Fronthaul Compression and Optimization for Cloud Radio Access Networks

Thang X. Vu[†], Hieu D. Nguyen[*], Tony Q.S. Quek[†], and Sumei Sun[*]
[†] Singapore University of Technology and Design, Singapore 487382, Singapore
Email: {xuanthang_vu, tonyquek}@sutd.edu.sg
[*] Institute of Inforcomm Research, A*STAR, Singapore 138632, Singapore
Email: {nguyendh, sunsm}@i2r.a-star.edu.sg

***Abstract*—In the present paper, we investigate the design and optimization for fronhaul links in cloud radio access networks (C-RAN). Existing C-RAN designs rely on the instantaneous network-wide channel state information (CSI), which might impose a significant overhead due to the potential large-scale of C-RAN. To overcome this limitation, we optimize C-RAN based on the average performance metrics which only require the second-order statistics of the fading channels. Firstly, a tight upper bound of the block error rate (BLER) over Rayleigh fading channels is derived in closed-form expression, through which some insights on C-RAN are drawn: i) full diversity order, which is equal to the number of RRHs, is achievable with respect to the signal to compression plus noise ratio; and ii) the BLER is limited below by either compression or Gaussian noises. Secondly, based on the derived bound, a compression optimization is proposed to minimize the fronthaul transmission rate while satisfying some predefined BLER constraints. The premise of the proposed optimization originates from practical scenarios where most applications tolerate a non-zero BLER. Finally, a fronthaul rate allocation scheme is proposed to minimize the system BLER. It is proved that the proposed allocation scheme, which imposes uniform compression noise across the RRHs, approaches the optimal allocation as the total fronthauls' bandwidth increases.**

I. Introduction

Future mobile networks are facing with the exponential data growth due to the proliferation of diverse mobile equipments and data-hungry applications. Among promising technology candidates for future mobile networks, cloud radio access network (C-RAN) has received much attention [1]. In C-RAN, one centralized processor or baseband unit (BBU) communicates with users via distributed remote radio heads (RRHs). The RRHs are connected to the BBU via high capacity, low latency fronthaul links and have minimal functioning since most baseband processing tasks are centralized at the BBU. C-RAN enables adaptive load balancing via virtual base station pool [2] and effective network-wide inter-cell interference management thanks to multi-cell processing [3]. However, since the baseband processing is executed at the BBU, it requires enormous transmission rate on the fronthaul links to transfer in-phase/quadrature-phase (I/Q) samples, which represent the physical signal obtained through the sampling of the complex baseband signals. Reducing this rate is extremely important in the implementation of C-RAN since the fronthaul links' capacity is limited.

Numerous research efforts have recently investigated the compression of C-RAN, from both information-theoretic perspective [5–8], which designs and optimizes the quantization noise to maximize the achievable sum rate, and practical system point of view [11], [12], which minimizes redundancy of control information in common public radio interface package. It is shown, in general, that the joint design of the precoding and quantization noise matrix can significantly improve the system sum rate over separate designs [5]. Additional performance gain can be achieved by optimizing the test channel [9] or performing hybrid compression and message-sharing strategy [6]. From the practical system perspective, efficient compression can be performed in both time- and frequency-domains (sub-carrier compression) [1]. Lossless compression is proposed to achieve a good statistical multiplexing gain with a negligible signal distortion [11], [12]. In this compression strategy, only data of active users are transferred on the fronthauls and a maximum amount of control informations are locally generated. We note that these papers understand C-RAN based on the instantaneous channel state information (CSI) of the whole network, which might impose a significant overhead and delay due to the potential large-scale of C-RAN.

In order to overcome the aforementioned limitation, we design and optimize C-RAN with only the second-order statistics of the fading channels, thus significantly reduce the signal overhead. Firstly, the system block error rate (BLER) over Rayleigh fading channels is analysed via union bound analysis. In particular, a tight upper bound of BLER is computed in closed-form expression, through which direct inspirations for insights of C-RAN are drawn: i) full diversity order of N (number of RRHs) is achievable with respect to the signal to compression plus noise ratio; and ii) the BLER is limited below by either compression or Gaussian noises. Secondly, given the BLER formula at hands, we propose an adaptive compression optimization to minimize the fronthaul transmission rate with an acceptable signal distortion such that the BBU can support a maximum number of RRHs. Our objective comes from practical demands where most applications can tolerate a non-zero BLER. Our design criterion differs from that in [9], which aims to fully occupy the fronthaul link capacity. Compared with the optimizations in [9], [13] which are based on the instantaneous CSIs, the proposed method only requires the second-order statistics of the fading channels, thus significantly reducing the signal overhead and computation time. Such reductions become more important in C-RAN systems, which are designed to support a large number of nodes. Thirdly, we propose a fronthaul rate allocation scheme to minimize the system BLER. Finally, the accuracy of our analysis and the effectiveness of the adaptive compression are verified via numerical results, which also show that a compression ratio of 330% can be achieved.

The rest of the paper is organized as follows. Section II describes in detail the system model and the receiver's structure. Section III analyses the BLER under Rayleigh fading channels. Section IV proposes a compression optimization to minimize the fronthauls' transmission rate. Section V proposes a rate allocation over the RRHs. Numerical results are discussed in Section VI. Finally, Section VII concludes the paper.

