Fronthaul Data Compression for Uplink CoMP in Cloud Radio Access Network (C-RAN)

Yinan Qi, Muhammad Zeeshan Shakir, Muhammad Ali Imran, Khalid A. Qaraqe, Atta Quddus, and Rahim Tafazolli

Abstract

The design of efficient wireless fronthaul connections for future heterogeneous networks incorporating emerging paradigms such as cloud radio access network (C-RAN) has become a challenging task that requires the most effective utilization of fronthaul network resources. In this paper, we propose to use distributed compression to reduce the fronthaul traffic in uplink Coordinated Multi-Point (CoMP) for C-RAN. Unlike the conventional approach where each coordinating point quantizes and forwards its own observation to the processing centre, these observations are compressed before forwarding. At the processing centre, the decompression of the observations and the decoding of the user message are conducted in a successive manner. The essence of this approach is the optimization of the distributed compression using an iterative algorithm to achieve maximal user rate with a given fronthaul rate. In other words, for a target user rate the generated fronthaul traffic is minimized. Moreover, joint decompression and decoding is studied and an iterative optimization algorithm is devised accordingly. Finally, the analysis is extended to multi-user case and our results reveal that, in both dense and ultra-dense urban deployment scenarios, the usage of distributed compression can efficiently reduce the required fronthaul rate and a further reduction is obtained with joint operation.

Index Terms

Heterogeneous network (HetNet), fronthaul, uplink Coordinated Multi-Point (CoMP), cloud radio access network (C-RAN), distributed compression.

I INTRODUCTION

The astronomical growth in mobile data traffic volume is making it increasingly difficult for operators to meet the traffic demands with the available resources and network architecture. This remarkable growing momentum of improving the quality of experience of the users will most likely continue due to the emerging needs of connecting people, machines, and applications through mobile infrastructure. Extrapolation of current growth trend predicts that the future networks are required to be prepared to support up to a thousand fold increase in total mobile broadband traffic by 2020 [1]. In light of this trend, heterogeneous network (HetNet), where small cells (SCs) are deployed to provide improved coverage and capacity to highly concentrated users, is envisaged as one of the key solutions to economic and sustainable future networks [2]-[4]. However, each added small cell has a profound effect on overall network performance, both with the increased capacity it brings as well as the large amount of co-channel interference it generates when sharing the same frequency [5]. One of the ways to improve the spectral efficiency of the system and mitigate the adverse effect of the huge interference is the use of advanced Coordinated Multi-Point (CoMP) topologies among the SCs in order to combat the generated interference using interference management or interference alignment-based transmission solutions [6]-[8].

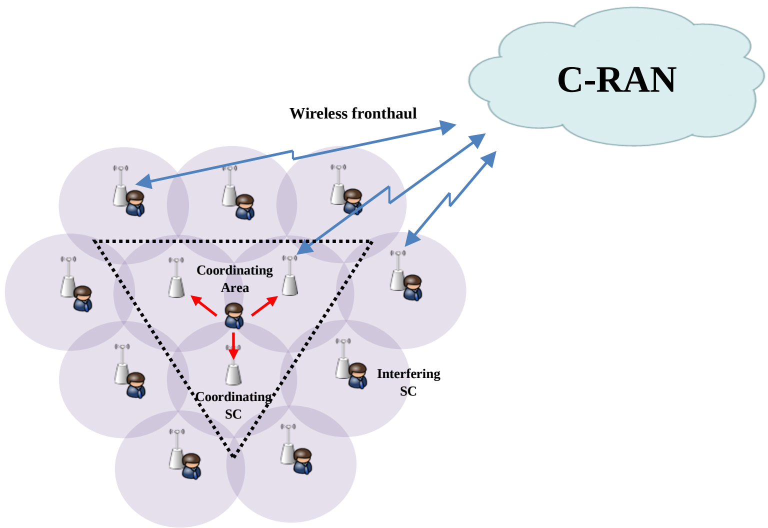


Figure. 1 A C-RAN Architecture with wireless front/fronthaul links

As one of the most promising potential solutions to efficiently conduct inter-cell coordination, Cloud radio access network (C-RAN) recently attracted a great deal of attention [7]-[13]. C-RAN consists of a large number of spatially separated SCs and a centralized data processing centre as shown in Fig. 1, where the SCs are simplified and activated wherever and whenever the demand arises and some or all other functionalities, e.g., digital baseband processing, are centralized [9]-[11]. In this paradigm, CoMP is the key to enable increased spectral resource usage as shown in Fig. 1 [12]-[13]. Nevertheless, this gain is expected to be harvested at the cost of high-capacity optical fibre fronthaul links (also called as fronthaul) connecting the SCs and the processing centre [9]-[10]. However, for those locations where existing fibre-access is unavailable and new fibre-access is difficult or expensive to be deployed, wired fronthaul links need to be replaced by wireless fronthaul solutions. However the wireless fronthaul links are usually with limited capacity, hence the design of fronthaul networks and efficient usage of fronthaul resources are of paramount importance to support ultra-dense SCs in a spectrum, cost, and energy efficient manner.

While most of the previous research on uplink CoMP assumes lossless fronthaul links with infinite capacity and the obtained results are only of theoretical interest, some researchers have brought attention to realistic fronthaul links with capacity limit [14]-[17]. However, in these works the coordinating SCs are assumed to simply quantize their own received signals before forwarding them to the processing centre via the fronthaul links. Considering the fact that the observations of all coordinating cells are actually correlated because they are broadcasted from the same source, distributed Wyner-Ziv compression can be applied to exploit this correlation and reduce the required fronthaul link rate. Distributed Wyner-Ziv compression is proposed in [18]-[19] and further extended to multiple sources by Gastpar in [20]. In distributed compression, each coordinating SC compresses its own received signal and at the processing centre the correlation between the receptions of all coordinating SCs will be exploited to decompress and reconstruct their observations which will then be used to decode the user message. Although the information-theoretic fundamentals behind the associated schemes are well known, the formulation of the optimization problem to optimize compression at each SC has not been addressed. In [21], distributed compression is investigated in a multi-relay scenario but the system assumptions are different since only one source is considered. In [22], multiple users transmit independently to a singling destination via a relay node. However, only one relay node is assumed whereas in our work we focus on multiple relays nodes or SCs. Side information is also explored for data dissemination in [23] but the considered system model and applicable scenarios are different from ours in the sense that no relay nodes are involved.

The main contributions of this work can be summarized as follows:

- An uplink CoMP framework incorporating capacity-limited multiple fronthaul links is considered, which can be easily employed in the C-RAN architecture;
- Under the proposed framework, distributed Wyner-Ziv compression is employed and optimized to maximize user rate with limited fronthaul capacity. On the other hand, the optimization can be interpreted as minimization of required fronthaul rate for a given target user rate. An iterative algorithm is designed to conduct the optimization task. Moreover, we consider joint operation of decompression and decoding and investigate its performance;
- In addition to the single user case, the analysis is further extended to multi-user case. Another optimization problem is formed and decoupled into a simpler formation. A new iterative optimization algorithm is proposed accordingly;
- Two typical uplink CoMP application scenarios with practical deployment and noise and interference assumptions are considered to yield insight into not only the efficiency of proposed schemes but also the applicability of uplink CoMP.

The rest of the paper is organized as follows: In the next section, the system model is presented. The achievable user rate is derived and optimized in Section III. The derivation is extended to joint decompression & decoding in section IV. Multi-user case is studied in section V. Numerical results and discussions are given in Section VI and the final section concludes the paper. **Notation:** The following notations and definitions are employed in this work: capital letters e.g., X, stand for random variables and lower case letters e.g., x, represent realization of these variables. Vectors are represented by bold letters, e.g., $\boldsymbol{X}$. $\boldsymbol{X}^{\mathrm{H}}$ stands for the conjugate transpose of $\boldsymbol{X}$. Equalities of vectors are element-wise, i.e., $\boldsymbol{X}>\boldsymbol{Y}$ means $X_i>Y_i$ for $\forall i$. A superscript, e.g., $\boldsymbol{X}^n$ represents the vector $(X_1,\ldots,X_n)$. Euclidean letters denote sets, e.g., $\mathrm{T}\equiv\{1,\ldots,T\}$, whose cardinality is $|\mathrm{T}|$. A subset Σ which fulfils $\Sigma\subseteq\mathrm{T}$ has a complementary set Σ^{C}, where $\Sigma\cup\Sigma^{\mathrm{C}}=\mathrm{T}$ and $\Sigma\cap\Sigma^{\mathrm{C}}=\emptyset$. A subset with element t removed is denoted as $\Sigma\backslash t$ and x_Σ refers to $\{x_1,\ldots,x_i,\ldots,x_{|\Sigma|}\}$. Mutual information and entropy are denoted as $I(\cdot)$ and $H(\cdot)$, respectively. Unless otherwise stated, all the logarithms are in base 2. $[x]^+=\max\{x,0\}$ and $\mathbf{E}\{\cdot\}$ denotes expectation.

II. SYSTEM MODEL

As shown in shown in Fig. 2, a user simultaneously transmits to N_κ coordinating SCs which are then connected to the processing centre of C-RAN via wireless fronthaul links. The channel between the user and the jth SC is denoted as h_j and assumed to be quasi-static, and the link capacity between the jth SC and the processing centre is C_j.

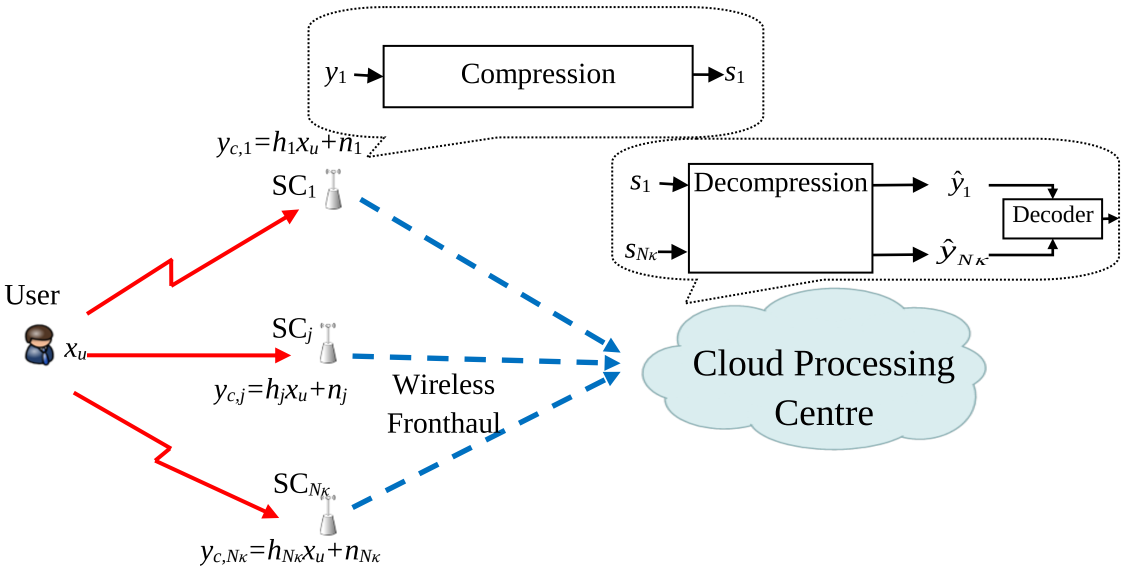


Figure. 2 System model

Assuming synchronized transmission across the entire network, the user transmits a message w to the SCs by sending a frame consisting of n symbols, each message belonging to a set $\Omega=\{1,\ldots,2^{nR}\}$. One codebook Ξ_u is defined for the user, where each element of its codeword $\boldsymbol{X}_u^n$ is assumed to be independently and identically distributed (*i.i.d.*) and modelled by circularly symmetric complex Gaussian (CSCG) distribution. The user transmits $\boldsymbol{X}_u^n$ to all the SCs and at the jth SC, the received signal is

$$y_{c,j}[t]=h_j x_u[t]+n_j[t], \qquad \text{for } 1\le j\le N_\kappa, 1\le t\le n, \tag{1}$$

where n_j is the additive Gaussian noise at SC$_j$, following CSCG distribution with zero mean and variance σ_j. Here we assume that $\sigma_j=\sigma_c$ for $1\le j\le N_\kappa$. Each SC compresses the received signal into the Wyner-Ziv bin index s_j using Wyner-Ziv lossy distributed source coding [18]-[20] and forwards s_j to the processing centre. Since the fronthaul transmission is assumed to operate at a different frequency band from the users, the coordinating SCs can be regarded as full-duplex nodes because they can forward the compressed index of the (n-1)th received frame receive the nth frame from the user at the same time as long as the duration of compression bin index transmission is shorter than the user frame, which is normally true because of the relatively high capacity of fronthaul links. In this work, we assume that the durations of frames from the user and SCs are the same as T. The total duration of transmitting n user frames is $nT+T$. If $n \gg 1$, it can be approximated as nT.

