Mixed Delay Constraints on a Fading C-RAN Uplink

Homa Nikbakht, Michèle Wigger, Walid Hachem, Shlomo Shamai Shitz

Mixed Delay Constraints on a Fading C-RAN Uplink

Homa Nikbakht and Michèle Wigger
LTCI, Télécom ParisTech
75013 Paris, France
{homa.nikbakht, michele.wigger}@telecom-paristech.fr

Walid Hachem
CNRS / LIGM (UMR 8049)
Universit Paris-Est Marne-la-Vallée
walid.hachem@u-pem.fr

Shlomo Shamai (Shitz)
Technion
sshlomo@ee.technion.ac.il

***Abstract*—A cloud radio access network (C-RAN) is considered where the first hop from the user equipments (UEs) to the basestations (BSs) is modeled by the fading Wyner soft-handoff model. The focus is on mixed-delay constraints where a set of messages (so called "slow" messages) are jointly decoded in the cloud unit (CU), whereas the remaining messages (called "fast" messages) have to be decoded immediately at the BSs. This paper presents inner and outer bounds on the capacity region for such a setup. Moreover, the multiplexing gain region is characterized exactly. The presented results show that for small fronthaul capacity it is beneficial to send both "fast" and "slow" messages. However, when the rate of "fast" messages is already large, then increasing it further, deteriorates the sum-rate of the system. In this regime, the stringent decoding delay on the "fast" messages penalizes the overall performance. Our results indicate that this penalty is larger at moderate SNR than at high SNR and it is also larger for random time-varying fading coefficients than for static ones.**

I. Introduction

The fifth generation (5G) of wireless cellular network has to accommodate different types of data traffics with different latency constraints. Delay-tolerant traffics allow for higher transmission rate by exploiting cooperation possibilities as in cloud radio access networks (C-RAN) [1]–[5]. In C-RAN, base stations (BSs) are connected to a cloud unit (CU) via finite-rate fronthaul links and data are typically decoded jointly at the CU to alleviate the effect of interference [1]. Delay-sensitive traffics are however not compatible with this new technology because they have to be decoded immediately at the BSs, and consequently cannot profit from joint processing.

In this paper, we consider transmission over a C-RAN with mixed delay constraints, i.e., where each mobile user can simultaneously send a delay-sensitive and a delay-tolerant stream. Throughout this paper, we call delay-sensitive traffic "fast" messages and delay-tolerant traffic "slow" messages. Mixed delay constraints in C-RANs have previously been studied in [6] where different decoding techniques are compared. In [6] UEs close to the BSs send only "fast" messages and it is assumed that these communications do not interfer. UEs located further away send "slow" messages and their interference pattern is modeled by Wyner's symmetric network [11], [12] with static channel coefficients.

In this work, each UE sends a pair of independent "fast" and "slow" messages. Communications between UEs and BSs are modeled by a Wyner's soft-handoff network with random time-varying channel coefficients. We present coding schemes for this setup and derive inner and outer bounds on the capacity region. Furthermore, we characterize its exact first-order high-SNR asymptotics, i.e., the multiplexing gain region. This result allows us to conclude that for moderate fronthaul capacities, the maximum "slow" multiplexing gain remains unchanged over a large regime of small and moderate "fast" multiplexing gains. The sum-multiplexing gain is thus improved if some of the messages can directly be decoded at the BSs. In contrast, for large fronthaul capacities or large "fast" multiplexing gains, this sum-multiplexing gain deteriorates by Δ if one further increases the "fast" multiplexing gain by Δ.

At moderate SNR the conclusion based on our inner bound are slightly different: If the "fast" rate is small or moderate, then the achievable sum-rate decreases by Δ when the rate of "fast" messages increases by Δ, and if the "fast" rate is large it decreases with a factor γ times Δ. The penalty factor γ is approximately 1 for static channel coefficients and typically higher for random coefficients. The stringent delay constraint on "fast" messages thus seems to be more harmful at moderate SNR and for time-varying channel conditions than at high SNR or for static channels.

Previous studies on mixed-delay constraints for cellular networks with cooperation [7]–[10] presented similar conclusions as our high-SNR results: For small or moderate "fast" rates the overall performance is not degraded by the stringent delay constraints. For large "fast" rates 1 bit of "fast" rate comes at the expense of 2 bits of "slow" rate.

Note that in this work, we use simple point-to-point compression techniques. Employing more sophisticated compression techniques [4] is left for our future work.

