Reliable mmWave Communication via Coordinated Multi-Point Connectivity

Dileep Kumar, Student Member, IEEE, Jarkko Kaleva, Member, IEEE, and Antti Tölli Senior Member, IEEE

Abstract—The fundamental challenge of mmWave frequency band is the sensitivity of the radio channel to blockage, which gives rise to unstable connectivity and also impacts the reliability of a system. In this paper, we explore the viability of using Coordinated Multi-Point (CoMP) connectivity for robust and resilient downlink communication. We provide a novel iterative algorithm for Weighted Sum-Rate Maximization (WSRM) leveraging Successive Convex Approximation (SCA) with low computational complexity, which admits a closed-form solution via iterative evaluation of the Karush-Kuhn-Tucker (KKT) optimality conditions. Unlike in the conventional Joint Transmission (JT)-CoMP schemes, the proposed precoder design has, per-iteration, computational complexity in the order of base-station (BS) antennas instead of system-wide joint transmit antennas. This is achieved by parallel beamformer processing across the coordinating BSs. Furthermore, for the downlink precoder design, a conservative estimate of the user-specific rates are obtained by considering possible combinations of potentially blocked links, thus providing greatly improved communication reliability. In the presence of random blockages, the proposed schemes are shown to significantly outperform several baseline scenarios in terms of effective throughput and outage performance.

Index Terms—Coordinate multi-point, Joint transmission, Convex optimization, Karush-Kuhn-Tucker conditions, mmWave, Blockage, Weighted sum-rate maximization.

I. INTRODUCTION

The proliferation of ever-increasing data-intensive wireless applications along with spectrum shortage motivates the investigation of millimeter-wave (mmWave) communication for 5th-generation (5G) and beyond cellular systems [1]–[3]. mmWave communication not only provides relatively large system bandwidth but also the possibility of packing a significant number of antenna elements for highly directional communication [1], which is important to ensure link availability as well as to control interference in the dense deployments [2]. Thus, mmWave mobile communication is anticipated to substantially increase the average system throughput. There are still many issues that need to be resolved before these technologies are ready for the commercial applications. The fundamental challenge is the sensitivity of the mmWave radio channel to blockages due to reduced diffraction, higher path and penetration losses [3]. These lead to rapid degradation of signal strength and give rise to unstable and unreliable connectivity. Furthermore, connection reliability is aggravated by the fact that, for instance, a human blocker can degrade the channel quality by 30 dB for up to hundreds of milliseconds [3]. Presence of such frequent and long duration blockages significantly reduce the quality-of-service. To overcome such challenges, use of CoMP schemes, where the users are concurrently connected to multiple BSs, are imperative for more robust communication [4]–[12]. It is envisioned that joint transmission and reception via spatially separated transceivers will be vital in upcoming 5G systems [13].

A. Prior Work

CoMP transmission and reception are typically used to increase the system throughput, particularly for the cell-edge users due to relatively long distance from the serving BS and adverse channel conditions (e.g., higher path-loss and interference from neighboring BSs). Such scenarios have been widely studied over the past decade under the context of 4th-generation (4G) systems [4]–[7]. Techniques, such as, JT, Coordinated Beamforming (CB) and Dynamic Point Selection (DPS) were standardized in 3rd Generation Partnership Project (3GPP) and are currently deployed in Long Term Evolution-Advanced (LTE-A) to enhance capacity and converge by efficiently utilizing the spatially separated transceivers [6]. For example, it is shown in [7] that JT-CoMP increases the coverage by, up to, 17% for general users and 24% for cell-edge users compared to non-cooperative scenarios.

Recent studies have considered the deployment of CoMP in the mmWave frequencies [8]–[12]. Also, it is proposed in 3GPP specifications as a key component for upcoming 5G systems [13]. In [8], [9], the authors showed significant coverage improvement by simultaneously serving a user with spatially distributed transmitters. Results were drawn from extensive real-time measurements for 73 GHz in the urban open square scenario in Brooklyn. The network coverage gain for the heterogeneous mmWave system was also confirmed in [10] using stochastic geometry tools. The work in [11] proposed a low complexity cooperation technique for the JT, wherein a subset of cooperating BSs is obtained by selecting the strongest BS in each tier. The authors also provided the impact of blockage density in multi-tier heterogeneous network. Similarly to earlier works on two-stage hybrid analog-digital beamforming design e.g. in [14], [15], authors in [12] extended for JT-CoMP in multi-user massive MIMO systems with a high-dimensional analog RF precoder followed
by a low-dimensional centralized digital baseband precoder. However, CoMP techniques in [4]–[12] were still devised to enhance the capacity and coverage by utilizing the spatially separated transceivers. Thus, they were not originally designed for the stringent reliability requirements of, e.g., industrial-grade critical applications.

It is well known that a system can provide any level of reliability by sequential data transmission, i.e., by retransmitting the same message at various protocol levels, until a receiver acknowledges correct reception over a dedicated feedback channel [16]. However, in the presence of random link blockage and high path-losses, mmWave feedback links are equally unreliable and thus results in redundant retransmission. On the other hand, allowable latency determines a strict upper limit on the number of retransmission attempts [17].

The loss of connection in the mmWave communication is mainly due to a sudden blockage of the dominant links caused by abrupt mobility, self-blockage or external blockers [3]. Accurate estimation of each blocker requires precise environment mapping and frequent Channel State Information (CSI) acquisition, which might result in significant coordination overhead and severe synchronization challenges. Furthermore, blockage events can create large latencies if a passive hand-off is inevitable [17]. Thus, the limitations of retransmission and the difficulty of accurate estimation of random blockers has motivated us to develop more robust and resilient transmission strategies that can retain a stable connectivity under the uncertainties of mmWave channels and random link blockages.

B. Contributions

Motivated by above concerns, we propose an innovative beamforming design for the JT-CoMP to improve the system throughput and retain stable connectivity in mmWave mobile access under random blockages. The key contributions of this paper include:

- We provide a novel beamformer design for WSRM with a strong emphasis on system-level reliability by exploiting multi-antenna spatial diversity and CoMP connectivity to optimize the downlink rate. Wherein, we consider a conservative estimate of the achievable Signal-to-Interference-Noise Ratio (SINR) over all possible combinations of potentially blocked links among the cooperating BSs. Managing a large set of link blockage combinations is considerably more difficult than conventional constrained optimization [18]–[20] due to the mutually coupled SINR constraints.
- We provide a novel low-complexity SCA based beamforming algorithm for the original non-convex and computationally challenging problem. More specifically, all coupled and non-convex constraints are conservatively approximated with a sequence of convex subsets and iteratively solved until the convergence. The underlying subproblems, for each SCA iteration, become second-order cone programs (SOCPs), and that are efficiently solvable by any standard-of-the-shelf solvers.
- We propose a novel framework that merges the SCA with dual [21] and best response [22] methods to admit parallel beamformer processing for the distributed BSs via iterative evaluation of the closed-form KKT conditions. The schemes proposed in [18], [22] cannot be used directly, thus our proposed KKT based solution is significantly more advanced, and provides a novel approach for solving highly coupled minimum SINR constraints. This leads to a practical, latency-conscious and computationally efficient implementation for multi-core cloud edge clusters.
- We provide detail implementation of proposed methods assuming digital beamforming architecture. Furthermore, for the completeness, a low-complexity two-stage hybrid analog-digital beamforming implementation is provided in the numerical section. Thus, our proposed methods are scalable to any arbitrary multi-point configuration and dense deployment. Finally, we present numerical examples to quantify the complexity and the performance advantages of proposed solutions in terms of achievable sum-rate and reliable connectivity.

The paper is an extended version of our previously published conference papers [23], [24]. In [23] we investigated WSRM design leveraging the SCA framework. Then in [24], we extend above system and provided an iterative algorithm. All of the aforementioned results have been further improved and extended here.