II. System model

We consider a C-RAN system consisting of M users denoted by $\mathrm{U}_1, ..., \mathrm{U}_M$, N RRHs denoted by $\mathrm{R}_1, ..., \mathrm{R}_N$, and one BBU (see [13, Fig. 1]). The users communicate with the RRHs via wireless medium, while the RRHs connect to the BBU by high-speed, low-latency optical fibres, which are known as fronthaul links [1]. In C-RAN, the RRHs can be seen as ”soft” relaying nodes because all baseband processing functions are immigrated to the BBU. The users and RRHs are equipped with a single antenna. In practical systems, a multiple-antenna RRH can be seen as a band of single-antenna RRHs which are subject to a sum rate constraint. Due to the limited capacity on the fronthaul links, I/Q signals need to be compressed before being sent to the processing center. The BBU decompresses the received signals from the RRHs and then performs further baseband processing.

We assume that all nodes are synchronous and all wireless channels are subject to block Rayleigh fading. The BBU is assumed to know the network-wide CSIs. Denote c_m as a modulated symbol emitted by user U_m. The modulated symbol $c_m, 1 \leq m \leq M$, belongs to the source codebook $\mathcal{S} = \{s_1, ..., s_{|\mathcal{S}|}\}$ which has average unit power, e.g., $\mathbb{E}_{s\in\mathcal{S}}|s|^2 = 1$, where $|.|$ denotes the cardinality of a set. The symbols transmitted by the sources are aggregated into a codeword $\mathbf{c} = [c_1, \ldots, c_M]^T$, where $(.)^T$ represents the vector/matrix transpose. The received signal at R_n is given by

$$y_n = \sum_{m=1}^{M} h_{nm}\sqrt{P_{nm}}c_m + z_n = \mathbf{h}_n\mathbf{\Lambda}_n\mathbf{c} + z_n, \qquad (1)$$

where $\mathbf{\Lambda}_n = \mathrm{diag}([\sqrt{P_{n1}}, \ldots, \sqrt{P_{nM}}])$, P_{nm} is the average received energy at R_n from U_m, including the path loss, h_{nm} is the channel fading coefficient between U_m and R_n, which is a complex Gaussian random variable with mean zero and unit variance, $\mathbf{h}_n = [h_{n1}, \ldots, h_{nM}]$ is the channel vector from all users to R_n, and z_n is independent identically distributed (i.i.d.) Gaussian noise with zero mean and variance σ^2.

A. Uniform compression scheme

Upon receiving analogue signals from the users, the RRHs quantize and compress them into digital bits and then forward these bits to the BBU. In this work, we consider uniform quantization as the compression method because of its low-complexity and practical implementation [13]. This compression method can be realized by flexibly tuning the analogue-to-digital converter (ADC) resolution. Therefore, a target compression ratio can be achieved by changing the resolution of the ADC. In the case where the resolution of the ADC is fixed due to some hardware constraints, this compression method can be performed by truncating some least important bits in the ADC's output. The compression is executed on the real and imaginary parts separately [1]. The received signal at the n-th RRH is first normalized as $\bar{y}_n = y_n/\eta_n = \mathcal{R}(\bar{y}_n) + i\mathcal{I}(\bar{y}_n)$, where $\mathcal{R}(x)$ and $\mathcal{I}(x)$ denote the real and imaginary parts of x, respectively, η_n is a scaling factor that restricts $\mathcal{R}(\bar{y}_n)$ and $\mathcal{I}(\bar{y}_n)$ within $[-1, 1]$ with high probability. The value of η_n can be calculated for a given codebook $\mathcal{S}$ and network topology. In this work, we employ the common three-sigma method in literature $\eta_n = 3\sqrt{\|\mathbf{\Lambda}_n\|^2 + \sigma^2}$, which is assumed to be known at the n-th RRH because its overhead is negligible compared with data.

In the next step, the normalized signal $\bar{y}_n$ is quantized into $\tilde{y}_n$ by a Q_n-bit uniform quantizer. The compressed signal can be calculated from the normalized signal as

$$\mathcal{R}(\tilde{y}_n) = \eta_n \frac{\lfloor \mathcal{R}(\bar{y}_n) \times 2^{Q_n} \rceil}{2^{Q_n}}, \ \mathcal{I}(\tilde{y}_n) = \eta_n \frac{\lfloor \mathcal{I}(\bar{y}_n) \times 2^{Q_n} \rceil}{2^{Q_n}},$$

where $\lfloor . \rceil$ stands for the rounding operation.

The quantization error at R_n is given as $q_n = y_n - \tilde{y}_n$. When the absolute value of y_n is large compared to quantization step, $\mathcal{R}(q_n)$ and $\mathcal{I}(q_n)$ can be well modelled as uniform random variables with the support $[-\delta_n, \delta_n]$, where $\delta_n = \eta_n 2^{-Q_n-1}$. We observe via intensive simulations that with the three-sigma rule, such assumption is still feasible even with a small number of quantization bits. After compression, $\tilde{y}_n$ is converted into a bit sequence which is later sent to the BBU via error-free by capacity-limited fronthaul links.

