Blockage-Robust Hybrid Beamforming Enabling High Sum Rate for Millimeter-Wave OFDM Systems

Sota Uchimura, *Graduate Student Member, IEEE*, Giuseppe Thadeu Freitas de Abreu, *Senior Member, IEEE*, and Koji Ishibashi, *Senior Member, IEEE*

***Abstract*—We propose a scheme for the concomitant design of hybrid beamforming and per-carrier transmit power allocation to mitigate the effect of random path blockages in coordinated multi-point (CoMP) systems using orthogonal frequency division multiplexing (OFDM) in millimeter-wave (mmWave) channels. In order to optimize both the beamformers and power allocation while dealing simultaneously with outage minimization and sum rate maximization (SRM) requirements, a regularized sum-of-outage minimization problem is formulated. The problem is then transformed into an empirical risk minimization (ERM) problem, solved via block stochastic learning and manifold optimization, with required learning rates derived and tuned to guarantee convergence. The method, which demands only a few radio frequency (RF) chains and relies only on knowledge of blockage probabilities, is shown via simulation results not only to outperform state-of-the-art (SotA) alternatives, but to actually achieve outage probabilities comparable to those a fully digital CoMP-SRM scheme with perfect knowledge of instantaneous blockages.**

***Index Terms*—Block stochastic learning, coordinated multi-point (CoMP), manifold optimization, millimeter-wave (mmWave) systems**

I. Introduction

HIGH-FREQUENCY bands, in particular in the spectrum referred to as mmWave bands, ranging from 24 [GHz] to 300 [GHz], play a key role in providing sufficient bandwidths to satisfy the requirements of fifth-generation mobile communication systems (5G) [1]–[3]. Similarly, in beyond 5G and sixth-generation mobile communication systems (6G), further utilization of the mmWave and sub-THz bands will be crucial to address spectrum shortage and meet even more sophisticated requirements [4]–[6].

Wireless communications at high-frequency bands suffer, however, from severe signal attenuation owing to larger free-space propagation losses [7], [8]. Fortunately, mmWave systems can be equipped with more antenna elements than microwave systems, due to the shorter wavelengths of carriers. Therefore, beamforming schemes operating with large antenna arrays that achieve high directivity can compensate for power losses, making multiple-input multiple-output (MIMO) technologies fundamental in future wireless communication systems [9]–[11].

The first contributions in this direction relied on fully digital beamforming approaches that require systems to be equipped with the same number of radio frequency (RF) chains as the number of antenna elements. In large MIMO systems, however, fully digital architecture leads to high cost and high power consumption demands, which limits the practicality of their implementation. In response to the latter, hybrid beamforming methods, which require fewer RF chains than antenna elements, have gained much attention as a practical alternative for mmWave MIMO [12]–[15]. However, the highly directive transmissions in sharp beams are prone to sudden and rapid attenuation due to blockages in propagation paths, caused by small objects such as pedestrians or vehicles [16]–[20].

In order to overcome this challenge, new blockage-robust transmission strategies have been actively discussed recently. An example of this is CoMP transmission, whereby multiple synchronized base stations (BSs) transmit data cooperatively, which was shown to maintain high data rates even in the presence of blockages [21], [22].

However, CoMP systems do not really resolve the path blockage problem, but rather avoid it by adding more diversity to the channel. The approach therefore detracts efficiency from the network, since power continues to be transmitted towards blocked paths. In contrast, blockage prediction and mitigation strategies have been proposed [23]–[25] to directly address the path blockage problem. To cite a few examples, methods based on the spatial correlation between mmWave and sub-6 [GHz] channels [23], visual information from cameras [24], and in-band signatures [25] were proposed, which predict instantaneous blockage occurrence or their probabilities.