C. Organization and Notations

The remainder of this paper is organized as follow. First, in Section II, we illustrate system architecture, channel and blockage model as well as provide the formulation of the problem. Section III provides theoretical analysis of blockage and evaluate throughput and reliability trade-offs. In Section IV, we describe the downlink precoder designs. The validation of our proposed methods with the numerical results are presented in Section V and finally we give our conclusions in Section VI.

Notations: In the following, we represent matrices and vectors with boldface uppercase and lowercase letters, respectively. The transpose, conjugate transpose and inverse operation are represented with the superscripts $\text{T}$, $\text{H}$ and $^{-1}$ respectively. |$\mathcal{X}$| indicates the cardinality of a set $\mathcal{X}$. $\Re\{\cdot\}$ and $|$ represent the real part and norm of a complex number, respectively. $\mathbf{I}_N$ indicates the $N \times N$ identity matrix. $\mathbb{C}^{M \times N}$ is a $M \times N$ matrix with elements in the complex field. $[\mathbf{a}]_n$ is the $n^{th}$ element of $\mathbf{a}$. Finally, $\nabla x y(x)$ denotes the gradient of $y(\cdot)$ with respect to variable $x$.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

We consider downlink in a mmWave based multi-user multiple-input-single-output (MU-MISO) communication system, consisting of $K$ single antenna users served by $B$ BSs. Each BS is equipped with $N_t$ transmit antennas, arranged in a uniform linear array (ULA) pattern. The antennas have 0 dBi gain and $s = \lambda/2$ spacing between any two adjacent elements, where $\lambda$ is the wavelength of carrier frequency.
We define \( B = \{ 1, 2, \ldots, B \} \) to be the set of all BS indices, \( K = \{ 1, 2, \ldots, K \} \) denotes the set active users, and the serving set of BSs for user \( k \) is represented with \( B_k \subseteq B \). We consider JT-CoMP, whereby, each active user \( k \) receives a coherently synchronous signal from the BSs in \( B_k \). Further, the downlink transmissions are performed using the same frequency and time channel. In this paper, if not mentioned otherwise, we restrict ourselves to the case where each antenna is connected to a dedicated baseband circuit, which effectively enables fully digital signal processing. In addition, we provide implementation for two-stage hybrid analog-digital beamforming architecture with coarse-level analog beamforming followed by a less-complex digital precoding. Finally, we assume a cloud (or centralized) radio access network (C-RAN), wherein all BSs are connected to the edge cloud by low-latency fronthaul links. It should be noted, in the C-RAN architecture, a common baseband processing unit (BBU) performs all the digital signal processing functionalities in a centralized manner, while the BSs implement only limited radio operations \[25\]. Such, fully centralized baseband processing provides more efficient BS coordination and, thus, enables more effective implementation for JT-CoMP \[25\]. However, in practice, fronthaul link capacity and signaling overhead will limit the maximum number coordinating BSs for each user. Furthermore, perfect estimation of available CSI is assumed at the BBU for the downlink precoder design and resource allocation, whereby, each BS receives related information for the serving users, such as control and data signals, using the fronthaul links.

The received signal \( y_k \) of user \( k \), \( y_k \) is defined as

\[
y_k = \sum_{b \in B_k} h_{b,k}^H f_{b,k} x_k + \sum_{u \in K \setminus b \in B_u} h_{b,k}^H f_{b,u} x_u + w_k, \tag{1}
\]

where \( h_{b,k} \in \mathbb{C}^{N_t \times 1} \) is the channel between a BS-user pair \( (b, k) \), \( w_k \sim \mathcal{CN}(0, \sigma^2) \) is circularly symmetric additive white Gaussian noise (AWGN) with power spectral density \( (PSD) \) of \( \sigma^2 \) and \( x_k \) is normalized data symbol, i.e., \( \mathbb{E}\{|x_k|^2\} = 1 \). Finally, \( f_{b,k} \in \mathbb{C}^{N_t \times 1} \) represents the portion of the joint precoder between a BS-user pair \( (b, k) \), designed by the BBU assuming perfect estimation of available CSI. Finally, the received SINR for each user \( k \) is formulated as

\[
\Gamma_k(B_k) = \frac{|\sum_{b \in B_k} h_{b,k}^H f_{b,k}|^2}{\sigma^2 + \sum_{u \in K \setminus b \in B_u} |\sum_{b \in B_u} h_{b,k}^H f_{b,u}|^2}. \tag{2}
\]

### B. Channel Model

Due to the relatively higher penetration and path-loss, low-scattering and reduced diffraction at the mmWave frequencies compared to sub-6 GHz frequency band, the channel can be considered to be spatially sparse \[1\], in which line-of-sight (LoS) is the dominant path and mainly contributor to the communications \[3\], \[8\], \[9\]. Thus, unblocked LoS link is highly desirable to initiate and maintain a stable mmWave communication \[26\]. The channel design, in this paper, is based on sparse geometric model using extended Saleh-Valenzuela model, which is widely adapted for the mmWave signal processing techniques \[14\], \[15\]. Specifically, we consider \( M_{b,k} \) paths for the channel \( h_{b,k} \) between a BS-user \( (b, k) \) pair, and expressed as

\[
h_{b,k} = \sqrt{\frac{N_t}{M_{b,k}}} \left\{ g_{b,k}^1 a_T^H(\phi_{b,k}^1) + \sum_{m=2}^{M_{b,k}} g_{b,k}^m a_T^H(\phi_{b,k}^m) \right\}, \tag{3}
\]

where \( \phi_{b,k}^1, \phi_{b,k}^m \) denote the angle-of-departure (AoD) for LoS (NLoS) path. Note that the AoD for each NLoS path \( m > 1 \) is assumed to be uniformly distributed, i.e., \( \phi_{b,k}^m \in [0, 2\pi] \), whereas, \( \phi_{b,k}^1 \) is related to actual position of BS-user \( (b, k) \) pair \[27\]. Finally, \( g_{b,k}^1 = v_{b,k} d_{b,k}^0 \) and \( g_{b,k}^m = v_{b,k} d_{b,k}^{-\eta} \), in which \( v_{b,k} \) is the BS-user distance, \( d_{b,k}^{-\eta} \) is random complex gain with zero mean and unit variance and \( \eta \) denotes the path-loss exponent for the LoS (NLoS) link. It has been shown empirically shown that \( \eta \) is much higher than the LoS path-loss exponent \( \eta \) \[3\], \[8\], \[9\]. The ULA transmit array steering vector is \( a_T(\phi) \in \mathbb{C}^{N_t \times 1} \) with

\[
[a_T(\phi)]_n = \frac{1}{\sqrt{N_t}} e^{j \frac{2\pi}{\lambda} (n-1) \sin(\phi)}, \quad n = 1, 2, \ldots, N_t. \tag{4}
\]

### C. Blockage Model

In the mmWave communication, the quality of wireless link between a BS-user pair, which could translates into a user’s achievable rate, mainly depends on the peculiarity of LoS path \[3\]. In this case, the major challenge stems from the fact that such LoS links may easily be blocked by obstacles. This results in intermittent connection, which severely impacts the quality of user-experience. In a typical mmWave scenario, channel modeling shows that outage on a mmWave link occurs with \( 20\% - 60\% \) probability \[28\] and that leads to over 10-fold decrease in the achievable throughput \[29\]. We consider a probabilistic blockage model, where each link undergoes independent blockage. This is reasonably accurate (especially when the blockers are not overly large) \[26\].
We simply assume blockage for the dominant LoS path while NLoS links are unobstructed. More specifically, the channel between any typical BS-user pair can have one of two states: fully-available or NLoS state. The NLoS state occurs when the dominant LoS link is blocked by any obstacle. It should be noted, even a mobile human blocker causes 30 dB attenuation and can obstruct the LoS path for hundreds of milliseconds [3]. This can be equivalently modeled as $0^{N \times 1}$ for the blocked LoS component. The fully-available state is defined in [3]. An additional inaccessible state may occur if the corresponding path-loss becomes sufficiently high to establish a communication link [29].