B. Decoding at the BBU

Because the fronthaul links are error-free, the BBU receives $\tilde{y}_n$ from the n-th RRH. The BBU employs a joint decompressing and detecting (JDD) algorithm, which exploits the structure of the quantizer and the codebook to perform decompression and detection for the source codeword simultaneously. Given the compressed signals and the CSI, the BBU optimally estimates the source codeword by using the maximum a posteriori (MAP) receiver as follows [13]:

$$\hat{\mathbf{c}} \ = \arg\max_{\mathbf{c}} \Pr\{\mathbf{c}\} \prod_{n=1}^{N} \Pr\{\tilde{y}_n|\mathbf{c}\}. \qquad (2)$$

In (2), $\Pr\{\tilde{y}_n|\mathbf{c}\}$ is the probability that the quantizer outputs $\tilde{y}_n$ from the channel observation $y_n = \mathbf{h}_n\mathbf{\Lambda}_n\mathbf{c} + z_n$. It is worth to mention that for real signal, the linear quantizer outputs y if the distance between the input and y is less than or equal to the quantization error. For the complex signal y_n, the quantizer outputs $\tilde{y}_n$ if both $|\mathcal{R}(y_n) - \mathcal{R}(\tilde{y}_n)|$ and $|\mathcal{I}(y_n) - \mathcal{I}(\tilde{y}_n)|$ are less than the quantization error. Because the quantization is performed independently for the real and the imaginary parts, we have $\Pr\{\tilde{y}_n|\mathbf{c}\} = \Pr\{\mathcal{R}(y_n) \in \Omega_1\} \times \Pr\{\mathcal{I}(y_n) \in \Omega_2\}$, where $\Omega_1 = [\mathcal{R}(\tilde{y}_n) - \delta_n, \mathcal{R}(\tilde{y}_n) + \delta_n]$ and $\Omega_2 = [\mathcal{I}(\tilde{y}_n) - \delta_n, \mathcal{I}(\tilde{y}_n) + \delta_n]$.

To derive the above probability, we remind that for the given codeword and fading channels, $\mathcal{R}(y_n)$ and $\mathcal{I}(y_n)$ are Gaussian random variables with the same variance $\sigma^2/2$ and means $\mathcal{R}(\mathbf{h}_n\mathbf{\Lambda}_n\mathbf{c})$ and $\mathcal{I}(\mathbf{h}_n\mathbf{\Lambda}_n\mathbf{c})$, respectively. By following similar techniques in [13] and applying the first-order Taylor's approximation on the $\operatorname{erfc}(.)$ function, $\Pr\{\tilde{y}_n|\mathbf{c}\}$ can be computed as

$$\Pr\{\tilde{y}_n|\mathbf{c}\} \simeq \frac{\delta_n}{\sqrt{2\pi}\sigma}\exp\left(-\frac{|\tilde{y}_n - \mathbf{h}_n\mathbf{\Lambda}_n\mathbf{c}|^2}{2\sigma^2}\right). \tag{3}$$

Substituting (3) into (2) we obtain a decoding rule for codeword $\hat{\mathbf{c}}$.

III. Performance Analysis over Rayleigh Fading Channels

This section analyses the BLER of C-RAN uplinks in Rayleigh fading channels. The BLER is defined as the probability of receiving codeword $\hat{\mathbf{c}}$ when a codeword $\mathbf{c} \neq \hat{\mathbf{c}}$ was transmitted. Note that a block error event occurs when at least one out of M symbols $\{c_m\}_{m=1}^{M}$ is decoded with error. Since the exact BLER is difficult to investigate, we instead resort to the union bound on the BLER and consider the average pairwise error probability (APEP) as follows:

$$\text{BLER} \leq \text{APEP} = \frac{1}{|\mathcal{S}|^M}\sum_{\mathbf{c}\in\mathcal{S}^M}\sum_{\tilde{\mathbf{c}}\in\mathcal{S}^M,\tilde{\mathbf{c}}\neq\mathbf{c}} \overline{\text{PEP}}_{\mathbf{c}\to\tilde{\mathbf{c}}}, \tag{4}$$

where $\overline{\text{PEP}}_{\mathbf{c}\to\tilde{\mathbf{c}}} = \mathbb{E}\{\text{PEP}_{\mathbf{c}\to\tilde{\mathbf{c}}}\}$ denotes the expectation over the channel fading coefficients of $\text{PEP}_{\mathbf{c}\to\tilde{\mathbf{c}}}$, which depends on the fading channels and is computed as [13, Eq.(9)]:

$$\text{PEP}_{\mathbf{c}\to\tilde{\mathbf{c}}} = \Pr\{Z > \psi\} = \frac{1}{2}\times \tag{5}$$
$$\operatorname{erfc}\!\left(\frac{\sum_{n=1}^{N}|\mathbf{h}_n\mathbf{\Lambda}_n(\tilde{\mathbf{c}}-\mathbf{c})|^2}{\sqrt{4\sigma^2\sum\limits_{n=1}^{N}|\mathbf{h}_n\mathbf{\Lambda}_n(\tilde{\mathbf{c}}-\mathbf{c})|^2+\frac{8}{3}\sum\limits_{n=1}^{N}\delta_n^2|\mathbf{h}_n\mathbf{\Lambda}_n(\tilde{\mathbf{c}}-\mathbf{c})|^2}}\right).$$

