

Latency-aware Reliable mmWave Communication via Multi-point Connectivity

Dileep Kumar, Satya Krishna Joshi, Antti Tölli

Centre for Wireless Communications, University of Oulu, Oulu, Finland

e-mail: {dileep.kumar, satya.joshi, antti.tolli}@oulu.fi

Abstract—The sensitivity of millimeter-wave (mmWave) radio channel to blockage is a fundamental challenge in achieving low-latency and reliable connectivity. In this paper, we explore the viability of using coordinated multi-point (CoMP) transmission for a delay bounded and reliable mmWave communication. We provide an iterative algorithm for the time-average sum-power-minimization problem by solving a system of Karush-Kuhn-Tucker (KKT) optimality conditions. We use the Lyapunov optimization framework and derive a dynamic control algorithm to transform a time-average stochastic problem into a sequence of deterministic subproblems. Furthermore, for the robust beamformer design, we consider a pessimistic estimate of the user-specific rate, assuming that a portion of CoMP links would be blocked during the data transmission phase, while ensuring the average latency requirements. The numerical examples illustrate that in the presence of random blockages, the proposed method outperforms baseline scenarios and results in energy-efficient, high-reliability and low-latency mmWave communication.

I. INTRODUCTION

Next-generation millimeter-wave (mmWave) based cellular systems will enable throughput-intensive and low-latency applications, such as tactile internet, augmented reality, and autonomous driving [1]. However, radio channel at the mmWave frequencies is inherently unreliable due to blockage, high penetration, and path losses [2]. These lead to rapid degradation of signal strength, and results in intermittent connectivity. For example, a mobile human blocker can obstruct the dominant communication paths (e.g., a line-of-sight (LoS) link) for hundreds of millisecond, and may disconnect the communication session [2]. On the other hand, adapting with unpredictable blockages and maintaining reliable communication session at mmWave frequencies is challenging, and demands critical latency overhead. Therefore, the use of macro-diversity via Coordinated Multi-Point (CoMP) connectivity is imperative for more robust and resilient communication [3]–[6].

In the Joint Transmission (JT)-CoMP schemes, each user equipment (UE) is concurrently served by multiple spatially distributed remote radio units (RRUs) [3]–[6]. The use of CoMP has been widely studied under the context of legacy 4G systems [3], [4] and lately also in the context of mmWave frequencies [5], [6]. CoMP techniques were mainly studied with the scope of enhancing capacity and cell-edge coverage (e.g., in [3]–[6]). Further, these schemes are studied for the

static case (i.e., resource allocation problem for a given time instance). Hence, these algorithms are not always applicable for a dynamic network with stringent latency and reliability requirements in the presence of random blockages.

In this paper, we develop a highly-reliable downlink transmission strategy for a mmWave based dynamic network, satisfying the user-specific latency requirements. Specifically, we consider a time-average sum-power minimization problem subject to maximum allowable queue length constraint in the presence of random blockages. The long-term time-average problem is transformed into a sequence of deterministic and independent subproblems using the Lyapunov optimization [7]. Furthermore, to improve communication reliability and avoid outage, we consider a pessimistic estimate of user-specific rate over the potential link blockage combinations. We obtain a robust beamformer by iteratively solving a system of Karush-Kuhn-Tucker (KKT) optimality conditions of the sum-power minimization problem. The proposed method provides insight into the trade-offs between required sum-power and achievable-rate, while ensuring average latency requirements.

II. SYSTEM MODEL

We consider a downlink transmission in a mmWave based cloud (or centralize) radio access network (C-RAN) architecture, where all RRUs are connected to the edge cloud by the fronthaul links, as illustrated in Fig. 1. Each RRU is equipped with N transmit antennas. We use $\mathcal{K} = \{1, 2, \dots, K\}$ to denote the set of all single antenna UEs, and $\mathcal{B} = \{1, 2, \dots, B\}$ to denote the set of all RRUs. The set of RRUs that serve k th UE is denoted by \mathcal{B}_k , such that $\mathcal{B}_k \subseteq \mathcal{B}$ for all $k \in \mathcal{K}$. We assume JT-CoMP connectivity, where k th UE receives a concurrently synchronous signal from its serving RRUs \mathcal{B}_k . Further, we assume that the network operates in time-slotted manner with each slot normalized to integer value $t \in \{0, 1, 2, \dots\}$, and all RRUs use same time-frequency resources for the downlink transmission.

Let $\mathbf{f}_{b,k}(t) \in \mathbb{C}^{N \times 1}$ denote the transmit beamformer from b th RRU to k th UE. Then the received signal $y_k(t)$ at k th UE during time slot t can be expressed as

$$y_k(t) = \sum_{b \in \mathcal{B}_k} \mathbf{h}_{b,k}^H(t) \mathbf{f}_{b,k}(t) d_k(t) + \sum_{u \in \mathcal{K} \setminus k} \sum_{g \in \mathcal{B}_u} \mathbf{h}_{g,k}^H(t) \mathbf{f}_{g,u}(t) d_u(t) + w_k(t), \quad (1)$$

This work was supported by the European Commission in the framework of the H2020-EUJ-02-2018 project under grant no. 815056 (5G-Enhance) and Academy of Finland under grants no. 318927 (6Genesis Flagship).