Since blockers are completely random, their position and orientation may change multiple times within the channel coherence interval. Further, due to the non-correlated blocking, blockage event and position of each blocker cannot be known a priori in a dynamic mobile environment. We assume a standard Time-Division-Duplex (TDD)-based CSI acquisition from reciprocal uplink followed by downlink data transmission phase. More specifically, BBU designs the transmit beamformer based on the available CSI acquired during the estimation phase [3]. Thus, a system can be in outage, if the dominant LoS link is not anymore available during the transmission phase due to random blocking. Similarly, a LoS link can also be in the blockage state during the channel estimation phase and available during data transmission phase. However, these links will not be included for the data transmission. Thus, from the reliability prospective, we have to consider the case when channel is available during the estimation phase but blocked during the transmission, which is not a priori known at the BBU for downlink precoder design.

### D. Problem Formulation

The major goal of this work is to develop a robust and resilient downlink transmission strategy that can retain a stable connectivity under the uncertainties of mmWave channel and random link blockages. To do that, we need to compute the optimal joint transmit beamformer $F = [\mathbf{f}_{1,1}, \mathbf{f}_{1,2}, \ldots, \mathbf{f}_{B,K}]$, while emphasizing the system-level reliability by exploiting the multi-antenna spatial diversity of the CoMP connectivity. The precoder design can be formulated as WSRM problem as

$$\text{maximize} \quad \sum_{k \in K} \delta_k \log (1 + \tilde{\gamma}_k) \quad (5a)$$

subject to \quad $\Gamma_k(B'_k) \geq \tilde{\gamma}_k \quad \forall k, \forall B'_k \subseteq B_k, \quad |B'_k| \geq L, \quad (5b)$

$\sum_{k \in K} \|\mathbf{f}_{b,k}\|^2 \leq P_b \quad \forall b, \quad (5c)$

where $\delta_k \geq 0 \forall k$ denotes the user-specific priority weights corresponding to the rate and $\Gamma_k(B'_k)$ is defined as in (2).

The total transmit power for each BS $b$ is bounded by $P_b$ in (5c). Finally, (5b) is the lower estimate of available SINR for each user $k$, which is computed over all possible subset combinations of size $|B'_k| \geq L$ of the serving set of BSs $B'_k (\subseteq B_k)$.

The resulting problem (5) is non-convex and intractable due to potentially large number of combinations of non-convex SINR constraint. Specifically, $\Gamma_k(B'_k)$ in (5b) cannot be formulated as a convex constraint. To this end, in Section IV we provide practical and computationally efficient iterative algorithms by exploiting the convex approximation techniques.

### III. ANALYSIS OF RELIABILITY AND THROUGHPUT TRADE-OFF

In this paper, we assume uncorrelated and randomly distributed blockers. Thus, the position of each blocker and/or blockage event is completely unknown. Therefore, to maintain a reliable and resilient connectivity, the BBU models the SINR over all possible subset combination of potentially blocked links and uses the worst case estimate of $\tilde{\gamma}_k$, in the objective, for the beamformer design. For example, referring to (5b) and Fig. 1 let $L = 3$ and $|B_k| = 4$, then the subset $B'_k \in B_k$ includes following BS indices

$$B'_k = \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}, \{1, 2, 3, 4\}.$$  

Consequently, at the solution $\tilde{\gamma}_k = \min(\Gamma_k(B'_k))$, where $c = [1, 2, \ldots, \sum_{i=1}^{L} |B'_k|]$. However, in actual, each user $k$ is always communicating with all $B_k (\subseteq B'_k)$ BSs, unless the dominant LoS link between a BS-user pair is not available because of the blockage. Therefore, a reliable connectivity for each user $k$ can be guaranteed, even if, $|B'_k| = L$ dominant links are not available during the transmission phase. In contrary, if more than $L$ links were available, the actual achievable rate would be somewhat larger than the assigned rate.

As an example, using the binomial distribution, we can approximately model the success probability $\tilde{q}_k$ for user $k$ as

$$\tilde{q}_k(|B_k|, q_{b,k}) = \sum_{t=0}^{(|B_k|)} \binom{|B_k|}{t} (1 - q_{b,k})^{(|B_k| - t)} q_{b,k}^t \forall k, \quad (6)$$

where $q_{b,k} \in [0, 1]$ is the probability of a channel is in the blockage state. Since all user are independent, the system is in outage if any of the $K$ users is in outage. It should be noted, this is a worst case assumption to enforce a strict system-level reliability. However, in practice, users with the unblocked links can still decode their received signal. Thus, with the worst case assumption, system outage is defined as

$$\bar{P}_{out} = 1 - \prod_{k=1}^{K} \tilde{q}_k(|B_k|, q_{b,k}). \quad (7)$$

It should be noted, closed-form expression in (7) models the case when the channel between a BS-user pair is either available or fully-blocked. However, we consider blocking of the dominant LoS link while keeping all NLoS components unobstructed. Thus, (7) only provides an approximation on the outage performance, as shown in Section IV.

Intuitively, we can observe the impact of (5b) on the system reliability and achievable sum-rate using (7). For example, by using the smaller subset size $L$, we can improve the system reliability. However, it leads to lower SINR estimate and, hence, lower rate to each user. Conversely, larger subset size $L$ can provide higher SINR and rate, but it is more susceptible
to outage, thus resulting in less stable connectivity for each user. Clearly, there is a trade-off between achievable sum-rate and reliable connectivity.

IV. PROPOSED PRECODER DESIGNS FOR RELIABLE COMMUNICATION

In this section, we elaborate on solving (5), which is intractable as-is, mainly due to non-convex SINR in (5b). Several approaches have been outlined in existing literature to handle such SINR non-convexity. However, in this paper, we employ a widely used convex approximations based on SCA framework [30], [31], wherein all non-convex constraints are approximated with the sequence of convex approximations. The underlying approximated subproblem is then iteratively solved until the convergence. SCA based solutions have been widely used in many practical applications, e.g., in satellite system [32], wire-line DSL network [33], small-cell heterogeneous network [34], energy efficiency [20], spectrum sharing [35] and decentralized multi-antenna interference coordination [19], [22]. In view of the prior works [19], [20], [22], there is a systemic approach for the design of downlink precoder in JT-CoMP scenario while accounting the uncertainties of mmWave channel and random blockers, thus motivating the current work.

A. Solution via Sequential Convex Approximation

The non-convex SINR constraint in (5b) can be handled by SCA technique, as shown in [19], [20], [22], where the authors provided SINR approximation method for CB [19], [20] and JT [22] assuming global CSI and no blockage. We extend these approaches to take into consideration coherent multi-point transmission and provide a novel grouping of multitude of potential SINR conditions that raise from the link blockage subsets. The individual SINR constraint for each \( B_k \subseteq B \) can be solved as in [19]. The exact details are omitted due to lack of space (we refer the reader to [19] for the details).