Before deriving $\overline{\text{PEP}}_{\mathbf{c}\to\tilde{\mathbf{c}}}$, we observe that $\left\{\mathbf{h}_n\mathbf{\Lambda}_n(\tilde{\mathbf{c}} - \mathbf{c})\right\}_{n=1}^{N}$ are i.i.d. random variables with $\mathcal{CN}(0, \|\mathbf{\Lambda}_n(\tilde{\mathbf{c}}-\mathbf{c})\|^2)$ distribution, where $\|\,.\,\|$ denotes the l_2-norm. Define $G_n^{\mathbf{c},\tilde{\mathbf{c}}}t_n \triangleq |\mathbf{h}_n\mathbf{\Lambda}_n(\tilde{\mathbf{c}}-\mathbf{c})|^2$, where $G_n^{\mathbf{c},\tilde{\mathbf{c}}} \triangleq \|\,\mathbf{\Lambda}_n(\tilde{\mathbf{c}}-\mathbf{c})\,\|^2$. Thus, $\left\{t_n\right\}_{n=1}^{N}$ are i.i.d. exponential random variables each with distribution e^{-t}, $0 < t < \infty$. We express the term inside (5) as

$$\frac{\sum_{n=1}^{N}G_n^{\mathbf{c},\tilde{\mathbf{c}}}t_n}{\sqrt{\sum_{n=1}^{N}\left(4\sigma^2+\frac{8}{3}\delta_n^2\right)G_n^{\mathbf{c},\tilde{\mathbf{c}}}t_n}} = \frac{\sum_{n=1}^{N}G_n^{\mathbf{c},\tilde{\mathbf{c}}}t_n}{\sqrt{\sum_{n=1}^{N}\beta_nG_n^{\mathbf{c},\tilde{\mathbf{c}}}t_n}},$$

where $\beta_n \triangleq 4\sigma^2 + \frac{8}{3}\delta_n^2$.

By changing variable t_n to $v_n = \beta_nG_n^{\mathbf{c},\tilde{\mathbf{c}}}t_n$, the average PEP over the fading channels is evaluated as follows:

$$\overline{\text{PEP}}_{\mathbf{c}\to\tilde{\mathbf{c}}} = \frac{1}{2\prod_{n=1}^{N}\beta_nG_n^{\mathbf{c},\tilde{\mathbf{c}}}}\int\limits_0^\infty\!...\!\int\limits_0^\infty \operatorname{erfc}\left(\frac{\sum_{n=1}^{N}\frac{v_n}{\beta_n}}{\sqrt{\sum_{n=1}^{N}v_n}}\right)$$
$$\times\exp\left(-\sum_{n=1}^{N}\frac{v_n}{\beta_nG_n^{\mathbf{c},\tilde{\mathbf{c}}}}\right)dv_1\dots dv_N. \tag{6}$$

The exact computation of (6) is challenging, especially over the set $\{v_n\}_{n=1}^{N}$ for arbitrary $\{\beta_n\}_{n=1}^{N}$. Therefore, we investigate an upper bound of (6) instead.

Denoting $\beta_{\max} = \max_{n\in\{1,\dots,N\}}\beta_n$, we have

$$\sum_{n=1}^{N}\frac{v_n}{\beta_n} \geq \sqrt{\sum_{n=1}^{N}\frac{v_n}{\beta_n}}\sqrt{\frac{\sum_{n=1}^{N}v_n}{\beta_{\max}}}. \tag{7}$$

Applying (7) to (6) while noting that $\operatorname{erfc}(x)$ is a decreasing function, we obtain

$$\overline{\text{PEP}}_{\mathbf{c}\to\tilde{\mathbf{c}}} \leq \frac{1}{2\prod_{n=1}^{N}\beta_nG_n^{\mathbf{c},\tilde{\mathbf{c}}}}\int\limits_0^\infty\!...\!\int\limits_0^\infty \operatorname{erfc}\left(\frac{1}{\sqrt{\beta_{\max}}}\sqrt{\sum_{n=1}^{N}\frac{v_n}{\beta_n}}\right)$$
$$= \frac{1}{12\prod_{n=1}^{N}\left(\frac{G_n^{\mathbf{c},\tilde{\mathbf{c}}}}{\beta_{\max}}+1\right)} + \frac{1}{4\prod_{n=1}^{N}\left(\frac{4G_n^{\mathbf{c},\tilde{\mathbf{c}}}}{3\beta_{\max}}+1\right)}. \tag{8}$$

where (8) results from the tight approximation of $\operatorname{erfc}(.)$ function $\operatorname{erfc}(x) \simeq \frac{1}{6}e^{-x^2} + \frac{1}{2}e^{-\frac{4x^2}{3}}$ [14].

Substituting (8) into (4), we obtain the upper bound of the BLER under Rayleigh fading channels. It is observed that C-RAN achieves full diversity of order N with respect to the signal to compression plus Gaussian noise ratio. The accuracy of the bounds obviously depends on how diverse the set $\{\beta_n\}_{n=1}^{N}$ is.