II. Problem Setup

Consider the uplink communication of a multi-cell C-RAN with K UEs and K BSs. UEs and BSs are indexed by $1, \ldots, K$. Each BS is connected to a CU via a separate fronthaul link of capacity C (see Fig. 1). At a given time $t \in \{1, \ldots, n\}$, the signal received at BS k is described as

$$Y_{k,t} = G_{k,t} X_{k,t} + F_{k,t} X_{k-1,t} + Z_{k,t}, \tag{1}$$

where $X_{k,t}$ and $X_{k-1,t}$ are the symbols sent by UE k and UE $k-1$ at time t; $\{Z_{k,t}\}$ are i.i.d circular Gaussian noises of

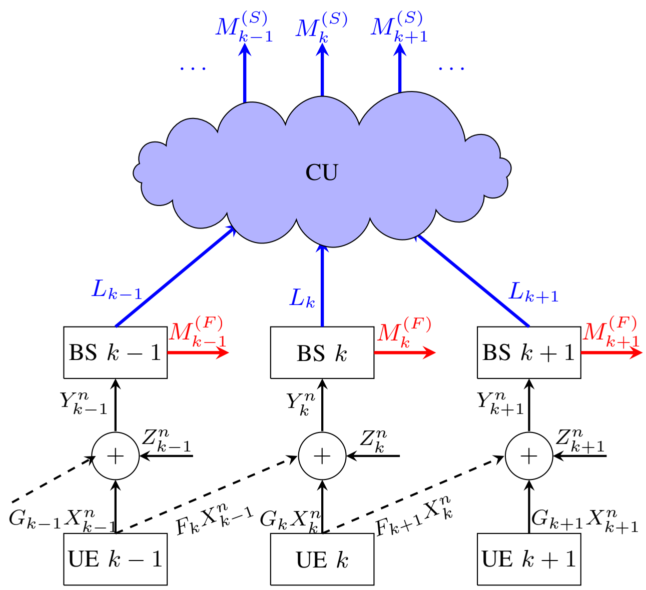

Fig. 1. System model

variance 1; and

$$G_{k,t}, F_{k,t} \in \mathbb{C} \tag{2}$$

represent the time-t fading coefficients. We assume that the sequence of channel coefficients

$$\{(G_{1,t}, G_{2,t}, \ldots, G_{K,t}, F_{1,t}, F_{2,t}, \ldots, F_{K,t})\}_{t=1}^{n} \tag{3}$$

is i.i.d. over time and distributed according to the K-tuple distribution

$$P_{G_1 \cdots G_K F_1 \cdots F_K} \tag{4}$$

of a given stationary and ergodic process $\{(G_i, F_i)\}_{i=-\infty}^{\infty}$ satisfying $\mathbb{E}\left[|G_0|^2\right] < \infty$ and $\mathbb{E}\left[|F_0|^2\right] < \infty$. Each BS k has perfect channel state information (CSI) about its own channel, i.e., it observes the realizations of $\{(G_{k,t}, F_{k,t})\}$ for all $t \in \{1, \ldots, n\}$. The UEs know only the statistics of the random channel coefficients and are said to have no CSI. Figure 1 shows an extract of our system model.

Each UE k wishes to convey the pair of independent messages $(M_k^{(F)}, M_k^{(S)})$ to BS k. The "fast" message $M_k^{(F)}$ is uniformly distributed over the set $\mathcal{M}_k^{(F)} := \left\{1, \ldots, \left\lfloor 2^{nR_k^{(F)}} \right\rfloor\right\}$ and has to be decoded by BS k as we explain shortly. The "slow" source message $M_k^{(S)}$ is uniformly distributed over $\mathcal{M}_k^{(S)} := \left\{1, \ldots, \left\lfloor 2^{nR_k^{(S)}} \right\rfloor\right\}$ and is decoded at the CU. Here, n denotes the blocklength of transmission and $R_k^{(F)}$ and $R_k^{(S)}$ are the rates of transmissions of the "fast" and the "slow" messages.

UE k computes its channel inputs $X_k^n := (X_{k,1}, \ldots, X_{k,n})$ as a function of the pair $(M_k^{(F)}, M_k^{(S)})$:

$$X_k^n = \phi_k^{(n)}\big(M_k^{(F)}, M_k^{(S)}\big), \tag{5}$$

for some function $\phi_k^{(n)}$ on appropriate domains so that the average block-power constraint

$$\frac{1}{n}\sum_{t=1}^{n} |X_{k,t}|^2 \le \mathsf{P}, \quad \text{a.s.,} \quad \forall\, k \in \{1, \ldots, K\}, \tag{6}$$

is satisfied.