In the following, the main steps are briefly reproduced. To begin with, the SINR expression in (5b) can be relaxed as

\[
\tilde{\gamma}_k \leq \Gamma_k(B_k^c) \triangleq \frac{|\sum_{b \in B_k^c} \mathbf{h}_{b,k}^H \mathbf{f}_{b,k}^*|^2}{I_k(B_k^c)}, \quad (8)
\]

where (8) is the lower bound for SINR as seen by user \( k \) \( \forall k \) and (9) is the total interference plus noise. It should be noted that each subset combination \( c \) in (5) will have a unique interference constraint, which includes all the interference terms corresponding to that BSs combination. For the conciseness of mathematical representation, (5) can be rewritten as

\[
1 + \tilde{\gamma}_k \leq \frac{\sigma^2 + \sum_{i \in \mathcal{K}} |\sum_{b \in B_k^c} \mathbf{h}_{b,k}^H \mathbf{f}_{b,u}^*|^2}{I_k(B_k^c)}, \quad (9)
\]

where (9) is the lower bound for SINR as seen by user \( k \) \( \forall k \) and (10) is the total interference plus noise. It should be noted that each subset combination \( c \) in (5) will have a unique interference constraint, which includes all the interference terms corresponding to that BSs combination. For the conciseness of mathematical representation, (5) can be rewritten as

\[
1 + \tilde{\gamma}_k \leq \frac{\sigma^2 + \sum_{i \in \mathcal{K}} |\sum_{b \in B_k^c} \mathbf{h}_{b,k}^H \mathbf{f}_{b,u}^*|^2}{I_k(B_k^c)}, \quad (10)
\]

Finally, we denote the stacked channel for user \( k \) corresponding to each subset combination \( c \) as \( \mathbf{H}_k \). For example, again referring to Fig. 1 let \( L = 3, |\mathcal{B}_k| = 4 \) and \( \mathcal{B}_k^c = \{1, 2, 4\} \) then the corresponding stacked channel \( \mathbf{H}_k \triangleq [\mathbf{h}_{1,k}, \mathbf{h}_{2,k}, 0, \mathbf{h}_{4,k}] \). Since, \( \gamma_k \forall k \) is computed over all possible subset combinations, the order of the BS indices is completely irrelevant.

Even with this relaxation of SINR in (2) by (11), problem (5) is still non-convex due to quadratic-over-linear constraint in (11), which cannot be handled directly [19]. Here, we resort to the SCA framework [30], [31], wherein all non-convex SINR constraints in (11) are approximated with a sequence of convex subsets and iteratively solved until the convergence [19], [20], [22]. We note that the RHS of (11) is a convex function, and thus can be bounded by the first order Taylor approximation as

\[
F(c, \mathbf{f}_k, z_k) \triangleq \sum_{k \in \mathcal{K}} \delta_k \log(1 + \tilde{\gamma}_k) \leq 2R \sum_{k \in \mathcal{K}} \frac{\tilde{\gamma}_k (\mathbf{f}_k^H \mathbf{h}_{b,k})^2}{1 + z_k} (1 - \frac{\tilde{\gamma}_k - \gamma_k}{\tilde{\gamma}_k + 1}) \quad (14)
\]

where (14) is the under-estimator for the RHS of (11), with equality, only at the operating points \( \{\bar{\gamma}_k, z_k\} \). After replacing (5b) with LHS of (14), an approximated subproblem for \( i^{th} \) SCA iteration is expressed in convex form along with the corresponding dual-variables as

\[
\max_{\mathbf{F}, \gamma_k} \sum_{k \in \mathcal{K}} \delta_k \log(1 + \gamma_k) \quad (15a)
\]

subject to

\[
a_{k,c} \rightarrow F(c, \mathbf{f}_k, \gamma_k) \geq I_k(B_k^c) \quad (15b)
\]

where \( \mathbf{a} = [a_{1,1}, a_{2,1}, \ldots, a_{K,C}]^T \) and \( \mathbf{z} = [z_1, z_2, \ldots, z_B]^T \) are the non-negative Lagrangian multipliers corresponding to the constraints (15b) and (15c), respectively. Dual variable \( z_b \) in (15c) is associated with the total transmit power constraint of BS \( b \) and for each user \( k, a_{k,c} \forall c \) is associated with each subset combinations of the SINR constraint. The role of the dual variables become clear in the following subsection. The above convex subproblem for each SCA iteration can be efficiently solved, in general, using existing convex optimization toolboxes, such as CVX [36]. The fixed operating points for the current iteration are updated from the solution of current SCA iteration. This is repeated until the convergence of the objective. Finally, the precoder design with proposed SCA relaxation has been summarized in Algorithm [1]
Algorithm 1: Successive Convex Approximation algorithm for Weighted Sum-Rate Maximization problem

1. Set $i = 1$ and choose any feasible initial points $\{f_k^{(0)}, \gamma_k^{(0)}\}$ $\forall b \in B$, $\forall k \in K$

2. repeat

3. Solve (15) with $\{f_k^{(i-1)}, \gamma_k^{(i-1)}\}$ and denote the local optimal values as $\{f_k^*, \gamma_k^*\}$

4. Update $f_k^{(i)} = f_k^*$ and $\gamma_k^{(i)} = \gamma_k^*$

5. $i = i + 1$

6. until convergence or for fixed number of iterations;

B. Low-Complexity Solution via KKT Conditions

Problem (15) can be more efficiently solved by parallel beamformer updates across the spatially distributed BSs. Unlike the approach presented in previous subsection, the precoder can be obtained by iteratively solving a system of KKT equations [37]. Thus, the KKT based solution provides closed-form steps for an algorithm that do not rely on generic convex solvers. Furthermore, iterative evaluation of KKT optimality conditions, for each SCA step, reveals conveniently parallel structure for the beamformer design with significantly lower computational complexity w.r.t. joint beamformer optimization across all distributed BS antennas. First, we briefly outline key challenges and our proposed solution.

The Lagrangian $L(F, \gamma_k, a_{k,c}, z_b)$ of (15) is given in (17). It should be noted that the KKT optimality conditions provide a first-order necessary and sufficient conditions for the solution of any convex optimization problem. Therefore, it is sufficient to show that the approximated convex subproblem satisfies the Slater’s conditions [37]. Thus, the solution for (15) can be obtained by iteratively solving a system of KKT optimality conditions, which include stationary, complementary slackness, and primal-dual feasibility requirements (for more detailed derivation see Appendix A).

It is worth pointing out that in (15b), user specific SINR constraints are highly interdependent over the link blockage combinations. This makes deriving an efficient solution via the KKT optimality conditions considerably more difficult than the case with only single SINR constraint per user [18], [22]. To overcome this challenge, Lagrangian multipliers $a_{k,c}$ $\forall (k, c)$ corresponding to each SINR constraint are iteratively solved using subgradient method. Another practical challenge is that the design of optimal precoder $F$ is inherently coupled between all distributed BS antennas due to coherent joint transmission to each user. The computational complexity of optimal precoder $\bar{F}_k$ $\forall k$ scales exponentially with length of joint beamformers (BNF) as in (21b), which quickly becomes intractable for dense deployments. Furthermore, BS specific dual variables $z_b$ $\forall b$ should be computed simultaneously (see (21b) in Appendix A). However, due to coupling and interdependence between $z_b$ $\forall b$ because of JT-CoMP, it is computationally challenging to obtain their exact values. To avoid this, we also incorporate a parallel optimization framework using best response [22] into the iterative optimization process (21c).

As a result, BS specific beamformers are solved in parallel for each iteration, while assuming that coupling from other cooperating BSs is fixed to previous iteration solution.

In the following, we provide a novel iterative algorithm by combining the SCA framework with dual and best response methods, which admits the closed-form solution in each step. Specifically, in each SCA iteration, the approximated convex subproblem is solved via the iterative evaluation of the KKT optimality conditions. Furthermore, to improve the rate of convergence, the SCA approximation point $\{f_k^{(i)}, \gamma_k^{(i)}\}$ is also updated in each iteration along with the dual variables (for more detailed see Appendix A).