As for mitigation approaches, although techniques based on handover management have been proposed [26], that strategy may lead to rapid throughput degradation and considerable delays due to need for re-establishing links when prediction fails, such that recent work has rather focused on the incorporation of blockage predictions into CoMP transmission methods [27]–[31]. As an example, a robust CoMP transmission method was proposed in [27], which is based on the design of hybrid beamforming with blockage probability, solved by a worst-case optimization approach, aiming at maximizing the total system data rate. Despite the elegance of the design, due to

S. Uchimura, and K. Ishibashi are with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, Tokyo 182-8285, Japan (e-mail: uchimura@awcc.uec.ac.jp, koji@ieee.org)

Giuseppe Thadeu Freitas de Abreu is with the School of Computer Science and Engineering, Constructor University, 28759 Bremen, Germany (e-mail: g.abreu@jacobs-university.de)

We would like to thank Dr. H. Iimori for helpful discussions. This work was supported by the Ministry of Internal Affairs and Communications in Japan under Grant JPJ000254. (*Corresponding author: Sota Uchimura*)

the high computational complexity, the method can mitigate blockage effects on line-of-sight (LOS) paths only, which results in frequent outages for some users and, consequently, sub-optimal quality of service (QoS).

In turn, cooperative beamforming designs to guarantee QoS under blockage occurrences on both LOS and non-line-of-sight (NLOS) paths were proposed in [28]–[31], which employ stochastic learning methods to minimize outage probabilities with respect to prescribed target rates, relying on knowledge of blockage probabilities. In particular, a fully digital outage minimization (OutMin) beamforming was first presented in [28], where a sum-of-outage minimization problem was formulated and cast into an empirical risk minimization (ERM) problem, efficiently solved via a mini-batch stochastic gradient descent (MSGD) approach. The extension of the latter to a hybrid design was then proposed in [29], and the approach was modified in [30] to also exploit reflected intelligence surfaces (RISs), by employing a block mini-batch stochastic gradient descent (BMSGD) technique to design beamforming vectors and reflection coefficients jointly.

A limitation of the aforementioned blockage-robust mitigation methods [28]–[30] is, however, that they all consider single carrier transmission over frequency flat channels, making them unsuitable to mmWave systems, which operate over much wider bandwidths than sub-6 [GHz] systems and are affected by frequency selectivity. While the effect of the frequency selectivity can be effectively mitigated by equalization over orthogonal frequency division multiplexing (OFDM) transmissions, conventional methods based on this approach exhibit high outage probabilities due to the per-carrier transmit power allocation that does not consider blockage probability and the distribution of the sum rate over subcarriers.

In order to address this limitation, in [31], a BMSGD-based scheme for the joint design of OutMin fully digital beamformers and optimal transmit power allocation was proposed for multi-carrier OFDM mmWave MIMO systems, which was shown to successfully combat both path blockage and frequency selective effects of the channel. Still, the method proposed in [31] has two drawbacks; requiring a fully digital architecture and exhibiting a decrease in total system data rate the same as single carrier approaches [29], [30].

From all the above, it is natural to consider the joint design of per-carrier power allocation, baseband beamforming, and analog beamforming to guarantee QoS requirement while maintaining a high total system data rate for mmWave OFDM systems, which is the novelty of this article compared with our previous work [31]. We therefore extend the latter approach to a flexible hybrid beamforming alternative, which is furthermore designed to minimizing both outage and loss in data rate. Simulation results confirm that the proposed scheme achieves, using only blockage probabilities and a few RF chains, outage probabilities comparable to those of a fully digital CoMP-sum rate maximization (SRM) transmission scheme under the ideal case where full knowledge of actual instantaneous blockage occurrences is available. These results also show that the proposed scheme achieves higher total system data rates than state-of-the-art (SotA) schemes while maintaining comparable outage probabilities.

The method is further optimized by the derivation and tuning of the convergence-guaranteeing learning rates, as well as with a discrete Fourier transform (DFT)-based initialization beamforming, which are also shown via simulations to be effective. The contributions of the article can be summarized as follows:

- A sum-outage-probability minimization problem is formulated, including per-user data rates aggregated over multiple subcarriers, manifold constraints, and a regularizer to increase the data rate, which ultimately enables joint hybrid robust beamforming design and power allocation, and balancing outage probability and total system data rate.
- A new BMSGD approach is developed to efficiently solve the aforementioned problem, yielding hybrid (both baseband and analog) beamformers for all subcarriers and users with optimal powers.
- The learning rates required to guarantee the convergence of the method are derived and tuned to obtain the lowest empirical risks, both for outage probabilities and rate losses.
- A simple initialization beamforming using the DFT matrix is introduced, which is also shown to be effective in improving the overall performance of the scheme.