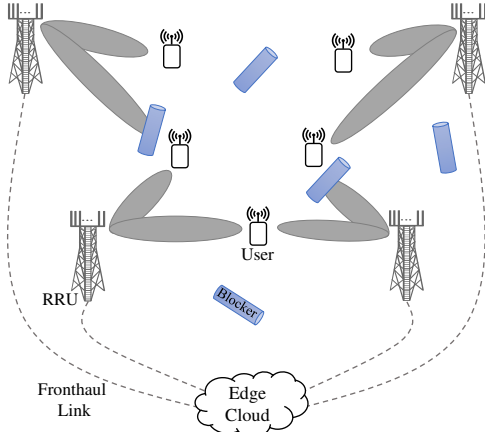


Fig. 1. C-RAN with B transmitters (RRUs) and K receivers (UEs) in the presence of randomly distributed blockers.

where $\mathbf{h}_{b,k}(t) \in \mathbb{C}^{N \times 1}$ is the channel vector between RRU-UE pair (b, k) , $w_k(t) \in \mathcal{CN}(0, \sigma_k^2)$ is circularly symmetric additive white Gaussian noise (AWGN), and $d_k(t)$ is data symbol associated with k th UE. Moreover, we assume that data symbols are normalized and independent, i.e., $\mathbb{E}\{|d_k(t)|^2\} = 1$ and $\mathbb{E}\{d_k(t)d_u^*(t)\} = 0$ for all $k, u \in \mathcal{K}$. The received signal-to-interference-plus-noise ratio (SINR) of k th UE during time slot t can be expressed as

$$\Gamma_k(\mathbf{F}(t)) = \frac{\left| \sum_{b \in \mathcal{B}_k} \mathbf{h}_{b,k}^H(t) \mathbf{f}_{b,k}(t) \right|^2}{\sigma_k^2 + \sum_{u \in \mathcal{K} \setminus k} \left| \sum_{g \in \mathcal{B}_u} \mathbf{h}_{g,k}^H(t) \mathbf{f}_{g,u}(t) \right|^2}, \quad (2)$$

where $\mathbf{F}(t) \triangleq [\mathbf{f}_{1,1}(t), \mathbf{f}_{1,2}(t), \dots, \mathbf{f}_{B,K}(t)]$.

A. Blockage Model & Achievable Rate

In mmWave frequency band, the radio channel is spatially sparse due to low-scattering, reduced diffraction, higher penetration, and path losses [2]. Hence, a mmWave link is inherently unreliable due to its susceptibility to blockages. In a typical mmWave communication, channel modeling shows that a link outage occurs with 20% – 60% probability [2]. Thus, we assume binary blockage such that the channel $\mathbf{h}_{b,k}(t)$ between RRU-UE pair (b, k) can have one of two states, i.e., it is either fully-available or completely blocked (when the channel is blocked we assume that $\mathbf{h}_{b,k}(t) = \mathbf{0}$). Furthermore, we consider random blockage, and the blocking of the channels $\{\mathbf{h}_{b,k}(t)\}_{b \in \mathcal{B}, k \in \mathcal{K}}$ for all t are independent¹ [8].

To improve system reliability and avoid outage under the uncertainties of mmWave channel, we preemptively underestimate the achievable SINR assuming that a portion of CoMP links would be blocked during the data transmission phase. This is specifically required in the mmWave channel, because of dynamic blockages which are not possible to track during the channel estimation phase. Let baseband processing unit (BBU) assume that each UE k have at least L_k available links (i.e., unblocked RRUs). Then we allow BBU to proactively model the SINR over all possible subset combinations, by

¹This is reasonably accurate assumption for a mmWave communication, especially when the blockers are not very large and closer to the UEs [8].

excluding the potentially blocked links, and allocate the rate to users such that transmission reliability is improved (i.e., to minimize the outage due to random blockages that appear during data transmission phase). As an example, let the set of RRUs that are used to serve k th UE with RRU indices $\mathcal{B}_k = \{1, 2, 3\}$. Then, with the assumption of at least $L_k = 2$ available links, the serving set of unblocked RRUs available to k th UE can be any one of following combinations:

$$\hat{\mathcal{B}}_k(L_k) = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}. \quad (3)$$

Let $C(L_k)$ denote the cardinality of set $\hat{\mathcal{B}}_k(L_k)$, defined as $C(L_k) = \sum_{l=L_k}^{|\mathcal{B}_k|} \frac{|\mathcal{B}_k|!}{l!(|\mathcal{B}_k|-l)!}$. We use \mathcal{B}_k^c to represent the c -th subset of $\hat{\mathcal{B}}_k(L_k)$, i.e., subset $\mathcal{B}_k^c \in \hat{\mathcal{B}}_k(L_k)$ such that cardinality $|\mathcal{B}_k^c| \geq L_k$ for all $c = 1, \dots, C(L_k)$. Then, the received SINR of k th UE for \mathcal{B}_k^c (i.e., c -th subset) is obtained excluding the potentially blocked links in expression (2), can be expressed as

$$\Gamma_k(\mathbf{F}(t), \mathcal{B}_k^c) = \frac{\left| \sum_{b \in \mathcal{B}_k^c} \mathbf{h}_{b,k}^H(t) \mathbf{f}_{b,k}(t) \right|^2}{\sigma_k^2 + \sum_{u \in \mathcal{K} \setminus k} \left| \sum_{g \in \mathcal{B}_u \setminus \mathcal{D}_k^c} \mathbf{h}_{g,k}^H(t) \mathbf{f}_{g,u}(t) \right|^2}, \quad (4)$$

where $\mathcal{D}_k^c = \mathcal{B}_k \setminus \mathcal{B}_k^c$ denotes a subset of potentially blocked RRUs which are excluded from the interfering links to k th UE. Thus, the pessimistic achievable rate² for k th UE during time slot t can be expressed as

$$r_k(t) = \log_2(1 + \gamma_k(t)), \quad (5)$$

where the SINR estimate $\gamma_k(t) = \min_{c=1, \dots, C(L_k)} \Gamma_k(\mathbf{F}(t), \mathcal{B}_k^c)$.

