Control-Data Separation with Decentralized Edge Control in Fog-Assisted Uplink Communications

Jinkyu Kang, Osvaldo Simeone, Joonhyuk Kang and Shlomo Shamai (Shitz)

Abstract

Fog-aided network architectures for 5G systems encompass wireless edge nodes, referred to as remote radio systems (RRSs), as well as remote cloud center (RCC) processors, which are connected to the RRSs via a fronthaul access network. RRSs and RCC are operated via Network Functions Virtualization (NFV), enabling a flexible split of network functionalities that adapts to network parameters such as fronthaul latency and capacity. This work focuses on uplink communications and investigates the cloud-edge allocation of two important network functions, namely the control functionality of rate selection and the data-plane function of decoding. Three functional splits are considered: *(i)* Distributed Radio Access Network (D-RAN), in which both functions are implemented in a decentralized way at the RRSs; *(ii)* Cloud RAN (C-RAN), in which instead both functions are carried out centrally at the RCC; and *(iii)* a new functional split, referred to as Fog RAN (F-RAN), with separate decentralized edge control and centralized cloud data processing. The model under study consists of a time-varying uplink channel in which the RCC has global but delayed channel state information (CSI) due to fronthaul latency, while the RRSs have local but more timely CSI. Using the adaptive sum-rate as the performance criterion, it is concluded that the F-RAN architecture can provide significant gains in the presence of user mobility.

Jinkyu Kang is with the School of Engineering and Applied Sciences (SEAS), Harvard University, Cambridge, MA 02138, USA (Email: jkkang@g.harvard.edu).

Osvaldo Simeone is with the Department of Informatics, King's College London, London, UK (Email: osvaldo.simeone@kcl.ac.uk).

Joonhyuk Kang is with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST) Daejeon, South Korea (Email: jhkang@ee.kaist.ac.kr).

Shlomo Shamai (Shitz) is with the Department of Electrical Engineering, Technion, Haifa, 32000, Israel (Email: sshlomo@ee.technion.ac.il).

Index Terms

Cloud-Radio Access Network (C-RAN), Fog-Radio Access Network (F-RAN), fronthaul, control data separation, 5G, Network Functions Virtualization (NFV).

I. INTRODUCTION

The evolution of the wireless network architecture traces a line from the decentralized implementation of control and data functionalities in conventional Distributed Radio Access Network (D-RAN) through the centralization of the protocol stack in Cloud-RAN (C-RAN) [2], [3] to the more recent fog-aided proposals with flexible functional splits between cloud and edge nodes [4]. An important motivation for the latest shift to fog-aided solutions is the realization that a fully centralized C-RAN system entails significant, and possibly prohibitive, requirements on the fronthaul connections between edge nodes and cloud, see, e.g., [5], [6] and references therein. Furthermore, the development of the Network Functions Virtualization (NFV) technology makes adaptive cloud-edge functional splits realizable via software [7].

The demarcation line between the functionalities to be implemented at the cloud and at the edge is typically drawn to include a given number of physical-layer functions at the edge nodes, such as synchronization, FFT/IFFT and resource demapping [5], [8]. The body of work concerned with edge-cloud functional splits generally aims at assessing the trade-off between performance and fronthaul capacity overhead of different demarcation lines.

In light of these developments, references [9]–[11] explore the application of the *data-control separation architecture* [12] as the guiding principle underlying the separation of functionalities between edge and cloud with the aim of addressing fronthaul latency limitations. Specifically, [9] puts forth the idea of performing the control decisions of the uplink hybrid automatic repeat request (HARQ) protocol at an edge node, while keeping the computationally expensive operation of data decoding at the cloud processor. As shown in [10], [11], [13], this approach can yield significant reductions in transmission latency thanks to the capability of the edge nodes to provide quick feedback to the mobile users with limited fronthaul

overhead.

An important lesson learned from [9]–[11], [13] is that the implementation of some control function-alities at the edge can be an enabler for the reduction of transmission latency even in the presence of significant delays on the fronthaul links. A work that provides related insights in the different set-up of a multi-hop network with orthogonal links is [14]. Reference [14] shows that centralized scheduling based on delayed channel state information (CSI) can be outperformed by local scheduling decisions, as long as each network node has more current CSI of its incoming and outgoing links with respect to the centralized scheduler.

In this work, as illustrated in Fig. 1, we study the optimal functional split of control and data plane functionalities at the edge nodes, referred to as remote radio systems (RRSs) [3], and at the cloud, referred to as remote cloud center (RCC) [3], for uplink communication. We specifically focus on the following functionalities: *(i)* the *control plane* functionality of the data rate selection, and *(ii)* the *data plane* functionality of data decoding. We aim at assessing the impact of fronthaul latency on the relative performance of different splits, whereby rate selection and data decoding may be performed separately at either cloud or edge.

As summarized in Fig. 1, we specifically consider three functional splits: *(i)* D-RAN, in which both rate selection and data decoding are implemented at each edge; *(ii)* C-RAN, whereby both rate selection and data decoding are instantiated at cloud; and *(iii)* Fog-RAN (F-RAN), whereby the control function of rate selection is performed at the edge, while data decoding is implemented at cloud. The latter functional split is studied here for the first time. The approach is motivated by the idea discussed above of leveraging decentralized control to counteract fronthaul delays. We remark that the label "F-RAN" has been used in works such as [15] to indicate systems with decentralized caching at the RRSs and centralized processing at the RCC. Here we suggest to use the term more generally to describe fog-based solutions involving both cloud and edge operations.

As seen in Fig. 2, the model under study consists of a time-varying uplink channel in which the RCC
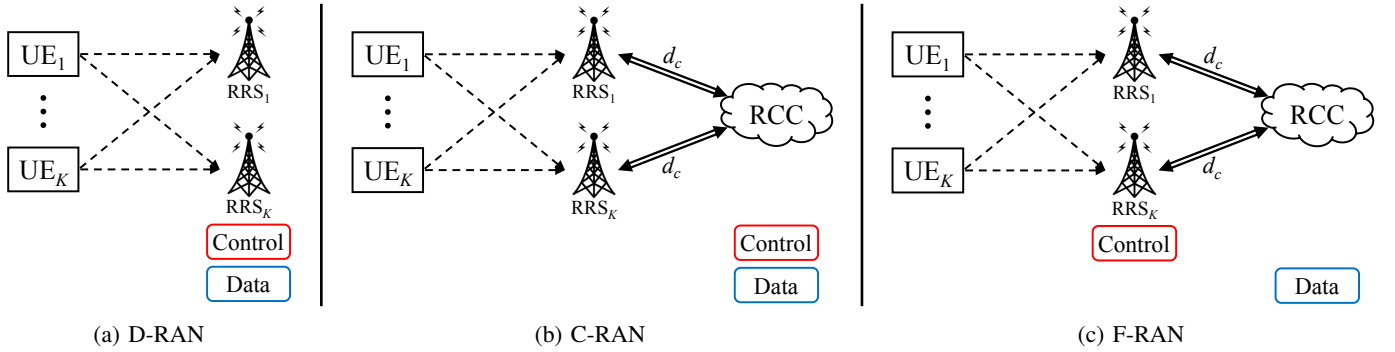
Fig. 1. Three radio access network architectures for different control-data functional splits between cloud and edge: (a) D-RAN, (b) C-RAN, and (c) F-RAN.

processor has global but delayed CSI due to fronthaul latency, while the RRSs have local CSI with a lower delay. Using the adaptive sum-rate as the performance criterion (see, e.g., [16]), the mentioned functional splits based on the control-data separation architecture are compared through analysis and numerical results.

The rest of the paper is organized as follows. We describe the system model and performance metric in Sec. II. We analyze the three radio access network architectures in Fig. 1 for different control-data functional splits between RCC and RRSs: D-RAN in Sec. III, C-RAN in Sec. IV, and F-RAN in Sec. V. In Sec. VI, numerical results are presented. Concluding remarks are summarized in Sec. VII.

II. SYSTEM MODEL AND PERFORMANCE METRIC

We consider the uplink of a fog-assisted system illustrated in Fig. 2, which consists of K remote radio systems (RRSs), a remote cloud center (RCC), and K active user equipments (UEs). We assume that user-cell association has been carried out, so that each UE i is associated to a given RRS i, and we have the same number of active UEs and RRSs. We denote the set of all UEs and RRSs as $\mathcal{K} = \{1, \ldots, K\}$. As further detailed below and illustrated in Fig. 1, we consider three different cloud-edge splits, namely: *(i)* D-RAN: The RCC is not present and both rate selection and data decoding for UE k are carried out at RRS k; *(ii)* C-RAN: The RCC implements both rate selection and data decoding for all UEs; *(iii)* F-RAN: In this novel functional split, the RRS k performs rate selection for UE k while data decoding for all UEs is performed at the RCC.

$$S_1(t-d_e), \{I_{1k}(t-d_e)\}$$

$$S_1(t)$$
$$I_{1K}(t)$$
RRS$_1$
$$d_c$$
RCC
$$I_{K1}(t)$$
$$S_K(t)$$
RRS$_K$
$$d_c \quad \{S_k(t-d)\}, \{I_{jk}(t-d)\}$$
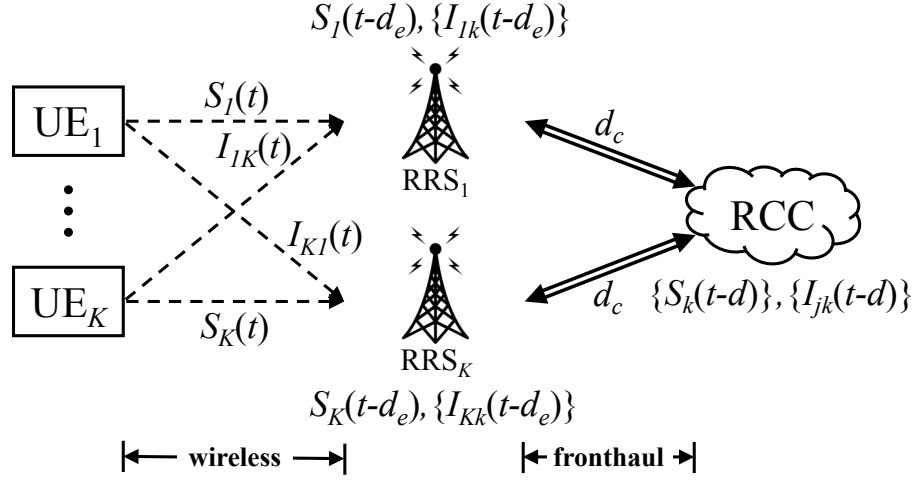$$S_K(t-d_e), \{I_{Kk}(t-d_e)\}$$

|← wireless →| |← fronthaul →|

Fig. 2. Uplink of the considered fog-assisted system. Communication on the fronthaul links entails a two-way latency of d_c time slots (of the wireless interface), while the time elapsed between CSI measurement at the edge and uplink scheduling is d_e time slots. The figure indicates the CSI available when selecting the rates for transmitting in slot t at the RRSs for D-RAN and F-RAN and at the RCC for C-RAN. Note that the CSI available for decoding can be assumed to be timely since it can be estimated from pilots included in the data frame, as seen in Fig. 3.