Notation: The following notation is used throughout the article. Matrices and vectors are denoted by upper- and lower-case bold letters, as in $\mathbf{X}$ and $\mathbf{x}$, respectively. The j-th column of a matrix $\mathbf{X}$ is denoted by $[\mathbf{X}]_j$. The sets of integers, real numbers, and complex numbers are represented by $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$, respectively. The operators $(\cdot)^{\mathrm{T}}$, $(\cdot)^*$, $(\cdot)^{\mathrm{H}}$, and $\mathrm{Tr}(\mathbf{X})$ respectively denote the transpose, conjugate, complex conjugate transpose, and trace of the argument. A diagonal matrix obtained from a vector $\mathbf{x}$, and a block diagonal matrix obtained from given matrices are respectively denoted by $\mathrm{diag}(\mathbf{x})$ and $\mathrm{blkdiag}(\cdots)$. An N-dimensional vector whose elements are all 1, and the identity matrix of size N, are respectively denoted by $\mathbf{1}_N$ and $\mathbf{I}_N$. The operators $\otimes$ and $\circ$ denote the Kronecker and Hadamard products, respectively. The functions $\mathrm{vec}(\cdot)$ and $\|\cdot\|_p$, respectively, denote vectorization and the l_p norm of the argument. The real part of a complex number is denoted by $\Re(\cdot)$, and the circularly symmetric complex normal distribution with mean μ and variance σ^2 is denoted by $\mathcal{CN}\left(\mu, \sigma^2\right)$.

II. System Model and Problem Formulation

A. Communication Scenario

Consider a CoMP downlink system employing OFDM transmission, as shown in Fig. 1, where multiple BSs, each equipped with an uniform planar array (UPA) comprising of N_{t} antenna elements and a fully connected structure with N_{RF} RF chains, cooperatively serve multiple single-antenna user equipments (UEs). The BSs are synchronized and connected to a common central processing unit (CPU), which designs baseband and analog beamformers via a fronthaul. It is assumed that space division multiple access (SDMA) enables access to multiple UEs, and all subcarriers are assigned to each UE.

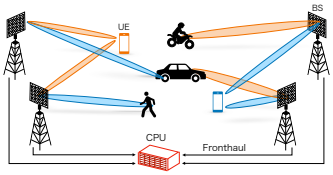


Fig. 1. A CoMP system, in which multiple BSs cooperatively serve multiple single-antenna UEs.

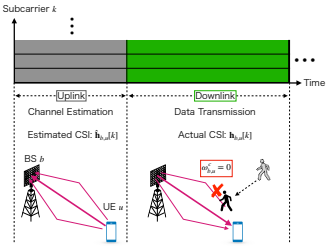


Fig. 2. A TDD system, in which channel parameters are perfectly estimated during uplink.

It is assumed that the uplink and downlink communications are separated via time-division duplexing (TDD), such that the uplink and downlink channels can be assumed reciprocal, except for path blockages[1], as illustrated in Fig. 2.

The path gains, angle of departures (AoDs), and multipath propagation time delays are assumed to be perfectly estimated from uplink signals. Following related literature, it is assumed that propagation paths may be suddenly blocked by surrounding obstacles, with probabilities ranging from 20 % to 60 % [16], [18], and that the blockage probability of each path can be perfectly estimated using any blockage prediction method [24], [25]. These assumptions imply that the available channel state information (CSI) may differ from the actual CSI during data transmissions, even under perfect channel estimation.

B. Channel Model

Let $B \in \mathbb{N}$ and $U \in \mathbb{N}$ denote the total number of BSs and UEs, while $b \in \mathcal{B} \triangleq \{1, 2, \ldots, B\}$ and $u \in \mathcal{U} \triangleq \{1, 2, \ldots, U\}$ denote the BS and UE indices, respectively. It is assumed that the mmWave channel between the b-th BS and the u-th UE contains a random number $C_{b,u}$ of clusters, with $C_{b,u}$ modeled as $C_{b,u} \sim \max(1, \text{Possion}(\lambda))$ with the intensity parameter λ [32].