B. Network Queueing & Delay

We assume that the BBU maintains a set of internal queues for storing network layer data of all UEs [7, Ch. 5]. Let $Q_k(t)$ denote the current queue backlog of k th UE during time slot t , and $A_k(t)$ represents the amount of data that exogenously arrive to it, with the mean arrival rate $\mathbb{E}[A_k(t)] = \lambda_k$. Then the dynamics of queue $Q_k(t)$ can be expressed as

$$Q_k(t+1) = [Q_k(t) - r_k(t) + A_k(t)]^+, \quad \forall k \in \mathcal{K}, \quad (6)$$

where $r_k(t)$ is transmission rate defined in expression (5). Furthermore, let \bar{Q}_k denote the time-averaged queue associated with k th UE, defined as

$$\bar{Q}_k \triangleq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[Q_k(t)], \quad (7)$$

where the expectation depends on the control policy, and is with respect to the random channel states and arrivals.

According to the Little's law, the average delay is directly proportional to the average queue length \bar{Q}_k [9, Ch. 1.4]. Hence, for k th UE, we can achieve a desired latency by imposing a constraint on its queue length at each time slot. Here, we use a probabilistic constraint on the queue length which is defined as

$$\Pr\{Q_k(t) \geq Q_k^{\text{th}}\} \leq \epsilon, \quad \forall t, \quad (8)$$

where Q_k^{th} is the maximum allowable queue length for k th UE and $\epsilon \ll 1$ is the tolerable queue length violation probability.

²We consider Gaussian signalling and each UE decodes its intended signal by treating all other interfering signal as noise (4).

C. Problem Formulation

Our objective is to develop a highly-reliable downlink transmission strategy for C-RAN based dynamic mmWave communication system, satisfying the latency requirement. Specifically, we consider a problem of time-average sum-power minimization for mmWave communication with random channel blockages, subject to the maximum allowable queue length constraint for each UE, and it can be expressed as

$$\min_{\mathbf{F}(t), \gamma_k(t), \forall t} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \left(\sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}} \mathbb{E}[\|\mathbf{f}_{b,k}(t)\|^2] \right) \quad (9a)$$

$$\text{s. t. } \Pr\{Q_k(t) \geq Q_k^{\text{th}}\} \leq \epsilon, \quad \forall k \in \mathcal{K}, \forall t \quad (9b)$$

$$\gamma_k(t) = \min_{c=1, \dots, C(L_k)} \Gamma_k(\mathbf{F}(t), \mathcal{B}_k^c), \quad (9c)$$

$$\forall k \in \mathcal{K}, \forall t,$$

where the function $\Gamma_k(\mathbf{F}(t), \mathcal{B}_k^c)$ is defined in (4). The constraint (9b) ensures that the queue backlog of each user is less than Q_k^{th} at each time slot with the probability $1 - \epsilon$, and thus ensure the desired probabilistic latency requirement. Note that for each user $k \in \mathcal{K}$, the constraint (9c) is pessimistic estimate of achievable SINR. More specifically, for a given L_k , BBU models the SINR over all possible subset combinations of potentially available RRUs from the serving set $\hat{\mathcal{B}}_k(L_k)$ (see Section II-A), and uses the pessimistic estimate of SINR in order to allocated the transmission rate for the users.

III. DYNAMIC ALGORITHM VIA LYAPUNOV OPTIMIZATION

Problem (9) consists of a long-term time-average objective function, non-linear probabilistic queue length constraint (9b), and a large number of coupled non-convex SINR expressions (9c). Hence, problem (9) is intractable. In this section, we derive a dynamic control algorithm for (9) by using Lyapunov optimization techniques [7].

We start by transforming the probabilistic constraint (9b) into a time-average constraint using the Markov's inequality, $\Pr\{Q_k \geq Q_k^{\text{th}}\} \leq \mathbb{E}[Q_k]/Q_k^{\text{th}}$ [9]. Thereby, problem (9) can be rewritten as

$$\min_{\mathbf{F}(t), \gamma_k(t), \forall t} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \left(\sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}} \mathbb{E}[\|\mathbf{f}_{b,k}(t)\|^2] \right) \quad (10a)$$

$$\text{s. t. } \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[Q_k(t)] \leq \epsilon Q_k^{\text{th}}, \quad (10b)$$

$$\forall k \in \mathcal{K}, \forall t$$

$$\gamma_k(t) \leq \Gamma_k(\mathbf{F}(t), \mathcal{B}_k^c), \quad (10c)$$

$$\forall k \in \mathcal{K}, c = 1, \dots, C(L_k), \forall t.$$

Note that we have relaxed (9c) while writing constraint (10c), and these constraints are equivalent at optimality.

Now we use Lyapunov optimization, specifically, a drift-plus-penalty method [7] to find a solution of problem (10). Here, we enforce time-average constraint (10b) by transforming it into a queue stability problem [7, Ch. 5]. Specifically, a virtual queue associated with (10b) for each UE $k \in \mathcal{K}$ is introduced, and the stability of these virtual queue implies that the constraint (10b) is satisfied.

Let $Z_k(t)$ be the virtual queue associated with (10b) for k th UE during time slot t , and we update $Z_k(t)$ as

$$Z_k(t+1) = [Z_k(t) + Q_k(t+1) - \epsilon Q_k^{\text{th}}]^+, \quad \forall k \in \mathcal{K}. \quad (11)$$

The expression (11) can be interpreted as a queue dynamics for k th UE with arrival rate $Q_k(t+1)$ and service rate ϵQ_k^{th} . Note that if the virtual queues $\{Z_k(t)\}_{k \in \mathcal{K}}$ are stable, then by using [7, Theorem 2.5] we can show that (10b) is satisfied.

