\text{rank}(\mathbf{V}_g) \le 1$, we obtain a difference-of-convex (DC) problem. Since the solution to the obtained problem may not satisfy the rank constraint, a projection process is needed as will be explained at the end of this section.

[1] We leave the problem of jointly optimizing the decoding order along with the other variables $\mathbf{v}$ and $\boldsymbol{\Omega}$ as a future work.

Algorithm 1 CCCP Algorithm for Problem (7)

1. Initialize the variables $\mathbf{V}^{(1)}$ and $\mathbf{\Omega}^{(1)}$ to arbitrary matrices that satisfy the constraints of problem (7) and set $t = 1$.
2. Update the variables $\mathbf{V}^{(t+1)}$ and $\mathbf{\Omega}^{(t+1)}$ as a solution of problem (8).
3. Stop if a convergence criterion is satisfied. Otherwise, set $t \leftarrow t + 1$ and go back to Step 2.

We find an efficient solution of the DC problem by adopting the CCCP based iterative algorithm [1] where the optimization variables are updated at each iteration as a solution of the convex problem by linearizing the terms that induce non-convexity of the problem. The detailed CCCP algorithm is described in Algorithm 1, where the superscript t ($t = 1, 2, \ldots$) of the variables $\mathbf{V}^{(t)}$ and $\mathbf{\Omega}^{(t)}$ indicates the iteration index, and the convex problem (8) is stated as

$$\begin{aligned} \max_{\mathbf{V},\, \mathbf{\Omega},\, R_{\min}} \quad & R_{\min} \qquad (8) \\ \text{s.t.} \quad & R_{\min} \leq \log_2\Big(1 + \mathbf{h}_k^\dagger \mathbf{\Omega} \mathbf{h}_k + \sum\nolimits_{g'=g}^{L_{\text{req}}} \mathbf{h}_k^\dagger \mathbf{V}_{g'} \mathbf{h}_k\Big) \\ & -\varphi_{1,g,k}\Big(\mathbf{V}, \mathbf{\Omega}, \mathbf{V}^{(t)}, \mathbf{\Omega}^{(t)}\Big) \\ & \text{for } k \in \cup_{g' \geq g} \mathcal{N}_{U,g'},\ \ g \in \mathcal{G}, \\ & \varphi_{2,i}\Big(\mathbf{V}, \mathbf{\Omega}, \mathbf{V}^{(t)}, \mathbf{\Omega}^{(t)}\Big) - \log_2 \det(\mathbf{\Omega}_i) \leq C_i \ \text{ for } i \in \mathcal{N}_R, \\ & \sum\nolimits_{g \in \mathcal{G}} \text{tr}\Big(\mathbf{E}_i^\dagger \mathbf{V}_g \mathbf{E}_i\Big) + \text{tr}(\mathbf{\Omega}_i) \leq P_i \ \text{ for } i \in \mathcal{N}_R, \end{aligned}$$

where we define the functions $\varphi_{1,g,k}(\mathbf{V}, \mathbf{\Omega}, \mathbf{V}^{(t)}, \mathbf{\Omega}^{(t)})$ and $\varphi_{2,i}(\mathbf{V}, \mathbf{\Omega}, \mathbf{V}^{(t)}, \mathbf{\Omega}^{(t)})$ as

$$\begin{aligned} & \varphi_{1,g,k}\Big(\mathbf{V}, \mathbf{\Omega}, \mathbf{V}^{(t)}, \mathbf{\Omega}^{(t)}\Big) \triangleq \varphi\Big(\Xi_{k,g},\ \Xi_{k,g}^{(t)}\Big), \\ & \varphi_{2,i}\Big(\mathbf{V}, \mathbf{\Omega}, \mathbf{V}^{(t)}, \mathbf{\Omega}^{(t)}\Big) \\ & \triangleq \varphi\Big(\sum\nolimits_{g \in \mathcal{G}} \mathbf{E}_i^\dagger \mathbf{V}_g \mathbf{E}_i + \mathbf{\Omega}_i,\ \sum\nolimits_{g \in \mathcal{G}} \mathbf{E}_i^\dagger \mathbf{V}_g^{(t)} \mathbf{E}_i + \mathbf{\Omega}_i^{(t)}\Big), \end{aligned}$$