Each BS k decodes the "fast" source message $M_k^{(F)}$ based on its own channel outputs $Y_k^n := (Y_{k,1}, \ldots, Y_{k,n})$. So, it produces:

$$\hat{M}_k^{(F)} = \psi_k^{(n)}\big(Y_k^n\big) \tag{7}$$

using some decoding function $\psi_k^{(n)}$ on appropriate domains.

It further produces the fronthaul message

$$L_k = q_k^{(n)}(Y_k^n), \tag{8}$$

using some encoding function

$$q_k^{(n)} : \mathbb{R}_k^n \to \left\{1, \ldots, \left\lfloor 2^{n\mathsf{C}} \right\rfloor\right\}. \tag{9}$$

The CU then decodes the set of "slow" messages as

$$\big(\hat{M}_1^{(S)}, \ldots, \hat{M}_K^{(S)}\big) := b^{(n)}\Big(L_1, \ldots, L_K\Big) \tag{10}$$

by means of a decoding function $b^{(n)}$.

The main focus of this paper is the achievable sum-rates of "fast" and "slow" messages. Given a maximum fronthaul link capacity C and a maximum allowed power P, the pair of (average) rates $(R^{(F)}, R^{(S)})$ is called *achievable* if for each positive integer K there exists a sequence (in n) of encoding and decoding functions $\{\phi_1^{(n)}, \ldots, \phi_K^{(n)}, \psi_1^{(n)}, \ldots, \psi_K^{(n)}, q_1^{(n)}, \ldots, q_K^{(n)}, b^{(n)}\}$ so that the probability of decoding error

$$P_e^{(n)} := \Pr\Bigg[\bigcup_{k\in\{1,\ldots,K\}} \Big\{\hat{M}_k^{(F)} \neq M_k^{(F)} \text{ or } \hat{M}_k^{(S)} \neq M_k^{(S)}\Big\}\Bigg]$$

tends to 0 as $n \to \infty$ and the rates satisfy

$$\overline{\lim_{K\to\infty}} \frac{1}{K}\sum_{k=1}^{K} R_k^{(F)} = R^{(F)}, \tag{11}$$

$$\overline{\lim_{K\to\infty}} \frac{1}{K}\sum_{k=1}^{K} R_k^{(S)} = R^{(S)}. \tag{12}$$

Note that the notation $\overline{\lim}$ refers to the limit superior.

Definition 1 (Capacity Region): The *capacity region* $\mathcal{C}(\mathsf{P}, \mathsf{C})$ is the closure of the set of all rate pairs $(R^{(F)}, R^{(S)})$ that are achievable with power P and fronthaul link capacity C.
We are particularly interested in the capacity in the asymptotic high-SNR regime. The *pair of multiplexing gains* $(\mathsf{S}^{(F)}, \mathsf{S}^{(S)})$ is called *achievable with frontaul multiplexing gain* μ, if there exists a sequence of rates $\{R^{(F)}(\mathsf{P}), R^{(S)}(\mathsf{P})\}_{\mathsf{P}>0}$ so that

$$\mathsf{S}^{(F)} := \overline{\lim_{\mathsf{P}\to\infty}} \frac{R^{(F)}}{\log(1+\mathsf{P})}, \tag{13}$$

$$\mathsf{S}^{(S)} := \overline{\lim_{\mathsf{P}\to\infty}} \frac{R^{(S)}}{\log(1+\mathsf{P})}, \tag{14}$$

and for each $\mathsf{P} > 0$ the pair $(R^{(F)}(\mathsf{P}), R^{(S)}(\mathsf{P}))$ is achievable with fronthaul capacity

$$\mathsf{C} = \mu \cdot \frac{1}{2}\log(1+\mathsf{P}). \tag{15}$$

Definition 2 (Multiplexing Gains): The closure of the set of all achievable multiplexing gains $(\mathsf{S}^{(F)}, \mathsf{S}^{(S)})$ is called *multiplexing gain region* and denoted $\mathcal{S}^{\star}(\mu)$.