We now define Lyapunov function and its drift, which is used to stabilize queues $\{Z_k(t)\}_{k \in \mathcal{K}}$. For a compact representation, let $\Psi(t) = [Z_1(t), \dots, Z_K(t), Q_1(t), \dots, Q_K(t)]^T$ denote a collection of virtual and actual queues. Then we define following quadratic Lyapunov function:

$$\mathcal{L}(\Psi(t)) \triangleq \frac{1}{2} \sum_{k \in \mathcal{K}} Z_k(t)^2. \quad (12)$$

Intuitively, we can observe that if $\mathcal{L}(\Psi(t))$ is small, then all queues $\{Z_k(t)\}_{k \in \mathcal{K}}$ are small. Contrarily, if $\mathcal{L}(\Psi(t))$ is large then at least one of the queues is large. Thus, by minimizing a drift of $\mathcal{L}(\Psi(t))$ from one time slot to another, queues $\{Z_k(t)\}_{k \in \mathcal{K}}$ can be stabilized.

The expected drift of Lyapunov function $\mathcal{L}(\Psi(t))$ can be expressed as [7]

$$\begin{aligned} \Delta(\Psi(t)) &= \mathbb{E}[\mathcal{L}(\Psi(t+1)) - \mathcal{L}(\Psi(t)) | \Psi(t)] \\ &= \frac{1}{2} \mathbb{E} \left[\sum_{k \in \mathcal{K}} (Z_k(t+1)^2 - Z_k(t)^2) | \Psi(t) \right]. \end{aligned} \quad (13)$$

Next, by using expressions (6) and (11) in (13), an upper bound of drift $\Delta(\Psi(t))$ can be expressed as³

$$\begin{aligned} \Delta(\Psi(t)) &\leq \zeta + \Phi(t) \\ &\quad - \mathbb{E} \left[\sum_{k \in \mathcal{K}} (Q_k(t) + A_k(t) + Z_k(t)) r_k(t) | \Psi(t) \right], \end{aligned} \quad (14)$$

where ζ and $\Phi(t)$ are positive constant term and satisfy the following condition⁴ for all time slots:

$$\zeta \geq \frac{1}{2} \mathbb{E} \left[\sum_{k \in \mathcal{K}} A_k(t)^2 + r_k(t)^2 | \Psi(t) \right],$$

$$\begin{aligned} \Phi(t) &= \sum_{k \in \mathcal{K}} \left[\frac{1}{2} (\epsilon Q_k^{\text{th}})^2 + \frac{1}{2} Q_k(t)^2 + Z_k(t) Q_k(t) \right. \\ &\quad \left. + (Q_k(t) + Z_k(t)) \lambda_k \right]. \end{aligned}$$

Now we define following drift-plus-penalty function [7] for problem (10):

$$\Delta(\Psi(t)) + V \mathbb{E} \left[\sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}} \|\mathbf{f}_{b,k}(t)\|^2 | \Psi(t) \right], \quad (15)$$

where $V \geq 0$ is a trade-off parameter. By using expression (14) in (15), and minimizing the upper bound of (15) subject to constraint (10c) at each time slot, we can stabilize queues $\{Z_k(t)\}_{k \in \mathcal{K}}$ and minimize the objective function of problem (10). Thus, we utilize the concept of *opportunistic minimization of an expectation* [7, Ch. 1.8] to minimize the drift-plus-penalty function (15), and obtain a dynamic control algorithm as detailed in Algorithm 1.

At each time slot of Algorithm 1, we need to solve problem (16) to find optimal transmit beamformers $\mathbf{F}(t)$. We derive an iterative algorithm for this in the next section.

³To obtain (14), we have used the fact that $([a+b-c]^+)^2 \leq (a+b-c)^2$ for any $a \geq 0$, $b \geq 0$, and $c \geq 0$.

⁴We have assumed that second moments of arrival and transmission processes are bounded [7].

Algorithm 1: Dynamic algorithm for problem (9)

- 1 For a given time slot t , observe queue backlogs $\{Q_k(t)\}$ and $\{Z_k(t)\}$, and solve following problem:

$$\min_{\mathbf{F}(t), \gamma_k(t)} V \sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}} \|\mathbf{f}_{b,k}(t)\|^2 - \sum_{k \in \mathcal{K}} (Q_k(t) + A_k(t) + Z_k(t)) \log_2(1 + \gamma_k(t)) \quad (16a)$$

$$\text{s. t. } \gamma_k(t) \leq \Gamma_k(\mathbf{F}(t), \mathcal{B}_k^c), \quad \forall k \in \mathcal{K}, c = 1, \dots, C(L_k). \quad (16b)$$

- 2 Update queues $Q_k(t+1)$ and $Z_k(t+1)$ by using (6) and (11), respectively, for all $k \in \mathcal{K}$
 3 Set $t = t + 1$, and go to step 1
-

IV. ITERATIVE ALGORITHM FOR PROBLEM (16)

Problem (16) is non-convex due SINR constraint (16b). In this section, we adopt Successive Convex Approximation (SCA) technique [10], [11] to find a solution for problem (16). In the following, we omit time index t to simplify the notation.

By using the expression of $\Gamma_k(\mathbf{F}, \mathcal{B}_k^c)$ (see (4)) and adding one on both sides of constraint (16b), we rewrite it as

$$1 + \gamma_k \leq \frac{\sigma_k^2 + \sum_{j \in \mathcal{K}} \left| \mathbf{h}_k^{\text{cH}} \mathbf{f}_j \right|^2}{\sigma_k^2 + \sum_{u \in \mathcal{K} \setminus k} \left| \mathbf{h}_k^{\text{cH}} \mathbf{f}_u \right|^2}, \quad (17)$$

where $\mathbf{f}_j \in \mathbb{C}^{|\mathcal{B}|N \times 1}$ is the stacked transmit beamformer and $\mathbf{h}_k^{\text{c}} \in \mathbb{C}^{|\mathcal{B}|N \times 1}$ is the stacked channel vector, defined as

$$\mathbf{f}_j \triangleq [\mathbb{1}_{\mathcal{B}_j}(1) \mathbf{f}_{1,j}^{\text{T}}, \dots, \mathbb{1}_{\mathcal{B}_j}(b) \mathbf{f}_{b,j}^{\text{T}}, \dots, \mathbb{1}_{\mathcal{B}_j}(B) \mathbf{f}_{B,j}^{\text{T}}]^{\text{T}},$$

$$\mathbf{h}_k^{\text{c}} \triangleq [\mathbb{1}_{\mathcal{G}_k^c}(1) \mathbf{h}_{1,k}^{\text{T}}, \dots, \mathbb{1}_{\mathcal{G}_k^c}(b) \mathbf{h}_{b,k}^{\text{T}}, \dots, \mathbb{1}_{\mathcal{G}_k^c}(B) \mathbf{h}_{B,k}^{\text{T}}]^{\text{T}}.$$