with the notations $\varphi(\mathbf{A}, \mathbf{B}) \triangleq \log_2 \det(\mathbf{B}) + \text{tr}(\mathbf{B}^{-1}(\mathbf{A} - \mathbf{B}))/\ln 2$ and $\Xi_{k,g} \triangleq 1 + \mathbf{h}_k^\dagger \mathbf{\Omega} \mathbf{h}_k + \sum_{g'=g+1}^{L_{\text{req}}} \mathbf{h}_k^\dagger \mathbf{V}_{g'} \mathbf{h}_k$.

In Step 1, we set a feasible initial point by randomly generating the beamforming matrices $\mathbf{V}$ and the quantization noise covariance matrices $\mathbf{\Omega}$ and then properly reducing $\mathbf{V}$ and $\mathbf{\Omega}$ so that the power and fronthaul capacity constraints are satisfied. When solving the convex problem in (8) at each iteration, we can use standard convex solvers such as CVX [17]. After the algorithm converges, each beamforming vector $\mathbf{v}_g$ is computed as $\mathbf{v}_g \leftarrow \sqrt{\lambda_{\max}(\mathbf{V}_g)}\mathbf{v}_{\max}(\mathbf{V}_g)$, where $\lambda_{\max}(\cdot)$ and $\mathbf{v}_{\max}(\cdot)$ take the principal eigenvalue and eigenvector, respectively, of an input matrix. The convergence of the CCCP algorithms was discussed in [1] and [18], and the complexity of the algorithm is given by the product of the number of iterations for convergence and the complexity of solving the convex problem (8) at each iteration. We have observed from simulations that the algorithm converges within a few tens of iterations, and the complexity of solving problem (8) is known to be polynomial in the problem size [19, Ch. 11], which is here given as $O(n_R^2 L_{\text{req}} + n_R^2)$.

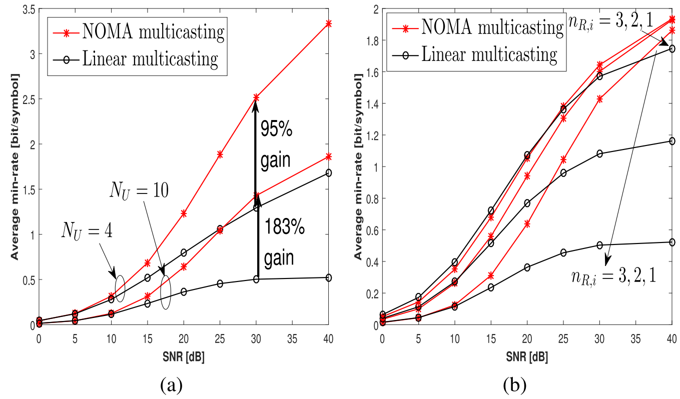

Fig. 2. Average minimum rate $R_{\min}$ with respect to the SNR for $N_R = 2$ and $C_i = 10$ bit/symbol (a) $N_U \in \{4, 10\}$ and $n_{R,i} = 1$ (b) $N_U = 10$ and $n_{R,i} = \{1, 2, 3\}$.

V. Numerical Results

In this section, we evaluate the proposed content-centric NOMA multicast scheme for the downlink of C-RAN by presenting numerical results. For simulation, we assume that the positions of the RRHs and the UEs are uniformly distributed within a circular area of radius[2] 100 m and that the elements of the channel vector $\mathbf{h}_{k,i}$ are independent and identically distributed (i.i.d.) as $\mathcal{CN}(0, \rho_{k,i})$, where the path-loss $\rho_{k,i}$ is given as $\rho_{k,i} = 1/(1 + (d_{k,i}/d_0)^\alpha)$. Here $d_{k,i}$ denotes the distance between RRH i and UE k, d_0 indicates the reference distance corresponding to the signal-to-noise ratio (SNR) of 0 dB, and α represents the path-loss exponent. In the simulation, we set $\alpha = 3$, $d_0 = 30$ m and $L = 10$ files in the library with the Zipf's popularity of $\gamma = 1$.