IV. Minimization of the Fronthaul Transmission Rate

In practical systems, different applications might require various quality of service (QoS) depending on specific individual. For example, a voice message usually requires a lower QoS compared to a video call. From the provider's perspective, it is always beneficial to minimize network resources as long as the required QoS is guaranteed. This motivates us to propose an adaptive compression scheme to minimize the fronthaul transmission rate (maximize the compression efficiency) under a certain target BLER so that a front-haul link can support a maximal number of RRHs. Such scheme is desirable for systems which support large front-haul feedback and/or require stringent BLER QoS. Note that the proposed adaptive compression is based on the average BLER, and thus relying only on the statistics but the instantaneous fading channels. Since the actual transmission rate at the n-th RRH is equal to the sampling resolution at that RRH, we refer Q_n to the transmission rate for convenience.

For a given QoS constraint ζ, we want to minimize the total fronthauls' transmission rate. The corresponding optimization is formulated as follows:

$$\underset{\{Q_n\,:\,Q_n\geq 1\}_{n=1}^{N}}{\texttt{minimize}} \quad \sum_{n=1}^{N}Q_n \tag{9}$$
$$\texttt{s.t.}\;\; \frac{1}{|\mathcal{S}|^M}\sum\sum_{\tilde{\mathbf{c}}\neq\mathbf{c}\in\mathcal{S}^M}\overline{\text{PEP}}_{\mathbf{c}\to\tilde{\mathbf{c}}} \leq \zeta.$$

In the following, we propose an optimization of problem (9) based on the upper bound of the APEP in (8). By changing variable to $\beta_n = 4\sigma^2 + \frac{2}{3}\eta_n^2 2^{-2Q_n}$, the optimization problem in this case is stated as follows:

$$\underset{\{\beta_n\}_{n=1}^{N}}{\texttt{maximize}} \quad \sum_{n=1}^{N} \log\left(\beta_n - 4\sigma^2\right) \tag{10}$$

$$\texttt{s.t.} \quad \frac{1}{|\mathcal{S}|^M}\sum_{l=1}^{L}\left(\frac{1}{12\prod_{n=1}^{N}\left(\frac{G_n^l}{\beta_{\max}}+1\right)} + \frac{1}{4\prod_{n=1}^{N}\left(\frac{4G_n^l}{3\beta_{\max}}+1\right)}\right) \leq \zeta,$$

$$0 < \beta_n - 4\sigma^2 \leq \frac{\eta_n^2}{6}, \ \forall n.$$

By introducing an auxiliary variable A, the above problem is written equivalently as follows:

$$\underset{\{\beta_n\}_{n=1}^{N},\, A>0}{\texttt{maximize}} \quad \sum_{n=1}^{N} \log\left(\beta_n - 4\sigma^2\right) \tag{11}$$

$$\texttt{s.t.} \quad \frac{1}{|\mathcal{S}|^M}\sum_{l=1}^{L}\log\left(\frac{1}{12\prod_{n=1}^{N}\left(G_n^l A+1\right)} + \frac{1}{4\prod_{n=1}^{N}\left(\frac{4}{3}G_n^l A+1\right)}\right) \leq \zeta,$$

$$\beta_n \leq \frac{1}{A}, \ \forall n,$$

$$0 < \beta_n - 4\sigma^2 \leq \frac{\eta_n^2}{6}, \ \forall n.$$

It is observed that given A, the problem (11) is convex and thus efficiently solvable. We therefore resort to bisection to solve (11) [15]. The steps are given in Table I. The integer quantization bit $Q_n^{integer}$ can be obtained from β_n simply by choosing the smallest following integer of $Q_n^{exact} = \frac{1}{2}\log_2(\frac{2}{3}\eta_n^2) - \frac{1}{2}\log_2(\beta_n - 4\sigma^2)$, i.e., $\lceil Q_n^{axact} \rceil$. In general, there is no bound for the optimality loss of such approximation. However, as the constraint threshold BLER tends to 0, the loss also tends to 0. The reasoning is that each Q_n becomes large in such case, which leads to a small $\frac{\lceil Q_n^{axact}\rceil - Q_n^{axact}}{Q_n^{axact}}$.

A. Optimality of identical quantization noise scheme

In this subsection, we will prove that a sampling scheme which leads to identical quantization noise, i.e., $\delta_1 = \delta_2 = \cdots = \delta_N$, is the optimal solution of the problem (9), as the BLER threshold ζ goes to zero. The formal statement of this result is given in the following proposition. The proof is omitted due to the space constraint.

Proposition 1: As the QoS threshold $\zeta \to 0$, the solution of problem (11) based on the upper bound of the APEP satisfies the identical compression noise.

Proposition 1 is not strong enough to state that the optimal solution of (9) approaches that of identical compression noise as the BLER threshold decreases, since(11) provides an upper bound for the original problem (9). Nevertheless, Proposition 1 provides a justification for implementing the sampling that imposes identical compression noise, especially under delay-constrained system where sophisticated adaptive sampling might not be welcomed.

TABLE I: ALGORITHM TO SOLVE (11)

1. Initialize A_H, A_L, and the accuracy ϵ.
2. $A_M = (A_H + A_L)/2$.
3. Given $A = A_M$, if (11) is feasible, then $A_L := A_M$. Otherwise $A_H := A_M$.
4. Repeat step 2 and 3 until $|A_H - A_L| \leq \epsilon$.