The indicator function $\mathbb{1}_{\mathcal{G}_k^c}(b)$ and $\mathbb{1}_{\mathcal{B}_j}(b)$ are defined as

$$\mathbb{1}_{\mathcal{G}_k^c}(b) = \begin{cases} 1 & \text{if and only if } b \in \mathcal{B} \setminus \mathcal{D}_k^c \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathbb{1}_{\mathcal{B}_j}(b) = \begin{cases} 1 & \text{if and only if } b \in \mathcal{B}_j \\ 0 & \text{otherwise.} \end{cases}$$

where $\mathcal{G}_k^c = \mathcal{B} \setminus \mathcal{D}_k^c$ for all $c = 1, \dots, C(L_k)$ and $k \in \mathcal{K}$. Note that we have added one on both sides of (16b) and expressed it equivalently as (17), in order to improve numerical stability of the algorithm. It will become clear in the following subsection.

For more compact representation, we introduce new functions $I_k(\mathbf{F}, c)$ and $G_k(\mathbf{F}, \gamma_k, c)$, defined as

$$I_k(\mathbf{F}, c) \triangleq \sigma_k^2 + \sum_{u \in \mathcal{K} \setminus k} \left| \mathbf{h}_k^{\text{cH}} \mathbf{f}_u \right|^2, \quad (18a)$$

$$G_k(\mathbf{F}, \gamma_k, c) \triangleq \frac{\sigma_k^2 + \sum_{j \in \mathcal{K}} \left| \mathbf{h}_k^{\text{cH}} \mathbf{f}_j \right|^2}{1 + \gamma_k}. \quad (18b)$$

Hence, expression (17) can be expressed as

$$I_k(\mathbf{F}, c) - G_k(\mathbf{F}, \gamma_k, c) \leq 0, \quad (19)$$

for all $c = 1, \dots, C(L_k)$ and $k \in \mathcal{K}$.

Note that function $G_k(\mathbf{F}, \gamma_k, c)$ is a *quadratic-over-linear*, which is a convex function [12, Ch. 3]. Hence, the left-hand-side of expression (19) is a difference of convex function. Thus, the best convex approximation of constraint function (19) can be obtained by replacing $G_k(\mathbf{F}, \gamma_k, c)$ with

its first-order approximation [10]. The linear first-order Taylor approximation of $G_k(\mathbf{F}, \gamma_k, c)$ around an arbitrary point $\{\mathbf{f}_k^{(i)}, \gamma_k^{(i)}\}$ can be expressed as

$$\tilde{G}_k(\mathbf{F}, \gamma_k, c) \triangleq 2 \sum_{j \in \mathcal{K}} \Re \left\{ \frac{\mathbf{f}_j^{(i)\text{H}} \mathbf{h}_k^{\text{c}} \mathbf{h}_k^{\text{cH}}}{1 + \gamma_k^{(i)}} (\mathbf{f}_j - \mathbf{f}_j^{(i)}) \right\} + \frac{\sigma_k^2 + \sum_{j \in \mathcal{K}} \left| \mathbf{h}_k^{\text{cH}} \mathbf{f}_j^{(i)} \right|^2}{1 + \gamma_k^{(i)}} \left(1 - \frac{\gamma_k - \gamma_k^{(i)}}{1 + \gamma_k^{(i)}} \right). \quad (20)$$

Hence, (16) can be approximated as the following convex problem:

$$\min_{\mathbf{F}, \gamma_k} V \sum_{k \in \mathcal{K}} \|\mathbf{f}_k\|^2 - \sum_{k \in \mathcal{K}} (Q_k + A_k + Z_k) \log_2(1 + \gamma_k) \quad (21a)$$

$$\text{s. t. } a_{k,c} : I_k(\mathbf{F}, c) - \tilde{G}_k(\mathbf{F}, \gamma_k, c) \leq 0, \quad \forall k \in \mathcal{K}, \forall c, \quad (21b)$$

where $\{a_{k,c}\}$ are non-negative Lagrangian multipliers associated with constraints (21b). Note that (21) approximates a solution for (16) in the vicinity of a point $\{\mathbf{f}_k^{(i)}, \gamma_k^{(i)}\}$. Thus, by iteratively solving (21) while updating $\{\mathbf{f}_k^{(i)}, \gamma_k^{(i)}\}$ with the solution of current iteration, we can find the best local optimal solution for (16).