All SCs forward the Wyner-Ziv bin index to the processing centre simultaneously via the wireless fronthaul links. We assume that the multiple fronthaul links do not interfere with each other. This is because in some of the fifth generation of cellular network (5G) proposals, the fronthaul links are designed to operate at very high carrier frequency, e.g., 60GHz [24] - [27], where the wave length is in millimetres thus the transmission is highly directional. The power leakage from one fronthaul link to another is small enough to be negligible. At the processing centre, the decoding of the message w is conducted in a successive manner: firstly, the compression indices are decompressed to reconstruct the observations of the coordinating SCs; secondly, the reconstructed signals are coherently combined to decode the messages w from the user.

III. ACHIEVABLE RATE

The achievable rate for such a system is a straightforward extension in [28] as

$$\begin{gathered} R \leq I\left(X_u; \hat{Y}_{c,\mathrm{J}}\right) \\ \text{subject to} \quad \forall \mathrm{S} \subseteq \mathrm{J} \ : I\left(\hat{Y}_{c,\mathrm{S}}; Y_{c,\mathrm{S}} \mid \hat{Y}_{c,\mathrm{S}^C}\right) \leq \sum_{j \in \mathrm{S}} C_j \end{gathered}, \tag{2}$$

where $\vartheta=\{1,\ldots,N_\kappa\}$ and $\hat{Y}_{c,j}$ can be interpreted as the estimation of $Y_{c,j}$. The constraint functions ensure that the compressed information can be successfully received by the processing centre and thus the SCs' observations are correctly reconstructed.

With CSCG distributed codebooks, the achievable rate is derived in Appendix A as

$$\begin{gathered} R = \log\left(1 + \sum_{j \in \mathrm{J}} \frac{\gamma_j P_X}{\sigma_c + \sigma_{w,j}}\right), \\ \text{subject to} \quad f_{\mathrm{S}}\left(\sigma_{w,\mathrm{J}}\right) \leq \sum_{j \in \mathrm{S}} C_j \end{gathered} \tag{3}$$

for all $\Sigma \subseteq \vartheta$, where

$$f_{\mathrm{S}}\left(\sigma_{w,\mathrm{J}}\right)=\log\left(\prod_{j\in\mathrm{J}}\left(\sigma_c+\sigma_{w,j}\right)+\sum_{j\in\mathrm{J}}\left(\gamma_j P_X\prod_{i\in\mathrm{J},i\neq j}\left(\sigma_c+\sigma_{w,i}\right)\right)\right)-\log\left(\prod_{j\in\mathrm{S}}\sigma_{w,j}\right)$$
$$-\log\left(\prod_{j\in\mathrm{S}^C}\left(\sigma_c+\sigma_{w,j}\right)+\sum_{j\in\mathrm{S}^C}\left(\gamma_j P_X\prod_{i\in\mathrm{S}^C,i\neq j}\left(\sigma_c+\sigma_{w,i}\right)\right)\right),$$

$\mathbf{E}\{X_u\}\leq P_X$ for $1\leq j\leq N_\kappa$, $\sigma_{w,j}$ is the variance of the compression noise introduced by distributed Wyner-Ziv compression at SC$_j$, and $\gamma_j=|h_j|^2$. Notice that the achievable rate (3) depends on the variances of the compression noises. The number of constraint equations, denoted as κ, depends on the number of coordinating SCs. For N_κ SCs, it is

$$\kappa=\sum_{l\in\mathrm{J}}\binom{N_\kappa}{l}=\sum_{l\in\mathrm{J}}\kappa_{N_\kappa,l}\ . \tag{4}$$

The optimization objective function can be formed as,

$$\begin{aligned}&\text{maximize} && R\left(\sigma_{w,\mathrm{J}}\right)=\log\left(1+\sum_{j\in\mathrm{J}}\frac{\gamma_j P_X}{\sigma_c+\sigma_{w,j}}\right),\\ &\text{subject to} && f_{\mathrm{S}^l}\left(\sigma_{w,\mathrm{J}}\right)\leq C_{\mathrm{S}^l},\ \sigma_{w,j}>0,\ j\in\mathrm{J}\end{aligned} \tag{5}$$

for $\forall\,\mathrm{S}^l$, where

$$C_{\mathrm{S}^l}=\sum_{i\in\mathrm{S}^l}C_i,\ \text{for}\ \forall\mathrm{S}^l\subseteq\mathrm{J}\ ,\ \text{and}\ \left|\mathrm{S}^l\right|=l,\ 1\leq l\leq N_\kappa\,.$$

For each l, there are $\kappa_{N\kappa,l}$ different Σ^l, each corresponding to one constraint function accordingly. Since the maximization objective function is not in standard concave form, we resort to its dual problem by forming its Lagrangian dual

$$L\left(\sigma_{w,\mathrm{J}},\boldsymbol{\lambda}^\kappa\right)=R\left(\sigma_{w,\mathrm{J}}\right)+\boldsymbol{\lambda}^{\kappa T}\left(\mathbf{C}^\kappa-\mathbf{f}^\kappa\left(\sigma_{w,\mathrm{J}}\right)\right), \tag{6}$$

where

$$\boldsymbol{\lambda}^\kappa=\left(\boldsymbol{\lambda}_{\mathrm{S}^1},\ldots,\boldsymbol{\lambda}_{\mathrm{S}^l},\ldots,\boldsymbol{\lambda}_{\mathrm{S}^{N_\kappa}}\right)^T,\ \boldsymbol{\lambda}_{\mathrm{S}^l}=\left(\lambda_1^l,\ldots,\lambda_{\kappa_{N_\kappa,l}}^l\right),\ \mathbf{C}^\kappa=\left(\mathbf{C}_{\mathrm{S}^1},\ldots,\mathbf{C}_{\mathrm{S}^l},\ldots,\mathbf{C}_{\mathrm{S}^{N_\kappa}}\right)^T,\ \mathbf{C}_{\mathrm{S}^1}=\left(C_1^l,\ldots,C_{\kappa_{N_\kappa,l}}^l\right),$$
$$\mathbf{f}^\kappa\left(\sigma_{w,\mathrm{J}}\right)=\left(\mathbf{f}_{\mathrm{S}^1}\left(\sigma_{w,\mathrm{J}}\right),\ldots,\mathbf{f}_{\mathrm{S}^l}\left(\sigma_{w,\mathrm{J}}\right),\ldots,\mathbf{f}_{\mathrm{S}^{N_\kappa}}\left(\sigma_{w,\mathrm{J}}\right)\right)^T,\mathbf{f}_{\mathrm{S}^l}\left(\sigma_{w,\mathrm{J}}\right)=\left(f_1^l\left(\sigma_{w,\mathrm{J}}\right),\ldots,f_{\kappa_{N_\kappa,l}}^l\left(\sigma_{w,\mathrm{J}}\right)\right).$$

The dual function is defined as a maximization function of (6)

$$\varphi\left(\boldsymbol{\lambda}^\kappa\right)=\max_{\sigma_{w,\mathrm{J}}}L\left(\sigma_{w,\mathrm{J}},\boldsymbol{\lambda}^\kappa\right). \tag{7}$$

Then the dual problem takes the following form:

$$\begin{aligned} &\text{minimization} \quad \varphi\left(\boldsymbol{\lambda}^{\kappa}\right) \\ &\text{subject} \quad \text{to} \quad \boldsymbol{\lambda}^{\kappa} \geq \mathbf{0}^{\kappa} \end{aligned}. \tag{8}$$

The dual objective function $\varphi(\boldsymbol{\lambda}^{\kappa})$ is a convex function regardless of the concavity of the primal function $R(\sigma_{w,\vartheta})$ [29]. If we can prove that the duality gap between the primal problem (5) and the dual problem (8) is zero, we can solve the primal problem by resorting to the dual problem because they have the same solution.

***Theorem* 1**: The duality gap of the primal problem (5) and dual problem (8) is zero.

Proof: See Appendix B for proof.■

With *Theorem* 1, the primal problem can be solved by searching for the solution of the dual problem. At first, we need to find the optimal $\sigma_{w,\vartheta}$ to maximize (6). Due to its high complexity, it is difficult to achieve a closed-form solution. A successive optimization algorithm can be applied, where only $\sigma_{w,j}$ is optimized at one time while other $\sigma_{w,\vartheta/i}$ are kept constant [30]-[31]. The optimization procedure starts from $\sigma_{w,1}$ and ends with $\sigma_{w,N\kappa}$ and the same procedure will be repeated until $\sigma_{w,\vartheta}$ converge.

We can define a set comprised by all subsets of ϑ and ordered according to the number of contained elements as $\mathrm{C}=\left\{\mathrm{S}_1^1,\ldots,\mathrm{S}_{N_\kappa}^1,\ldots,\mathrm{S}_1^l,\ldots,\mathrm{S}_{\kappa_{N_\kappa,l}}^l,\ldots,\mathrm{S}_1^{N_\kappa}\right\}$. Note that each element of set X is also a set. Then we define a subset of X as X_j comprised only by subsets containing the jth SC. Equation (4) gives the cardinality of X_j as

$$|\mathrm{X}_j|=(\kappa+1)/2,\ |\mathrm{X}_j^{\,C}|=(\kappa-1)/2. \tag{9}$$

It is proven in Appendix C that optimizing Lagrangian dual function (7) with respect to only $\sigma_{w,j}$ can be simplified as maximizing the following function:

$$\Phi\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)=\frac{\left(1+\underline{\sigma}_{w,j}+A_j\right)^{(1-\lambda_\Sigma)}\prod\limits_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(1+\underline{\sigma}_{w,j}+B_{l,i}^j\right)^{\lambda_i^l}\left(\underline{\sigma}_{w,j}\right)^{\lambda_{\mathrm{C}_j}}}{1+\underline{\sigma}_{w,j}}, \tag{10}$$

where

$$\underline{\sigma}_{w,j}=\frac{\sigma_{w,j}}{\sigma_c},\lambda_\Sigma=\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_\kappa,l}}\lambda_i^l,\lambda_{\mathrm{C}_j}=\sum_{\mathrm{S}_i^l\in\mathrm{C}_j}\lambda_i^l,$$

and A_j and $B^j_{l,i}$ are given in Appendix C. Clearly, (10) is a continuous function when $\underline{\sigma}_{w,j}\geq 0$ and thus maximizing (10) can be solved by resorting to its derivatives in the following proposition.

***Proposition* 1**: The optimal $\underline{\sigma}_{w,j}{}^{*}$ that maximizes (10) is chosen from a set Δ_{sub} containing all the positive roots of a $(\kappa+1)/2$ degree polynomial $q(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa})$,

$$\underline{\sigma}_{w,j}{}^{*}=\underset{\underline{\sigma}_{w,j}\in\Delta_{sub}}{\arg\max}\,\Phi\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right), \tag{11}$$

where the expression of $q(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa})$ is given in Appendix D.

It is difficult to derive closed-form expressions for the roots of $q(\underline{\sigma}_{w,j},\lambda_{\kappa})$ but they can be easily calculated via some scientific software[1].

In one iteration, $\underline{\sigma}_{w,\vartheta}$ will be optimized one by one from $\underline{\sigma}_{w,1}$ to $\underline{\sigma}_{w,N\kappa}$ based on (11) until $\underline{\sigma}_{w,\vartheta}$ eventually converges to the optimal $\underline{\sigma}_{w,\vartheta}{}^{*}$. Then the dual minimization problem (8) can be solved by successively optimizing elements of $\boldsymbol{\lambda}^{\kappa}$. However, the searching domain of $\boldsymbol{\lambda}^{\kappa}$ is too large to be feasible. The following proposition reduces the size of the searching domain.