An important role in the analysis is played by the timeliness of the CSI available at the RRS and RCC at the time of rate selection. In particular, as illustrated in Fig. 3, for D-RAN and F-RAN, we assume that the latency between uplink training transmission or CSI feedback and the time slot allocated for uplink transmission equals d_e time slots of the wireless channel. As an example, the latency contributions for uplink transmission in LTE Release 14 [17] are Scheduling Request (SR) periodicity, uplink scheduling delay and uplink grant transmission. For a transmission time interval (TTI) of, say, $0.5-1$ ms, the latency d_e can be large as $1-2$ slots [17]. For C-RAN, in addition to the delay d_e, one needs to consider the two-way communication between RRSs and RCC on the fronthaul. This entails a latency equal to d_c time slots. The fronthaul transport latency is reported to be around 0.25 ms in [18] for single-hop fronthaul links and can amount to multiple milliseconds in the presence of multihop fronthauling, while fronthaul-related processing at the RCC can take fractions to a few milliseconds [19]. As a result, for TTI of $0.5-1$ ms, the RCC CSI latency d_c can be as large as $3-5$ slots.
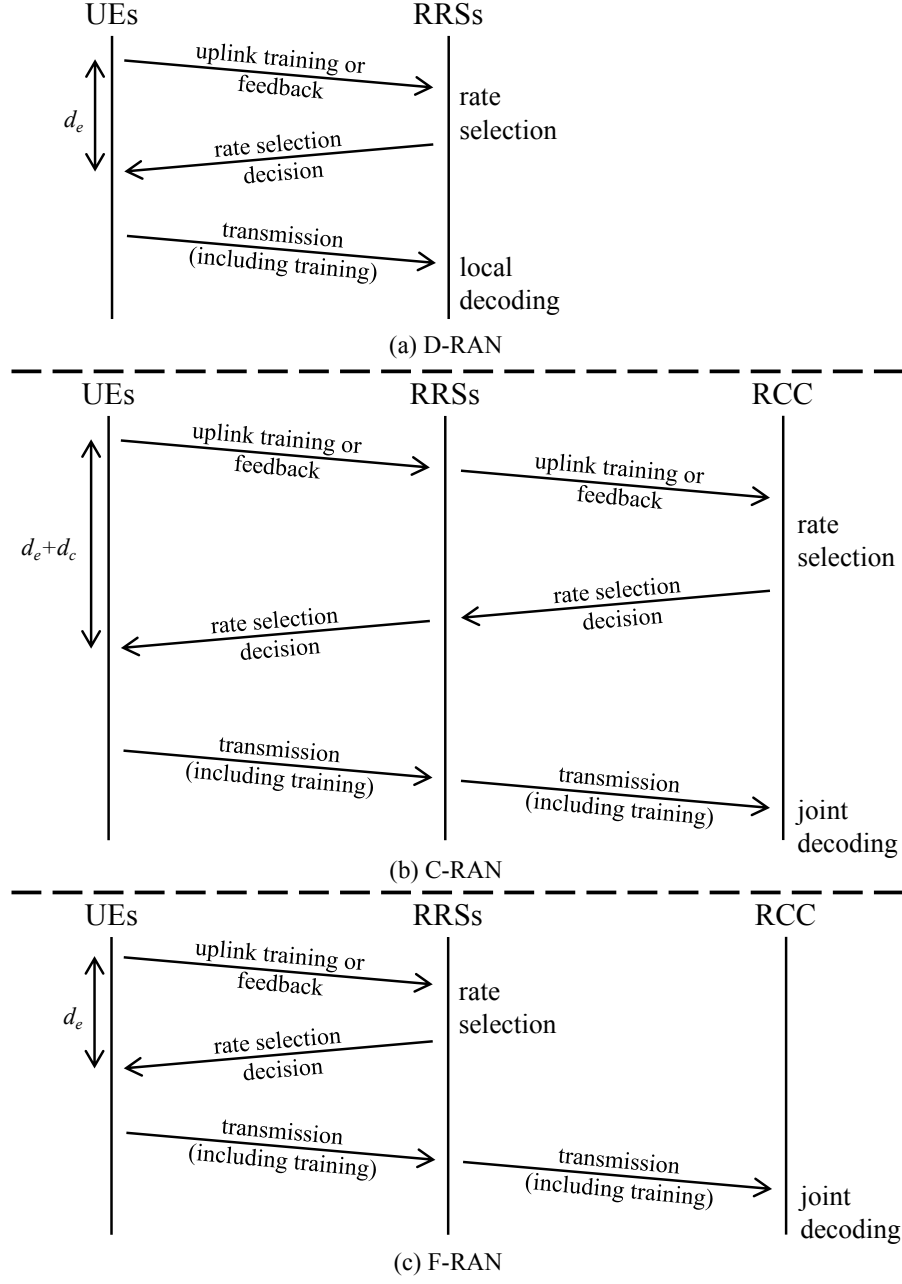
Fig. 3. Protocol timeline (time increasing from top to bottom) for: (a) D-RAN; (b) C-RAN; and (c) F-RAN. For D-RAN (a) and F-RAN (c), the latency between uplink training transmission or CSI feedback and the time slot allocated for uplink transmission equals d_e time slots of the wireless channel. For C-RAN (b), in addition to the scheduling delay d_e, one needs to consider the latency associated with two-way fronthaul communication between RRSs and RCC, which equals d_c time slots.

A. Channel Model

At each time slot t, the instantaneous power received at RRS i from UE i is denoted as $S_i(t)$, while the received power for the cross-channel between an UE i and the RRS $j \neq i$ is denoted as $I_{ji}(t)$. These are assumed to vary across the time index $t = 1, 2, \ldots, T$, which runs over the transmission intervals,
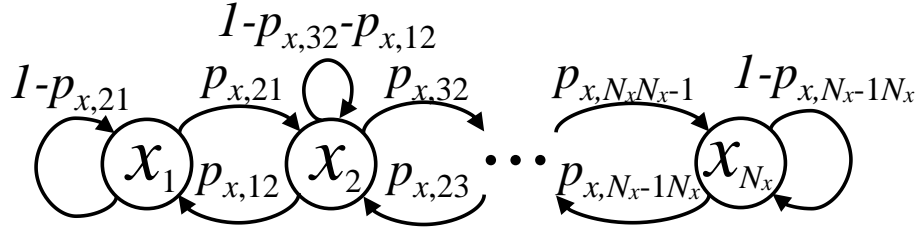
Fig. 4. Markov model for the direct channel and cross-channel instantaneous power processes with N_x states for $x \in \{S, I\}$ [20].

according to a Markov model. This model can be obtained, for instance, by approximating the standard Clarke's model via quantization, see, e.g., [20]. The channel matrix between the UEs and the RRSs at any channel use k of the transmission interval $t = 1, 2, \ldots, T$ can be written as

$$\mathbf{H}(t, k) = [\mathbf{h}_1^T(t, k), \ldots, \mathbf{h}_K^T(t, k)]^T, \tag{1}$$

with $\mathbf{h}_j(t, k) = [\sqrt{I_{j1}(t)}e^{j\theta_{j1}(t,k)}, \ldots, \sqrt{S_j(t)}e^{j\theta_{jj}(t,k)}, \ldots, \sqrt{I_{jK}(t)}e^{j\theta_{jK}(t,k)}]$, where the phases $\theta_{ji}(t, k)$ are uniformly distributed in the interval $[0, 2\pi)$, mutually independent as per the standard Rayleigh fading model, and vary in an ergodic manner over the channel use index k within each transmission interval t.

As illustrated in Fig. 4, the direct instantaneous fading power $S_i(t)$ can take N_S values, indexed in ascending order as $\{S_1, \ldots, S_{N_S}\}$, and is governed by a Markov chain with transition probabilities $p_{S,mn} = \Pr[S_i(t+1) = S_m | S_i(t) = S_n]$. In a similar manner, the cross-channel power $I_{ji}(t)$ can take N_I values, indexed in ascending order as $\{I_1, \ldots, I_{N_I}\}$, and varies according to a Markov chain with transition probabilities $p_{I,mn} = \Pr[I_{ji}(t+1) = I_m | I_{ji}(t) = I_n]$. We denote the set of all states for the direct channel as $\mathcal{N}_S = \{S_1, \ldots, S_{N_S}\}$ and for the cross-channel as $\mathcal{N}_I = \{I_1, \ldots, I_{N_I}\}$. We recall that Markovian models are widely adopted for the evaluation of the performance of wireless systems (see, e.g., [20]–[22]). Note that, as in [20], channel variations can only occur between adjacent states, i.e., $p_{x,mn} = 0$ if $|m - n| > 1$ for $x \in \{S, I\}$. Details on the quantization process from Clarke's model, which is assumed for the numerical results presented in Sec. VI, can be found in Appendix A.

We conclude this subsection by introducing some useful notation. According to the adopted Markov model, the probability that the direct channel state changes from state S_n to the state S_m, for $S_m, S_n \in \mathcal{N}_S$,

after d transmission intervals can be written as

$$\Pr\left[S_i(t) = S_m | S_i(t-d) = S_n\right] = \beta_S^{m|n}(d), \tag{2}$$

where the probability $\beta_S^{m|n}(d)$ is obtained as the (m, n) entry of the matrix $\mathbf{T}_S^d$, with the transition matrix $\mathbf{T}_S$ having $p_{S,mn}$ as the (m, n) entry, i.e., $[\mathbf{T}_S]_{m,n} = p_{S,mn}$. Moreover, the stationary probability $\pi_{S,m}$ for the state S_m is obtained by solving the linear system as (see, e.g., [23])

$$\pi_{S,m} = \sum_{S_n \in \mathcal{N}_S} \pi_{S,n} p_{S,mn}, \tag{3}$$

for $S_m \in \mathcal{N}_S$. Analogously, we define $\beta_I^{m|n}(d)$ as the d-step transition probability for the interference process, i.e., $\Pr\left[I_{ji}(t) = I_m | I_{ji}(t-d) = I_n\right] = \beta_I^{m|n}(d)$ with $m, I_n \in \mathcal{N}_I$, and $\pi_{I,m}$ as the steady-state probability of the interference process, i.e., $\pi_{I,m} = \sum_{I_n \in \mathcal{N}_I} \pi_{I,n} p_{I,mn}$ for $I_m \in \mathcal{N}_I$. We also use the notation $\pi_{\mathbf{S}} = \prod_{m \in \mathcal{K}} \pi_{S,m}$ for the joint stationary probability of any given direct channel vector $\mathbf{S} \in \mathcal{N}_S^K$ and we also define $\pi_{\mathbf{I}} = \prod_{i \in \mathcal{K}} \prod_{j \in \mathcal{K}, j \neq i} \pi_{I,I_{ji}}$ for the joint stationary probability of any given cross channel vector $\mathbf{I} \in \mathcal{N}_S^{K-1 \times K}$.

B. Cloud-Edge Functional Splits

As discussed, we focus on the *control* functionality of rate adaptation, or adaptive modulation and coding, that is, the selection of the transmission rates R_j (bit/s/Hz) for any UE $j \in \mathcal{K}$, and on the *data* plane functionality of data decoding. The three control-data functional splits under study (see Fig. 1) are formalized below.