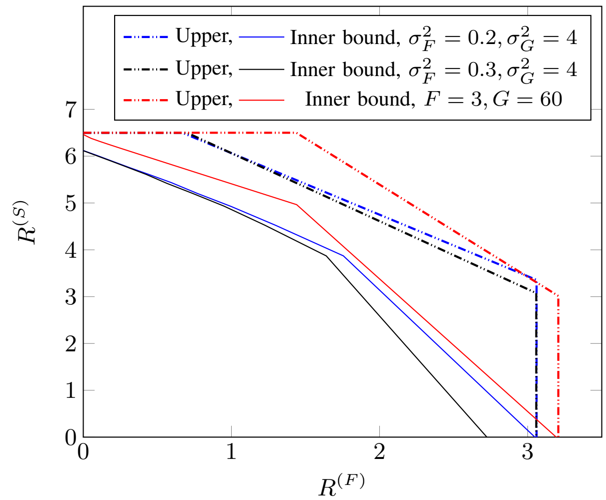

Fig. 2. Capacity inner bound in Theorem 1 and capacity outer bound in Theorem 2 for $\mathsf{P} = 100$, $\mathsf{C} = 6.5$, and different variances σ_F^2 and σ_G^2 of F and G, and for static G and F.

III. Main Results

For each $\beta \in [0, 1]$ define $\sigma_\beta^2 \geq 0$ as the unique positive real number satisfying

$$\mathbb{E}\left[\log\left(1 + \frac{1 + (1-\beta)\mathsf{P}|G_0|^2 + |F_0|^2\mathsf{P})}{\sigma_\beta^2}\right)\right] = \mathsf{C} \quad (16)$$

and the random process $\{W_i^\beta\}_{i=-\infty}^\infty$ as the unique stationary process satisfying:

$$W_i^\beta = \left(1 + (1-\beta)\mathsf{P}|G_i|^2\left(1 + \frac{(1-\beta)\mathsf{P}|F_i|^2 W_{i-1}^\beta}{1 + \sigma_\beta^2 + \beta\mathsf{P}|F_i|^2}\right)^{-1}\right)^{-1}. \quad (17)$$

Notice that the joint process $\{(F_i, G_i, W_i^\beta)\}_{i=-\infty}^\infty$ is also stationary and ergodic.

For a given $\beta \in [0, 1]$, let $\mathcal{R}(\beta) \subseteq \mathbb{R}^2$ be the set of all non-negative pairs $(R^{(F)}, R^{(S)})$ that satisfy

$$R^{(F)} \leq \mathbb{E}\left[\log\left(1 + \frac{\beta\mathsf{P}|G_0|^2}{(1-\beta)\mathsf{P}|G_0|^2 + \mathsf{P}|F_0|^2 + 1}\right)\right] \quad (18a)$$

and

$$R^{(S)} \leq \mathbb{E}\Big[\log\left(1 + \frac{(1-\beta)\mathsf{P}|F_0|^2 W_{-1}^\beta}{1 + \sigma_\beta^2 + \beta\mathsf{P}|F_0|^2}\right) - \log W_0^\beta\Big]. \quad (18b)$$

Theorem 1 (Capacity Inner Bound): The convex closure of the sets $\{\mathcal{R}(\beta) \colon \beta \in [0, 1]\}$ is achievable:

$$\text{conv cl}\left(\bigcup_{\beta\in[0,1]} \mathcal{R}(\beta)\right) \subseteq \mathcal{C}(\mathsf{P}, \mathsf{C}). \quad (19)$$

Proof: See Section IV. ■

Theorem 2 (Capacity Outer Bound): Assuming $\log|G_0|$ and $\log|F_0|$ are integrable near 0, any rate pair $(R^{(F)}, R^{(S)})$ in the capacity region $\mathcal{C}(\mathsf{P}, \mathsf{C})$ satisfies the following four constraints:

$$\begin{aligned} 2R^{(F)} + R^{(S)} \leq{} & \mathbb{E}\big[\log(1 + (|G_0|^2 + |F_0|^2)\mathsf{P})\big] \\ & + \mathbb{E}\left[\log\left(1 + \frac{|F_0|^2}{|G_0|^2}\right)\right] \\ & + \max\left\{\mathbb{E}\left[\log\left(\frac{|G_0|^2}{|F_0|^2}\right)\right], 0\right\}, \end{aligned} \quad (20a)$$

$$\begin{aligned} R^{(F)} + R^{(S)} \leq{} & \frac{1}{2}\mathbb{E}\big[\log(1 + (|G_0|^2 + |F_0|^2)\mathsf{P})\big] \\ & + \frac{1}{2}\max\left\{\mathbb{E}\left[\log\left(\frac{|G_0|^2}{|F_0|^2}\right)\right], 0\right\} \\ & + \frac{1}{2}\mathbb{E}\big[\log(1 + (|F_0|^2)^{-1})\big] + \frac{\mathsf{C}}{2}, \end{aligned} \quad (20b)$$

$$R^{(F)} \leq \frac{1}{2}\mathbb{E}\big[\log(1 + |G_0|^2\mathsf{P})\big], \quad (20c)$$