***Proposition* 2**: If $\forall\,\lambda_i^l\geq 1$ for $1\leq i\leq\kappa_{N\kappa,l}$ and $1\leq l\leq N_\kappa$, $\Phi(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa})\leq 1$ and the maximum is achieved only when $\underline{\sigma}_{w,i}$ approaches $+\infty$.

Proof: See Appendix E for proof. ■

If we choose $\boldsymbol{\lambda}^{\kappa}<\mathbf{1}^{\kappa}$, the maximum is achieved with finite $\underline{\sigma}_{w,i}$. According to (6), less compression noise means higher achievable rate. Hence, *Proposition* 2 actually defines the feasible range of each element of $\boldsymbol{\lambda}^{\kappa}$ as [0, 1). With this feasible range, the searching domain of $\boldsymbol{\lambda}^{\kappa}$ is greatly reduced to $\mathbf{0}^{\kappa}\leq\boldsymbol{\lambda}^{\kappa}<\mathbf{1}^{\kappa}$. Due to the convexity of $\varphi\left(\boldsymbol{\lambda}^{\kappa}\right)$, subgradient method can be used for solving minimization problem (8) [29],[31]. The searching direction of λ_i^l is given as

$$g_i^l\left(\sigma_{w,\mathrm{J}}{}^{*}\right)=\frac{\partial\varphi\left(\boldsymbol{\lambda}^{\kappa}\right)}{\partial\lambda_i^l}=C_i^l-f_i^l\left(\sigma_{w,\mathrm{J}}{}^{*}\right). \tag{12}$$

---

[1]Examples are function *roots* in Matlab and function *fsolve* in Maple, where some bisection methods, e.g., VCA method, are employed to compute the roots and the worst case complexity is the same as Sturm algorithm as identified in [32].

The searching criterion is: if $g_i^l\left(\sigma_{w,\mathrm{J}}{}^*\right)\leq 0$, increase λ_i^l; otherwise decrease λ_i^l. The overall algorithm is given as

Algorithm 1

Step 1: Initialize $\boldsymbol{\lambda}^{\kappa}{}_{\min}= \mathbf{0}^{\kappa}$ and $\boldsymbol{\lambda}^{\kappa}{}_{\max}= \mathbf{1}^{\kappa}$;

Step 2: Let $\boldsymbol{\lambda}^{\kappa}=(\boldsymbol{\lambda}^{\kappa}{}_{\min}+\boldsymbol{\lambda}^{\kappa}{}_{\max})/2$;

Step 3: Let t=1, initialize $\underline{\sigma}_{w,j}{}^{(t)}=+\infty$[2] from j=1 to N_{κ};

Step 4: From j=1 to N_{κ}, update $\underline{\sigma}_{w,j}{}^{(t+1)}$ based on (11);

Step 5: If $\sum_{j\in\mathrm{J}}\left|\sigma_{w,j}{}^{(t+1)}-\sigma_{w,j}{}^{(t)}\right|\leq\varepsilon_{\sigma}$, $\sigma_{w,\mathrm{J}}{}^*=\sigma_{w,\mathrm{J}}{}^{(t+1)}$ and go to next step; otherwise, t=t+1 and go to Step 4.

Step 6: For $1\leq i\leq\kappa_{N\kappa,l}$ and $1\leq l\leq N_{\kappa}$, if $g_i^l\left(\sigma_{w,\mathrm{J}}{}^*\right)\leq 0$, $\lambda_{i,\min}^l=\lambda_i^l$; otherwise $\lambda_{i,\max}^l=\lambda_i^l$;

Step 7: If $\sum_{l=1}^{N_{\kappa}}\sum_{i=1}^{\kappa_{N_{\kappa},l}}\left|\lambda_{i,\max}^l-\lambda_{i,\min}^l\right|\leq\varepsilon_{\lambda}$, stop the algorithm and $\boldsymbol{\lambda}^{\kappa}=(\boldsymbol{\lambda}^{\kappa}{}_{\min}+\boldsymbol{\lambda}^{\kappa}{}_{\max})/2$; otherwise, go back to Step 2.

IV. ACHIEVABLE RATE WITH JOINT DECOMPRESSION AND DECODING

The aforementioned constraints of (2) and (3) guarantee that the received signals compressed at the coordinating SCs can be correctly decompressed at the processing centre. The decoding of the message w is successive, starting from decompression and then followed by decoding of w using reconstructed signals. However, since the main objective of the processing centre is to decode the user message w rather than successful decompression, these constraints are actually unnecessarily imposed. Even if the decompressed information has errors and the observation reconstruction is not correct, it is still possible for the processing centre to correctly decode w, i.e., the reception of the compressed information can tolerate errors. In this regard, the decompression and decoding procedure at the processing centre should be modified accordingly. It is pointed out that rather than a

[2] Ideally, the initial value of $\underline{\sigma}_{wi}^2$ to start the iteration should be $+\infty$. Practically we choose a large enough value 10^{10} as the initial value. ε_{σ}, ε_{λ} and ε are very small values.

successive decoding process the decompression and decoding of message w can be conducted in a joint manner [33], [36]-[37].

When only the errors of the user message is taken into consideration with joint decompression and decoding, the achievable rate is given in [33] as

$$R = \min_{\mathrm{S} \subseteq \mathrm{J}} \left\{ \sum_{j \in \mathrm{S}} \left[C_j - I\left(\hat{Y}_{c,j}; Y_{c,j} \mid X_u \right) \right] + I\left(\hat{Y}_{c,\mathrm{S}^C}; X_u \right) \right\}. \tag{13}$$

With Gaussian distributed codebooks, it is proven in Appendix A that (13) can be written as the following expression:

$$R = \min_{\mathrm{S} \subseteq \mathrm{J}} \left\{ \sum_{j \in \mathrm{S}} \left[C_j - \log\left(1 + \frac{\sigma_c}{\sigma_{w,j}} \right) \right] + \log\left(1 + \sum_{j \in \mathrm{S}^C} \left(\frac{\gamma_j P_X}{\sigma_c + \sigma_{w,j}} \right) \right) \right\}. \tag{14}$$

Similar to (3), the achievable rate still depends on the variances of the compression noises but the difference is that the constraints are no longer imposed, i.e., (14) should be optimized as

$$R = \max_{\sigma_{w,\mathrm{J}}} \left\{ \min_{\mathrm{S} \subseteq \mathrm{J}} \left\{ \sum_{j \in \mathrm{S}} \left[C_j - \log\left(1 + \frac{\sigma_c}{\sigma_{w,j}} \right) \right] + \log\left(1 + \sum_{j \in \mathrm{S}^C} \left(\frac{\gamma_j P_X}{\sigma_c + \sigma_{w,j}} \right) \right) \right\} \right\}. \tag{15}$$

Notice that (15) is not in standard concave form with respect to $\underline{\sigma}_{w,\vartheta}$. However, by using an intermediate variable z_j defined as

$$z_j = \log\left(1 + \frac{\sigma_c}{\sigma_{w,j}} \right) = \log\left(1 + \frac{1}{\underline{\sigma}_{w,j}} \right), \tag{16}$$

(15) can be expressed as

$$R = \max_{z_{\mathrm{J}}} \left\{ \min_{\mathrm{S} \subseteq \mathrm{J}} \left\{ \sum_{j \in \mathrm{S}} \left[C_j - z_j \right] + \log\left(1 + \sum_{j \in \mathrm{S}^C} \left(\frac{\gamma_j P_X \left(1 - 2^{-z_j} \right)}{\sigma_c} \right) \right) \right\} \right\}. \tag{17}$$

The equation inside the minimization function is clearly concave. The minimum of two concave functions is also a concave function because the intersection of two concave sets is also concave. It follows that the equation inside the maximization function in (17) is a concave function and similar to the previous case, can be solved by using the subgradient method. The direction of the subgradient search is

$$g_j(z_j)=\begin{cases}-1, & \text{if } j\in \mathrm{S}_{\min} \\ \dfrac{\ln(2)\gamma_j P_X 2^{-z_j}}{\sigma_c\left(1+\sum\limits_{j\in \mathrm{S}^C}\left(\dfrac{\gamma_j P_X\left(1-2^{-z_j}\right)}{\sigma_c}\right)\right)}, & \text{if } j\notin \mathrm{S}_{\min}\end{cases}, \tag{18}$$

where $\Sigma_{\min}$ is the solution subset to (14) and can be obtained by exhaustive searching, i.e., comparing all subsets of ϑ at z_ϑ. The searching criterion is: if $g_j(z_j)\leq 0$, decrease z_j; otherwise increase z_j. When $\underline{\sigma}_{w,j}$ is in the region $(0,+\infty)$, according to (16), z_j is also in the range of $(0,+\infty)$. When the optimal $\underline{\sigma}_{w,j}^*$ is found using **Algorithm 1** proposed in the previous section, it can be used as the initial value for the optimization.

The overall optimization algorithm is given below.

Algorithm 2

Step 1: Using **Algorithm 1** to obtain $\underline{\sigma}_{w,\vartheta}^*$;

Step 2: Initialize $z_j= z_j(\underline{\sigma}_{w,j}^*)$ according to (20) for $1\leq j\leq N_\kappa$;

Step 3: Let k=1 and $z_\vartheta^{(1)}=(z_1(\underline{\sigma}_{w,j,\max}),\ldots, z_{Nc}(\underline{\sigma}_{w,N\kappa,\max}))$;

Step 4: Let j=1, t=1;

Step 5: Let $z_{j,\max}^{(t)}= \varepsilon_{z,max}$ and $z_{j,\min}^{(t)}= \varepsilon_{z,min}$[3];

Step 6: Let $z_j^{(t+1)} =(z_{j,\min}^{(t)}+z_{j,\max}^{(t)})/2$;

Step 7: Calculate $g_j(z_j^{(t+1)})$ according to (18). If $g_j(z_j^{(t+1)})\leq 0$, $z_{j,\max}^{(t+1)}=z_j^{(t+1)}$; otherwise $z_{j,\min}^{(t+1)}=z_j^{(t+1)}$.

Step 8: If $|z_j^{(t+1)}-z_j^{(t)}|>\varepsilon_z$, go to step 6. If $|z_j^{(t+1)}-z_j^{(t)}|\leq\varepsilon_z$ and $j=N_\kappa$, go to next step. If $|z_j^{(t+1)}-z_j^{(t)}|\leq \varepsilon_z$ and $j<N_\kappa$, $j=j+1$ and go to step 5;

Step 9: Let $z_\vartheta^{(k+1)}=(z_1^{(t+1)},\ldots, z_{N\kappa}^{(t+1)})$. If $|z_\vartheta^{(k+1)}- z_\vartheta^{(k)}| \leq\varepsilon_\vartheta$, stop; otherwise, go back to step 4.

V. SUM-RATE FOR THE MULTI-USER CASE

[3] Idealy, $\varepsilon_{z,max}$ should be $+\infty$ and $\varepsilon_{z,min}$ should be 0. Practically we choose a large enough value 10^{50} as the initial value of $\varepsilon_{z,max}$ and a small enough value 10^{-50} as the initial value of $\varepsilon_{z,min}$. ε_z and ε_ϑ are very small values

In previous sections, we investigated the single user scenario. In this section, we study the multi-user case to supplement the analysis. We consider N_u-user and N_κ-SC case and for simplicity we assume N_u=N_κ=2. The received signals at SCs are expressed as

$$\mathbf{y}_c = \mathbf{h}\mathbf{x}_u + \mathbf{n}, \tag{19}$$

where

$$\mathbf{y}_c = \begin{bmatrix} y_{c,1} \\ y_{c,2} \end{bmatrix}, \mathbf{h} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix}, \mathbf{x}_u = \begin{bmatrix} x_{u,1} \\ x_{u,2} \end{bmatrix}, \text{ and } \mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}.$$

According to [33], the achievable sum-rate can be expressed as

$$\max_{\sigma_{w,i} \geq 0} I\left(\mathbf{X}_u ; \hat{Y}_{c,\mathcal{J}}\right) \tag{20}$$

$$\text{s.t. } \forall \mathcal{S} \subseteq \mathcal{J} \;\; : I\left(Y_{\boldsymbol{\kappa},\mathcal{S}} ; \hat{Y}_{c,\mathcal{S}} \mid \hat{Y}_{c,\mathcal{S}^C}\right) \leq \sum_{j \in \mathcal{S}} C_j \,,$$

where ϑ={1,2}.