• *Distributed Radio Access Network* (D-RAN): D-RAN amounts to the most conventional cellular implementation in which control and data plane functionalities are carried out at the RRSs, that is, at the edge. Accordingly, for each time slot t, each RRS j selects rate R_j for UE j on the basis of *local* delayed CSI about the direct channel $S_j(t - d_e)$ and about the cross channel $\mathbf{I}_j(t - d_e)$ from all other UEs to the RRS j. This information can be obtained, e.g., by means of uplink training in a Time Division Duplex system or via feedback with Frequency Division Duplex (FDD). Moreover, each RRS j individually performs

decentralized local data decoding of the signal transmitted by UE j by treating interference as noise. Since data packets are assumed to include training signals, we assume that channel decoding at each RRS can leverage current CSI about the data packet.

• *Cloud Radio Access Network* (C-RAN): In the C-RAN architecture, the RCC carries out both control and data processing. Specifically, the RCC selects jointly all rates $\{R_j\}_{j\in\mathcal{K}}$ on the basis of *global* delayed CSI $\{S_j(t-d)\}$ and $\{\mathbf{I}_j(t-d)\}$ for $j \in \mathcal{K}$ about the channels from all UEs to the all RRSs. Note that the delay d includes the additional fronthaul delay d_c between RRSs and RCC as well as the scheduling delay d_e, i.e., $d = d_c + d_e$. Moreover, upon reception of the signals received by the RRSs on the fronthaul links, the RCC performs centralized joint data decoding. Again, CSI for date decoding can be estimated from the training sequences in the packet and hence timely CSI can be assumed for decoding.

• *Fog Radio Access Network* (F-RAN): The novel F-RAN solution is a hybrid implementation with control processing at the edge and data processing at the cloud. In the proposed solution, each RRS j selects the rate R_j based on local delayed CSI as for D-RAN, while the RCC performs centralized joint data decoding on behalf of the RRSs as in C-RAN.

C. Performance Metric

To compare the different functional splits in Fig. 1, we will use the performance metric of the *adaptive sum-rate* (with no power control) used in [16] and references therein. This is defined as the average sum-rate that can be achieved while guaranteeing no outage in each transmission slot. Note that an outage event corresponds to the case that the signal of *at least* one user is not decoded correctly. The average is taken here with respect to the steady-state distribution (7) of the random channel gains $\{S_i(t)\}$ and $\{I_{ji}(t)\}$ for $i, j \in \mathcal{K}$ and $j \neq i$. To ensure that no outage occurs, in each transmission interval, transmission rates $\{R_j\}_{j\in\mathcal{K}}$ for all users are chosen by the RCC or by the RRSs, depending on the functional splits, so that successful decoding can be guaranteed. The adaptive sum-rate is the corresponding achievable average of the sum rates $\sum_{j=1}^{K} R_j$.

More generally, we will consider the *ϵ-outage adaptive sum-rate*, which is defined as the maximum adaptive sum-rate under the constraints a (small) outage probability ϵ is allowed in each slot. We emphasize that an outage event is caused by the imperfect knowledge of the CSI at the time of rate selection.

In the following sections, we analyze the performance in terms of the ϵ-outage adaptive sum-rates of the mentioned control-data functional splits between RCC and RRSs in the presence of the scheduling delay d_e and the fronthaul transmission delay d_c.

III. DISTRIBUTED RADIO ACCESS NETWORK (D-RAN)

In this section, we study the conventional cellular implementation based on D-RAN. Accordingly, for each slot t, each RRS j selects the transmission rate $R_j(t)$ for the user j in its cell based on the available delayed direct channel $S_j(t - d_e)$ and cross-channels $I_{ji}(t - d_e)$ for $i \in \mathcal{K} \setminus \{j\}$. Furthermore, it performs local data decoding by treating interference from the out-of-cell user as noise. As a result, in a D-RAN, an outage event for the j-th RRS/UE pair occurs at time t if the selected rate $R_j(t)$ is larger than the current available rate $C_j(S_j(t), \{I_{ji}(t)\}) \triangleq \log_2(1 + S_j(t)/(1 + \sum_{i=1,i\neq j}^{K} I_{ji}(t))$.

The adaptive outage sum-rate can then be expressed as a function of a conditional CDF of the achievable rates $C_j(S_j(t), \{I_{ji}(t)\})$ for each UE $j \in \mathcal{K}$, where the conditioning is over the delayed CSI $S_j(t - d_e)$ and $I_{ji}(t - d_e)$. This CDF is defined as

$$F_{d_e}(R_j|S_j, \mathbf{I}_j) \triangleq \Pr\left[C_j(S_j(t), \{I_{ji}(t)\}) < R_j \mid S_j(t - d_e) = S_j, I_{ji}(t - d_e) = I_{ji}\right], \qquad (4)$$

where $\mathbf{I}_j = \{I_{ji}\}_{i\in\mathcal{K}}$ is the collection of the states for the cross-channels from all UEs to RRS. The conditional CDF (4) can be computed in terms of the conditional probabilities $\beta_S^{m|S_j}(d_e)$ and $\beta_I^{m|I_{ji}}(d_e)$ as

$$F_{d_e}(R_j|S_j, \mathbf{I}_j) = \sum_{\mathbf{I}_j(t), S_j(t): C_j(t) < R_j} \beta_S^{S_j(t)|S_j}(d_e) \prod_{i\in\mathcal{K}\setminus\{j\}} \beta_I^{I_{ji}(t)|I_{ji}}(d_e), \qquad (5)$$

where we have used the short-hand notation $C_j(t) = C_j(S_j(t), \{I_{ji}(t)\})$.

Proposition 1: With D-RAN, an achievable ϵ-outage adaptive sum-rate is given by

$$R^{\text{D-RAN}}(d_e, \epsilon) = \sum_{j\in\mathcal{K}} \mathbb{E}_{\mathbf{S},\mathbf{I}}\left[F_{d_e}^{-1}(\bar{\epsilon}|S_j, \mathbf{I}_j)\right], \qquad (6)$$

where $F_{d_e}^{-1}(\bar{\epsilon}|S_j, \mathbf{I}_j)$ is the inverse of the conditional CDF (5), the average is taken with respect to the product distribution $\pi_\mathbf{S}\pi_\mathbf{I}$ and $\bar{\epsilon} = 1 - (1 - \epsilon)^{1/K}$.

Proof: If each RRS j chooses rate $R_j = F_{d_e}^{-1}(\bar{\epsilon}|S_j, \mathbf{I}_j)$, it is by construction guaranteed that, when $S_j(t - d_e) = S_j$ and $I_{ji}(t - d_e) = I_{ji}$, the individual probability of outage is no larger than $\bar{\epsilon}$. Since outage events of different users are independent, overall outage probability is no larger than $\epsilon = 1 - (1 - \bar{\epsilon})^K$. ∎

IV. CLOUD RADIO ACCESS NETWORK (C-RAN)

In a C-RAN, at any transmission interval t, the RCC performs rate adaptation in a centralized manner based on the available global and delayed CSI, namely $\{S_j(t - d)\}$ and $\{I_{ji}(t - d)\}$ for all $i, j \in \mathcal{K}$ with $i \neq j$, where the delay $d = d_e + d_c$ includes the edge and fronthaul delays. Furthermore, the RCC performs centralized joint data decoding on behalf of the connected RRSs. Given the complexity of the problem of analyzing the ϵ-outage adaptive sum-rate for C-RAN, we first consider a simplified scenario with two RRS-UE pairs, in which the direct links have fixed fading power and the cross-channel have two states, i.e., $K = 2$, $N_S = 1$, and $N_I = 2$. We then tackle the general case with multiple RRS-UE pairs and multiple channel states.

A. Analysis with two RRSs and UEs

Here, we focus on a simplified scenario with two RRSs and UEs, namely $K = 2$; fixed direct channels $S_i(t) = S$ for $i \in \mathcal{K}$, which may be realized in practice via power control; and cross-channels $I_{12}(t) \triangleq I_1(t)$ and $I_{21}(t) \triangleq I_2(t)$ taking values in a binary set $\mathcal{N}_I = \{I_L, I_H\}$ with $I_H \geq I_L$. Note that the latter assumption implies that the cross-channels can take either a "low" value I_L or a "high" value I_H. To simplify the notation, we set the transition probabilities for the Markov chain describing the variation of the cross-channels as $p_{I,HL} \triangleq p$ and $p_{I,LH} \triangleq q$. Accordingly, the stationary probabilities for the "low" and "high" states of the cross-channels are obtained as

$$\pi_L = \frac{q}{p + q} \quad \text{and} \quad \pi_H = \frac{p}{p + q}, \tag{7}$$

respectively.

To proceed, we define $C(I_1, I_2)$ as

$$C(I_1, I_2) \triangleq \mathbb{E}[\log_2 \det(\mathbf{I} + \mathbf{H}(I_1, I_2)\mathbf{H}^\dagger(I_1, I_2))], \tag{8}$$

where $\mathbf{H}(I_1, I_2) = [\sqrt{S}e^{j\theta_{11}} \ \sqrt{I_1}e^{j\theta_{12}}; \sqrt{I_2}e^{j\theta_{21}} \ \sqrt{S}e^{j\theta_{22}}]$. The expectation in (8) is taken over the random phases $\boldsymbol{\theta} = [\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}]$, which are mutually independent and uniformly distributed in the interval $[0, 2\pi]$. The quantity in (8) is the maximum achievable sum-rate in a time-slot with $I_1(t) = I_1$ and $I_2(t) = I_2$ if joint data decoding is performed at the RCC (see, e.g., [24, Ch. 4]). We will also use the notation C_{xy} for $C(I_x, I_y)$ when $I_1 = I_x$ and $I_2 = I_y$ for $x, y \in \{L, H\}$. We finally observe that $C_{LH} = C_{HL}$.