$$R^{(S)} \leq \mathsf{C}. \quad (20d)$$

Proof: Omitted due to space limitations. ■

Corollary 1 (Multiplexing Gain Region): The multiplexing gain region $\mathcal{S}^\star(\mu)$ is the set of all nonnegative pairs $(\mathsf{S}^{(F)}, \mathsf{S}^{(S)})$ satisfying

$$2\mathsf{S}^{(F)} + \mathsf{S}^{(S)} \leq 1, \quad (21a)$$

$$\mathsf{S}^{(S)} \leq \mu. \quad (21b)$$

Proof: The converse holds by Theorem 2 and the achievability by Theorem 1. Specifically, for the achievability part, it suffices to prove that the two pairs

$$\big(\mathsf{S}^{(S)} = 0, \quad \mathsf{S}^{(F)} = 1/2\big), \quad (22)$$

$$\big(\mathsf{S}^{(S)} = \min\{\mu, 1\}, \quad \mathsf{S}^{(F)} = \max\{0, 1/2 - \mu/2\}\big) \quad (23)$$

are achievable. The multiplexing gain pair in (22) can be achieved by silencing every second UE, which decomposes the network into $K/2$ non-interfering point-to-point links. If $\mu \geq 1$, the multiplexing gain pair $(\mathsf{S}^{(S)} = 1, \mathsf{S}^{(F)} = 0)$ is achieved by a scheme where each BS quantizes its observed outputs to noise level and the CU decodes all the transmitted "slow" messages based on these quantized outputs. If $\mu < 1$, then the multiplexing gain pair $(\mathsf{S}^{(S)} = \mu, \mathsf{S}^{(F)} = 1/2 - \mu/2)$ in (23) is achieved by a scheme that time-shares the schemes above over fractions of $1 - \mu$ and μ of the time. ■

Figure 2 illustrates the proposed inner and outer bounds on the capacity region for independent random processes $\{G_i\}$ and $\{F_i\}$, where each F_i is circularly Gaussian of variance σ_F^2 and each G_i is circularly Gaussian of variance σ_G^2. Numerical simulations are performed for different values of σ_F^2. The figure also presents inner and outer bounds on the capacity region assuming static channel coefficients (regions in red). As can be seen from the figure, for small values of $R^{(F)}$, the slope $\frac{\delta R^{(F)}}{\delta R^{(S)}}$ of the inner bound is approximately -1 both for static and random channel coefficients. This means increasing the rate of "fast" messages by Δ, decreases the rate of "slow" messages by Δ and thus the sum-rate remains constant. For large values of $R^{(F)}$ and random time-varying channel coefficients, the slope of the inner bound is around -3.5 for $\sigma_F^2 = 0.2$ and around -4 for $\sigma_F^2 = 0.3$. In contrast, this slope is around -2.7 for static channel coefficients. Increasing an already large "fast" rate $R^{(F)}$ thus penalizes the sum-rate of the system and is more pronounced under random fading.

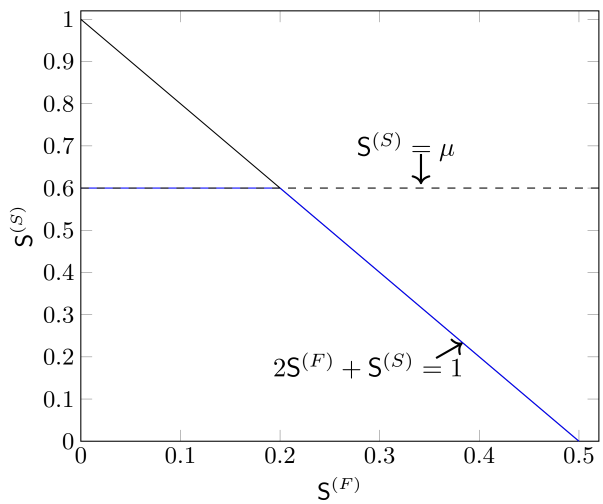

Fig. 3. Region $\mathcal{S}^\star(\mu)$ for $\mu = 0.6$ in blue and $\mu \geq 1$ in black.

Figure 3 shows the multiplexing gain region for different values of μ. We notice that when $\mu < 1$, for $\mathsf{S}^{(F)} \leq \frac{1}{2} - \frac{\mu}{2}$, the multiplexing gain of "slow" messages is constant and solely limited by the fronthaul capacity. In this regime, the sum-multiplexing gain of the system is increased by decoding parts of the messages directly at the BSs. When $\mu < 1$ and $\mathsf{S}^{(F)} > \frac{1}{2} - \frac{\mu}{2}$, or when $\mu \geq 1$, the slope of the boundary of the region is -2. In these regimes the maximum sum-multiplexing gain is decreased by Δ when the "fast" multiplexing gain increases by Δ.