In [34] a successive refinement Wyner-Ziv code design has been proposed to achieve the entire Berger-Tung rate region for the remote Gaussian multi-terminal Central Estimating Officer (CEO) problem. On that code design, a multi-hypothesis assumption has been made for successively decoding the received data at SCs, i.e., information successively decoded from SCs is used as multiple side information for decoding data of the remaining SCs. In this regard, the achievable sum-rate can be rewritten as

$$\max_{\sigma_{w,i} \geq 0} I\left(\mathbf{X}_u ; \hat{Y}_{c,\mathcal{J}}\right) \tag{21}$$

$$\text{s.t. } I\left(Y_{\boldsymbol{\kappa},\mathcal{S}_{(j)}} ; \hat{Y}_{c,\mathcal{S}_{(j)}} \mid \hat{Y}_{c,\mathcal{S}_{\pi\{j-1\}}}\right) \leq C\,,$$

where $\Sigma_{(j)}$={π(j)}, $\Sigma_{\pi(j-1)}$={ π(1),…,π(j-1)} for j = 1, 2, and π is a permutation of the index of SCs. It is clear that resolving (21) requires exhaustive searching for all possible permutations of the SC index, i.e., a search of N_u! permutations. Eq. (21) can be rewritten, based on chain rule, as

$$I\left(\mathbf{X}_u;\hat{Y}_{c,\mathrm{J}}\right)=\sum_{j=1}^{N_s} I\left(\mathbf{X}_u;\hat{Y}_{c,\mathrm{S}_{(j)}} \mid \hat{Y}_{c,\mathrm{S}_{\pi\{j-1\}}}\right). \quad (22)$$

Then for N_κ=N_u=2 case, we propose to decouple the optimization problem (21) into two sub-problems as

Problem 1:

$$\max_{\sigma_{w,i}\geq 0} I\left(\mathbf{X}_u;\hat{Y}_{c,\pi(1)}\right) \quad (23)$$

$$\text{s.t. } I\left(Y_{c,\pi(1)};\hat{Y}_{c,\pi(1)}\right)\leq C,$$

and

Problem 2:

$$\max_{\sigma_{w,i}\geq 0} I\left(\mathbf{X}_u;\hat{Y}_{c,\pi(2)} \mid \hat{Y}_{c,\pi(1)}\right) \quad (24)$$

$$\text{s.t. } I\left(Y_{c,\pi(2)};\hat{Y}_{c,\pi(2)} \mid \hat{Y}_{c,\pi(1)}\right)\leq C,$$

Considering Gaussian inputs, expressions of (21) and (23)-(24) are given in Appendix A and the two optimization problems can also be solved by using Lagrangian method. Based on [34] and [35], the optimal solutions to (23) and (24) can be achieved as

$$\frac{1}{\sigma_{w,\pi(1)}}=\left[\frac{1}{\lambda}\left(\frac{1}{\sigma_r}-\frac{1}{R_{y_{\pi(1)}}}\right)-\frac{1}{\sigma_r}\right]^+, \quad (25)$$

and

$$\frac{1}{\sigma_{w,\pi(2)}}=\left[\frac{1}{\lambda}\left(\frac{1}{\sigma_r}-\frac{1}{R_{y_{\pi(2)}|\hat{y}_{\pi(1)}}}\right)-\frac{1}{\sigma_r}\right]^+, \quad (26)$$

respectively, where

$$R_{y_{\pi(1)}}=P_X\mathbf{h}_{\pi(1)}\mathbf{h}_{\pi(1)}^H+\sigma_r, \text{ and } R_{y_{\pi(2)}|\hat{y}_{\pi(1)}}=P_X\mathbf{h}_{\pi(2)}\left(\mathbf{I}+P_X\mathbf{I}\left(\frac{\mathbf{h}_{\pi(1)}{}^H\mathbf{h}_{\pi(1)}}{\sigma_{w,\pi(1)}\left(1+\sigma_r/\sigma_{w,\pi(1)}\right)}\right)\right)^{-1}\mathbf{h}_{\pi(2)}{}^H+\sigma_r.$$

Iterative algorithm similar to **Algorithm 1** can be used to achieve the optimal solutions and the direction of the subgradient search is given as

$$g_{\pi(1)}^{l}\left(\sigma_{w,\pi(1)}^{\;\;*}\right)=C-\log\left(1+R_{y_{\pi(1)}}/\sigma_{w,\pi(1)}\right), \tag{27}$$

and

$$g_{\pi(2)}^{l}\left(\sigma_{w,\pi(2)}^{\;\;*}\right)=C-\log\det\left(\mathbf{I}+R_{y_{\pi(2)}|\hat{y}_{\pi(1)}}\begin{bmatrix}\frac{1}{\sigma_{w,\pi(1)}} & 0\\ 0 & \frac{1}{\sigma_{w,\pi(2)}}\end{bmatrix}\right), \tag{28}$$

respectively.

VI. SIMULATION RESULTS AND DISCUSSIONS

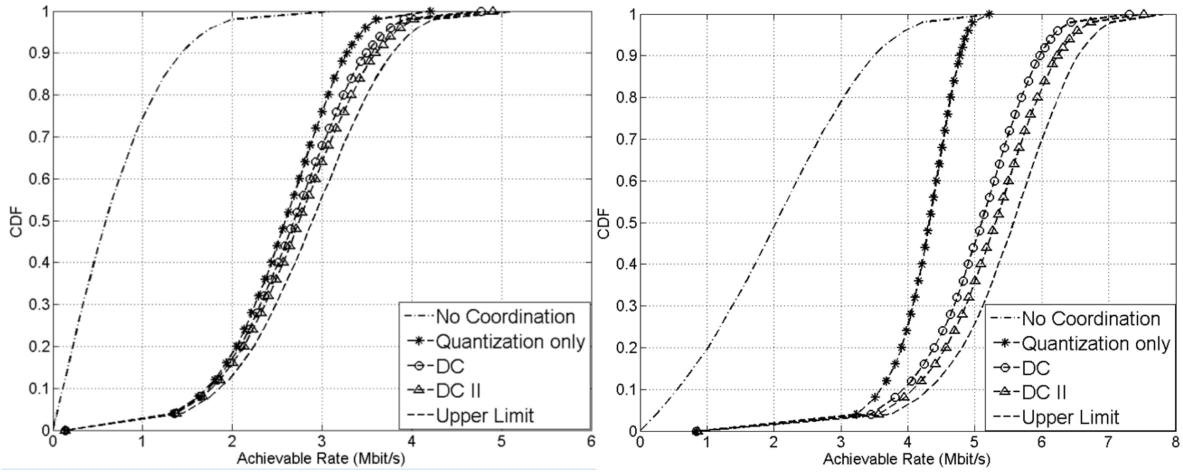
In this section, we present numerical results to evaluate the gain brought by the distributed compression schemes. Realistic assumptions are considered to provide insight into the practical performance of the proposed schemes. As shown in Fig. 1, the SCs are connected with the processing centre via capacity limited wireless fronthaul links. The cell edge users are associated with multiple SCs and a coordinating area is formed. All fronthaul links are assumed to have the same capacity C for simplicity. The carrier frequency f_c is 2GHz and the frequency reuse factor is 1, i.e., all SCs use the same frequency but the users within one coordinating area occupy different frequency bands based on certain scheduling mechanisms. Therefore, there is no intra-cell interference from the users within the same coordinating area but other users in adjacent cells outside the area generate inter-cell interference. We consider the worst case that there is one user occupying exactly the same frequency band in each interfering SC, transmitting with maximal power 23dBm. For simplicity, we assume that the interfering users are always located at the centre of the cells.

Two different scenarios are studied as follows:

1) Dense urban deployment: the radius of SCs is assumed to be 200 meters. Both the target user and the interfering users are assumed to be located outdoor and there are no obstacles between the SCs and the users and therefore the users do not suffer penetrating loss;

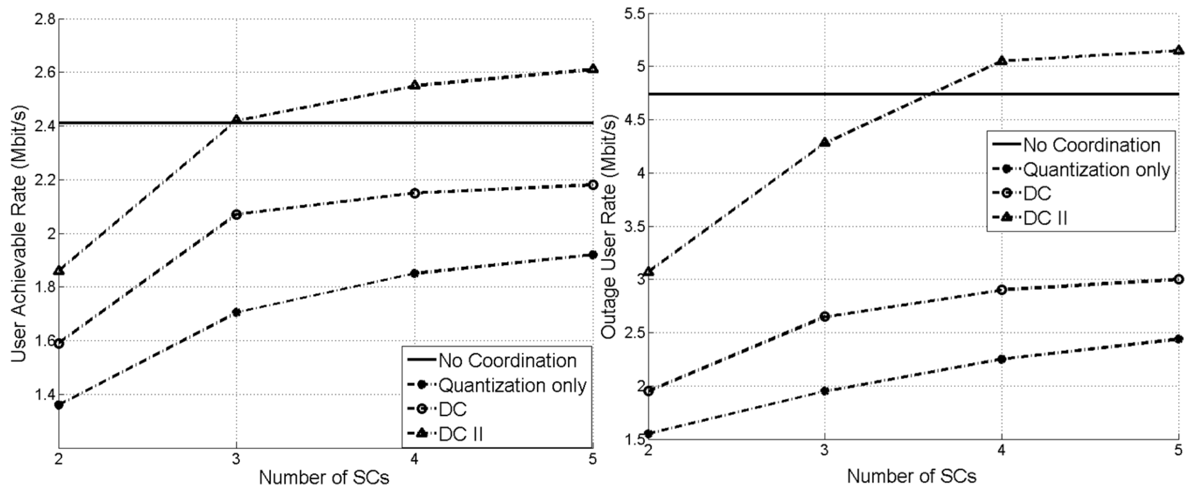
2) Ultra-dense urban deployment: the radius of SCs is only 20 meters and the target user is assumed to be located indoor but the interfering users are assumed to be outdoor. Therefore penetrating loss needs to be considered.

The parameters such as pathloss, penetrating loss, and antenna gain will be identified in the next paragraph.



(a) Dense deployment (b) Ultra-dense deployment

Fig. 3 CDF of the achievable user rate (N_κ=3, C=4Mbit/s)



(a) Dense deployment (b) Ultra-dense deployment

Fig. 4 Outage user rate (P_{out}=0.99, C=1Mbit/s)

The first scenario is a more general case and the second scenario is more applicable in some specific situations, where ultra-dense SCs are deployed to cover hotspot events like exhibitions. Path loss and Rayleigh fast fading are taken into account for both scenarios. For the first scenario, the path loss model is given as PL=22.7+36.7log$_{10}d$+26.0log$_{10}f_c$ in [38], where d is the distance in meter and f_c is in GHz. For the second scenario, the path loss model within the coordinating area is given as PL=37+30.0log$_{10}d$, and the path loss model outside the coordinating area is PL=15.3+37.6log$_{10}d$ +A_d, where A_d= 20dB is the penetrating loss [38]. We only consider the interference from the users in the first tier interfering cells. Both the users and the SCs are equipped with omni-directional antennas with 0dB antenna gain. The channel model is given as $h'=10^{-PL(d)/20}h$, where h ~ XN(0,1). We assume that the noise power density N_0=-171dBm/Hz and the bandwidth B=1MHz. The overall noise level should be the summation of thermal noise and interference.