As discussed above, with C-RAN, the transmission rates $R_1 = R_1(I_1, I_2)$, $R_2 = R_2(I_1, I_2)$ are selected by the RCC based on the available delayed CSI $\{I_1(t-d) = I_1, I_2(t-d) = I_2\}$ and joint data decoding is performed at the RCC. The set of achievable rate pairs (R_1, R_2) with joint decoding at the RCC is given by the capacity region $\mathcal{C}(I_1(t), I_2(t))$ of the ergodic multiple access channel between the two users at the two RRSs. Using standard results in network information theory (see, e.g., [24, Ch. 4]), we have

$$\mathcal{C}(I_1(t), I_2(t)) = \left\{ (R_1, R_2) \left| \begin{array}{l} R_1 \leq \log_2(1 + S + I_2(t)) \\ R_2 \leq \log_2(1 + S + I_1(t)) \\ R_1 + R_2 \leq C(I_1(t), I_2(t)) \end{array} \right. \right\}. \tag{9}$$

The capacity regions $\mathcal{C}_{LL}$, $\mathcal{C}_{LH}$, $\mathcal{C}_{HL}$, $\mathcal{C}_{HH}$ are illustrated in Fig. 5 under different conditions on the channels (S, I_L, I_H). Observing the capacity regions in Fig. 5, we note that, in the case $C_{LH} \leq C_{LL}/2 + \log_2(1 + S + I_L)$, there are achievable rate pairs that maximize the sum-rate in both capacity regions $\mathcal{C}_{LH}$ and $\mathcal{C}_{HL}$, namely the points marked as b and c in Fig. 5(a), while this is not true for case $C_{LH} > C_{LL}/2 + \log_2(1 + S + I_L)$ as can be seen in Fig. 5(b) and Fig. 5(c). This will play a role in the derivation of an achievable ϵ-outage adaptive sum-rate below.
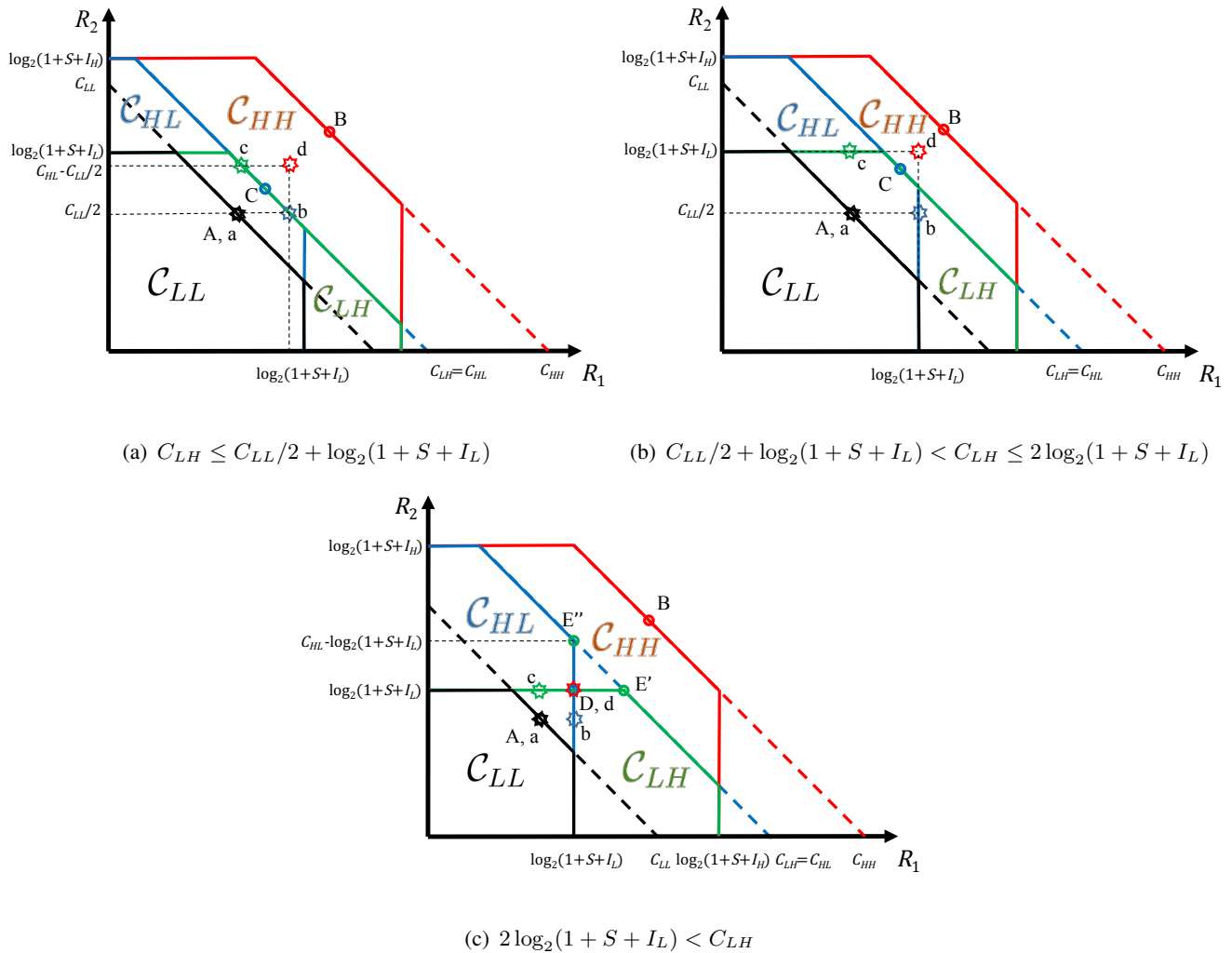
(a) $C_{LH} \leq C_{LL}/2 + \log_2(1 + S + I_L)$

(b) $C_{LL}/2 + \log_2(1 + S + I_L) < C_{LH} \leq 2\log_2(1 + S + I_L)$

(c) $2\log_2(1 + S + I_L) < C_{LH}$

Fig. 5. Capacity regions $\mathcal{C}_{xy}$ in (9) when the interference realizations are $I_1 = I_x$ and $I_2 = I_y$ if (a) $C_{LH} \leq C_{LL}/2 + \log_2(1 + S + I_L)$; (b) $C_{LL}/2 + \log_2(1 + S + I_L) < C_{LH} \leq 2\log_2(1 + S + I_L)$; and (c) $2\log_2(1 + S + I_L) < C_{LH}$. The points A, B, C, D, E, and E' denote the rates selected by the RCC in the C-RAN scheme discussed in Sec. IV-A and the points a, b, c, and d indicate the four rate pairs (R_x, R_y) selected by the F-RAN scheme in Sec. V-A.

An outage occurs at time t if the selected rate pair (R_1, R_2) is outside the capacity region $\mathcal{C}_{I_1(t)I_2(t)}$. Accordingly, the outage probability in a time slot t for which the CSI available at the RCC is $I_1(t-d) = I_1$ and $I_2(t-d) = I_2$ can be computed as

$$\Pr[(R_1, R_2) \notin \mathcal{C}(I_1(t), I_2(t))|I_1(t-d) = I_1, I_2(t-d) = I_2]. \tag{10}$$

An achievable ϵ-outage adaptive sum-rate is summarized in the next lemma, where we defined the probabilities $P_{xy}^{HH} = \beta^{H|x}(d)\beta^{H|y}(d)$, $P_{xy}^{LH} = \beta^{L|x}(d)\beta^{H|y}(d)$, $P_{xy}^{HL} = \beta^{H|x}(d)\beta^{L|y}(d)$, and $P_{xy}^{LL} = \beta^{L|x}(d)\beta^{L|y}(d)$. The notation $P_{xy}^{\bar{x}\bar{y}}$ indicates the probability of transitioning from delayed states $\{I_1(t-d) = I_x, I_2(t-d) = $

$I_y\}$ to current states $\{I_1(t) = I_{\bar{x}}, I_2(t) = I_{\bar{y}}\}$.

Proposition 2: With C-RAN, an achievable ϵ-outage adaptive sum-rate $R^{\text{C-RAN}}(d, \epsilon)$ is given as

$$R^{\text{C-RAN}}(d, \epsilon) = \pi_L^2 R_{LL} + 2\pi_L \pi_H R_{LH} + \pi_H^2 R_{HH}, \tag{11}$$

with R_{xy} being defined as

$$R_{xy} = \begin{cases} C_{LL} & \text{if } \epsilon \leq P_{xy}^{LL}, \\ C_{LH} & \text{if } P_{xy}^{LL} < \epsilon \leq 1 - P_{xy}^{HH}, \\ C_{HH} & \text{if } 1 - P_{xy}^{HH} < \epsilon \leq 1, \end{cases} , \tag{12}$$

if $C_{LH} \leq 2\log_2(1 + S + I_L)$, and as

$$R_{xy} = \begin{cases} C_{LL} & \text{if } \epsilon \leq P_{xy}^{LL}, \\ 2\log_2(1 + S + I_L) & \text{if } P_{xy}^{LL} < \epsilon \leq \widetilde{P}_{xy}, \\ C_{LH} & \text{if } \widetilde{P}_{xy} < \epsilon \leq 1 - P_{xy}^{HH}, \\ C_{HH} & \text{if } 1 - P_{xy}^{HH} < \epsilon \leq 1, \end{cases} \tag{13}$$

if $C_{LH} > 2\log_2(1 + S + I_L)$, with $\widetilde{P}_{xy} = \min(P_{xy}^{LH}, P_{xy}^{HL}) + P_{xy}^{LL}$.

Proof: See Appendix B for the proof. ∎

Remark 1: An outage event can be generally avoided only if transmitting always at the minimum sum-rate C_{LL}, since the latter yields rate pairs that are within the capacity region in all other states (see Fig. 5). Therefore, with $d > 0$ and $\epsilon = 0$, the adaptive sum-rate of C-RAN is given by $R^{\text{C-RAN}}(d, 0) = C_{LL}$.

Remark 2: In the absence of CSI delay, i.e., with $d = 0$, the outage adaptive sum-rate $R^{\text{C-RAN}}(0, \epsilon)$ in (11) with any $\epsilon \neq 1$ can be simplified as $R^{\text{C-RAN}}(0, \epsilon) = \pi_L^2 C_{LL} + 2\pi_L \pi_H C_{LH} + \pi_H^2 C_{HH}$.

B. General Case

We now consider the general case with multiple RRS/UE pair and multiple channel states. To this end, we define the following rate expression for any subset $\mathcal{L} = \{l_1, \ldots, l_L\} \subseteq \mathcal{K}$ of RRSs

$$C(\{S_j\}_{j \in \mathcal{L}}, \{\mathbf{I}_j\}_{j \in \mathcal{L}}) \triangleq \mathbb{E}\left[\log_2 \det\left(\mathbf{I} + \mathbf{H}(\{S_j\}_{j \in \mathcal{L}}, \{\mathbf{I}_j\}_{j \in \mathcal{L}})\mathbf{H}^\dagger(\{S_j\}_{j \in \mathcal{L}}, \{\mathbf{I}_j\}_{j \in \mathcal{L}})\right)\right], \tag{14}$$

where we have introduced the channel matrix

$$\mathbf{H}(\{S_j\}_{j\in\mathcal{L}}, \{\mathbf{I}_j\}_{j\in\mathcal{L}}) = [\mathbf{h}_1(S_1, \mathbf{I}_1), \ldots, \mathbf{h}_{l_L}(S_{l_L}, \mathbf{I}_{l_L})], \tag{15}$$

with $\mathbf{h}_j(S_j, \mathbf{I}_j) = [\sqrt{I_{jl_1}}e^{j\theta_{jl_1}}, \ldots, \sqrt{S_j}e^{j\theta_{jj}}, \ldots, \sqrt{I_{jl_L}}e^{j\theta_{jl_L}}]^T$. The expectation in (14) is taken over the random phases $\{\theta_{ji}\}$ for $i, j \in \mathcal{L}$. Under joint data decoding at the RCC with full receiver CSI, when the CSI is $S_j(t) = S_j$ and $I_{ji}(t) = I_{ji}$, for all $i, j \in \mathcal{K}$ and $i \neq j$, the capacity region $\mathcal{C}(\{S_j(t)\}, \{I_{ji}(t)\})$ of the ergodic multiple access channel between the UEs and the RCC is given as (see, e.g., [24, Ch. 4])

$$\mathcal{C}(\mathbf{S}, \mathbf{I}) = \left\{ (R_1, \ldots, R_K) : \sum_{j\in\mathcal{L}} R_j \leq C(\{S_j\}_{j\in\mathcal{L}}, \{\mathbf{I}_j\}_{j\in\mathcal{L}}), \ \forall \mathcal{L} \subseteq \mathcal{K} \right\}. \tag{16}$$

Note that (16) is an extension of (9) to the general scenario studied here.