IV. Proof of Theorem 1

Fix $\beta \in [0, 1]$.

Recall that σ_β^2 was defined so as to verify (see (16))

$$\mathbb{E}\left[\log\left(1 + \frac{1 + (1-\beta)\mathsf{P}|G_0|^2 + |F_0|^2\mathsf{P})}{\sigma_\beta^2}\right)\right] = \mathsf{C}. \quad (24)$$

For each $k \in \{1, \ldots, K\}$, define

$$X_k = U_k + V_k, \quad (25a)$$
$$Y_k = G_k X_k + F_k X_{k-1} + Z_k, \quad (25b)$$
$$C_k = G_k V_k + F_k X_{k-1} + Z_k, \quad (25c)$$
$$\hat{C}_k = C_k + Q_k, \quad (25d)$$

for $\{Q_k\}$, $\{U_k\}$, $\{V_k\}$, and $\{Z_k\}$ independent zero-mean Gaussian random variables with variances σ_β^2, $\beta\mathsf{P}$, $(1-\beta)\mathsf{P}$, and 1.

Random Code Construction: For each $k \in \{1, \ldots, K\}$, generate codebooks $\mathcal{C}_{u,k}$, $\mathcal{C}_{v,k}$, and $\mathcal{C}_{w,k}$ randomly. Codebook

$$\mathcal{C}_{u,k} := \left\{u_k^n(i)\colon \quad i = 1, \ldots, \left\lfloor 2^{nR_k^{(F)}} \right\rfloor\right\} \quad (26)$$

is generated by picking all entries i.i.d. circularly Gaussian of variance βP. Independently thereof, codebook

$$\mathcal{C}_{v,k} := \left\{v_k^n(j)\colon \quad j = 1, \ldots, \left\lfloor 2^{nR_k^{(S)}} \right\rfloor\right\} \quad (27)$$

is generated by picking all entries i.i.d. circularly Gaussian of variance $(1-\beta)P$. Quantization codebook

$$\mathcal{C}_{c,k} := \left\{\hat{c}_k^n(\ell)\colon \quad \ell = 1, \ldots, \left\lfloor 2^{n\mathcal{C}} \right\rfloor\right\} \quad (28)$$

is generated by picking all entries i.i.d. according to $P_{\hat{C}}$.

Reveal all codebooks to all terminals. We explain the encoding and decoding operations assuming that

$$G_k^n = g_k^n \qquad \text{and} \qquad F_k^n = f_k^n. \quad (29)$$

(Recall that BS k and the CU know these realizations.)

UE k: Sends

$$x^n = u_k^n\big(M_k^{(F)}\big) + v_k^n\big(M_k^{(S)}\big). \quad (30)$$

BS k: Decodes its "fast" message $M_k^{(F)}$ based on its own channel outputs $Y_k^n = y_k^n$. It then looks for a unique $\hat{i}_k$ such that

$$(u_k^n(\hat{i}_k), y_k^n, g_k^n, f_k^n) \in \mathcal{A}_\epsilon^{(n)}(P_{UYG_0F_0}), \quad (31)$$

where $\mathcal{A}_\epsilon^{(n)}(\cdot)$ refers to the jointly typical set as defined in [14] and where given $G_0 = g$ and $F_0 = f$ the pair (U, Y) is a centered bivariate Gaussian vector of covariance matrix

$$\mathsf{K}_{UY|g,f} = \begin{pmatrix} \beta\mathsf{P} & g\beta\mathsf{P} \\ g\beta\mathsf{P} & (g^2 + f^2)\mathsf{P} + 1 \end{pmatrix}. \quad (32)$$

If none or more than one such indices $\hat{i}_k$ exist, BS k declares an error. Otherwise it declares

$$\hat{M}_k^{(F)} = \hat{i}_k. \quad (33)$$

Subsequently it forms the difference

$$c_k^n := y_k^n - u_k^n(\hat{i}_k), \quad (34)$$

and looks for an index ℓ_k such that

$$(\hat{c}_k^n(\ell_k), c_k^n, f_k^n, g_k^n) \in \mathcal{A}_\epsilon^{(n)}(P_{\hat{C}CFG}). \quad (35)$$

If none or more than one such indices ℓ_k exist, BS k declares an error. Otherwise it sends $L_k = \ell_k$.