A. Iterative Algorithm via KKT Conditions

The Karush-Kuhn-Tucker (KKT) optimality conditions provide necessary and sufficient conditions for the solution of a convex problem. Hence, we tackle problem (21) by iteratively solving a system of KKT optimality conditions [12, Ch. 5.5]. The Lagrangian $\mathcal{L}(\mathbf{F}, \gamma_k, a_{k,c})$ of problem (21) is given in (24). The stationary conditions for (21) can be expressed as

$$\nabla_{\gamma_k} : \frac{Q_k + A_k + Z_k}{1 + \gamma_k} - \sum_{c=1}^{C(L_k)} a_{k,c} \frac{\sigma_k^2 + \sum_{j \in \mathcal{K}} \left| \mathbf{h}_k^{\text{cH}} \mathbf{f}_j^{(i)} \right|^2}{(1 + \gamma_k^{(i)})^2}, \quad (25a)$$

$$\nabla_{\mathbf{f}_k} : \mathbf{f}_k^{\text{H}} \left(V \mathbf{1} + \sum_{u \in \mathcal{K} \setminus k} \sum_{c=1}^{C(L_u)} a_{u,c} \mathbf{h}_u^{\text{c}} \mathbf{h}_u^{\text{cH}} \right) = \sum_{j \in \mathcal{K}} \sum_{c=1}^{C(L_j)} a_{j,c} \left(\frac{\mathbf{f}_k^{(i)\text{H}} \mathbf{h}_j^{\text{c}} \mathbf{h}_j^{\text{cH}}}{1 + \gamma_j^{(i)}} \right). \quad (25b)$$

In addition to (25) and primal-dual feasibility constraints, the KKT conditions also include the complementary slackness as $a_{k,c} \{I_k(\mathbf{F}, c) - \tilde{G}_k(\mathbf{F}, \gamma_k, c)\} = 0, \forall k \in \mathcal{K}, c = 1, \dots, C(L_k)$.

Note that the user-specific SINR constraints (21b) are coupled and interdependent over the link blockage combinations (see Section II-A). Hence, obtaining a closed-form solution for Lagrangian multipliers $\{a_{k,c}\}$ in expression (25) is significantly more challenging than the case with single SINR constraint per-user [10], [11]. To overcome this challenge, we resort to the subgradient approach, where all non-negative Lagrangian multipliers $\{a_{k,c}\}$ are iteratively solved

$$\begin{aligned} \mathfrak{L}(\mathbf{F}, \gamma_k, a_{k,c}) = & \sum_{k=1}^K \left[V \|\mathbf{f}_k\|^2 - (Q_k + A_k + Z_k) \log_2(1 + \gamma_k) + \sum_{c=1}^{C(L_k)} a_{k,c} \sigma_k^2 + \sum_{u \in \mathcal{K} \setminus k} \sum_{c=1}^{C(L_u)} a_{u,c} |\mathbf{h}_u^{\text{cH}} \mathbf{f}_k|^2 \right. \\ & \left. - 2 \sum_{j \in \mathcal{K}} \sum_{c=1}^{C(L_j)} a_{j,c} \Re \left\{ \frac{\mathbf{f}_k^{(i)\text{H}} \mathbf{h}_j^{\text{c}} \mathbf{h}_j^{\text{cH}}}{1 + \gamma_j^{(i)}} (\mathbf{f}_k - \mathbf{f}_k^{(i)}) \right\} - \sum_{c=1}^{C(L_k)} a_{k,c} \frac{\sigma_k^2 + \sum_{j \in \mathcal{K}} |\mathbf{h}_k^{\text{cH}} \mathbf{f}_j^{(i)}|^2}{1 + \gamma_k^{(i)}} \left(1 - \frac{\gamma_k - \gamma_k^{(i)}}{1 + \gamma_k^{(i)}} \right) \right]. \quad (24) \end{aligned}$$

using the subgradient method [13]. The steps in the iterative algorithm are:

$$\mathbf{f}_k^{(i)\text{H}} = \sum_{j \in \mathcal{K}} \sum_{c=1}^{C(L_j)} a_{j,c}^{(i-1)} \left(\frac{\mathbf{f}_k^{(i-1)\text{H}} \mathbf{h}_j^{\text{c}} \mathbf{h}_j^{\text{cH}}}{1 + \gamma_j^{(i-1)}} \right) \times \quad (26a)$$

$$\left\{ V \mathbf{1} + \sum_{u \in \mathcal{K} \setminus k} \sum_{c=1}^{C(L_u)} a_{u,c}^{(i-1)} \mathbf{h}_u^{\text{c}} \mathbf{h}_u^{\text{cH}} \right\}^{-1},$$

$$\gamma_k^{(i)} = (Q_k + A_k + Z_k) \times \quad (26b)$$

$$\left\{ \sum_{c=1}^{C(L_k)} a_{k,c}^{(i-1)} \frac{\sigma_k^2 + \sum_{j \in \mathcal{K}} |\mathbf{h}_k^{\text{cH}} \mathbf{f}_j^{(i-1)}|^2}{(1 + \gamma_k^{(i-1)})^2} \right\}^{-1} - 1,$$

$$\Gamma_k^{(i)}(\mathbf{F}, \mathcal{B}_k^{\text{c}}) = \frac{|\mathbf{h}_k^{\text{cH}} \mathbf{f}_k^{(i)}|^2}{\sigma_k^2 + \sum_{u \in \mathcal{K} \setminus k} |\mathbf{h}_k^{\text{cH}} \mathbf{f}_u^{(i)}|^2}, \quad (26c)$$

$$a_{k,c}^{(i)} = a_{k,c}^{(i-1)} + \beta [\gamma_k^{(i)} - \Gamma_k^{(i)}(\mathbf{F}, \mathcal{B}_k^{\text{c}})], \quad (26d)$$

where β is small positive step-size⁵. In expression (26d), the dual variables $\{a_{k,c}\}$ are iteratively updated based on the violation of SINR constraint (10c), using the subgradient method [13]. Note that due to reformulation of constraint (16b) as in (17), we get $\{1 + \gamma_j\}_{j \in \mathcal{K}}$ in the denominator of (26a) and (26b), and these are invertible, even if some UEs have zero SINR. Thus, the proposed iterative method for problem (16) is numerically stable, and it is summarized in Algorithm 2.