To summarize, the steps in the iterative algorithm are

$$f_k^{(i)} = \left( \bar{F}_k^* + \sum_{c \in C} \sum_{u \in K} a_{k,c}^{(i-1)} h_{b,u}^T h_{b,u} \right)^{-1} t_k^{(i-1)}; \tag{16a}$$

$$f_k^{(i)} = f_k^{(i-1)} + \psi \left[ f_k^{(i-1)} - f_k^{(i-1)} \right]; \tag{16b}$$

$$\Gamma_k^{(i)}(B_k^i) = \frac{\bar{h}_k^H f_k^{(i)}}{\sigma^2 + \sum_{u \in K \setminus k} |h_{b,u}^T f_k^{(i)}|^2}; \tag{16c}$$

$$z_k^{(i)} = \delta_k \left( \sum_{c \in C} \sum_{u \in K} a_{k,c}^{(i-1)} \frac{\sigma^2 + \sum_{i,c} \bar{h}_k^H f_k^{(i-1)} (1 + \gamma_k^{(i-1)})^{-1}}{(1 + \gamma_k^{(i-1)})^2} - 1 \right) \tag{16d};$$

$$a_{k,c}^{(i)} = \left( a_{k,c}^{(i-1)} + \beta \left( z_k^{(i)} - \Gamma_k^{(i)}(B_k^i) \right) \right) ^+; \tag{16e}$$

where

$$t_k^{(i-1)} = \left\{ \sum_{c \in C} a_{k,c}^{(i-1)} \left[ \frac{f_k^{(i-1)} h_{b,u}^T h_{b,u}}{1 + \gamma_k^{(i-1)}} \right] - \sum_{c \in C} \sum_{u \in K \setminus k} a_{k,c}^{(i-1)} \sum_{g \in B} f_k^{(i-1)} h_{g,u}^T h_{b,k} \right\};$$

$$\psi > 0 \text{ and } \beta > 0 \text{ are small positive step sizes, and } x^+ \triangleq \max(0, x).$$

Thus, the KKT expressions in (16) are solved in an iterative manner, starting with initializing the variables $\{f_k^{(0)}, \gamma_k^{(0)}, a_{k,c}^{(0)}\}$ to some feasible values, such that SINR and total transmit power constraint for each distributed BS is satisfied. Finally, the beamformer design via iterative evaluation of KKT optimality conditions has been summarized in Algorithm 2.

Algorithm 2: Low-Complexity Iterative algorithm for Weighted Sum-Rate Maximization problem

1. Set $i = 1$ and choose any feasible initial point $\{f_k^{(0)}, \gamma_k^{(0)}, a_{k,c}^{(0)}\}$ $\forall b \in B$, $\forall k \in K$, $\forall c \in C$

2. repeat

3. Solve $f_k^{(i)}$ from (16a) with $\{f_k^{(i-1)}, \gamma_k^{(i-1)}, a_{k,c}^{(i-1)}\}$

4. Update $f_k^{(i)}$ using (16b)

5. Calculate $\gamma_k^{(i)}$ from (16d)

6. Update $a_{k,c}^{(i)}$ using (16e) with $\{\gamma_k^{(i)}, \Gamma_k^{(i)}(B_k^i)\}$

7. $i = i + 1$

8. until convergence or for fixed number of iterations;
\[ L(F, \hat{\gamma}_k, a_{k,c}, z_b) = -\sum_{k \in K} \delta_k \log (1 + \hat{\gamma}_k) + \sum_{k \in B} z_b \left( \sum_{k \in K} \|f_{a,k}\|^2 - P_b \right) + \sum_{k \in K} \sum_{c \in C} a_{k,c} \left( \sigma^2 + \sum_{u \in K \setminus k} \|h_{k}^H f_{u}\|^2 - 2\Re \{h_{k}^H f_{k}^H (1 + \hat{\gamma}_k) (f_{k}^H - f_{k}^{(i)}) \} \right) - \sigma^2 + \sum_{i \in K} \|h_{i}^H f_{i}^{(i)}\|^2 \left( \frac{1 + \hat{\gamma}_k}{1 + \hat{\gamma}_k} \right) \]. \tag{17}

1) Lagrangian Multipliers: Dual variables \(a_{k,c} \forall (k, c)\) corresponding to \([15]\) are interdependent and highly coupled due to the common SINR constraint over the link blockage combinations. Their exact values for each SCA iteration can not be obtained as a close-form expression. Therefore, we resort to widely used subgradient approach, such as the constrained ellipsoid method, which converges to the local optimal solution for a convex optimization problem \([21]\). It should be noted that choice of the step size \(\beta\) in \((16e)\) depends on the system model, as it directly affects the convergence rate as well as control the oscillation in WSRM objective. There have been several studies in the literature on the convergence properties of subgradient approach, with the different step size rules \([21], [38], [39]\). More precisely, monotonic convergence can not be guaranteed, in general, for the constrained ellipsoid method, and thus, one has to track and adjust the step size accordingly. In the proposed iterative approach in Algorithm 2, the dual variables \(a_{k,c} \forall (k, c)\) are updated based on the violation of the SINR \((16e)\) with a small positive step size.

Dual variables \(z_b \forall b\) are chosen to satisfy the total transmit power constraints \((15c)\), using the bisection search method. It should be noted, in a multi-cell scenario, the total sum-power constraint may not necessarily holds with the equality. Specifically, for each BS \(b\) if \(\sum_{k \in K} \|f_{k,b}\|^2 < P_b\) then \(z_b = 0\), i.e., non-negative dual variable \(z_b\) is set zero in order to satisfy the corresponding complementary slackness condition \([37]\). Otherwise, there exist \(a_{k,c} > 0\) such that \(\sum_{k \in K} \|f_{k,b}\|^2 = P_b\). This value is unique and provides the solution \([37]\), which can be calculated by the bisection search with respect to the total power constraint.

2) Best Response: The precoder \(F\) is inherently coupled between all distributed BS antennas \([16]\), because of the coherent joint transmission to each user. One possible approach is based on updating the beamformers sequentially, i.e., using the Gauss-Seidel type update process, which provides monotonic convergence for a WSRM optimization problems. However, it is shown in \([40]\) that the convergence rate drastically reduces even with a slight increase in number of cooperating BSs. Here, instead, we implement a parallel optimization framework \([41]\), which efficiently parallelizes the beamformer updates across the distributed BS antennas and hence significantly reduces the per-iteration computational complexity. For a given iteration, BS specific beamformers are solved in parallel while assuming the coupling from other BSs fixed to the previous iteration solution, as in \((16a)\). The objective function can be shown to converge if we allow sufficiently large number of subgradient iterations per fixed SCA approximation (until increased objective) for each BS before making the best response step with sufficiently small step sizes \([22]\). However, here we are more interested in fast and robust rate of convergence, for which we allow only a single subgradient update per best response iteration. It is shown by numerical examples that this provides good performance with small number of iterations. More details on the convergence behavior and choice of step size \(\psi \in (0, 1)\) with the best response based parallel optimization approach are provided in \([41]\). For time invariant channels, convergent \(\psi\) can be bounded to Lipschitz constant of the objective \([41]\). It should be noted that the transmit power constraints are convex, therefore, regularized update with \(\psi < 1\) will strictly preserve the feasibility of total transmit power.

3) Feasible Initial Point: In the SCA framework, all non-convex constraint are approximated with sequence of convex subset and then iteratively solved until the convergence \([30], [31]\). Thus, it is very important to initialize the iterative algorithm with some feasible starting points. To this end, one possible solution for the feasible initial \(f_{k,0}\) is any beamformer satisfying the sum-power constraint \((5c)\), which can easily be obtained by scaling any randomly generated beamforming vector. Then, calculating the lower bound on achievable SINR from \((3)\), i.e., \(\hat{z}_{k,0} = \min(\Gamma_k(B_k))\). However, it should be noted that depending on the resulting SINR, the randomly generated initial solution can be very far from an solution and may require significant number of iterations until the convergence. As an example, for a system model with \(N_t \geq K\), an efficient initial point can be obtained by simply matching \(f_{k,0}\) to the corresponding channel \(h_{b,k} \forall (b, k)\), i.e., based on maximum ratio transmission (MRT), by neglecting the potential in-cell and inter-cell interference. In addition, non-negative dual variables \(a_{k,c}^{(0)} \in [0, 1] \forall (k, c)\) are randomly chosen such that LHS of \((15)\) is strictly positive. It is worth highlighting that initializing the algorithm with different values does not, on the average, affect the final solution of \((5)\), provided sufficient number of iterations, and that initial point is feasible and satisfies the constraints.