Fig. 3 depicts the cumulative density function (cdf) of the achievable user rate in dense and ultra-dense deployment, respectively, with N_κ=3 SCs and fronthaul rate C=4 Mbit/s using:

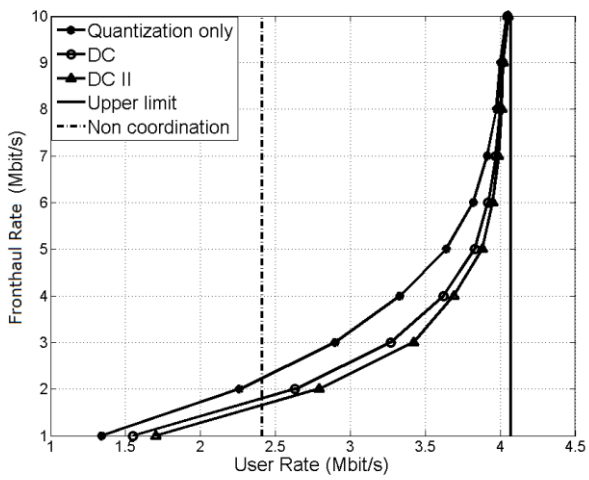
- Quantization only, where the SCs simply quantize the received signal and send the quantized information to the processing centre. The achievable rate for such a system can be easily extended from (2) with constraints $I\left(\hat{Y}_{\kappa,j};Y_{c,j}\right)\leq C_j$ for $\forall j\in\vartheta$.
- DC: all SCs conduct distributed compression and the decompression and decoding of message w is successive at the processing centre.
- DC II: all SCs conduct distributed compression but the decompression and decoding of message w are carried out in a joint manner at the processing centre.
- Upper Limit: the capacity of the fronthaul links are assumed to be infinity so that the compression noise approaches zero and the achievable rate approaches

$$R = \log\left(1 + \sum_{j \in \mathrm{J}} \frac{\gamma_j P_X}{\sigma_c}\right). \tag{29}$$

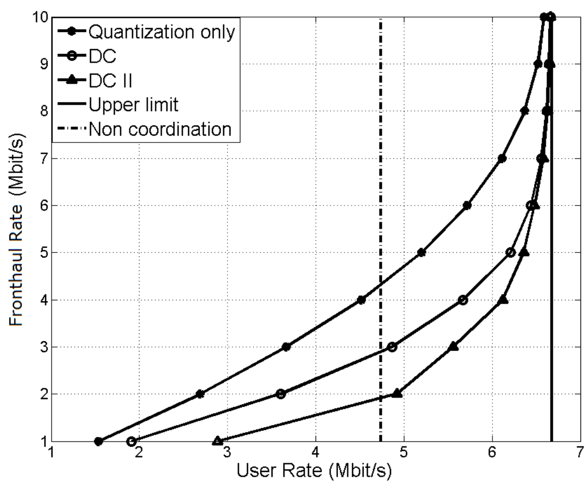
For the comparison purpose, no coordination case is also plotted, where the user is still assumed to be located at the cell edge. It is clearly shown that distributed compression outperforms the quantization only scheme and the performance can be further enhanced by applying joint decompression and decode. In the dense deployment scenario, the performance of all schemes is close to the upper limit. It is shown that at probability 0.9, the gain of distributed compression over quantization is only 0.22Mbit/s. On the contrary, in the ultra-dense deployment scenario, compared with the quantization noise or compression noise, the co-channel interference is relatively small because of the penetrating loss. Hence the quantization noise or compression noise dominates the performance and under such circumstances, the impact of reduced compression noise using optimized distributed compression introduces more gains according to (3). It is shown that the gap between quantization only and distributed compression is 1.2Mbit/s at probability 0.9.

Fig. 4 plots the outage user rate with a low fronthaul rate C=1Mbits/s for different number of SCs and the outage probability P_{out}=10^{-2}. Two important observations are obtained. Firstly, the gain of adding a new SC diminishes as the size of the coordinating network grows. It is shown in Fig. 4(a) that the joint operation achieves almost 0.6Mbit/s gain from 2 to 3 SCs but only less than 0.2Mbit/s gain from 3 to 4 SCs. Secondly, if the fronthaul link capacity is low it is evident that coordination does not always outperform non-coordination. Actually, only the joint operation can outperform non-coordination when coordinating SCs are increased to 3 and 4 for dense and ultra-dense deployment, respectively. This is because with low fronthaul rate the variances of the noise introduced by quantization or compression are so large that the potential coordinating gain cannot be harvested.

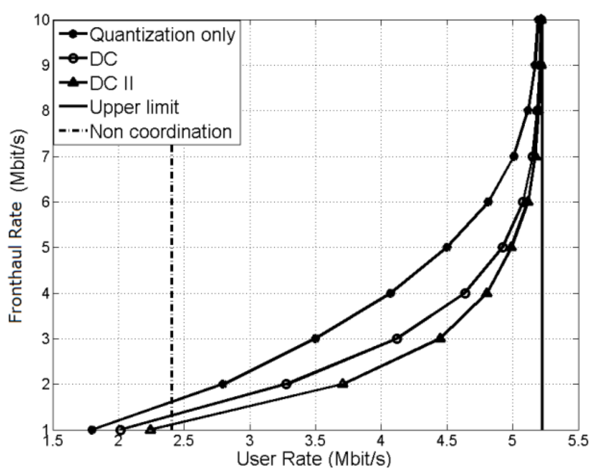
Fig. 5 illustrates the required fronthaul link rate C for a given outage user rate at P_{out}=10^{-2} for different N_κ in the dense deployment scenario. The two vertical lines are non-coordination and upper limit when C approaches infinity, respectively. The non-coordination lines intersect with the required fronthaul rate curves and only beyond these intersection points does coordination outperform non-coordination. It is clear shown that distributed compression effectively reduces the required fronthaul rate C to achieve a target user rate. For 2 SCs, at user rate=2Mbit/s the required C reduces 0.3Mbit/s from quantization to distributed compression and a further 0.14Mbit/s to joint decompression & decoding so that totally around 26% reduction is obtained. With a higher target user rate, the distributed compression and joint decompression & decoding schemes are more efficient than quantization. For instance, at R=3.3Mbit/s, the rate reduction is around 30% from

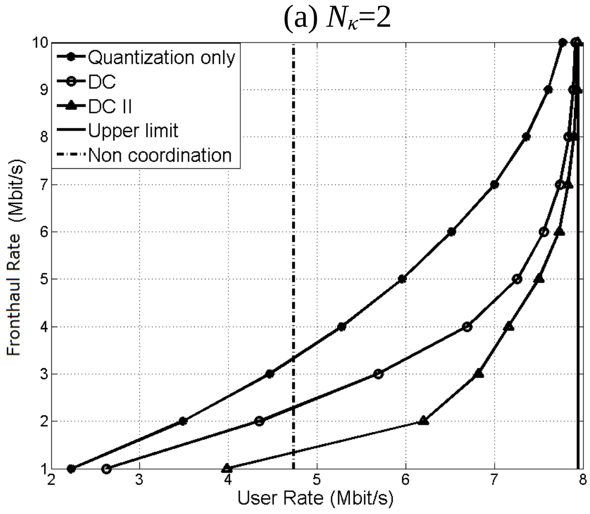


(a) N_κ=2



(a) N_κ=2





(b) N_κ=4

Fig. 5 Required C for dense deployment

(b) N_κ=4

Fig. 6 Required C for ultra-dense deployment

quantization to joint decompression & decoding. However, when the target user rate is further increased to approach the upper limit, the gain diminishes eventually and the required fronthaul rate C is approaching infinity. Fig. 5 also indicates that the performance gap between quantization and distributed compression becomes more remarkable with increased number of coordinating SCs. Fig. 6 shows the ultra-dense deployment indoor scenario. Because of the relatively weak interference, the user rate is more quantization or compression noise dominant and the gaps between three curves are more remarkable due to enhanced efficiency of the proposed optimization algorithms. The same trend as the outdoor dense deployment scenario is observed that the gains of distributed compression and joint

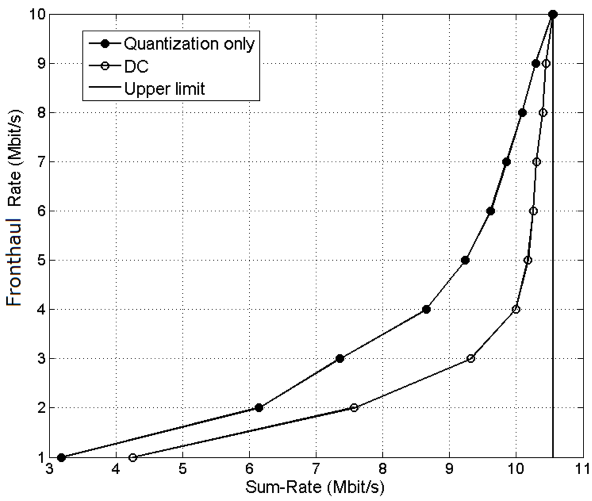


Fig. 7 Required C for ultra-dense deployment with multi-user (N_u=N_κ=2)

decompression & decoding increase with the target user rate firstly but eventually diminishes when approaching the upper limit.

The achievable sum-rate for multi-user case is illustrated in Fig.7 for ultra-dense scenario. Here we try all possible permutations of set ϑ = {1,2} to achieve the best sum-rate, namely,

π={1,2} and {2,1}, respectively. Compared with the quantization only scheme, distributed compression requires much less fronthaul rate for a give target sum-rate.

VII. CONCLUSIONS

In this paper, we propose to use distributed Wyner-Ziv compression to reduce the demanded fronthaul traffic to forward the received signals in uplink CoMP, tailored for future cellular network architectures such as C-RAN, which makes use of numerous wireless fronthaul links with capacity limit and thus demands efficient usage of the fronthaul network resources.

The distributed compression is wisely designed to maximize the user achievable rate with a given fronthaul rate using iterative algorithms and the analysis is extended from successive to joint decompression and decoding. Numerical results are generated for two typical scenarios: outdoor dense deployment and indoor ultra-dense deployment. Our results reveal that, in both scenarios, the fronthaul rate is essential to the coordinating gain obtained. With a low fronthaul link rate, introducing more coordinating cells merely achieves marginal improvement. The gain of coordination can only be harvested in the existence of medium to high fronthaul capacity.

More importantly, we compare the distributed compression with the conventional quantization only scheme and find that for a given target user rate (single user case), the required fronthaul link rate can be effectively reduced by distributed compression, and with joint operation of decompression and decoding further improvement can be obtained. The analysis is further extended to multi-user case where the required fronthaul rate for a target sum-rate is investigated and similarly, distributed compression significantly outperforms the conventional scheme. The achieved fronthaul traffic reduction becomes more significant with the increased target user rate firstly but eventually diminishes when approaching upper limit where the required fronthaul rate approaches infinity. Compared with the outdoor coordination case, distributed compression yields greater reduction in the indoor case where the compression noise dominates the performance.

APPENDIX A

The mutual information in equation (2) can be expressed as

$$I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right)=H\left(\hat{Y}_{c,\mathrm{J}}\right)-H\left(\hat{Y}_{c,\mathrm{J}}\mid X_u\right), \tag{A1}$$

$$\begin{aligned} I\left(\hat{Y}_{\mathbf{r},\mathrm{S}};Y_{c,\mathrm{S}}\mid\hat{Y}_{c,\mathrm{S}^C}\right)&=H\left(\hat{Y}_{c,\mathrm{S}}\mid\hat{Y}_{c,\mathrm{S}^C}\right)-H\left(\hat{Y}_{\mathbf{r},\mathrm{S}}\mid Y_{c,\mathrm{S}},\hat{Y}_{c,\mathrm{S}^C}\right)\\ &=H\left(\hat{Y}_{c,\mathrm{J}}\right)-H\left(\hat{Y}_{c,\mathrm{S}^C}\right)-H\left(\hat{Y}_{c,\mathrm{S}}\mid Y_{c,\mathrm{S}}\right)\end{aligned}. \tag{A2}$$

According to [18]-[20], the estimation $\hat{Y}_{c,j}$ can be expressed as $\hat{Y}_{c,j}=Y_{c,j}+W_j$, where W_j is a Gaussian variable independent of $Y_{c,j}$. Hence

$$H\left(\hat{Y}_{c,\mathrm{J}}\right)=\log\left(\prod_{j\in\mathrm{J}}\left(\sigma_c+\sigma_{w,j}\right)+\sum_{j\in\mathrm{J}}\left(\gamma_j P_X\prod_{i\in\mathrm{J},i\neq j}\left(\sigma_c+\sigma_{w,i}\right)\right)\right). \tag{A3}$$

$$H\left(\hat{Y}_{c,\mathrm{S}^C}\right)=\log\left(\prod_{j\in\mathrm{S}^C}\left(\sigma_c+\sigma_{w,j}\right)+\sum_{j\in\mathrm{S}^C}\left(\gamma_j P_X\prod_{i\in\mathrm{S}^C,i\neq j}\left(\sigma_c+\sigma_{w,i}\right)\right)\right), \tag{A4}$$

$$H\left(\hat{Y}_{c,\mathrm{J}}\mid X_u\right)=\log\left(\prod_{j\in\mathrm{J}}\left(\sigma_c+\sigma_{w,j}\right)\right), \tag{A5}$$

$$H\left(\hat{Y}_{\mathbf{r},\mathrm{S}}\mid Y_{c,\mathrm{S}}\right)=\log\left(\prod_{j\in\mathrm{S}}\sigma_{w,j}\right). \tag{A6}$$