When selecting the transmission rates for slot t, the RCC has delayed CSI, namely $\{S_j(t - d)\}$ and $\{I_{ji}(t - d)\}$ for $i, j \in \mathcal{K}$ and $i \neq j$, and is hence not informed about the current capacity region $\mathcal{C}(\{S_j(t)\}, \{I_{ji}(t)\})$ in (16). To evaluate an achievable the ϵ-outage adaptive sum-rate, we introduce an outage sum-rate region $\mathcal{C}^\epsilon(\mathbf{S}, \mathbf{I})$ that has the property that, conditioned on $S_j(t-d) = S_j$ and $I_{ji}(t-d) = I_{ji}$, the set of rates $(R_1, \ldots, R_K) \in \mathcal{C}^\epsilon(\mathbf{S}, \mathbf{I})$ belong to the capacity region $\mathcal{C}(\{S_j(t)\}, \{I_{ji}(t)\})$ with probability no smaller than $1 - \epsilon$. As a result of this definition, choosing a rate pair in $\mathcal{C}^\epsilon(\{S_j(t - d)\}, \{I_{ji}(t - d)\})$ guarantees a probability of outage smaller than or equal to ϵ.

We specifically propose to define

$$\mathcal{C}_d^\epsilon(\mathbf{S}, \mathbf{I}) \triangleq \left\{ (R_1, \ldots, R_K) : \sum_{j\in\mathcal{L}} R_j \leq C(\{F_{S_j,d}^{-1}(\bar{\epsilon}|S_j)\}_{j\in\mathcal{L}}, \{F_{I_{ji},d}^{-1}(\bar{\epsilon}|I_{ji})\}_{j,i\in\mathcal{L}, i\neq j}), \ \forall \mathcal{L} \subseteq \mathcal{K} \right\}, \tag{17}$$

where $F_{S_j,d}^{-1}(\bar{\epsilon}|S_j)$ is defined as the state of $\bar{\epsilon}$-percentile of conditional distributions $\beta_S^{\cdot|S_j}(d)$, that is, the maximum state value $x \in \{S_1, \ldots, S_{N_S}\}$ such that $\Pr[S_j(t) \leq x | S_j(t - d) = S_j] \leq \bar{\epsilon}$; $F_{I_{ji},d}^{-1}(\bar{\epsilon}|I_{ji})$ is analogously defined as the state of $\bar{\epsilon}$-percentile of conditional distributions $\beta_I^{\cdot|I_{ji}}(d)$; and $\bar{\epsilon}$ is the individual outage probability of each state such that $1 - (1 - \bar{\epsilon})^{K^2} = \epsilon$ as in Proposition 1.

The problem of maximizing the resulting ϵ-outage adaptive sum-rate over the choice of the sum-rates $\{R(\mathbf{S}, \mathbf{I})\}$ for $\mathbf{S} \in \mathcal{N}_S^K$ and $\mathbf{I} \in \mathcal{N}_I^{K-1 \times K}$ can be then formulated as

$$R^{\text{C-RAN}}(d, \epsilon) = \underset{\{R(\mathbf{S},\mathbf{I})\} \geq 0}{\text{maximize}} \quad \sum_{\mathbf{S} \in \mathcal{N}_S^K} \sum_{\mathbf{I} \in \mathcal{N}_I^{K-1 \times K}} \pi_{\mathbf{S}} \, \pi_{\mathbf{I}} \, R(\mathbf{S}, \mathbf{I}) \tag{18a}$$

$$\text{s.t.} \quad R(\mathbf{S}, \mathbf{I}) \in \mathcal{C}_d^\epsilon(\mathbf{S}, \mathbf{I}), \tag{18b}$$

where the constraint (18b) applies to all values $\mathbf{S} \in \mathcal{N}_S^K$ and $\mathbf{I} \in \mathcal{N}_I^{K-1 \times K}$. The problem (18) is a linear program (LP) and can be tackled using standard solvers. Note that as in Remark 1, an outage event in case of the positive delay d can be avoided if transmitting at the sum-rate $C(\{F_{S_j,d}^{-1}(\bar{\epsilon}|S_j)\}, \{F_{I_{ji},d}^{-1}(\bar{\epsilon}|I_{ji})\})$.

V. FOG RADIO ACCESS NETWORK (F-RAN)

With F-RAN, each RRS individually performs rate adaptation based on the available CSI, while the RCC performs joint data decoding in a centralized manner.

A. Analysis with two RRSs and UEs

In this section, we consider the system model in Section IV-A with two RRSs and UEs, constant direct channel over T intervals, and two-state cross-channels. Accordingly, we will use the same notation introduced in Section IV-A for the transition and stationary probabilities of the cross-channels as well as for the sum-rate $C(I_1, I_2)$ when the cross-channels equal $I_1(t) = I_1$ and $I_2(t) = I_2$.

With F-RAN, the transmission rate R_j for user j is selected by each RRS j based on the available local CSI $I_j(t - d_e)$, which is subject to the scheduling delay d_e as for D-RAN, while the RCC performs centralized joint data decoding on behalf of the RRSs. We define as R_L and R_H the rates selected by each RRS j when $I_j(t) = I_L$ and $I_j(t) = I_H$, respectively. As for C-RAN, the outage probability is the probability that the rate pair (R_1, R_2) does not belong to the capacity region $\mathcal{C}_{I_1(t)I_2(t)}$.

Proposition 3: With F-RAN, an achievable ϵ-outage adaptive sum-rate $R^{\text{F-RAN}}(d_e, \epsilon)$ is given as

$$R^{\text{F-RAN}}(d_e, \epsilon) = 2\pi_L^2 R_L(d_e, \bar{\epsilon}) + 2\pi_L \pi_H (R_L(d_e, \bar{\epsilon}) + R_H(d_e, \bar{\epsilon})) + 2\pi_H^2 R_H(d_e, \bar{\epsilon}), \tag{19}$$

where $\bar{\epsilon} = 1 - (1 - \epsilon)^{1/K}$ and $R_x(d_e, \bar{\epsilon})$ is defined as

$$R_x(d_e, \bar{\epsilon}) = \begin{cases} R_L & \text{if } \bar{\epsilon} \leq \beta^{L|x}(d_e), \\ R_H & \text{if } \beta^{L|x}(d_e) < \bar{\epsilon} \leq 1, \end{cases} \tag{20}$$

with

$$R_L = \begin{cases} C_{LL}/2 & \text{if } C_{LH} > C_{LL}/2 + \log_2(1 + S + I_L), \\ C_{LL}/2 & \text{if } C_{LH} \leq C_{LL}/2 + \log_2(1 + S + I_L) \text{ and } \pi_L^2 > \pi_H^2, \\ C_{LH} - \log_2(1 + S + I_L) & \text{if } C_{LH} \leq C_{LL}/2 + \log_2(1 + S + I_L) \text{ and } \pi_L^2 \leq \pi_H^2, \end{cases} \tag{21}$$

and

$$R_H = \begin{cases} \log_2(1 + S + I_L) & \text{if } C_{LH} > C_{LL}/2 + \log_2(1 + S + I_L), \\ C_{LH} - C_{LL}/2 & \text{if } C_{LH} \leq C_{LL}/2 + \log_2(1 + S + I_L) \text{ and } \pi_L^2 > \pi_H^2, \\ \log_2(1 + S + I_L) & \text{if } C_{LH} \leq C_{LL}/2 + \log_2(1 + S + I_L) \text{ and } \pi_L^2 \leq \pi_H^2. \end{cases} \tag{22}$$

respectively.

Proof: As for Proposition 1, we impose that the individual outage probability for each UE-RRS pair be no larger than $\bar{\epsilon} = 1 - (1 - \epsilon)^{1/K}$ so that the overall outage probability is no larger than ϵ by construction. To this end, we first evaluate the rates R_L and R_H selected to guarantee no outage, i.e., $\bar{\epsilon} = 0$, for each RRS j when $I_j(t) = I_L$ and $I_j(t) = I_H$, respectively.

In order to guarantee no outage, the rate pair (R_L, R_L) must be inside the capacity region $\mathcal{C}_{LL}$, and hence, from Fig. 5, the rate R_L should be selected in the interval $[0, C_{LL}/2]$. In a similar manner, the rate pair (R_L, R_H) (or (R_H, R_L)) should be inside the capacity region $\mathcal{C}_{LH}$ (or $\mathcal{C}_{HL}$). Therefore, the rate R_H can be no larger than $\min(C_{LH} - R_L, \log_2(1 + S + I_L))$. Finally, the rate pair (R_H, R_H) must belong to the capacity region $\mathcal{C}_{HH}$, which is guaranteed by the conditions derived above. Based on these considerations, the adaptive sum-rate can be computed by solving the problem

$$\begin{aligned} \underset{R_L}{\text{maximize}} \quad & 2(\pi_L^2 + \pi_L\pi_H)R_L + 2(\pi_L\pi_H + \pi_H^2)\min\left(C_{LH} - R_L, \log_2(1 + S + I_L)\right) \\ \text{s.t.} \quad & 0 \leq R_L \leq C_{LL}/2, \end{aligned} \tag{23}$$

where the objective is obtained by averaging the achievable rate. Solving the linear max-min program [25] yields (20).

Now, if the rate $R_x(d_e, \bar{\epsilon})$ in (20) is selected by the RRS j, the individual outage probability for each UE j does not exceed $\bar{\epsilon}$ by definition. In fact, with this choice, an outage occurs if the transmission rate R_H for UE j is chosen by each RRS j when the local delayed CSI is $I_j(t - d_e) = I_x$ for $x \in \{L, H\}$ and the current CSI is $I_j(t) = I_L$. This probability is equal to $\beta^{L|x}(d_e)$. ∎

B. General Case

With F-RAN, the transmission rate R_j for user j is selected by each RRS j based on the available delayed CSI $S_j(t - d_e)$ and $I_{ji}(t - d_e)$ for $i \in \mathcal{K} \setminus \{j\}$ as for D-RAN, while the RCC performs centralized joint data decoding on behalf of the RRSs. The set of achievable rate pairs $(R_1, \ldots, R_K)$ with joint decoding at the RCC when the channel states are $\{S_i(t)\}$ and $\{I_{ji}(t)\}$ is given by the capacity region $\mathcal{C}(\{S_i(t)\}, \{I_{ji}(t)\})$ in (16).