4) Complexity Analysis: The approximated convex sub-problem \([15]\) can be solved in a generic convex optimization solver as a sequence of second-order cone programs (SOCP) \([36], [42]\). The complexity of the problem scales exponentially with the length of the joint beamformers \((BN_t)\) and number of constraints \([42]\). Thus, particularly, for dense mmWave deployments with large \(N_t\) and \(B\), the complexity quickly becomes intractable in practice. The complexity of the proposed KKT solution is dominated by \((16a)\), which mainly consist of matrix multiplications and inversion operations, and that scale in BS specific beamformer \((N_t)\). Furthermore, the dimensions of all inverse operation in \((16)\) are directly proportional to number of transmit antennas \((N_t)\) at each
BS. In addition, the complexity of matrix inversion operation in \[16\] can be alleviated by solving \(f_{b,k}^*\) from a system of linear equations. Thus, providing significant reduction in the computational complexity for even modestly sized systems.

V. SIMULATION RESULTS

This section provides numerical results to validate the performance of proposed methods. In particular, we analyze the impact of subset size \(L\) on outage performance and achievable throughput, as well as evaluate the trade-off between these performance metrics. We further elaborate on convergence behavior and performance gap between proposed algorithms.

A. Simulation Setup

We consider a downlink mmWave MU-MISO system with \(B = K = 4\), where the BSs are equipped with ULAs of \(N_t = 16\) antennas. Each BS is placed at the corner of a square cell with the inter-site spacing of 100 meters and connected to common BBU in the edge cloud. We consider JT-CoMP with full-coordination, i.e., \(B_k = B\ \forall k\), such that each active user is coherently served by all BSs (excluding the potentially blocked links). For simplicity, all BSs are assumed with the same maximum transmit power, i.e., \(P_b = P_t \ \forall b\), noise variance of \(\sigma^2 = 1\), carrier-frequency \(f_c = 28\) GHz and the cell-edge SNR = 15 dB. More precisely, SNR = \(P_t/d_e^{\psi/2} \ \forall b\), where \(d_e^{\psi/2}\) denotes the cell-edge distance and \(\psi = 2\). The user priorities are set to be equal (i.e., \(\delta_k = 1 \ \forall k\)). All users are assumed to be randomly dropped within the coverage region, hence having different path gain and AoD with respect to each BS. If not mentioned otherwise, each antenna is equipped with a dedicated baseband circuit to enable fully digital signal processing. We use \(\beta = 0.005\) and \(\psi = 0.05\) for the subgradient and best response step sizes, respectively.

B. Outage Probability

The first result, illustrated in Fig. 2 shows the outage performance as a function of increasing link blockage probability \(q_{b,k}\) between each BS-user pair \((b,k)\) for the WSRM problem. Outage event occurs if the assigned transmit rate \(R_k\) exceeds the achievable link rate \(C_k\), for any user \(k = 1,\ldots,K\), given all available links i.e.,

\[
P_{out} \triangleq \mathbf{P}\{R_k > C_k\} \ \forall k \in K.
\]

It can be concluded from Fig. 2 that the outage performance is greatly improved by decreasing the subset size \(L\). Clearly, lower \(L\) provides more stable and robust communication. The beamformer design with \(L = 1\) can provide reliable connectivity even if all but one LoS links are blocked. For example, with the link specific blockage probability of 0.2 the outage probability is decreased from 0.9 to (close to) zero by changing \(L\) from 4 to 1. Thus, the specific allocation with \(L = 1\) can withstand blockage up to a single active link.

Comparing with the theoretical approximation in \[7\], for a specific scenario with \(L = 1\), the outage performance appears slightly more, as the optimization problem solved at the BBU may end up assigning non-zero powers to only a subset of users, while all remaining users are assigned zero rate. In such a scenario, missing a LoS link results in somewhat different blockage than what is predicted by the theoretical formula \[7\]. An increase in the subset size \(L\) also increases the SINR estimate. Thus, it is likely that all active users are assigned with some non-zero downlink rate, and, therefore, the simulated results closely match to theoretical results obtained using \[7\]. Furthermore, it can be seen that the proposed method significantly outperforms the conventional full-JT \((L = B)\), CB and MRT based precoder designs.

C. Effective Throughput

Fig. 3 illustrates the trade-off between achievable throughput and outage performance with identical user-specific weights (i.e., \(\delta_k = 1 \ \forall k\)) for WSRM problem. Here, the effective throughput \(T_e\) is defined as

\[
T_e \triangleq (1 - P_{out})R,
\]

Fig. 2. Outage performance as function of increasing blockage probability with theoretical formula \(\square\) (solid) and simulations (dash-dotted).

Fig. 3. Effective throughput as a function increasing blockage probability.
where \( R = \sum_{k \in K} \delta_k \log \left(1 + \frac{\gamma_k}{\delta_k}\right) \), i.e., when each active user successfully receives the transmit data. It can be observed that with the smaller subset size \( L \), throughput is significantly reduced due to pessimistic estimate of the aggregated SINR. However, it remains stable, even at much higher link blockage probabilities. On the contrary, with the conventional JT (\( L = B \)), the throughput quickly approaches zero due to higher outage. Clearly, there is trade-off between achievable system throughput and outage performance. More specifically, for a given outage threshold, we can guarantee minimum achievable throughput and vice versa. In addition, parameter \( L \) can be considered as an optimization variable for a given outage performance. Our proposed method provides more robust and resilient connectivity under uncertainties of mmWave channel, whereas, with the conventional full-JT (\( L = B \)), CB and MRT methods, sum-rate rapidly decreases towards zero, even if the blockage probability is slightly increased.

D. Effect of Initialization and Step Sizes

First, in Fig. 4 and Fig. 5 we examine the convergence behavior of iterative dual method with \( L = 3 \). It can be concluded that convergence is very sensitive to choice of the step sizes \((\psi, \beta)\). For example, with the larger value of step sizes, algorithm can converge with fewer number of iterations, but might results in more fluctuations to the sum-rate objective. It should be noted that convergence may not necessarily be monotonic, which is inherent feature of subgradient update (16e).

In addition, ingenious choice of the feasible initial points \( \{\hat{w}_{b,k}, \hat{\gamma}_k\} \) also impact the rate of convergence. For the considered scenario with \( N_t \geq K \), a simple MRT based initialization for \( \hat{w}_{b,k} \forall (b,k) \) significantly improves the rate of convergence and attains near optimal solution with fewer number of iterations, as shown in Fig. 5. However, irrespective of the choice of initialization point and step sizes \((\beta, \psi)\), both algorithm converges to same local optimal solution on the average, provided sufficient number of iterations.

Finally, we compare the sum-rate performance of low-complexity method based on iteratively solving set of optimal KKT conditions with the solution obtained directly by the optimization toolbox [36]. Moreover, with the assumption of full-CSL, (on average) the achievable performance approaches to the theoretical upper bound. It can be concluded, from Fig. 6 that Algorithm 2 achieves near optimal performance and the resulting gap in the sum-rate is mainly due to insufficient convergence because of the fixed number of maximum iterations. Therefore, the proposed KKT based iterative method provides a low-complexity solution for practical implementations without significant degradation in the achievable performance.

E. Hybrid Analog-Digital Beamforming Implementation

While the problem formulation and proposed solutions are generic, they can be easily extended to any multi-point configuration. Until now, we have restricted ourselves to the case where each antenna is equipped with a dedicated baseband circuit that enables fully digital signal processing. In this subsection, we provide an implementation for two-stage hybrid analog-digital architecture with a coarse-level analog beamforming and limited number of RF circuits followed by a less-complex digital precoding in the digital baseband domain. Generally, due to high power consumption and cost of mixed signal components in mmWave system, the analog beamforming is performed using a network of phase-shifters [14], [15]. To this end, one of the common solutions in the literature is to select the analog beams from a fixed predefined codebook [14].