V. MINIMIZATION OF THE BLER

In this section, we aim at minimizing the BLER for a given total fronthaul bandwidth of Q_{sum} bits. Particularly, we will allocate sampling resolution Q_n to the n-th RRH to achieve the smallest BLER. The optimization problem is formulated as follows:

$$\underset{\{Q_n : Q_n \geq 1\}_{n=1}^{N}}{\texttt{minimize}} \quad \mathrm{BLER} \tag{12}$$

$$\texttt{s.t.} \quad \sum_{n=1}^{N} Q_n \leq Q_{sum},$$

where BLER is given in (4).

In order to guarantee the effective of the optimization in (12), the BLER is calculated based on the upper bound of the PEP given in (8). By changing variable to $\beta_n = 4\sigma^2 + \frac{2}{3}\eta_n^2 2^{-2Q_n}$, the resulting optimization problem is given as

$$\underset{\{\beta_n\}_{n=1}^{N}}{\texttt{minimize}} \quad \sum_{l=1}^{L}\left(\frac{1}{12}\prod_{n=1}^{N}\frac{1}{\frac{G_n^l}{\beta_{\max}}+1} + \frac{1}{4}\prod_{n=1}^{N}\frac{1}{\frac{4G_n^l}{3\beta_{\max}}+1}\right) \tag{13}$$

$$\texttt{s.t.} \quad \sum_{n=1}^{N}\log_2(\beta_n - 4\sigma^2) \geq \sum_{n=1}^{N}\log_2\left(\frac{2}{3}\eta_n^2\right) - 2Q_{sum},$$

$$0 < \beta_n \leq \frac{1}{6}\eta_n^2, \ \forall n.$$

By introducing an arbitrary variable A, the problem (13) is equivalent to the following problem:

$$\underset{\substack{\{\beta_n:\beta_n>0\}_{n=1}^{N}\\ A>0}}{\texttt{minimize}} \quad \sum_{l=1}^{L}\left(\frac{1}{12}\prod_{n=1}^{N}\frac{1}{G_n^l A+1} + \frac{1}{4}\prod_{n=1}^{N}\frac{1}{\frac{4}{3}G_n^l A+1}\right) \tag{14}$$

$$\texttt{s.t.} \quad \beta_n \leq \frac{1}{A}, \ \forall n \tag{15a}$$

$$\sum_{n=1}^{N}\log_2(\beta_n - 4\sigma^2) \geq \sum_{n=1}^{N}\log_2\left(\frac{2}{3}\eta_n^2\right) - 2Q_{sum}, \tag{15b}$$

$$\beta_n \leq 4\sigma^2 + \frac{1}{6}\eta_n^2, \ \forall n. \tag{15c}$$

We observe that problem (14) is a convex optimization problem on $\{\beta_n\}$ for a given A. Therefore (14) can efficiently be solved by standard methods, i.e., bisection [15]. The steps to solve (14) are given in Table II.

TABLE II: ALGORITHM TO SOLVE (14)

```
1. Initialize $A_H$, $A_L$, and the accuracy $\epsilon$.
2. $A_M$ = $(A_H + A_L)/2$.
3. Given $A = A_M$, if (14) is feasible, then $A_L := A_M$.
   Otherwise $A_H := A_M$.
4. Repeat step 2 and 3 until $|A_H - A_L| \leq \epsilon$.
```

In general, it is difficult to obtain the exact formula for solution of 12. Under certain circumstances, the closed-form expression of the solution of 12 can be derived as following proposition. The proof is omitted due to the space constraint.

Proposition 2: If there exist $\{q_n : q_n \geq 1\}_{n=1}^N$ such as $\eta_1 2^{-q_1} = \eta_2 2^{-q_2} = \cdots = \eta_N 2^{-q_N}$ and $\sum_{n=1}^N q_n = Q_{sum}$, then the solution of problem (14) satisfies the identical quantization noise and is given as

$$\beta_n^\star = 2^{\frac{1}{N}\left(\sum_{n=1}^N \log_2(\frac{2}{3}\eta_n^2) - 2Q_{sum}\right)} + 4\sigma^2, \ \forall n.$$

Consequently, the optimal rate allocation $\{Q_n^\star\}_{n=1}^N$ is given as

$$Q_n^\star = \frac{1}{2}\log_2\left(\frac{2}{3}\eta_n^2\right) + \frac{1}{N}Q_{sum} - \frac{1}{2N}\sum_{n=1}^N \log_2\left(\frac{2}{3}\eta_n^2\right).$$

Corollary 1: For symmetric C-RAN systems, i.e., $\eta_1 = \eta_2 = \cdots = \eta_N$, which employ the quantization and the receiver as in Section II, the uniform sampling $\{Q_n = Q_{sum}/N\}_{n=1}^N$ achieves the minimum BLER.