Insert equations of (A3) and (A5) to (A1),

$$I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right)=\log\left(\frac{\prod_{j\in\mathrm{J}}\left(\sigma_c+\sigma_{w,j}\right)+\sum_{j\in\mathrm{J}}\left(\gamma_j P_X\prod_{i\in\mathrm{J},i\neq j}\left(\sigma_c+\sigma_{w,i}\right)\right)}{\prod_{j\in\mathrm{J}}\left(\sigma_c+\sigma_{w,j}\right)}\right)=\log\left(1+\sum_{j\in\mathrm{J}}\frac{\gamma_j P_X}{\sigma_c+\sigma_{w,j}}\right), \tag{A7}$$

Insert (A3), (A4) and (A6) into (A2),

$$
\begin{aligned}
&I\left(\hat{Y}_{\mathbf{r},\mathrm{S}};Y_{c,\mathrm{S}}\mid\hat{Y}_{c,\mathrm{S}^C}\right)\\
&=\log\left(\prod_{j\in\mathrm{J}}\left(\sigma_c+\sigma_{w,j}\right)+\sum_{j\in\mathrm{J}}\left(\gamma_j P_X\prod_{i\in\mathrm{J},i\neq j}\left(\sigma_c+\sigma_{w,i}\right)\right)\right)\\
&-\log\left(\prod_{j\in\mathrm{S}^C}\left(\sigma_c+\sigma_{w,j}\right)+\sum_{j\in\mathrm{S}^C}\left(\gamma_j P_X\prod_{i\in\mathrm{S}^C,i\neq j}\left(\sigma_c+\sigma_{w,i}\right)\right)\right)-\log\left(\prod_{j\in\mathrm{S}}\sigma_{w,j}\right)
\end{aligned}
,\qquad (A8)
$$

and the left side of the constraint equation is obtained.

The mutual information terms in (13) are expressed as

$$
\begin{aligned}
I\left(\hat{Y}_{c,j};Y_{c,j}\mid X_u\right)&=H\left(\hat{Y}_{c,j}\mid X_u\right)-H\left(\hat{Y}_{c,j}\mid Y_{c,j},X_u\right)\\
&=H\left(\hat{Y}_{c,j}\mid X_u\right)-H\left(\hat{Y}_{c,j}\mid Y_{c,j}\right)
\end{aligned}
,\qquad (A9)
$$

$$
I\left(\hat{Y}_{c,S^C};X_u\right)=H\left(\hat{Y}_{c,S^C}\right)-H\left(\hat{Y}_{c,S^C}\mid X_u\right).\qquad (A10)
$$

Based on (A4)-(A6), (A9) and (A10) can be rewritten as

$$
I\left(\hat{Y}_{c,j};Y_{c,j}\mid X_u\right)=\log\left(1+\frac{\sigma_c}{\sigma_{w,j}}\right),\qquad (A11)
$$

$$
I\left(\hat{Y}_{c,S^C};X_u\right)=\log\left(1+\sum_{j\in\mathrm{S}^C}\left(\frac{\gamma_j P_X}{\sigma_c+\sigma_{w,j}}\right)\right).\qquad (A12)
$$

Inserting (A11) and (A12) to (13) gives (14).

The mutual information expressions in (21), (23) and (24) are given as

$$
I\left(\mathbf{X}_u;\hat{Y}_{c,\mathrm{J}}\right)=\log\det\left(I+P_X\sum_{j=1}^{2}\frac{\mathbf{h}_j^H\mathbf{h}_j}{\sigma_{w,j}+\sigma_r}\right),\qquad (A13)
$$

$$
I\left(\mathbf{X}_u;\hat{Y}_{c,\pi(1)}\right)=\log\det\left(\mathbf{I}+P_X\frac{\mathbf{h}_{\pi(1)}^H\mathbf{h}_{\pi(1)}}{\sigma_{w,\pi(1)}+\sigma_r}\right),\qquad (A14)
$$

$$
I\left(\mathbf{X}_u;\hat{Y}_{c,\pi(2)}\right)=I\left(\mathbf{X}_u;\hat{Y}_{c,\mathrm{J}}\right)-I\left(\mathbf{X}_u;\hat{Y}_{c,\pi(1)}\right).\qquad (A15)
$$

APPENDIX B

Let $\mathbf{C}_X^{\kappa}$ and $\mathbf{C}_Y^{\kappa}$ be vectors of constraints. Considering another case where

$$
\mathbf{C}_Z^{\kappa}=v\mathbf{C}_X^{\kappa}+\left(1-v\right)\mathbf{C}_Y^{\kappa},\qquad (B1)
$$

with $0\le v\le 1$. Let R_X and R_Y be the optimal solution to the primal problem with constraints $\mathbf{C}_X^{\kappa}$ and $\mathbf{C}_Y^{\kappa}$, respectively. We assume that codebooks $\Xi_{X,j}$ and $\Xi_{Y,j}$ with coding rate $R_{X,j}\le C_{X,j}$

and $R_{Y,j} \le C_{Y,j}$ for $\forall j \in \vartheta$, respectively, are used in the fronthaul link between SC_j and the processing point. Considering the idea of time-sharing, we also assume a case Z that at SC_j codebook $\Xi_{Z,j}$ is constructed by using the first vn symbols of the first codebook $\Xi_{X,j}$ and the last $(1-v)n$ symbols of the second code book $\Xi_{Y,j}$ with constraint $C_{Z,j}=vC_{X,j}+(1-v)C_{X,j}$ for some $0 \le v \le 1$. The rate of this new codebook is

$$R_{Z,j}=vR_{X,j}+(1-v)R_{Y,j} \le vC_{X,j}+(1-v)C_{Y,j}=C_{Z,j}, \tag{B2}$$

and the constraint is satisfied for case Z. Clearly, with the new code book, the rate can be achieved up to

$$R_Z=vR_X+(1-v)R_Y. \tag{B3}$$

It is pointed out in [39] that time-sharing cannot decrease the compression noise. Since the compression noises are at the denominators of (3), it also means that (3) cannot be increased by time-sharing. Let R_Z^* be the optimal solution to constraint $\mathbf{C}_Z^{\kappa}$. Then we have

$$R_Z^* \ge R_Z = vR_X + \left(1-v\right)R_Y. \tag{B4}$$

In addition, if we increase the constraint $\mathbf{C}^{\kappa}$, the distortion, in nature, should be decreased and hence (3) can be increased. Therefore, it can be concluded that $R(\sigma_{w,\vartheta})$ is a non-decreasing concave function with regard to the constraint $\mathbf{C}^{\kappa}$ and thus satisfies the time sharing property. According to [40], with time sharing property, the duality gap is zero.

APPENDIX C

The Lagarangian function (6) can be expressed as

$$\begin{aligned} L\left(\sigma_{w,\mathrm{J}},\boldsymbol{\lambda}^{\kappa}\right) = & \left(1-\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_{\kappa},l}}\lambda_i^l\right)H\left(\hat{Y}_{c,\mathrm{J}}\right)-H\left(\hat{Y}_{c,\mathrm{J}} \mid X_u\right)+\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_{\kappa},l}}\lambda_i^l H\left(\hat{Y}_{c,\mathrm{S}_i^{lC}}\right) \\ & +\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_{\kappa},l}}\lambda_i^l H\left(\hat{Y}_{\mathbf{r},\mathrm{S}_i^l} \mid Y_{c,\mathrm{S}_i^l}\right)+\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_{\kappa},l}}\lambda_i^l C_i^l \end{aligned}. \tag{C1}$$

If we fix $\sigma_{w,\vartheta/j}$, maximizing $L(\sigma_{w,j},\boldsymbol{\lambda}^c)$ is equivalent to maximizing

$$\begin{aligned} & \left(1-\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_{\kappa},l}}\lambda_i^l\right)\log\left(\prod_{k\in\mathrm{J}}\left(\sigma_c+\sigma_{w,k}\right)+\gamma_j P_X\prod_{k\in\mathrm{J},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)+\sum_{l\in\mathrm{J},l\neq j}\left(\gamma_l P_X\prod_{k\in\mathrm{J},k\neq l}\left(\sigma_c+\sigma_{w,k}\right)\right)\right)-\log\left(\sigma_c+\sigma_{w,j}\right) \\ & +\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_{\kappa},l}}\lambda_i^l\log\left(\prod_{k\in\mathrm{S}_i^{lC}}\left(\sigma_c+\sigma_{w,k}\right)+\sum_{l\in\mathrm{S}_i^{lC}}\left(\gamma_l P_X\prod_{k\in\mathrm{S}_i^{lC},k\neq l}\left(\sigma_c+\sigma_{w,k}\right)\right)\right)+\sum_{l\in\mathrm{L}}\sum_{i=1}^{\kappa_{N_{\kappa},l}}\lambda_i^l\log\left(\prod_{k\in\mathrm{S}_i^l}\sigma_{w,k}\right) \end{aligned}$$

(C2).

Since X_j^C does not contain element j, maximizing the last term of (C2) is equivalent to maximizing $\sum_{\mathrm{S}_i^l \in \mathrm{C}_j} \lambda_i^l \log\left(\sigma_{w,j}\right)$ with fixed $\sigma_{w,\vartheta/j}$. The first and third terms of (C2) can be re-organized as (C3) and (C4), respectively.

$$\left(1-\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_\kappa,l}}\lambda_i^l\right)\log\left(\sigma_c+\sigma_{w,j}+\frac{\gamma_j P_X \prod_{k\in\mathrm{J},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)}{\prod_{k\in\mathrm{J},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)+\sum_{l\in\mathrm{J},l\neq j}\left(\gamma_l P_X\prod_{k\in\mathrm{J},k\neq j,k\neq l}\left(\sigma_c+\sigma_{w,k}\right)\right)}\right)$$
$$+\left(1-\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_\kappa,l}}\lambda_i^l\right)\log\left(\prod_{k\in\mathrm{J},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)+\sum_{l\in\mathrm{J},l\neq j}\left(\gamma_l P_X\prod_{k\in\mathrm{J},k\neq j,k\neq l}\left(\sigma_c+\sigma_{w,k}\right)\right)\right), \tag{C3}$$

$$\sum_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\lambda_i^l\log\left(\sigma_c+\sigma_{w,j}+\frac{\gamma_j P_X\prod_{k\in\mathrm{S}_i^{lC},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)}{\prod_{k\in\mathrm{S}_i^{lC},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)+\sum_{l\in\mathrm{S}_i^{lC},l\neq j}\left(\gamma_j P_X\prod_{k\in\mathrm{S}_i^{lC},k\neq j,k\neq l}\left(\sigma_c+\sigma_{w,k}\right)\right)}\right)$$
$$+\sum_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\lambda_i^l\log\left(\prod_{k\in\mathrm{S}_i^{lC},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)+\sum_{l\in\mathrm{S}_i^{lC},l\neq j}\left(\gamma_j P_X\prod_{k\in\mathrm{S}_i^{lC},k\neq j,k\neq l}\left(\sigma_c+\sigma_{w,k}\right)\right)\right). \tag{C4}$$