We assume that analog beamforming vector \( w_{b,k} \) between a BS-user pair \((b,k)\) is obtained from a fixed beam steering codebook \( \mathcal{W} \) with cardinality \(|\mathcal{W}| = 32\). Furthermore, we assume that each BS \( b \) independently decides analog beamformers to maximize the local signal power, i.e., based on the criterion

\[
\text{maximize} \quad w_{b,k} = |b_{b,k}^H v_m|^2
\]  

(20)

Case-1: For example, let \( N_{RF} = K \) be the number of RF circuits at each BS \( b \), then, from (20) we obtain...
W_b = [w_{b,1}, w_{b,2}, \ldots, w_{b,K}] \in \mathbb{C}^{N_b \times K}. After fixing the analog beams, each BS b estimates the effective channel, i.e., \( \tilde{H}_b = H_b^H W_b \) and computes the digital precoder, as explained in Section [IV].

Case-2: For example, if we consider \( N_{RF} = 1 \) and \( K = 4 \) then each BS b will have at most one active analog beam in a given direction. Therefore, aligning such directional beam towards a specific user will degrade the achievable SNR for all other active users. However, to efficiently utilize the JT-CoMP gain, one needs to provide a comparable SNR to all the users. To do that, we first obtain a compromise transmit beam by appropriate phase-shifts and amplitude scaling to each antenna element, i.e., by superimposing best beam of each user \( k \), i.e., \( \tilde{w}_b = \sum_k w_{b,k} \forall b \) [27]. It should be noted, in general, \( \tilde{w}_b \) may not satisfy the uni-modulus constraints [15] on beamforming coefficients (we refer the reader to [27] for the details). Optimization of the compromise transmit beam with uni-modulus constraints is left for the future work. After fixing the compromise transmit beam, each BS estimates the effective channel, i.e., \( \tilde{H}_b = H_b^H \tilde{W}_b \) and obtain the digital precoders, as explained in Section [IV].

Fig. 7 shows the sum-rate performance assuming perfect-CSI and no blockage. It can be seen that achievable sum-rate with the two-stage hybrid beamforming architecture is in general lower than full-digital beamforming. This is mainly due to dimensionality reduction of the digital precoder brought by fixed analog beamformers in the two-stage hybrid architecture. When \( N_{RF} = K \), each BS implements user-specific analog beam selection and achieves comparable performance to the full-digital beamforming. However, when \( N_{RF} = 1 \), each BS implements a compromise transmit beam which is aligned to all \( K \) users, thus significantly reducing the achievable analog beamforming gain because of relatively wide beams. In addition, the overall system is degree-of-freedom (DoF) limited, i.e., \( L < K \), which leads to significant decrease in achievable sum-rate performance in high-SNR conditions.

In the hybrid architecture, analog beamformers are obtained from a predefined beamforming codebook. Thus, the computational complexity of the proposed methods mostly stem from the computation of digital precoder. It can be seen from [16] that computations are dominated by matrix multiplications and inversion operations in (16a). Hence, complexity depends on the dimensions of matrix and it is of \( O(N_b^2) \) for inverse operation [37]. Therefore, hybrid analog-digital beamforming architecture provides a dimensionality reductions for digital precoder from \( N_t \) to \( N_{RF} \). Thus significantly reducing the computational complexity of matrix inverse operation. As an example, for \( N_t = 16 \) and \( N_{RF} = \{K, 1\} \), the hybrid architecture may provides a complexity reduction for the matrix inversion by \{98.44\%, 99.98\%\}, respectively.

VI. CONCLUSION

In this paper, we studied robust and reliable downlink transmission in mmWave wireless access by exploiting multi-antenna spatial diversity via coordinated multi-point connectivity. We proposed a novel and computationally efficient iterative algorithm based on SCA framework and parallelization of the corresponding KKT solutions, while accounting for the uncertainties of mmWave channel in terms of link blockages. We devised a low complexity scheme, which is tractable for practical implementations, based on the best response and subgradient methods, wherein, each BS-specific variables are optimized in parallel. This provided significant reduction in the computational complexity with respect to joint optimization over the BS-specific precoders. Our proposed methods are scalable to any arbitrary multi-point configuration and dense deployment. Simulation results manifested the robustness of proposed beamformer design in presence of random link blockages. The outage performance and achievable sum-rate with the proposed method significantly outperform the baseline scenarios and results in a more stable connectivity for highly reliable communication.
Considering the Lagrangian in (17), the KKT optimality conditions for dual variables \(a_{k,c}\) and \(z_b\) are obtained by differentiating with respect to associated primal optimization variables \(\gamma_k\) and \(\Gamma_k\). After some algebraic manipulations, the stationary conditions are given as

\[
\nabla \gamma_k : \sum_{c \in C} a_{k,c} \sigma^2 + \sum_{i \in K} |h_k^H f_{i}^H|^2 = \frac{\delta_k}{1 + \gamma_k} \quad (21a)
\]

\[
\nabla \Gamma_k : f_{i}^H = \left( \sum_{b \in B} z_b E_b + \sum_{c \in C} \sum_{u \in K \backslash k} a_{u,c} h_u^H h_{b,u}^H \right)^{-1} + \sum_{c \in C} a_{k,c} f_{k}^H h_{k,1}^H h_{b,k}^H \gamma_k^2 + \sum_{c \in C} a_{k,c} \frac{f_{k}^H h_{k,1}^H h_{b,k}^H}{1 + \gamma_k^2} \quad (21b)
\]

where \(E_b \triangleq \text{diag} \{0, \ldots, \|N_b \cap (B \setminus b)\|, \ldots, 0\}\) is a block diagonal matrix, with all entries are zeros except \(\|N_b \cap (B \setminus b)\|\) for BS \(b\). From (21b), it can be noticed that the computational complexity scales exponentially with the length of joint beamformers \((B \setminus N_b)\), mainly due to matrix inversion. Furthermore, all coupled and interdependent dual variables \(z_b\) \(\forall b\) are calculated simultaneously, which makes computationally challenging to obtain their exact values in the closed-form expressions. Thus, particularly, for the dense deployments with large \(N_b\) and \(B\), the complexity of iterative algorithm may become intractable in practice. Alternatively, here, instead, we implement a parallel optimization framework using the best response approach, which efficiently parallelize the beamformer updates across the distributed BS with significantly reduced complexity as

\[
\nabla b_{k}: f_{b,k}^H = \left( \sum_{c \in C} \sum_{u \in K \backslash k} a_{u,c} h_u^H h_{b,u}^H \right)^{-1} + \sum_{c \in C} a_{k,c} f_{k}^H h_{k,1}^H h_{b,k}^H \gamma_k^2 + \sum_{c \in C} a_{k,c} \frac{f_{k}^H h_{k,1}^H h_{b,k}^H}{1 + \gamma_k^2} \quad (21c)
\]

In addition to (21) and the feasibility constraints, the KKT conditions also include the complementary slackness conditions as

\[
\inf_{k, c} \{ f_k(B_k) - f_{k}^i(c, \Gamma_k, \gamma_k; f_{i}^i) \} = 0 \quad \forall (k, c),
\]

\[
\sum_{k \in K} \|f_{k,b}^H \|^2 - P_b = 0 \quad \forall b.
\]

Let us assume the user-specific priority weights \(\delta_k > 0 \quad \forall k\) and we know \(\gamma_k \geq 0 \quad \forall k\), then from (21a), we can observe

\[
\sum_{c \in C} a_{k,c} \sigma^2 + \sum_{i \in K} |h_k^H f_{i}^H|^2 = \frac{\delta_k}{1 + \gamma_k} > 0 \quad \forall k.
\]

In other words, at least one of dual-variables \(a_{k,c}\) \(\forall c\) for each user \(k\) is strictly positive and LHS of (23) is zero if and only if \(\delta_k = 0 \quad \forall k\). For simplicity, (23) can be written as

\[
\gamma_k = \delta_k \left\{ \sum_{c \in C} a_{k,c} \sigma^2 + \sum_{i \in K} |h_k^H f_{i}^H|^2 \right\} - 1 \quad \forall k.
\]

The dual-variables \(a_{k,c}\) \(\forall c\) are highly coupled and interdependent due to the common SINR constraint, as also seen from (21a) and (22a). Therefore, we cannot calculate the exact values of these variables in closed-form expressions. However, all the coupled non-negative Lagrangian multipliers \(a_{k,c}\) \(\forall (k, c)\) can be iteratively solved using the subgradient method, such as based on constrained ellipsoid method. For iteration \(i\), the update for the dual variable \(a_{k,c}\) with a small positive step size \(\beta\) can be formulated as

\[
a_{k,c}^{(i)} = \left( a_{k,c}^{(i-1)} + \beta \left[ \delta_k \left( \frac{1}{1 + \gamma_k^2} \right) \right] \right) \quad \forall (k, c).
\]

The dual-variables \(a^{(0)} = \left( a_{1,1}^{(0)}, a_{2,2}^{(0)}, \ldots, a_{K,C}^{(0)} \right)^T\) are initialized with positive small values. From (21c), we obtain the transmit precoder as in (16a). Finally, the dual variables \(z_b\) \(\forall b\) are chosen to satisfy the total power constraints (15c), using the bisection search method.