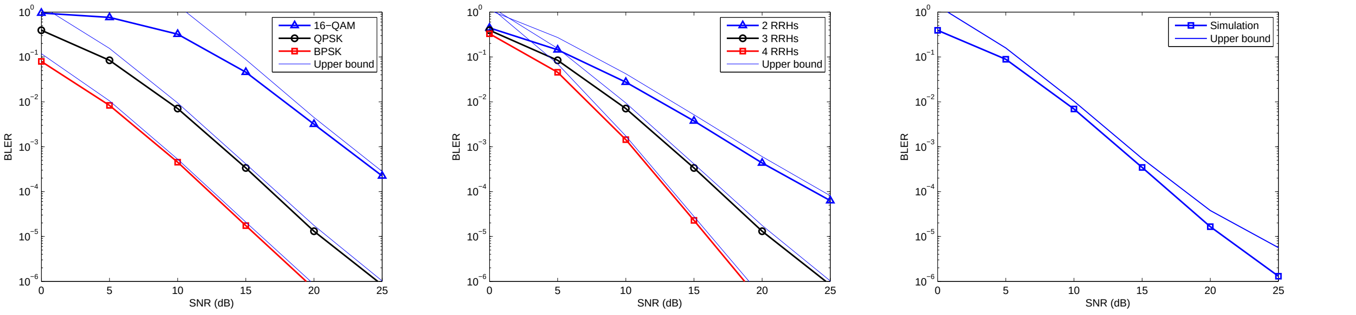
Proof: The proof is obtained straightforward from Proposition 2 by using $\eta_1 = \eta_2 = \cdots = \eta_N$. □

VI. NUMERICAL RESULTS

The preliminary simulation is evaluated for a C-RAN system under block Rayleigh fading channel, i.e., h_{mn}'s are independent identically distributed (i.i.d.) random variables, each is distributed as $\mathcal{CN}(0,1)$. Unless otherwise stated, $M = 3, N = 3$ C-RAN is considered and QPSK modulation with the codebook $\mathcal{S} = \{-1-1i, -1+1i, 1-1i, 1+1i\}/\sqrt{2}$ is used. The BBU is assumed to know the CSI of all the wireless channels.

Fig. 1a and 1b present the BLER performance of C-RAN in symmetric network topology, i.e., $\eta_1 = \eta_2 = \cdots = \eta_N$. The sampling rate is equally allocated, i.e., $Q_1 = Q_2 = Q_3 = 6$ bits. As the result, the quantization noise at every RRH is identical. Fig. 1a compares the bounds with simulation results for $N = 3$ RRHs and three different modulations, i.e., BPSK, QPSK and 16-QAM. For all cases, the derived bounds closely match the simulation results. It is observed that the bound is closer to the simulation for BPSK than 16-QAM because in the later the signal constellation is more diverse. Fig. 1b presents the theoretical and simulation results for different number of RRHs with QPSK modulation. Similar conclusion is observed: the derived bounds closely match the simulations. The same conclusion is also true for Fig. 1c, which shows the simulation results and the corresponding bound under non-identical quantization noise scenario.

Fig. 2 presents the performance of the proposed adaptive compressions versus SNR. For a given BLER target, we want to maximize the compression efficiency, or equivalently to minimize the actual fronthaul transmission rate. For reference, the scheme without QoS constraint which fully occupies the fronthaul bandwidth is also plotted. In addition, to provide full details about the proposed optimization, a curve corresponding to exact optimum solutions of (11) is drawn (without integer condition of Q_n). This curve is mark as "- exact" in the figure. A curve corresponding to integer $\{Q_n\}$ are marked as "- integer". The threshold ζ is equal to 10^{-2}. Note that the optimization is carried out for each operating SNR only once. The BLER performance is shown in Fig. 2a and the actual fronthaul rate is presented in Fig. 2b. Under the small SNR regime, the optimization does not satisfy the target QoS because the channel is too poor. Even using all 10 bits for quantization does not satisfy the target QoS. Therefore, the optimization is inactive for these SNRs. As a result, all schemes consume full fronthaul bandwidth, as shown in Fig. 2b. Under the high SNR regime, the optimization is activated (from 10 dB in the figure). As expected, the optimization scheme meets the target QoS while significantly improves the compression

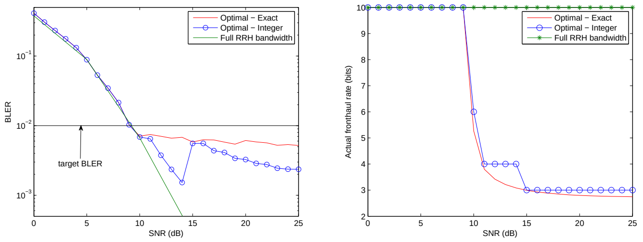


(a)

(b)

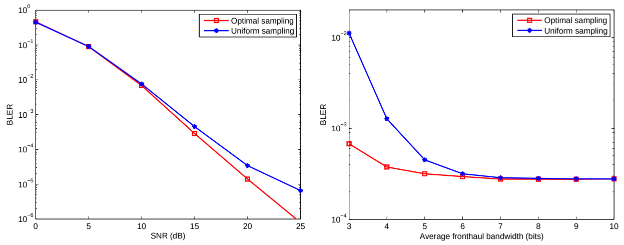
(c)

Fig. 1: Performance of the C-RAN in Rayleigh fading channels for various settings. Markers show simulation results; solid curves without markers correspond to the upper bounds. $M = N = 3$; (a) and (b): symmetric parameters with $Q_1 = Q_2 = Q_3 = 6$ bits; (c): asymmetric parameters with $\{Q_1, Q_2, Q_3\} = \{5, 6, 7\}$.



(a) (b)

Fig. 2: Performance of the optimization with QoS constraint proposed in Section IV. QPSK modulation. The target BLER ζ is equal to 10^{-2}. The total fronthaul bandwidth $Q_{sum} = 30$ bits, $M = N = 3$.