Clearly, the second terms of both (C3) and (C4) do not contain $\sigma_{w,j}$, therefore they can be ignored when maximizing (C1) with respect to $\sigma_{w,j}$. Let

$$A_j=\frac{\gamma_j P_X\prod_{k\in\mathrm{J},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)}{\sigma_c\left(\prod_{k\in\mathrm{J},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)+\sum_{l\in\mathrm{J},l\neq j}\left(\gamma_l P_X\prod_{k\in\mathrm{J},k\neq j,k\neq l}\left(\sigma_c+\sigma_{w,k}\right)\right)\right)},$$
$$B_{l,i}^j=\frac{\gamma_j P_X\prod_{k\in\mathrm{S}_i^{lC},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)}{\sigma_c\left(\prod_{k\in\mathrm{S}_i^{lC},k\neq j}\left(\sigma_c+\sigma_{w,k}\right)+\sum_{l\in\mathrm{S}_i^{lC},l\neq j}\left(\gamma_j P_X\prod_{k\in\mathrm{S}_i^{lC},k\neq j,k\neq l}\left(\sigma_c+\sigma_{w,k}\right)\right)\right)}, \tag{C5}$$

for $j\in\mathrm{J}$. Based on (C3)-(C5), maximizing (C2) is equivalent to maximizing

$$\left(1-\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_\kappa,l}}\lambda_i^l\right)\log\left(\sigma_c+\sigma_{w,j}+\sigma_c A_j\right)-\log\left(\sigma_c+\sigma_{w,j}\right)+\sum_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\lambda_i^l\log\left(\sigma_c+\sigma_{w,j}+\sigma_c B_{l,i}^j\right)+\sum_{\mathrm{S}_i^l\in\mathrm{C}_j}\lambda_i^l\log\left(\sigma_{w,j}\right). \tag{C6}$$

(C6) can be normalized by σ_c to give

$$L\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{c}\right)$$
$$=\max_{\underline{\sigma}_{w,j}>0}\left\{\left(1-\lambda_{\Sigma}\right)\log\left(1+\underline{\sigma}_{w,j}+A_{j}\right)+\sum_{\mathrm{S}_i^l\in \mathrm{C}_j^C}\lambda_i^l\log\left(1+\underline{\sigma}_{w,j}+B_{l,i}^{j}\right)+\lambda_{\mathrm{C}_j}\log\left(\underline{\sigma}_{w,j}\right)-\log\left(1+\underline{\sigma}_{w,j}\right)\right\}, \quad \text{(C7)}$$

where

$$\underline{\sigma}_{w,j}=\frac{\sigma_{w,j}}{\sigma_c},\lambda_{\Sigma}=\sum_{l\in\mathrm{J}}\sum_{i=1}^{\kappa_{N_{K},l}}\lambda_i^l,\lambda_{\mathrm{C}_j}=\sum_{\mathrm{S}_i^l\in\mathrm{C}_j}\lambda_i^l .$$

Considering that the logarithm function is monotonically increasing, (C7) can be further simplified as maximizing

$$\Phi\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{c}\right)=\frac{\left(1+\underline{\sigma}_{w,j}+A_{j}\right)^{\left(1-\lambda_{\Sigma}\right)}\prod_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(1+\underline{\sigma}_{w,j}+B_{l,i}^{j}\right)^{\lambda_i^l}\left(\underline{\sigma}_{w,j}\right)^{\lambda_{\mathrm{C}_j}}}{1+\underline{\sigma}_{w,j}}, \quad \text{(C8)}$$

subject to $\underline{\sigma}_{wi}^{2}\geq 0$.

APPENDIX D

The derivative of (10) is given as

$$\frac{\partial\Phi\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}{\partial\underline{\sigma}_{w,j}}=\frac{\left(1+\underline{\sigma}_{w,j}\right)y'\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)-y\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}{\left(1+\underline{\sigma}_{w,j}\right)^{2}}, \quad \text{(D1)}$$

where

$$y\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)=\left(1+\underline{\sigma}_{w,j}+A_{j}\right)^{\left(1-\lambda_{\Sigma}\right)}\prod_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(1+\underline{\sigma}_{w,j}+B_{l,i}^{j}\right)^{\lambda_i^l}\left(\underline{\sigma}_{w,j}\right)^{\lambda_{\mathrm{C}_j}},$$
$$\begin{aligned}y'\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)=&\left(1-\lambda_{\Sigma}\right)\left(1+\underline{\sigma}_{w,j}+A_{j}\right)^{-\lambda_{\Sigma}}\prod_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(1+\underline{\sigma}_{w,j}+B_{l,i}^{j}\right)^{\lambda_i^l}\left(\underline{\sigma}_{w,j}\right)^{\lambda_{\mathrm{C}_j}}\\&+\left(1+\underline{\sigma}_{w,j}+A_{j}\right)^{\left(1-\lambda_{\Sigma}\right)}\sum_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(\lambda_i^l\left(1+\underline{\sigma}_{w,j}+B_{l,i}^{j}\right)^{\lambda_i^l-1}\prod_{\mathrm{S}_k^l\in\mathrm{C}_j^C,k\neq i}\left(1+\underline{\sigma}_{w,j}+B_{l,k}^{j}\right)^{\lambda_k^l}\right)\left(\underline{\sigma}_{w,j}\right)^{\lambda_{\mathrm{C}_j}}\\&+\lambda_{\mathrm{C}_j}\left(1+\underline{\sigma}_{w,j}+A_{j}\right)^{\left(1-\lambda_{\Sigma}\right)}\prod_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(1+\underline{\sigma}_{w,j}+B_{l,i}^{j}\right)^{\lambda_i^l}\left(\underline{\sigma}_{w,j}\right)^{\lambda_{\mathrm{C}_j}-1}\end{aligned} .$$

We rewrite (D1) as

$$\frac{\partial\Phi\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}{\partial\underline{\sigma}_{w,j}}=\frac{\hat{y}\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)\left(\left(1+\underline{\sigma}_{w,j}\right)\frac{y'\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}{\hat{y}\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}-\frac{y\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}{\hat{y}\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}\right)}{\left(1+\underline{\sigma}_{w,j}\right)^{2}}, \quad \text{(D2)}$$

where

$$\hat{y}\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)=\left(1+\underline{\sigma}_{w,j}+A_j\right)^{-\lambda_\Sigma}\prod_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(1+\underline{\sigma}_{w,j}+B_{l,i}^j\right)^{\lambda_i^l-1}\left(\underline{\sigma}_{w,j}\right)^{\lambda_{\mathrm{C}_j}-1}>0\,.$$

Thus letting (D1) be 0 is equivalent to

$$q\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)=\left(1+\underline{\sigma}_{w,j}\right)\frac{y'\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}{\hat{y}\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}-\frac{y\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}{\hat{y}\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)}=0\,. \tag{D3}$$

Combine (D1) and (D2), (D3) can be given as

$$\begin{aligned}&q\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)\\&=\left(1+\underline{\sigma}_{w,j}\right)\left(\begin{aligned}&\left(1-\lambda_\Sigma\right)\underline{\sigma}_{w,j}\prod_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(1+\underline{\sigma}_{w,j}+B_{l,i}^j\right)+\lambda_{\mathrm{C}_j}\left(1+\underline{\sigma}_{w,j}+A_j\right)\prod_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(1+\underline{\sigma}_{w,j}+B_{l,i}^j\right)\\&+\underline{\sigma}_{w,j}\left(1+\underline{\sigma}_{w,j}+A_j\right)\sum_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(\lambda_i^l\prod_{\mathrm{S}_k^l\in\mathrm{C}_j^C,k\neq i}\left(1+\underline{\sigma}_{w,j}+B_{l,k}^j\right)\right)\end{aligned}\right)\\&-\underline{\sigma}_{w,j}\left(1+\underline{\sigma}_{w,j}+A_j\right)\prod_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\left(1+\underline{\sigma}_{w,j}+B_{l,i}^j\right)\end{aligned}\,. \tag{D4}$$

Clearly, $q(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa})$ is a polynomial function of $\underline{\sigma}_{w,j}$ and the term with the largest degree is

$$\left(1-\lambda_\Sigma\right)\left(\underline{\sigma}_{w,j}\right)^{\left|\mathrm{C}_j^C\right|+2}+\sum_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\lambda_i^l\left(\underline{\sigma}_{w,j}\right)^{\left|\mathrm{C}_j^C\right|+2}+\lambda_{\mathrm{C}_j}\left(\underline{\sigma}_{w,j}\right)^{\left|\mathrm{C}_j^C\right|+2}-\left(\underline{\sigma}_{w,j}\right)^{\left|\mathrm{C}_j^C\right|+2}=0\,. \tag{D5}$$

Therefore, $q(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa})$ is actually a $(\kappa+1)/2$ degree polynomial function of $\underline{\sigma}_{w,j}$ having $(\kappa+1)/2$ roots, denoted as a set $\Delta=\{\sigma_1,\ldots,\sigma_{(\kappa+1)/2}\}$, by letting $q(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa})=0$. However, not all members of root set Δ are viable solutions. Since $\underline{\sigma}_{w,j}>0$, only positive roots should be considered. We define $\underline{\Delta}_{\mathrm{sub}}$ as a subset of Δ including only positive elements.

Considering a special case that $\underline{\sigma}_{w,j}\rightarrow+\infty$, we have

$$\begin{aligned}&\lim_{\underline{\sigma}_{w,j}\rightarrow+\infty}q\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right)\\&=\underline{\sigma}_{w,j}\left(\left(1-\lambda_\Sigma\right)\left(\underline{\sigma}_{w,j}\right)^{\left|\mathrm{C}_j^C\right|+1}+\lambda_{\mathrm{C}_j}\left(\underline{\sigma}_{w,j}\right)^{\left|\mathrm{C}_j^C\right|+1}+\left(\sum_{\mathrm{S}_i^l\in\mathrm{C}_j^C}\lambda_i^l\right)\left(\underline{\sigma}_{w,j}\right)^{\left|\mathrm{C}_j^C\right|+1}\right)-\underline{\sigma}_{w,j}^{\left|\mathrm{C}_j^C\right|+2}\\&=0\end{aligned}\,, \tag{D6}$$

which means $+\infty$ is also a local maximum. Therefore, a new set is defined as $\Delta_{\mathrm{sub}}=\{\underline{\Delta}_{\mathrm{sub}},+\infty\}$.

Note that the members of Δ_{sub} only guarantee that the first derivative of $\Phi(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa})$ is zero at those points thus can be regarded as local maximum or minimum. In order to make sure that only maximum is found, the optimal $\underline{\sigma}_{w,j}{}^{*}$ should be chosen as

$$\underline{\sigma}_{w,j}{}^{*} = \underset{\underline{\sigma}_{w,j} \in \Delta_{sub}}{\arg\max}\, \Phi\left(\underline{\sigma}_{w,j}, \boldsymbol{\lambda}^{\kappa}\right). \tag{D7}$$

APPENDIX E

With given $\boldsymbol{\lambda}^{c}$, maximizing the Lagrangian dual (6) is equivalent to maximizing

$$I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right) - \sum_{l\in \mathrm{J}} \sum_{i=1}^{\kappa_{N_\kappa,l}} \lambda_i^l I\left(\hat{Y}_{\mathbf{r},\mathrm{S}_i^l};Y_{c,\mathrm{S}_i^l} \mid \hat{Y}_{c,\mathrm{S}_i^{lC}}\right), \tag{E1}$$

where $\lambda_i^l \geq 0$. For a particular λ_j^l we have

$$I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right) - \sum_{l\in \mathrm{J}} \sum_{i=1}^{\kappa_{N_\kappa,l}} \lambda_i^l I\left(\hat{Y}_{\mathbf{r},\mathrm{S}_i^l};Y_{c,\mathrm{S}_i^l} \mid \hat{Y}_{c,\mathrm{S}_i^{lC}}\right) \leq I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right) - \lambda_j^l I\left(\hat{Y}_{\mathbf{r},\mathrm{S}_j^l};Y_{c,\mathrm{S}_j^l} \mid \hat{Y}_{c,\mathrm{S}_j^{lC}}\right). \tag{E2}$$

If $\lambda_j^l \geq 1$, we have

$$I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right) - \sum_{l\in \mathrm{J}} \sum_{i=1}^{\kappa_{N_\kappa,l}} \lambda_i^l I\left(\hat{Y}_{\mathbf{r},\mathrm{S}_i^l};Y_{c,\mathrm{S}_i^l} \mid \hat{Y}_{c,\mathrm{S}_i^{lC}}\right) \leq I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right) - I\left(\hat{Y}_{\mathbf{r},\mathrm{S}_j^l};Y_{c,\mathrm{S}_j^l} \mid \hat{Y}_{c,\mathrm{S}_j^{lC}}\right). \tag{E3}$$