Algorithm 2: Iterative algorithm for problem (16)

- 1 Set $i = 1$ and initialize $\{\mathbf{f}_k^{(0)}, \gamma_k^{(0)}, a_{k,c}^{(0)}\}, \forall c, \forall k,$
 - 2 **repeat**
 - 3 Solve $\mathbf{f}_k^{(i)}$ from (26a) with $\{\mathbf{f}_k^{(i-1)}, \gamma_k^{(i-1)}, a_{k,c}^{(i-1)}\}$
 - 4 Obtain $\gamma_k^{(i)}$ from (26b)
 - 5 Calculate $\Gamma_k^{(i)}(\mathbf{F}, \mathcal{B}_k^{\text{c}})$ from (26c) with updated $\mathbf{f}_k^{(i)}$
 - 6 Obtain $a_{k,c}^{(i)}$ using (26d) with $\{\gamma_k^{(i)}, \Gamma_k^{(i)}(\mathbf{F}, \mathcal{B}_k^{\text{c}})\}$
 - 7 Set $i = i + 1$
 - 8 **until** convergence or for fixed iterations;
-

V. SIMULATION RESULTS

We consider a mmWave based downlink transmission with UEs $K = 2$, RRUs $B = 2$, and each RRU is equipped with a uniform linear array (ULA) of $N = 16$ antennas. We assume a fully-connected JT-CoMP scenario, i.e., $\mathcal{B}_k = \mathcal{B}$ for all $k \in \mathcal{K}$ such that all RRUs coherently serve each UE. RRUs are placed with the spacing of 50 meters and connected to a common BBU. All single antenna UEs are randomly dropped, thus each UE have different path-gain and angle with the RRUs.

⁵The step size $\beta > 0$ in (26d) depends on the system model, as it directly affects the convergence rate and controls the oscillation in the objective [13].

The radio channel $\mathbf{h}_{b,k}(t)$ between a RRU-UE pair (b, k) is based on sparse geometric model [14], and defined as

$$\mathbf{h}_{b,k}(t) = \sqrt{\frac{N}{M}} \sum_{m=1}^M v_{b,k}(t) d_{b,k}^{-\psi_m}(t) \mathbf{a}_{\mathcal{T}}^{\text{H}}(\phi_{b,k}^m(t)), \quad (27)$$

where $M = 3$ is the number of independent paths, and $v_{b,k}(t)$ is random complex gain with zero mean and unit variance. The distance between RRU-UE pair is represented with $d_{b,k}$, and the notation ψ_m denote a random path-loss exponent. In the simulation, we set $\psi_m \in [2, 6] \forall m$. The array response vector for ULA is represented with $\mathbf{a}_{\mathcal{T}}(\phi_{b,k}^m(t)) \in \mathbb{C}^{N \times 1}$, where angle-of-departure (AoD) for each path is uniformly distributed, i.e., $\phi_{b,k}^m(t) \in [-\pi/2, \pi/2] \forall m, \forall t$. Similar to [8], we assume a probabilistic blockage model, where the channel between RRU-UE pair is either fully available, i.e., as in (27) or completely blocked i.e., $\mathbf{h}_{b,k}(t) = \mathbf{0}$, with the probability of $q_{b,k} \in [0, 1]$ for all $b \in \mathcal{B}$ and $k \in \mathcal{K}$.

For each UE k , we assume identical subset size $L_k = L$, arrival $A_k \sim \text{Pois}(\lambda)$ with $\lambda = 3.5$ bits/slot and allowable queue length $Q_k^{\text{th}} = 5$ bits with tolerable probability $\epsilon = 0.1$ in problem (9). In the simulations, we set the subgradient step size $\beta = 0.5$ in (26d), and the frequency $f_c = 28$ GHz.

The outage event occurs if the instantaneous transmit rate $r_k(t)$ exceeds the supported rate⁶ $c_k(t)$ for all $k \in \mathcal{K}$. Then, the queue dynamics $Q_k(t)$ in (6) can be expressed as

$$Q_k(t+1) = [Q_k(t) - r_k(t) \mathbb{1}_{\{r_k(t) \leq c_k(t)\}} + A_k(t)]^+ \forall k, \quad (28)$$

where $\mathbb{1}_{\{\cdot\}}$ is an indicator function. Specifically, expression (28) implies that queue backlog also increases with each unsuccessful downlink transmission due to random blockages.

For the baseline methods, we consider Coordinated Beamforming (CB) [3] and full-JT (i.e., $L = B$) [11] beamformer design. Fig. 2 illustrates the average queue delay performance with $V = 1$ and blockage of 20%. Result shows that our proposed and baseline methods satisfy the queue backlogs of each user k (i.e., $Q_k^{\text{th}} = 5$) within the allowable tolerance level $\epsilon = 0.1$. Thus, all methods strictly achieve the desired average latency requirements (i.e., constraint (9b)). However, our proposed method substantially improves the sum-power performance while ensuring the average latency requirements, as shown in Fig. 3. For example, with $V = 1$ blockage of 20%, our proposed method improves the sum-power performance by 7dB and 9dB compared to CB and full-JT, respectively.

⁶For a give time slot t , let $\mathcal{S}_k(t) = \{\gamma_k^*(t), \mathbf{f}_k^*(t)\}_{\forall k \in \mathcal{K}}$ denote solution obtained from Algorithm (2). Then, for each user k , the transmission rate is given by $r_k = \log_2(1 + \gamma_k^*)$. However, the actual supported rate for k th user depends on the beamformers $\{\mathbf{f}_k^*(t)\}_{\forall k \in \mathcal{K}}$ and channel state $\{\mathbf{h}_{b,k}(t)\}_{b \in \mathcal{B}, k \in \mathcal{K}}$, which can not be exactly known to the BBU during data transmission phase due to random blockages. The supported rate can be calculated using the actual SINR values (2), i.e., $c_k = \log_2(1 + \Gamma_k(\mathbf{F}^*(t)))$, and these rates are unknown to the BBU.