Based on the definition of outage sum-rate region in (17), in the presence of the scheduling delay d_e, a probability of outage smaller than or equal to ϵ is guaranteed if the rate tuple $(R_1, \ldots, R_K)$ is selected in $\mathcal{C}_{d_e}^\epsilon(\mathbf{S}, \mathbf{I})$ in (17), when $S_j(t - d_e) = S_j$ and $I_{ji}(t - d_e) = I_{ji}$ for $i, j \in \mathcal{K}$ and $i \neq j$. The problem of maximizing the ϵ-outage adaptive sum-rate over the choice of the rates $\{R(S_j, \mathbf{I}_j)\}$ for $S_j \in \mathcal{N}_S$ and $\mathbf{I}_j \in \mathcal{N}_I^{K-1}$ under the outage sum-rate region $\mathcal{C}_{d_e}^\epsilon(\mathbf{S}, \mathbf{I})$ can then be written as

$$R^{\text{F-RAN}}(d_e, \epsilon) = \underset{\{R(S_j, \mathbf{I}_j)\} \geq 0}{\text{maximize}} \quad \sum_{\mathbf{S} \in \mathcal{N}_S^K} \sum_{\mathbf{I} \in \mathcal{N}_I^{K-1 \times K}} \pi_\mathbf{S} \pi_\mathbf{I} \sum_{j \in \mathcal{K}} R(S_j, \mathbf{I}_j) \tag{24a}$$

$$\text{s.t.} \quad (R(S_1, \mathbf{I}_1), \ldots, R(S_K, \mathbf{I}_K)) \in \mathcal{C}_{d_e}^\epsilon(\mathbf{S}, \mathbf{I}), \tag{24b}$$

where the constraint (24b) applies to all values $\mathbf{S} \in \mathcal{N}_S^K$ and $\mathbf{I} \in \mathcal{N}_I^{K-1 \times K}$. As for problem (18), problem (24) is an LP and can be solved using standard solvers.

VI. NUMERICAL RESULTS

In this section, we evaluate the performance of D-RAN, C-RAN, and F-RAN in terms of the ϵ-outage adaptive sum-rate as a function of key system parameters such as fronthaul and scheduling delays and mobile velocity. We set the carrier frequency to 1 GHz ($c/\lambda = 1$ GHz with $c = 3 \times 10^8$ m/s) and the
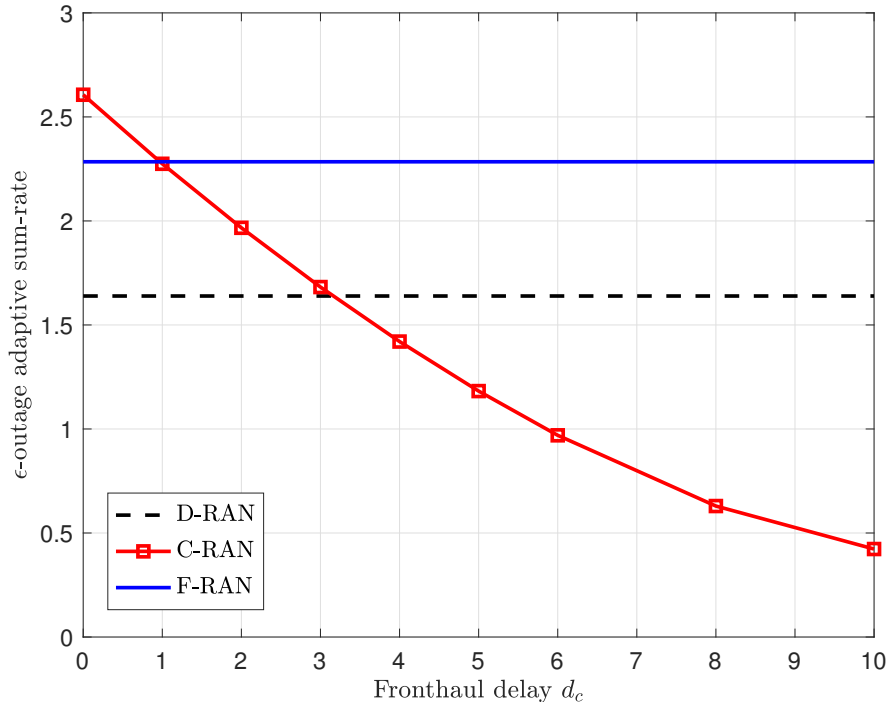
Fig. 6. ϵ-outage adaptive sum-rate vs. fronthaul delay d_c for D-RAN, C-RAN, and F-RAN ($\epsilon = 0$, $d_e = 2$, $N_S = N_I = 15$, $\gamma_S = 5$ dB, $\gamma_I = 0$ dB, and $v = 100$ km/h).

transmission interval duration to $T_p = 0.1$ ms. Unless stated otherwise, we set the number of direct and cross channel states to $N_S = N_I = 15$, the mobile velocity to $v = 100$ km/h, and the SNRs of the desired and cross channel signal to $\gamma_S = 5$ dB and $\gamma_I = 0$ dB, respectively.

We first investigate the impact of the fronthaul delay d_c when the scheduling delay is $d_e = 2$ and the outage level is $\epsilon = 0$. We recall that the fronthaul delay affects only C-RAN, whereby rate selection is performed based on CSI outdated by $d_e + d_c$ slots. From Fig. 6, we observe that the centralized data decoding and control performed by C-RAN provides significantly performance gains over the decentralized data and control carried out by D-RAN, but only if the fronthaul delay d_c is sufficiently small. Instead, when the fronthaul latency d_c is large enough, the outdating of the CSI used by C-RAN to perform rate selection causes a significant performance degradation. This shows that centralized control based on global but delayed CSI can yield a degraded performance as compared to decentralized control based on local but more timely CSI. F-RAN is able to leverage the gains of centralized decoding of C-RAN, while at the same time also being robust to fronthaul delays thanks to decentralized control as in D-RAN.
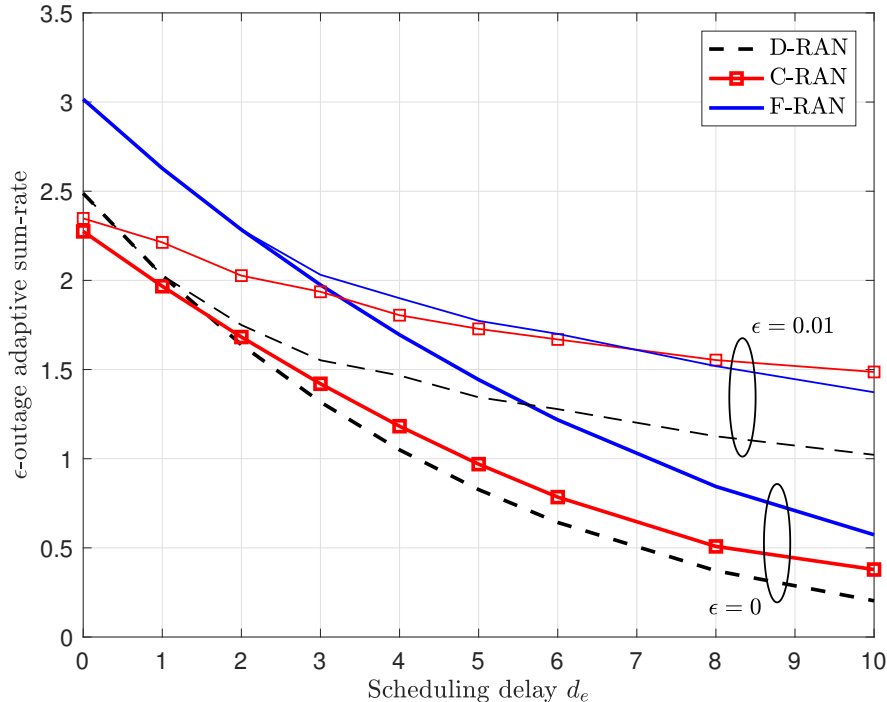
Fig. 7. ϵ-outage adaptive sum-rate vs. scheduling delay d_e under the finite-state Markov model [20] ($d_c = 3$, $N_S = N_I = 15$, $\gamma_S = 5$ dB, $\gamma_I = 0$ dB, and $v = 100$ km/h).

The impact of the scheduling delay d_e is studied in Fig. 7, where the ϵ-outage adaptive sum-rate is plotted versus d_e with a fronthaul delay $d_c = 3$. We recall that, while the fronthaul latency only affects the performance of C-RAN, rate selection for all schemes operates on an increasingly outdated CSI as d_e becomes larger. We consider both $\epsilon = 0$ and $\epsilon = 0.001$. At the given value of d_c, F-RAN is seen to outperform both C-RAN and D-RAN for all values of d_e, with decreasing absolute gains as d_e increases. We also see that C-RAN can outperform D-RAN for sufficiently large scheduling delay d_e, especially if one allows for a positive outage ϵ. This is because the performance of decentralized control is degraded as d_e grows larger.

The impact of the outage level ϵ is further investigated in Fig. 8, where we set $d_c = 5$ and $d_e = 2$. F-RAN is again seen to outperform both D-RAN and C-RAN, unless the allowed outage probability ϵ becomes large enough, here $\epsilon > 0.05$, in which case C-RAN can improve over F-RAN. In a similar way, if one accepts a sufficiently large outage probability, here $\epsilon > 0.0001$, C-RAN can perform better than D-RAN.
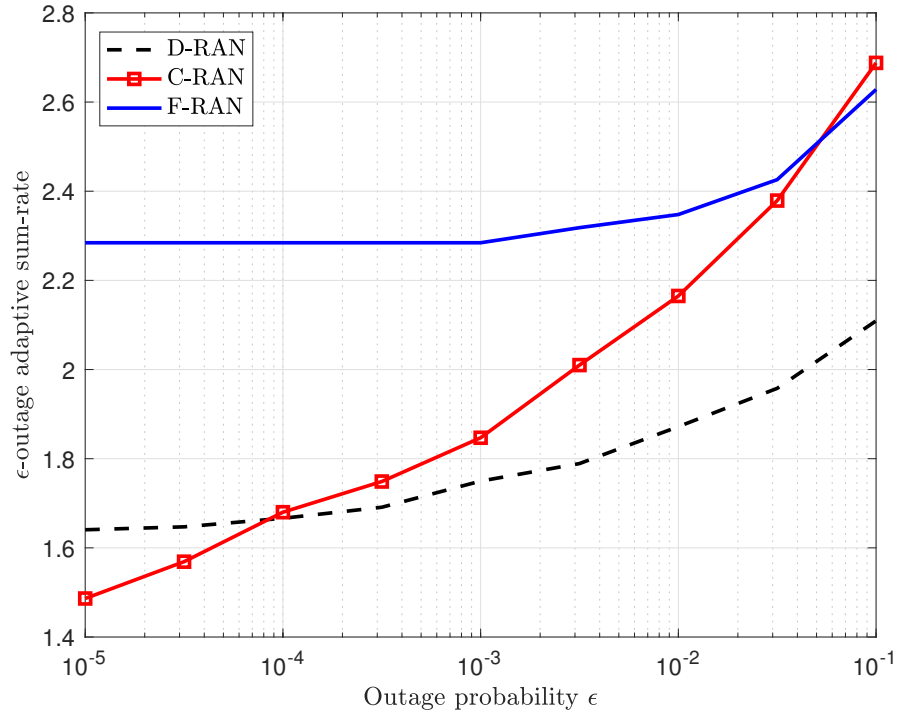
Fig. 8. ϵ-outage adaptive sum-rate vs. outage probability ϵ for D-RAN, C-RAN, and F-RAN ($d_c = 5$, $d_e = 2$, $N_S = N_I = 15$, $\gamma_S = 5$ dB, $\gamma_I = 0$ dB, and $v = 100$ km/h).
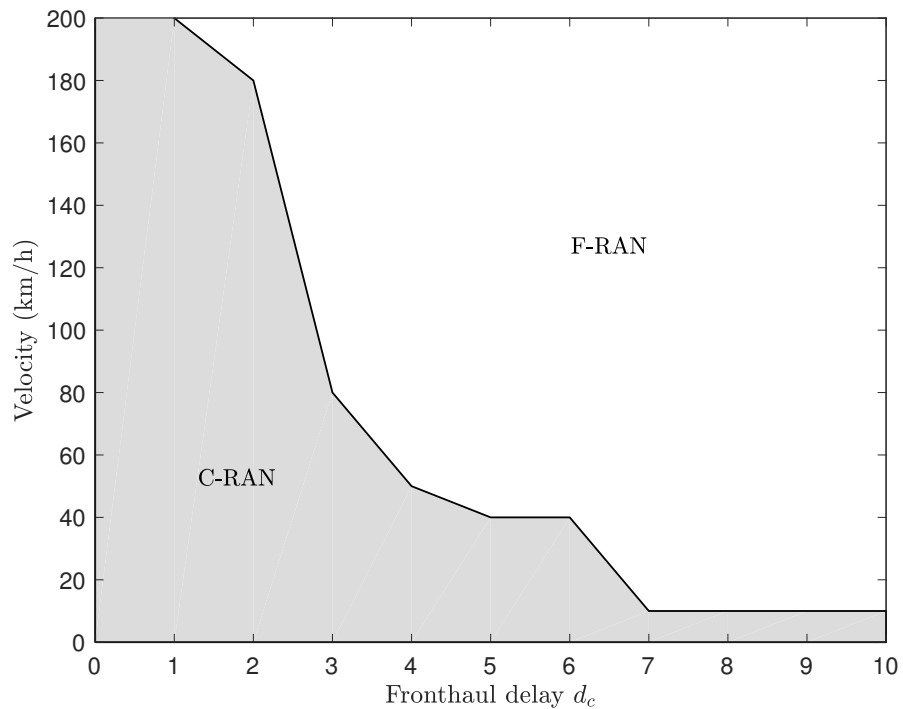