The right term can be expressed as

$$\begin{aligned}
I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right) - I\left(\hat{Y}_{\mathbf{r},\mathrm{S}_j^l};Y_{c,\mathrm{S}_j^l} \mid \hat{Y}_{c,\mathrm{S}_j^{lC}}\right) &= H\left(\hat{Y}_{c,\mathrm{J}}\right) - H\left(\hat{Y}_{c,\mathrm{J}} \mid X_u\right) - \left(H\left(\hat{Y}_{c,\mathrm{J}}\right) - H\left(\hat{Y}_{c,\mathrm{S}_j^{lC}}\right) - H\left(\hat{Y}_{c,\mathrm{S}_j^l} \mid Y_{c,\mathrm{S}_j^l}\right)\right) \\
&= H\left(\hat{Y}_{c,\mathrm{S}_j^{lC}}\right) + H\left(\hat{Y}_{c,\mathrm{S}_j^l} \mid Y_{c,\mathrm{S}_j^l}\right) - H\left(\hat{Y}_{c,\mathrm{J}} \mid X_u\right) \\
&= \log\left(\prod_{j\in \mathrm{S}_i^{lC}} \left(\sigma_c + \sigma_{w,j}\right) + \sum_{j\in \mathrm{S}_i^{lC}} \left(\gamma_j P_X \prod_{k\in \mathrm{S}_i^{lC}, k\neq j} \left(\sigma_c + \sigma_{w,k}\right)\right)\right) + \log\left(\prod_{j\in \mathrm{S}_i^l} \sigma_{w,j}\right) \\
&\quad - \log\left(\sigma_c + \sigma_{w,j}\right)
\end{aligned}. \tag{E4}$$

If we maximize (E4) with respect to $\underline{\sigma}_{w,j}$ only, following the similar derivation as Appendix C, it is given as

$$\begin{aligned}
I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right) - I\left(\hat{Y}_{\mathbf{r},\mathrm{S}_j^l};Y_{c,\mathrm{S}_j^l} \mid \hat{Y}_{c,\mathrm{S}_j^{lC}}\right) &= \log\left(1+\underline{\sigma}_{w,j} + B_{l,i}^j\right) + \log\left(\underline{\sigma}_{w,j}\right) - \log\left(1+\underline{\sigma}_{w,j}\right) \\
&= \log\left(\frac{\left(1+\underline{\sigma}_{w,j} + B_{l,i}^j\right)\left(\underline{\sigma}_{w,j}\right)}{1+\underline{\sigma}_{w,j}}\right) = \log u\left(\underline{\sigma}_{w,j}\right)
\end{aligned}. \tag{E5}$$

The first derivative of $u(\underline{\sigma}_{w,j})$ is given as

$$\frac{\partial u\left(\underline{\sigma}_{w,j}\right)}{\partial \underline{\sigma}_{w,j}} = \frac{\left(1+\underline{\sigma}_{w,j}\right)^2 + B_{l,i}^j}{\left(1+\underline{\sigma}_{w,j}\right)^2}. \tag{E6}$$

Since $\underline{\sigma}_{w,j} \geq 0$ and $B_{l,i}^j > 0$, it follows that $\frac{\partial u\left(\underline{\sigma}_{w,j}\right)}{\partial \underline{\sigma}_{w,j}} > 0$, which means that $u(\underline{\sigma}_{w,j})$ is a monotonically increasing function of $\underline{\sigma}_{w,j}$ and its maximum is achieved only when $\underline{\sigma}_{w,j} \to +\infty$ as

$$\max_{\underline{\sigma}_{w,j}} u\left(\underline{\sigma}_{w,j}\right) = \lim_{\underline{\sigma}_{w,j}\to+\infty} \frac{\left(1+\underline{\sigma}_{w,j}+A_i\right)^{1-\lambda_j^l} \left(\underline{\sigma}_{w,j}\right)^{\lambda_j^l}}{1+\underline{\sigma}_{w,j}} = 1. \tag{E7}$$

Hence

$$I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right) - \sum_{l\in\mathrm{J}} \sum_{i=1}^{\kappa N_{\kappa,l}} \lambda_i^l I\left(\hat{Y}_{\mathbf{r},\mathrm{S}_i^l};Y_{c,\mathrm{S}_i^l} \mid \hat{Y}_{c,\mathrm{S}_i^{lC}}\right) \leq I\left(X_u;\hat{Y}_{c,\mathrm{J}}\right) - I\left(\hat{Y}_{\mathbf{r},\mathrm{S}_j^l};Y_{c,\mathrm{S}_j^l} \mid \hat{Y}_{c,\mathrm{S}_j^{lC}}\right) \leq \log 1 = 0. \tag{E8}$$

In the meantime,

$$\Phi\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right) = \frac{\left(1+\underline{\sigma}_{w,j}+A_j\right)^{(1-\lambda_\Sigma)} \prod_{\mathrm{S}_i^l\in\mathrm{C}_j^C} \left(1+\underline{\sigma}_{w,j}+B_{l,i}^j\right)^{\lambda_i^l} \left(\underline{\sigma}_{w,j}\right)^{\lambda_{\mathrm{C}_j}}}{1+\underline{\sigma}_{w,j}}, \tag{E9}$$

which implies that the maximum value 0 can be achieved at $+\infty$. Based on (E8) and (E9), we can conclude that

$$\arg\max_{\underline{\sigma}_{w,j}} \Phi\left(\underline{\sigma}_{w,j},\boldsymbol{\lambda}^{\kappa}\right) = +\infty, \quad \forall\ \lambda_j^l \geq 1. \tag{E10}$$

ACKNOWLEDGEMENT

The research leading to these results has received funding from the NPRP project. The authors would like to acknowledge the contributions of their colleagues in the NPRP project, although the views expressed are those of the authors and do not necessarily represent the projects.

REFERENCES

[1] NSN white paper, 2020: Beyond 4G Radio Evolution for the Gigabit Experience, 2011.

[2] Ericsson white paper, *5G Radio Access*, Jun. 2013.

[3] O. Stanze and A. Weber, "Heterogeneous networks with LTE-Advanced technologies," *Bell Labs Technical Journal*, vol. 18, pp. 41-58, Jun. 2013.
[4] F. R. Yu, X. Zhang, and V. Leung, *Green Communications and Networking*, CRC Press, 2012.
[5] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE Journal on Selected Areas in Communications* , vol. 28, no. 9, pp. 1479-1489, Dec. 2010.
[6] P. Marsch and G. Fettweis, *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*, Cambridge University Press., Jun. 2011.
[7] A. Davydov, G. Morozov, I. Bolotin, and A. Parathanassiou, "Evaluation of Joint Transmission CoMP in C-RAN based LTE-A HetNets with large coordination areas," in *Proc. of IEEE Globecom workshop*, pp. 801-806, Atlanta, Dec. 2013.
[8] L. Li, J. Liu, K. Xiong, and P. Butovitsch, "Field test of uplink CoMP joint processing with C-RAN testbed," in *Proc. of the 7th International ICST Conference on Communications and Networking in China*, pp.753-757, KunMing, Aug. 2012.
[9] China Mobile Research institute, *C-RAN White paper V2.5*, Oct 2011.
[10] China Mobile Research Institute, "C-RAN - Road Towards Green Radio Access Network," in *Proc. of Wireless World Research Forum, Meeting 27 (WWRF 27)*, pp. 73–80, Düsseldorf, 2011.
[11] P. Rost, C. J. Bernardos, A. De Domenico, M. Di Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wübben, "Cloud Technologies for Flexible 5G Radio Access Networks", accepted by *IEEE Communications Magazine*.
[12] K. Sakaguchi, S. Sampei, H. Shimodaira, and R. Rezagah, "Cloud cooperated heterogeneous cellular networks", in *Proc. of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 787-791, Naha, Nov. 2013.
[13] Z. Ghebretensae, K. Laraqui, S. Dahlfort, and J. Chen, "Transmission solutions and architectures for heterogeneous network built as C-RANs," in *Proc. of the 7th International ICST Conference on Communications and Networking in China*, pp.748-752, KunMing, Aug. 2012.
[14] P. Marsch, "Coordinated multi-point under a constrained fronthaul and imperfect channel knowledge," Vogt Verlag, Jun. 2010, Ph.D. thesis, ISBN 978-3-938860-35-9.
[15] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: a deterministic approach," *IEEE Trans. on Information Theory*, vol. 57, no. 4, pp. 1872-1905, Apr. 2011.
[16] P. Marsch, and G. Fettweis, "Uplink CoMP under a constrained fronthaul and imperfect channel knowledge," *IEEE Trans. on Wireless Comm.*, vol. 10, no. 6, pp. 1730-1742, Jun. 2011.
[17] S. Wang, L. Chen, Y. Yang, and G. Wei, "Quantization in uplink multi-cell processing with fixed-order successive interference cancellation scheme under fronthaul constraint," accepted by 2014 IEEE 79th Vehicular Technology Conference (VTC 2014 Spring), May 2014, Seoul, Korea.
[18] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Information Theory*, vol. 22, no. 1, pp. 1-10, Jan. 1976.
[19] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: general sources," *Information and control*, pp. 60-80, 1978.
[20] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. on Information Theory*, vol. 50, no. 11, pp. 2762-2768, Nov. 2004.
[21] Y. Qi, M. Imran, R. Souza, and R. Tafazolli, "On the Optimization of Distributed Compression in Multi-Relay Cooperative Networks," accepted by *IEEE Trans. on Vehicular Technology*, 2015.
[22] M. Dai, H. Y. Kwan, and C. W. Sung, "Linear network coding strategies for the multiple-access relay channel with packet erasures," *IEEE Trans. on Wireless Commun.*, vol. 12, no. 1, pp. 218-227, Jan. 2013.
[23] M. Dai, K. W. Shum, and C. W. Sung, "Data dissemination with side information and feedback," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 9, pp. 4708-4720, Sept. 2014.
[24] F. Boccardi, J. Heath, R.W., A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
[25] HUAWEI white paper, *5G: a technical vision*, Dec. 2013.
[26] R. J. Weiler (et al), "Enabling 5G fronthaul and access with millimeter-waves," in *Proc. EuCNC 2014*, Bologna Italy, Jun, 2014.
[27] AVIAT network, "60GHz band: a solution for small cell fronthaul?" online at http://blog.aviatnetworks.com/2013/10/08/60ghz-band-a-solution-for-small-cell-fronthaul/
[28] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, No. 9, pp. 3037-3063, Sep. 2005.
[29] C. A. Floudas, *NonLinear and mixed-integer optimization*, Oxford Univ. Press, 1995.
[30] T. A. Porsching, "Jacobi and Gauss-Seidel methods for nonlinear network problems," *SIAM Journal*, vol. 6, no. 3, 1969.
[31] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, second edition, 1999.

[32] F. Rouillier and P. Zimmerman, “Efficient isolation of polynomial's real roots,” *Journal of Computational and Applied Mathematics*, vol. 162, pp. 33-50, Jan. 2004.
[33] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *IEEE Trans. Inform. Theory*, vol. 54, No. 7, pp. 3008-3023, Jul. 2008.
[34] J. Chen and T. Berger, "Distributed source coding via successive refinement," 42nd Annual Allerton Conference on Communication, Control, and Computing 2006, online at www.ece.mcmaster.ca/~junchen/Allerton2004.pdf.
[35] C. Tian and J. Chen, “Remote vector Gaussian source coding with decoder side information,” IEEE Trans. Inf. Theory, vol. 55, no. 10, pp. 4676-4680, Oct. 2009.
[36] S. H. Lim, Y. Kim, A. E. Gamal, and S. Chung, "Noisy network coding," *IEEE Trans. on Information Theory*, vol. 57, no. 5, pp. 3132-3152, May 2011.
[37] X. Wu and L Xie, "On the optimal compressions in the compress-and-forward relay schemes," *IEEE Trans. on Information Theory*, vol. 59, no. 5, pp.2613-2628, May 2013.
[38] H. Holtkamp (et al), “D2.2: Definition and parameterization of reference systems and scenarios,” Tech. Rep. of FP7 project EARTH, Jun. 2010, available online at https://www.ict-earth.eu/publications/deliverables/deliverables.html.
[39] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1st ed., 1991.
[40] W. Yu and R. Liu, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. on Commun.*, vol. 52, no. 7, pp. 1310-1322, Jul. 2006.