We compare the performance of the proposed NOMA multicasting scheme with the conventional linear multicast precoding scheme [1], [3], [4], where each UE performs single-user detection to obtain the requested file information while treating the interference signals as additive noise. For this linear precoding scheme, the achievable rate R_g for the UEs in the gth multicasting group is given as

$$R_g = \min_{k \in \mathcal{N}_{U,g}} \log_2\left(1 + \frac{\left|\mathbf{h}_k^\dagger \mathbf{v}_g\right|^2}{1 + \mathbf{h}_k^\dagger \mathbf{\Omega} \mathbf{h}_k + \sum_{g' \in \mathcal{G} \setminus \{g\}} \left|\mathbf{h}_k^\dagger \mathbf{v}_{g'}\right|^2}\right),$$

and hence the optimization can be addressed by tackling problem (7) with the rate constraint modified based on the above rate expression.

In Fig. 2, we plot the average minimum rate $R_{\min}$ with respect to the SNR for a C-RAN downlink with $N_R = 2$ and $C_i = 10$ bit/symbol ((a) $N_U \in \{4, 10\}$ and $n_{R,i} = 1$; (b) $N_U = 10$ and $n_{R,i} = \{1, 2, 3\}$). It is observed from Fig. 2-(a) that a performance gain of the proposed NOMA scheme increases with the SNR since the performance of both schemes is limited by the additive noise or fronthaul quantization noise signals when the SNR is small. Also, we can see that the performance gain of the NOMA scheme grows

[2] In this letter, we focus on the case of fixed radius. We have observed in our simulations that the advantage of the NOMA scheme becomes pronounced for a smaller radius, since the network becomes more interference-coupled.

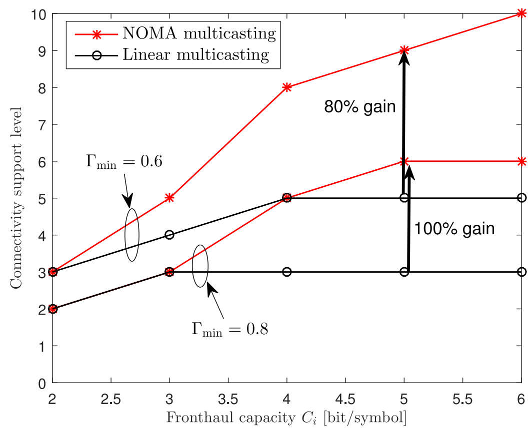

Fig. 3. Connectivity support level with respect to the fronthaul capacity C_i ($N_R = 2$, $n_{R,i} = 1$ and 20 dB SNR).

when more UEs are simultaneously supported due to the increased number of interference signals. Fig. 2-(b) shows that the performance loss caused by decreasing the number of RRH antennas is more significant for the case of the linear multicasting scheme. This is because the proposed NOMA scheme can manage the interference signals more efficiently by means of not only cooperative beamforming across the RRHs but also SIC decoding at the UEs.

We now investigate the performance of the proposed NOMA scheme in terms of the connectivity support level. Fig. 3 exhibits the connectivity support level with respect to the fronthaul capacity C_i for a C-RAN downlink with $N_R = 2$, $n_{R,i} = 1$ and 20 dB SNR. Here the connectivity support level is defined as the maximum number of UEs that can be served simultaneously while satisfying given quality-of-service (QoS) constraint. For the QoS condition, we impose the average minimum rate $\mathbb{E}[R_{\min}]$ to be larger than or equal to a given threshold level $\Gamma_{\min}$. From Fig. 3, we can check that the performance gap between the proposed NOMA multicasting scheme and the conventional linear approach increases as the fronthaul capacity grows. Specifically, the gain becomes more pronounced as the QoS constraint gets stronger. For instance, the NOMA scheme shows a performance gain of 100% compared to the linear scheme for $C_i = 5$ bit/symbol and $\Gamma_{\min} = 0.8$ bit/symbol.