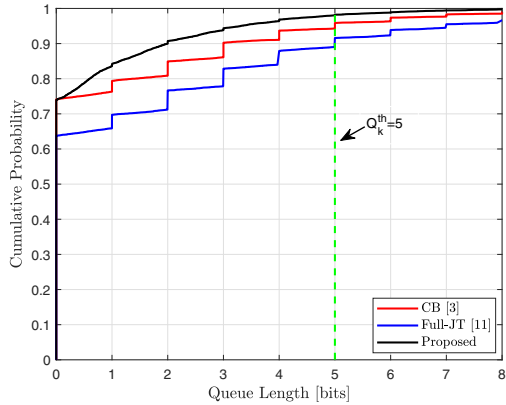


Fig. 2. Queue backlog with $V = 1$ and blockage probability of 20%.

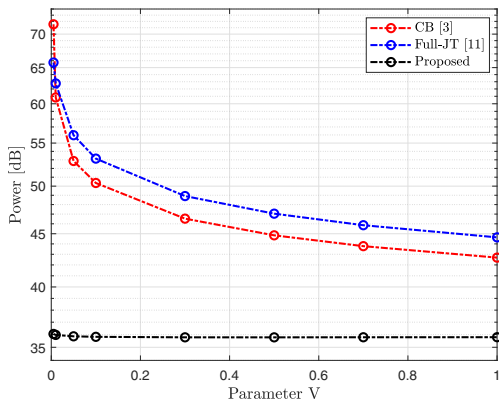


Fig. 3. Sum-power with increasing V and blockage probability of 20%.

Hence, the proposed method significantly outperforms the conventional full-JT and CB design.

Fig. 4 illustrates the impact of parameter L in constraint (9c) on achievable rate and reliable connectivity. For example, the use of smaller subset size L provides a lower SINR estimate (see (4)), and hence lower rate to each user. However, pessimistic SINR estimate over the link blockage combinations greatly improves the outage performance. Thus, it leads to more stable and resilient connectivity in the presence of random blockages. For example, with $V = 1$ and blockage of 30%, the outage is decreased from 46% to 9% by changing parameter L from 2 (full-JT) to 1 (proposed) in problem (9). Clearly, there is a trade-off between achievable rate, latency, and reliable connectivity. More specifically, for a given queue length (i.e., latency requirements), we can guarantee a minimum user-rate and vice-versa. However, compared to the baseline schemes, the proposed method provides energy-efficient, high-reliability and low-latency mmWave communication.

VI. CONCLUSION

In this paper, we have studied achievable latency and reliable connectivity in mmWave systems by exploiting the multi-antenna spatial diversity and CoMP transmission. We considered a time-average sum-power minimization problem subject to maximum allowable queue length constraint. We have adapted the Lyapunov optimization framework to derive a dynamic control algorithm for the long-term time-average

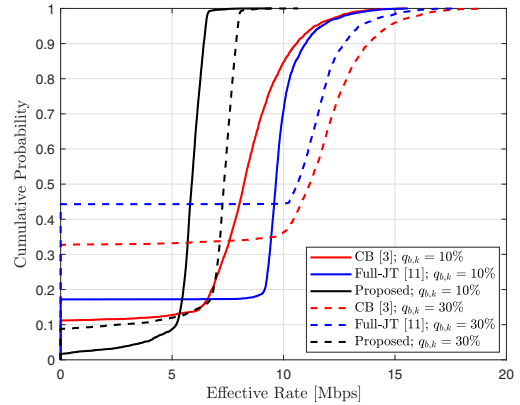


Fig. 4. Effective user-rate with $V = 1$, blockage probability $q_{b,k} = 10\%$ (solid line) and blockage probability $q_{b,k} = 30\%$ (dotted line).

stochastic problem. We have proposed an iterative algorithm for the sum-power minimization problem based on the SCA framework and KKT optimality conditions, while accounting for the uncertainties of mmWave radio channel in terms of blockages. The numerical results manifested the robustness of the proposed beamformer design in the presence of random blockage. Specifically, the achievable rate and sum-power performance with our proposed method outperforms the baseline scenarios and results in energy-efficient, high-reliability and low-latency mmWave communication.

REFERENCES

- [1] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proc. IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.
- [2] G. R. MacCartney, T. S. Rappaport, and S. Rangan, "Rapid fading due to human blockage in pedestrian crowds at 5G millimeter-wave frequencies," in *Proc. IEEE Global Commun. Conf.*, Dec. 2017.
- [3] A. Tölli, H. Pennanen, and P. Komulainen, "On the value of coherent and coordinated multi-cell transmission," in *Proc. IEEE Int. Conf. Commun. Workshop*, Jun. 2009, pp. 1–5.
- [4] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.
- [5] D. Maamari, N. Devroye, and D. Tuninetti, "Coverage in mmWave cellular networks with base station co-operation," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2981–2994, Apr. 2016.
- [6] G. R. MacCartney and T. S. Rappaport, "Millimeter-wave base station diversity for 5G coordinated multipoint (CoMP) applications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3395–3410, Jul. 2019.
- [7] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*, ser. Synthesis Lectures on Communication Networks. Morgan & Claypool, 2010, vol. 7.
- [8] M. Di Renzo, "Stochastic geometry modeling and analysis of multi-tier millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5038–5057, Sep. 2015.
- [9] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of queueing theory*. John Wiley & Sons, 2008.
- [10] G. Venkatraman, A. Tölli, M. Juntti, and L. Tran, "Traffic aware resource allocation schemes for multi-cell MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2730–2745, Jun. 2016.
- [11] J. Kaleva, A. Tölli, M. Juntti, R. A. Berry, and M. L. Honig, "Decentralized joint precoding with pilot-aided beamformer estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2330–2341, May 2018.
- [12] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [13] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *lecture notes of EE392o, Stanford University*, 2003.
- [14] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, 2002.