Fig. 9. Regions of the plane (d_c, v) in which F-RAN or C-RAN yield a larger ϵ-outage adaptive sum-rate when allowing an outage of $\epsilon = 0.01$ ($d_e = 3$, $N_S = N_I = 12$, $\gamma_S = 5$ dB, and $\gamma_I = 0$ dB).
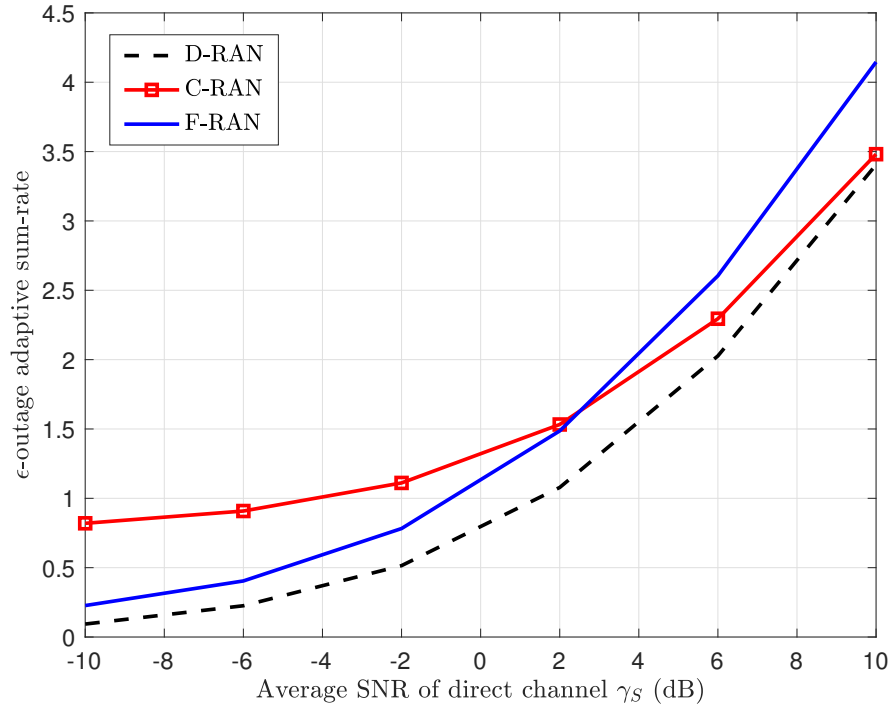
Fig. 10. ϵ-outage adaptive sum-rate vs. average SNR of direct channel states γ_S for D-RAN, C-RAN, and F-RAN ($d_c = 3$, $d_e = 2$, $\epsilon = 0.001$, $N_S = N_I = 15$, $\gamma_I = 0$ dB, and $v = 100$ km/h).
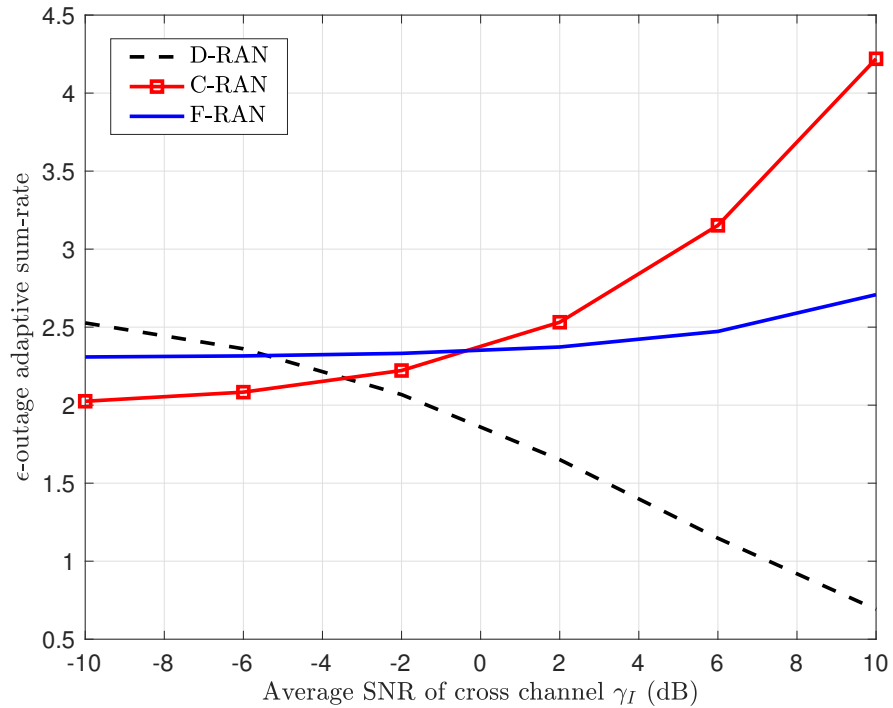

Fig. 11. ϵ-outage adaptive sum-rate vs. average SNR of interference channel states γ_I for D-RAN, C-RAN, and F-RAN ($d_c = 3$, $d_e = 2$, $\epsilon = 0.01$, $N_S = N_I = 15$, $\gamma_S = 5$ dB, and $v = 100$ km/h).

In order to obtain additional quantitative insight into the operating regimes in which different functional splits are to be preferred, Fig. 9 shows the regions of the plane with coordinates given by the fronthaul delay d_c and mobile velocity v in which each scheme offers the best ϵ-outage adaptive sum-rate. We set $d_e = 3$, $N_S = N_I = 12$, and $\epsilon = 0.01$. F-RAN is seen to be advantageous when the mobile velocity and the fronthaul delay are large enough. Note that F-RAN always outperforms D-RAN (not shown). The boundary line in Fig. 9 provides the maximum fronthaul delay d_c that can be tolerated by C-RAN for a given value of the velocity v, while still yielding gains as compared to F-RAN (and hence also D-RAN).

Finally, in Fig. 10 and 11, the ϵ-outage adaptive sum-rate is plotted versus average SNR of direct and interference channel states, respectively, for $d_c = 3$, $d_e = 2$. It is seen that F-RAN is able to outperform C-RAN under the given conditions unless SNR γ_S of direct channel is large or the average interfering channel gain γ_I is small. This is because the centralized decoding performed by C-RAN is effective in compensating for low direct CSI channels by leveraging the cross-channel signal paths. In fact, C-RAN is able to treat the cross-channels as useful signals rather than as interference.

VII. Conclusions

The control-data separation architecture offers a promising guiding principle for the implementation of functional splits between edge and cloud, as enabled by NFV, in fog-aided 5G systems. In this paper, we have analyzed the relative merits of functional splits whereby rate selection and data decoding are carried out either at the edge or at the cloud by adopting the criterion of ϵ-outage adaptive sum-rate. Among the main conclusions, this paper showed that the fully centralized architecture favored in the original instantiation of the C-RAN architecture is to be preferred only if the fronthaul latency is small or the time-variability of the channel is limited. Otherwise, a fog-based solution, whereby the control functionality of rate selection is carried out at the edge while joint data decoding is performed at the cloud, yields potentially significant gains. This conclusion demonstrates the value of decentralized but more timely CSI as compared to centralized but delayed CSI for the purpose of scheduling.

Among interesting open problems, we mention here the study of models that allow for a more general definition of functional splits including a flexible demarcation line at the physical layer. Another relevant open aspect is the impact of outage events due to quasi-static fading, both in terms of coding strategies at the physical layer, such as the broadcast approach [26], and of retransmission policies at the data link layer. Yet another issue is the modeling of both capacity and delays on the fronthaul links (see, e.g., [27]). Finally, it would be interesting to study downlink communication under the same assumptions on the heterogeneity of CSI available at edge and cloud considered in this paper.

APPENDIX A

Here, we follow [20] to define the N_S-state Markov model for the direct channel gains. Defining as γ_S the average SNR of the direct channel states, the values $\{S_1, \ldots, S_{N_S}\}$ of the direct channel gains are obtained by selecting each value S_m to be equal to the middle point in the quantization interval $[\Gamma_{S,m}, \Gamma_{S,m+1})$, which is identified by solving the equal-probability conditions $1/N_S = \exp(-\Gamma_{S,m}/\gamma_S) - \exp(-\Gamma_{S,m+1}/\gamma_S)$ with $\Gamma_{S,1} = 0$ and $\Gamma_{S,N_S+1} = \infty$ for $m = 1, \ldots, N_S$. In a similar manner, defining as γ_I the average SNR of the cross-channel states, the value I_m is equal to the middle point in each quantization interval $[\Gamma_{I,m}, \Gamma_{I,m+1})$, which is obtained by solving the equations $1/N_I = \exp(-\Gamma_{I,m}/\gamma_I) - \exp(-\Gamma_{I,m+1}/\gamma_I)$ with $\Gamma_{I,1} = 0$ and $\Gamma_{I,N_I+1} = \infty$ for $m = 1, \ldots, N_I$. For a mobile moving with velocity v and transmitting with a carrier of wavelength λ, the transition probabilities are given as

$$
p_{x,mn} = \begin{cases} \frac{N(\Gamma_{x,m})T_p}{\pi_{x,n}} & \text{if } m = n+1, \\[2mm] \frac{N(\Gamma_{x,n})T_p}{\pi_{x,n}} & \text{if } m = n-1, \\[2mm] 0 & \text{if } |m-n| > 1, \end{cases} \tag{25}
$$

for $x \in \{S, I\}$, where $\Gamma_{x,m}$ is $N(\Gamma_{x,m}) = \sqrt{2\pi\Gamma_{x,m}/\gamma_x}v/\lambda \exp(\Gamma_{x,m}/\gamma_x)$ is the crossing rate of state $\Gamma_{x,m}$ for Clarke's model [20] and T_p is the duration of a transmission interval.