VI. Conclusion

In this letter, we have studied the NOMA transmission for a content-centric multicast downlink of a C-RAN system equipped with finite-capacity fronthaul links. Under the assumption of content-based communication in which the UEs request files from a library with a given popularity, we have solved the problem of maximizing the minimum of delivery rates of the requested files while satisfying the per-RRH fronthaul capacity and transmit power constraints. Since the problem is an instance of the DC problems, we have derived a CCCP-based iterative algorithm. Through numerical results, we have validated the performance gains of the proposed NOMA scheme in terms of the delivery rate and the connectivity support level. As future work, we mention the development of more efficient optimization algorithms based on direct CCCP [9, Sec. IV] and the weighted minimum mean squared error (WMMSE) approaches [10], [12], and a hybrid design of NOMA and linear precoding systems based on rate splitting and superposition coding [20].

References

[1] M. Tao, E. Chen, H. Zhou, and W. Yu, “Content-centric sparse multicast beamforming for cache-enabled cloud RAN,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118–6131, Sep. 2016.

[2] W. Lee, H. Park, H.-B. Kong, J. S. Kwak, and I. Lee, “A new beamforming design for multicast systems,” *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 4093–4097, Oct. 2013.

[3] S.-H. Park *et al.*, “Robust interference management via linear precoding and linear/non-linear equalization,” *J. Signal Process. Syst.*, vol. 83, no. 2, pp. 133–149, May 2016.

[4] S.-H. Park, O. Simeone, and S. S. Shitz, “Joint optimization of cloud and edge processing for fog radio access networks,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7621–7632, Nov. 2016.

[5] Y. Saito *et al.*, “Non-orthogonal multiple access (NOMA) for cellular future radio access,” in *Proc. IEEE Veh. Technol. Conf.*, Jun. 2013, pp. 1–5.

[6] J. Choi, “Non-orthogonal multiple access in downlink coordinated two-point systems,” *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 313–316, Feb. 2014.

[7] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, “A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems,” *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76–88, Jan. 2016.

[8] W. Shin *et al.*, “Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges,” *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 176–183, Oct. 2017.

[9] E. Chen, M. Tao, and Y.-F. Liu, “Joint base station clustering and beamforming for non-orthogonal multicast and unicast transmission with backhaul constraints,” *arXiv:1712.01508*, Dec. 2017.

[10] J. Kim, S.-H. Park, O. Simeone, I. Lee, and S. S. Shitz, “Joint design of digital and analog processing for downlink C-RAN with large-scale antenna arrays,” in *Proc. IEEE SPAWC*, Sapporo, Japan, Jul. 2017, pp. 1–5.

[11] S.-H. Park, O. Simeone, O. Sahin, and S. S. Shitz, “Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory,” *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 69–79, Nov. 2014.

[12] Y. Zhou and W. Yu, “Fronthaul compression and transmit beamforming optimization for multi-antenna uplink C-RAN,” *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4138–4151, Aug. 2016.

[13] H. Haci, H. Zhu, and J. Wang, “Performance of non-orthogonal multiple access with a novel asynchronous interference cancellation technique,” *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1319–1335, Mar. 2017.

[14] O. Simeone, O. Somekh, H. V. Poor, and S. S. Shitz, “Downlink multicell processing with limited-backhaul capacity,” *EURASIP J. Adv. Signal Process.*, vol. 2009, Feb. 2009, Art. no. 3.

[15] S.-H. Park, K.-J. Lee, C. Song, and I. Lee, “Joint design of fronthaul and access links for C-RAN with wireless fronthauling,” *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1657–1661, Nov. 2016.

[16] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge Univ. Press, 2011.

[17] M. Grant and S. Boyd. (Sep. 2013). *CVX: MATLAB Software for Disciplined Convex Programming, Ver 2.0 Beta*. [Online]. Available: http://cvxr.com/cvx

[18] B. K. Sriperumbudur and G. R. Lanckriet, “A proof of convergence of the concave-convex procedure using Zangwill’s theory,” *Neural Comput.*, vol. 24, no. 6, pp. 1391–1407, Jun. 2012.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004

[20] Y. Mao, B. Clerckx, and V. O. K. Li, “Rate-splitting multiple access for downlink communication systems: Bridging, generalizing and outperforming SDMA and NOMA,” *arXiv:1710.11018v3*, Oct. 2017.