**REFERENCES**

[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!” IEEE Access, vol. 1, pp. 335–349, 2013.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5G be?” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] 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, pp. 1–7.

[4] A. Tölli, H. Pennanen, and P. Kontulainen, “On the value of coherent and coordinated multi-cell transmission,” in Proc. IEEE Int. Conf. Commun. Workshops, Jun. 2009, pp. 1–5.

[5] A. Tölli, M. Codreanu, and M. Juntti, “Cooperative MIMO-OFDM cellular system with soft handover between distributed base station antennas,” IEEE Trans. Wireless Commun., vol. 7, no. 4, pp. 1428–1440, Apr. 2008.

[6] R. Irmer, H. Drose, P. Marsch, M. Grieber, G. Fettweis, S. Brueck, H. Mayer, L. Thiele, and V. Jungnickel, “Coordinated multipoint: Concepts, performance, and field trial results,” IEEE Commun. Mag., vol. 49, no. 2, pp. 102–111, Feb. 2011.

[7] G. Nigam, P. Minero, and M. Haenggi, “Coordinated multipoint joint transmission in heterogeneous networks,” IEEE Trans. Commun., vol. 62, no. 11, pp. 4134–4146, Nov. 2014.

[8] G. R. MacCartney, T. S. Rappaport, and A. Ghosh, “Base station diversity propagation measurements at 73 GHz millimeter-wave for 5G coordinated multipoint (CoMP) analysis,” in Proc. IEEE Global Commun. Conf. Workshops, Dec. 2017, pp. 1–7.

[9] 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.

[10] D. Maamari, N. Devroye, and D. Tsimihety, “Coverage in mmWave cellular networks with base station co-operation,” IEEE Trans. Wireless Commun., vol. 15, no. 4, pp. 2981–2994, Apr. 2016.

[11] C. Skouroumounis, C. Psomas, and I. Krikidis, “Low-complexity base station selection scheme in mmWave cellular networks,” IEEE Trans. Commun., vol. 65, no. 9, pp. 4049–4064, Sep. 2017.

[12] C. Lee and W. Chung, “Hybrid RF-baseband precoding for cooperative multilayer massive MIMO systems with limited RF chains,” IEEE Trans. Commun., vol. 65, no. 4, pp. 1575–1589, Apr. 2017.
[13] 3GPP TSG RAN Meeting #82, TS 37.340, “NR: Multi-connectivity; Overall description (Release 15),” Rel. 15, Jun 2018.

[14] A. Alkhateeb, G. Leus, and R. W. Heath, “Limited feedback hybrid precoding for multi-user millimeter wave systems,” IEEE Trans. Wireless Commun., vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[15] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, “Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems,” IEEE J. Sel. Areas Commun., vol. 10, no. 3, pp. 485–500, Apr. 2016.

[16] N. A. Johansson, Y. E. Wang, E. Eriksson, and M. Hessler, “Radio access for ultra-reliable and low-latency 5G communications,” in Proc. IEEE Int. Conf. Commun. Workshop, Jun. 2015, pp. 1184–1189.

[17] H. Shariatmadari, Z. Li, M. A. Uusitalo, S. Iraji, and R. Jntti, “Link adaptation design for ultra-reliable communications,” in Proc. IEEE Int. Conf. Commun., May 2016, pp. 1–5.

[18] J. Kaleva, A. Tölli, and M. Junnti, “Decentralized sum rate maximization with QoS constraints for interfering broadcast channel via successive convex approximation,” IEEE Trans. Signal Process., vol. 64, no. 11, pp. 2788–2802, Jun. 2016.

[19] G. Venkatraman, A. Tölli, M. Junnti, 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.

[20] O. Tervo, M. Junnti, and L. Tran, “Energy-efficient beam coordination strategies with rate-dependent processing power,” IEEE Trans. Signal Process., vol. 65, no. 22, pp. 6097–6112, Nov. 2017.

[21] J. Kaleva, R. Berry, and M. Honig, “Decentralized joint precoding with pilot-aided beamformer estimation,” IEEE Trans. Signal Process., vol. 66, no. 9, pp. 2330–2341, May 2018.

[22] D. Kumar, J. Kaleva, and A. Tölli, “Multi-point connectivity for reliable mmWave,” in Proc. European Sign. Process. Conf., Sep. 2019.

[23] D. Kumar, J. Kaleva, and A. Tölli, “Rate and reliability trade-off for reliable mmWave communication via multi-point connectivity,” in Proc. IEEE Global Commun. Conf., Dec. 2019.

[24] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, “Cloud RAN for mobile networks—a technology overview,” IEEE Communications Surveys Tutorials, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[25] 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.

[26] J. Zhang, M. Mezavilla, R. Ford, S. Rangan, S. Panwar, E. Mellios, D. Kong, A. Nix, and M. Zorzi, “Transport layer performance in 5G mmWave cellular,” in Proc. IEEE Int. Conf. on Comp. Commun. Workshop, Apr. 2016, pp. 730–735.

[27] A. Beck, A. Ben-Tal, and L. Tetruashvili, “A sequential parametric convex approximation method with applications to nonconvex truss topology design problems,” Journal of Global Optimization, vol. 47, no. 1, pp. 29–51, 2010.

[28] J. Wang, L. Zhou, K. Yang, X. Wang, and Y. Liu, “Multicast precoding for multigateway multibeam satellite systems with feeder link interference,” IEEE Trans. Wireless Commun., vol. 18, no. 3, pp. 1637–1650, Mar. 2019.

[29] J. Papandriopoulos and J. S. Evans, “SCALE: A low-complexity distributed protocol for spectrum balancing in multiser DSL networks,” IEEE Trans. Inf. Theory, vol. 55, no. 8, pp. 3711–3724, Aug. 2009.

[30] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, “Joint subchannel assignment and power allocation for OFDMA femtocell networks,” IEEE Trans. Commun., vol. 13, no. 1, pp. 342–355, Jan. 2014.

[31] T. Wang and L. Vandendorpe, “On the SCALE algorithm for multiser multicarrier power spectrum management,” IEEE Trans. Signal Process., vol. 60, no. 9, pp. 4992–4998, Sep. 2012.

[32] S. Boyd and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” http://cvxr.com/cvx, Mar. 2014.

[33] J. Kaleva, A. Tölli, M. Honig, A. Tölli, and M. Junnti, “Decentralized sum MSE minimization for coordinated multi-point transmission,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, May 2014, pp. 469–473.

[34] G. Scutari, F. Facchinei, P. Song, D. Palomar, and J. Pang, “Decomposition by partial linearization: Parallel optimization of multi-agent systems,” IEEE Trans. Signal Process., vol. 62, no. 3, pp. 641–656, Feb. 2014.

[35] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, “Applications of second-order cone programming,” Linear Algebra and Applications, vol. 284, pp. 193–228, Nov. 1998.