APPENDIX B

The RCC chooses the rates R_1 and R_2 when $I_1(t-d) = I_1$ and $I_2(t-d) = I_2$ in such a way that the probability that the chosen rates are outside the capacity region $\mathcal{C}_{I_1(t)I_2(t)}$ for the current channel states $I_1(t)$ and $I_2(t)$ in (9) is less than ϵ. Specifically, referring to Fig. 5 for an illustration, when $I_1(t-d) = I_x$ and $I_2(t-d) = I_y$:

- If $\epsilon \leq P_{xy}^{LL}$, the RCC selects $R_1 = R_2 = C_{LL}/2$ (point A in Fig. 5);

- If $1 - P_{xy}^{HH} < \epsilon \leq 1$, the RCC selects $R_1 = R_2 = C_{HH}/2$ (point B in Fig. 5);

- If $P_{xy}^{LL} < \epsilon \leq 1 - P_{xy}^{HH}$ and $C_{LH} \leq 2\log_2(1 + S + I_L)$, the RCC selects $R_1 = R_2 = C_{LH}/2$ (point C in Fig. 5(a) and Fig. 5(b));

- If $P_{xy}^{LL} < \epsilon \leq \widetilde{P}_{xy}$ and $C_{LH} > 2\log_2(1 + S + I_L)$, the RCC selects $R_1 = R_2 = \log_2(1 + S + I_L)$ (point D in Fig. 5(c));

- If $\widetilde{P}_{xy} < \epsilon \leq 1 - P_{xy}^{HH}$ and $C_{LH} > 2\log_2(1 + S + I_L)$, the RCC selects either $R_1 = C_{LH} - \log_2(1 + S + I_L)$ and $R_2 = \log_2(1 + S + I_L)$ (point E$'$ in Fig. 5(c)), or $R_1 = \log_2(1 + S + I_L)$ and $R_2 = C_{LH} - \log_2(1 + S + I_L)$ (point E$''$ in Fig. 5(c)), where the first rate pair is selected when $P_{xy}^{HL} + P_{xy}^{LL} < P_{xy}^{LH} + P_{xy}^{LL}$ and the other pair otherwise.

We will argue next that these choices guarantee a probability of outage (10) no larger than ϵ.

A. When $R_1 + R_2 = C_{LL}$ (point A in Fig. 5), the probability of outage can be easily seen to be zero, as discussed before, because the capacity region $\mathcal{C}_{LL}$ is included in a capacity region $\mathcal{C}_{I_1(t)I_2(t)}$ with any current channel states $\{I_1(t), I_2(t)\}$.

B. If the rates are selected so that $R_1 + R_2 = C_{HH}$ (point B in Fig. 5), the upper bound (10) on the outage probability is easily seen to be $1 - P_{xy}^{HH}$, which, in the relevant regime, does not exceed ϵ, since any interference state other than $I_1(t) = I_H$ and $I_2(t) = I_H$ causes an outage. It can also be noted that the upper bound (10) is in fact tight, since the outage events for the two users coincide.

C. If $C_{LH} \leq 2\log_2(1+S+I_L)$, the capacity regions (9) are shown in Fig. 5(a) and Fig. 5(b). If the rates are selected to be $R_1 = R_2 = C_{LH}/2$ (point C in Fig. 5(a) and Fig. 5(b)), the upper bound on the probability

of outage can be calculated as P_{xy}^{LL}, since only the interference state $\{I_1(t) = I_L, I_2(t) = I_L\}$ causes an outage. Again, this probability is, by definition of the scheduling scheme, less than ϵ, and the upper bound is in fact tight.

D. If $C_{LH} > 2\log_2(1 + S + I_L)$, the capacity regions (9) are shown in Fig. 5(c). If the rates are selected such that $R_1 = R_2 = \log_2(1 + S + I_L)$ (point D in Fig. 5(c)), the upper bound (10) on the outage probability is easily seen to be tight and equal to P_{xy}^{LL}, which is smaller than ϵ in the relevant regime.

E. If $C_{LH} > 2\log_2(1 + S + I_L)$ and the rate pair $(R_1, R_2) = (C_{LH} - \log_2(1 + S + I_L), \log_2(1 + S + I_L))$ at E′ is selected, the upper bound (10) on the probability of outage is equal to $P_{xy}^{HL} + P_{xy}^{LL}$ and tight. This is because an outage for both users is caused by the states $(I_1(t), I_2(t)) = (I_H, I_L)$ and $(I_1(t), I_2(t)) = (I_L, I_L)$. In a similar manner, if the rate pair $(R_1, R_2) = (\log_2(1 + S + I_L), C_{LH} - \log_2(1 + S + I_L))$ at E″ is selected, the probability of outage is given as $P_{xy}^{LH} + P_{xy}^{LL}$. Therefore, by selecting between the rate pairs at E′ and E″, we obtain the probability of outage $\widetilde{P}_{xy} = \min(P_{xy}^{HL} + P_{xy}^{LL}, P_{xy}^{LH} + P_{xy}^{LL})$. This outage probability is also smaller than ϵ by construction of the scheduling scheme.

REFERENCES

[1] J. Kang, O. Simeone, J. Kang, and S. Shamai, "Control-data separation across edge and cloud for uplink communications in C-RAN," *Proc. IEEE Wireless Comm. and Net. Conf.*, pp. 1–6, San Francisco, CA, USA, Mar. 2017.

[2] J. Huang and R. Duan, "C-RAN: the road towards green RAN," White Paper, ver. 3.0, China mobile Research Institute, Oct. 2013.

[3] J. Huang and Y. Yuan, "White paper of next generation fronthaul interface," *[Online]. Available: labs.chinamobile.com/cran*, ver. 1.0, China mobile Research Institute, Jun. 2015.

[4] EU H2020 5G NORMA, "D3.2, 5G NORMA network architecture – intermediate report," *[Online] Available: https://5gnorma.5g-ppp.eu*, Jan. 2017.

[5] D. Wubben, P. Rost, J. Bartelt, M. Lalam, V. Savin, M. Gorgoglione, A. Dekorsy, and G. Fettweis, "Benefits and impact of cloud computing on 5G signal processing: Flexible centralization through cloud-RAN," *IEEE Sig. Proc. Mag.*, vol. 31, no. 6, pp. 35–44, Nov. 2014.

[6] O. Simeone, A. Maeder, M. Peng, O. Sahin, and W. Yu, "Cloud radio access network: Virtualizing wireless access for dense heterogeneous networks," *Journal of Communications and Networks*, vol. 18, no. 2, pp. 135–149, Apr. 2016.

[7] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega, D. Aziz, and H. Bakker, "Network slicing to enable scalability and flexibility in 5G mobile networks," *arXiv:1704.02129v1*, Apr. 2017.

[8] C.-Y. Chang, R. Schiavi, N. Nikaein, T. Spyropoulos, and C. Bonnet, "Impact of packetization and functional split on C-RAN fronthaul performance," *Proc. IEEE Int. Conf. on Comm.*, pp. 1–7, Kuala Lumpur, Malaysia, May 2016.

[9] U. Dotsch, M. Doll, H. P. Mayer, F. Schaich, J. Segel, and P. Sehier, "Quantitative analysis of split base station processing and determination of advantageous architectures for LTE," *Bell Labs Technical Journal*, vol. 18, no. 1, pp. 105–128, Jun. 2013.

[10] P. Rost and A. Prasad, "Opportunistic hybrid ARQ: Enabler of centralized-RAN over nonideal backhaul," *IEEE Wireless Comm. Lett.*, vol. 3, no. 5, pp. 481–484, Jul. 2014.

[11] S. Khalili and O. Simeone, "Inter-layer per-mobile optimization of cloud mobile computing: A message passing approach," *Transactions on Emerging Telecommunication Technologies*, vol. 27, no. 6, pp. 814–827, Feb. 2016.

[12] A. Mohamed, O. Onireti, M. Imran, A. Imran, and R. Tafazolli, "Control-data separation architecture for cellular radio access networks: A survey and outlook," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1, pp. 446–465, Jun. 2015.

[13] S. Gulati, B. Natarajan, S. Kalyanasundaram, and R. Agrawal, "Performance analysis of centralized RAN deployment with non-ideal fronthaul in LTE-advanced networks," *Proc. IEEE Veh. Technol. Conf.*, pp. 1–5, May 2016.

[14] M. Johnston and E. Modiano, "A new look at wireless scheduling with delayed information," *Proc. IEEE Int. Symp. Info. Th.*, pp. 1407–1411, Hong Kong, Jun. 2015.

[15] A. Sengupta, R. Tandon, and O. Simeone, "Fog-aided wireless networks for content delivery: Fundamental latency trade-offs," *arXiv:1605.01690*, May 2016.

[16] S. Sreekumar, B. Dey, and S. Pillai, "Distributed rate adaptation and power control in fading multiple access channels," *IEEE Trans. Info. Th.*, vol. 61, no. 10, pp. 5504–5524, Oct. 2015.

[17] 3rd Generation Partnership Project (3GPP) TR 36.881, "Study on latency reduction techniques for LTE (Release 14)," 2016.

[18] NGMN Alliance, "Further study on critical C-RAN technologies," *[Online] Available: https://www.ngmn.org*, Mar. 2015.

[19] N. Nikaein, "Processing radio access network functions in the cloud: Critical issues and modeling," *in Proc. of Int. Workshop on Mobile Cloud Computing and Services*, pp. 36–42, Paris, France, Sep. 2015.

[20] H. S. Wang and N. Moayeri, "Finite-state Markov channel-a useful model for radio communication channels," *IEEE Trans. on Veh. Technol.*, vol. 44, no. 1, pp. 163 –171, Feb. 1995.

[21] Y. Wei, F. R. Yu, and M. Song, "Distributed optimal relay selection in wireless cooperative networks with finite-state Markov channels," *IEEE Trans. on Veh. Technol.*, vol. 59, no. 5, pp. 2149–2158, Feb. 2010.

[22] K. Zheng, F. Liu, L. Lei, C. Lin, and Y. Jiang, "Stochastic performance analysis of a wireless finite-state Markov channel," *IEEE Trans. Wireless Comm.*, vol. 12, no. 2, pp. 782–793, Jan. 2013.

[23] J. R. Norris, *Markov Chains*. Cambridge University Press, 1998.

[24] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[25] J. E. Falk, "A linear max-min problem," *Math. Program.*, vol. 5, no. 1, pp. 169–188, 1973.

[26] S. Shamai and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Info. Th.*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.

[27] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Fronthaul compression for cloud radio access networks: signal processing advances inspired by network information theory," *IEEE Sig. Proc. Mag.*, vol. 31, no. 6, pp. 69–79, Nov. 2014.