(a) (b)

Fig. 3: Performance of the optimal rate allocation. QPSK modulation, $M = N = 3$. Left: performance versus SNR, $Q_{sum} = 15$ bits. Right: performance versus average per-RRH bandwidth, SNR = 15 dB.

efficiency. Because the full-bandwidth scheme always uses 10 bits for quantization, its fronthaul rate is 10 bits per sample for all SNRs. On the other hand, a compression efficiency of 330% is observed, which only requires 3 bits per sample to achieve a BLER less than 10^{-2}. Note that the integer constraint on $\{Q_n\}_{n=1}^{N}$ results in lower BLER than the exact value of Q_n because $Q_n^{integer} = \lceil Q_n^{exact} \rceil$ is the smallest integer that is larger than or equal to Q_n^{exact}. The sharp step in BLER curves of integer Q_n results from the $\lceil . \rceil$ operation.

Fig. 3a shows the BLER of the proposed rate allocation scheme described in Section V and a reference uniform rate allocation, e.g., $Q_1 = Q_2 = Q_3$, for asymmetric network with scaling factors $\{\eta_1, \eta_2, \eta_3\} = \{3, 6, 9\}$. The total fronthauls' bandwidth is $Q_{sum} = 15$ bits. As expected, the optimum rate allocation achieves smaller BLER than the reference scheme. Under the small SNR regime, both uniform and optimal rate allocations get similar performance because at this SNR the thermal noise is dominant. As the SNR increases, the performance gain provided by the optimal rate allocation is larger. This is because in the high SNR regime, the quantization noise has more impacts on the system performance.

Fig. 3b shows the BLER performance of the optimal rate allocation versus different fronthaul's bandwidth. Uniform rate allocation is also plotted. The operating SNR is equal to 15 dB. It is observed that the optimal scheme is more effective for small fronthaul's bandwidth. As the fronthaul's bandwidth increases, the gap between two schemes degrades and two curves eventually coincide. Such observation can be explained by Theorem 2, which proves the optimality of uniform rate allocation scheme at large total fronthaul's bandwidth.

VII. Conclusions

We have studied the performance of cloud radio access networks in Rayleigh fading channels. Analytical results have been derived for the system block error rate by using pairwise error analysis. It has been shown that the system BLER is limited below by either Gaussian or compression noise, but it achieves full diversity order with respect to the signal to Gaussian plus compression noise ratio. Based on the analysed bounds, a optimization has been proposed to improve the compression efficiency while satisfying the predefined QoS constraint. The proposed optimization problem comes from practical scenarios where most applications tolerate a target QoS, e.g., a non-zero BLER. Furthermore, the optimal rate allocation scheme has been proposed for the given the total fronthauls' bandwidth to minimize the BLER.

References

[1] ChinaMobile, "C-RAN: the road towards green RAN," 2011, white paper.

[2] Z. Zhu, P. Gupta, Q. Wang, S. Kalyanaraman, Y. Lin, H. Franke, and S. Sarangi, "Virtual base station pool: Towards a wireless network cloud for radio access networks," in *Proc. ACM Int. Conf. on Computing Frontiers*, New York, Mar. 2011, pp. 34:1–34:10.

[3] P. Marsch and G. Fettweis, "Uplink CoMP under a constrained backhaul and imperfect channel knowledge," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1730–1742, Jun. 2011.

[4] Y. Zhou, W. Yu, and D. Toumpakaris, "Uplink multi-cell processing: Approximate sum capacity under a sum backhaul constraint," in *Proc. IEEE Information Theory Workshop*, Sevilla, Sep. 2013, pp. 1–5.

[5] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5646–5658, Aug. 2013.

[6] P. Patil and W. Yu, "Hybrid compression and message-sharing strategy for the downlink cloud radio-access network," in *Proc. IEEE Information Theory and Application workshop*, San Diego, CA, Feb. 2014, pp. 1–6.

[7] Y. Wang, H. Wang, and L. Scharf, "Optimum compression of a noisy measurement for transmission over a noisy channel," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1279–1289, Mar. 2014.

[8] A. del Coso and S. Simoens, "Distributed compression for MIMO coordinated networks with a backhaul constraint," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4698–4709, Sep. 2009.

[9] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint decompression and decoding for cloud radio access networks," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 503–506, Mar. 2013.

[10] –,"Common public radio intreface (CPRI): interface specification," 2013, CPRI Specification V6.0.

[11] J. Lorca and L. Cucala, "Lossless compression technique for the fronthaul of LTE/LTE-advanced Cloud-RAN architectures," in *Proc. IEEE Int. Symp. and Workshops on a World of Wireless, Mobile and Multimedia Networks*, Madrid, Jun. 2013, pp. 1–9.

[12] A. Nanba, S. nd Agata, "A new IQ data compression scheme for fronthaul link in centralized RAN," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, London, Sep. 2013, pp. 210–214.

[13] T. X. Vu, H. D. Nguyen, and T. Q. S. Quek, "Adaptive Compression and Joint Detection for Fronthaul Uplinks in Cloud Radio Access Networks," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4565–4575, 2015.

[14] M. Chiani, D. Dardari, and M. K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Trans. Commun.* vol. 2, no. 4, pp. 840–845, Jul. 2003.

[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.