CU: Assume it receives $L_1 = l_1, \ldots, L_K = l_K$. Then, it looks for a unique set of indices $\hat{j}_1, \ldots, \hat{j}_K$ that satisfy

$$\big(\hat{c}_1^n(\ell_1), \ldots, \hat{c}_K^n(\ell_K), v_1^n(\hat{j}_1), \ldots, v_K^n(\hat{j}_K), g_1^n, \ldots, g_K^n, f_1^n, \ldots, f_K^n\big) \in \mathcal{A}_\epsilon^{(n)}(P_{\hat{C}_1,\ldots,\hat{C}_K,V_1,\ldots,V_K,\mathbb{G},\mathbb{F}}), \quad (36)$$

where

$$\mathbb{G} = \{G_{1,t}, \ldots, G_{K,t}\}_{t=1}^n, \quad \mathbb{F} = \{F_{1,t}, \ldots, F_{K,t}\}_{t=1}^n. \quad (37)$$

If none or multiple such indices $\hat{j}_1, \ldots, \hat{j}_K$ exist, the CU declares an error. Otherwise, it declares

$$\hat{M}_k^{(S)} = \hat{j}_k, \qquad k \in \{1, \ldots, K\}. \quad (38)$$

Analysis: For any $k \in \{1, \ldots K\}$, decoding in (31) is successful with probability tending to 1 as $n \to \infty$, if

$$R_k^{(F)} < I(U_k; Y_k|G_k, F_k) = \mathbb{E}\left[\log\left(1 + \frac{\beta\mathsf{P}|G_k|^2}{(1-\beta)\mathsf{P}|G_k|^2 + \mathsf{P}|F_k|^2 + 1}\right)\right]. \quad (39)$$

Assuming that this decoding was successful, quantization in (35) also succeeds with probability tending to 1 as $n \to \infty$, because by the choice of the quantization noise σ_β^2:

$$\begin{aligned}\mathcal{C} &\geq I(\hat{C}_k; C_k | G_k, F_k) \\ &= \mathbb{E}\left[\log\left(1 + \frac{1 + (1-\beta)\mathsf{P}|G_k|^2 + |F_k|^2\mathsf{P})}{\sigma_\beta^2}\right)\right]. \end{aligned} \tag{40}$$

Now assuming that both the decoding in (31) and the quantization in (35) were successful, the decoding in (36) also succeeds with probability tending to 1 as $n \to \infty$, if

$$\begin{aligned}&\frac{1}{K}\sum_{k=1}^{K} R_k^{(S)} \\ &< \frac{1}{K} I(V_1, \ldots, V_K; \hat{C}_1, \ldots, \hat{C}_K | \mathbb{G}, \mathbb{F}) \\ &= \frac{1}{K} I\Big(\{S_k\}_{k=1}^K\ ;\ \{\tilde{G}_k S_k + \tilde{F}_k S_{k-1} + \Xi_k\}_{k=1}^K \Big| \mathbb{G}, \mathbb{F}\Big), \end{aligned} \tag{41}$$

where $\{\tilde{Z}_k\}$ and $\{\Xi_k\}$ are sequences of i.i.d. circularly symmetric Gaussian noises of variances $1 + \sigma_\beta^2 + f_k^2 \beta P$ and 1 and

$$S_k := \sqrt{(1-\beta)^{-1/2} P} \cdot V_k, \tag{42a}$$

$$\tilde{G}_k := \frac{\sqrt{(1-\beta)P}}{\sqrt{1 + \sigma_\beta^2 + \beta P |F_k|^2}} \cdot G_k, \tag{42b}$$

$$\tilde{F}_k := \frac{\sqrt{(1-\beta)P}}{\sqrt{1 + \sigma_\beta^2 + \beta P |F_k|^2}} \cdot F_k, \quad k \in \{1, \ldots, K\}. \tag{42c}$$

By these definitions and assumptions on the channel, the sequences $\{S_k\}$, $\{(\tilde{G}_k, \tilde{F}_k)\}$, $\{\Xi_k\}$ satisfy the conditions in [13, Assumption 1] (where we associate G_k, F_k, and V_k in [13] with $\tilde{G}_k$, $\tilde{F}_k$, and Ξ_k in this manuscript). Therefore in the limit when $K \to \infty$, the mutual information in (41) can be evaluated using [13, Theorem1] to obtain:

$$R^{(S)} \leq \mathbb{E}\left[\log\left(1 + \frac{(1-\beta)\mathsf{P}|F_0|^2 W_{-1}^\beta}{1 + \sigma_\beta^2 + \beta\mathsf{P}|F_0|^2}\right) - \log W_0^\beta\right] \tag{43}$$

where $\{W_i\}_{i=-\infty}^{\infty}$ is the unique stationary process satisfying

$$W_i^\beta = \left(1 + (1-\beta)\mathsf{P}|G_i|^2 \Big(1 + \frac{(1-\beta)\mathsf{P}|F_i|^2 W_{i-1}^\beta}{1 + \sigma_\beta^2 + \beta\mathsf{P}|F_i|^2}\Big)^{-1}\right)^{-1}. \tag{44}$$

Achievability of the pairs (18) follows then from (39) and (43).

V. Conclusion

We presented inner and outer bounds on the capacity region of a C-RAN under mixed delay constraints and characterized the multiplexing gain region of this network. We obtained the following conclusions. When the fronthaul capacities are small, then the overall performance of the system can be improved if some of the data streams (the delay sensitive streams) are directly decoded at the BSs. The stringent delay constraint on these streams however becomes harmful when their rate is too large. In this regime the total sum-rate has to be decreased by a penalty factor γ times Δ when the delay-sensitive rate is increased by Δ. The penalty factor $\gamma \approx 1$ for static channel coefficients or in the high-SNR regime, and it can be significantly larger for random channel coefficients and at moderate SNRs.

To reduce the gap between the proposed inner and outer bounds on capacity, in future works we plan to include more sophisticated multi-user compression techniques [4], [5] at the BSs and the cloud processor.

Acknowledgement

The work of H. Nikbakth and M. Wigger was supported by the ERC under grant agreement no. 715111. The work of S. Shamai has been supported by the European Union's Horizon 2020 Research And Innovation Programme, grant agreement no. 694630.

References

[1] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges, *IEEE Wireless Communications*, vol. 22, no. 2, pp. 152–160, Apr. 2015.

[2] M. Shafi et al., "5G: A tutorial overview of standards, trials, challenges, deployment, and practice, *IEEE Journ. Sel. Areas in Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[3] T. Quek, M. Peng, O. Simeone, and W. Yu, "Cloud radio access networks: principles, technologies, and applications," *Cambridge University Press*, 2017.

[4] I. E. Aguerri, A. Zaidi, G. Caire, and S. Shamai (Shitz), "On the capacity of cloud radio access networks with oblivious relaying, *arXiv:1710.09275v3*, Jan. 2019.

[5] S. Park, O. Simeone, O. Sahin and S. Shamai (Shitz) , "Robust and Efficient Distributed Compression for Cloud Radio Access Networks, *IEEE Trans. Vehicular. Tech.*, vol. 62, no. 2, pp. 692–703, Feb. 2013.

[6] R. Kassab, O. Simeone and P. Popovski, "Coexistence of URLLC and eMBB services in the C-RAN uplink: an information-theoretic study," *arXiv:1804.06593*, Apr. 2018.

[7] H. Nikbakht, M. Wigger and S. Shamai (Shitz), "Mixed delay constraints in Wyner's soft-handoff network," in *Proc. IEEE ISIT 2018*, Vail, CO, USA, June 17–22, 2018, pp. 1171–1175.

[8] H. Nikbakht, M. Wigger and S. Shamai (Shitz), "Mixed delay constraints at maximum sum-multiplexing gain," in *Proc. IEEE ITW 2018*, Guangzhou, China, Nov 25–29, 2018, pp. 285–289.

[9] H. Nikbakht, M. Wigger and S. Shamai (Shitz), "Multiplexing gain region of sectorized cellular networks with mixed delay constraints," *arXiv:1902.11231*, Feb. 2019.

[10] K. M. Cohen, A. Steiner, and S. Shamai (Shitz) "The broadcast approach under mixed delay constraints," in *Proc. IEEE ISIT 2012*, Cambridge (MA), USA, July 1–6, 2012, pp. 209–213.

[11] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713–1727, Nov. 1994.

[12] O. Simeone, N. Levy, A. Sanderovich, O. Somekh, B. M. Zaidel, H. V. Poor and S. Shamai (Shitz), "Cooperative wireless cellular systems: an information-theoretic view," *Found. and Trends in Comm. and Inf. Theory (FnT),* Vol. 8, no. 1–2, 2011, pp. 1–177, Now Publishers, 2012.

[13] W. Hachem, A. Hardy, and S. Shamai (Shitz), "Mutual information of wireless channels and block-Jacobi ergodic operators," *arXiv:1811.04734v2*, Mar. 2019.

[14] T. M. Cover and J. Y. Thomas. *Elements of Information Theory*, Wiley-Interscience, New York, 2006.