Consequently, the estimated channel $\hat{\mathbf{h}}_{b,u}[d] \in \mathbb{C}^{N_\text{t}}$ between the b-th BS and the u-th UE at the d-th delay tap can be modeled as [33]

$$\hat{\mathbf{h}}_{b,u}[d] = \frac{1}{\sqrt{C_{b,u}}} \left[\sum_{c=1}^{C_{b,u}} g_{b,u}^c p(dT_\text{s} - \tau_c) \mathbf{a}_{N_\text{t}}(\theta_{b,u}^c, \phi_{b,u}^c) \right], \quad (1)$$

where $d \in \{0, 1, \ldots, D-1\}$ denotes the delay tap index, and $D \in \mathbb{N}$ denotes the total number of delay taps; the path gain of the c-th cluster is modeled as $g_{b,u}^c \sim \mathcal{CN}(0, 10^{-\text{PL}_{b,u}^c/10})$, with the associated path loss $\text{PL}_{b,u}^c \in \mathbb{R}$ calculated as $\text{PL}_{b,u}^c = \alpha + 10\beta \log_{10}(d_{b,u}) + \gamma$, in which $d_{b,u}$ denotes distance between the b-th BS and the u-th UE, and the parameters α, β, γ are listed in [32, Table I]; the function $p(dT_\text{s} - \tau_c)$ represents the equivalent pulse response at the transmitter and receiver, calculated by the sampling duration T_s and time delay at the c-th cluster τ_c. Without loss of generality, the index $c = 1$, and the time delay $\tau_1 = 0$ represent the LOS components.

[1]Although path blockage at the uplink also happens, modeling the phenomenon is unnecessary for design purposes since uplink blockages merely prevent the corresponding paths to be known/exploited at the downlink [28].

In the above, the vector $\mathbf{a}_{N_\text{t}}(\theta_{b,u}^c, \phi_{b,u}^c) \in \mathbb{C}^{N_\text{t}}$ denotes the array response as a function of the elevation $\theta_{b,u}^c$ and azimuth $\phi_{b,u}^c$ of AoD of the c-th cluster from the b-th BS toward the u-th UE and is given by

$$\begin{aligned} &\mathbf{a}_{N_\text{t}}(\theta_{b,u}^c, \phi_{b,u}^c) \\ &\quad = \mathbf{c}_{N_\text{t}^\text{h}}\left(\tfrac{1}{2}\sin(\theta_{b,u}^c)\cos(\phi_{b,u}^c)\right) \otimes \mathbf{c}_{N_\text{t}^\text{v}}\left(\tfrac{1}{2}\cos(\theta_{b,u}^c)\right), \end{aligned} \quad (2)$$

where $N_\text{t}^\text{h} \in \mathbb{N}^+$ and $N_\text{t}^\text{v} \in \mathbb{N}^+$ respectively denote the number of antenna elements in the horizontal and vertical directions, satisfying $N_\text{t} = N_\text{t}^\text{h} N_\text{t}^\text{v}$, and $\mathbf{c}_N \in \mathbb{C}^N$ is the uniform linear array (ULA) response, given by [33]

$$\mathbf{c}_N(x) \triangleq \frac{1}{\sqrt{N}} \left[1,\ \mathrm{e}^{j2\pi x}, \cdots, \mathrm{e}^{j2\pi(N-1)x}\right]^\text{T} \in \mathbb{C}^N. \quad (3)$$

Let $k \in \mathcal{K} \triangleq \{0, 1, \ldots, K-1\}$ denote the subcarrier indices, where $K \in \mathbb{N}^+$ is the total number of available subcarriers. Using the relationship between the delay and frequency domains described by the Fourier transform, the mmWave channel between the b-th BS and the u-th UE at the k-th subcarrier can be modeled as

$$\hat{\mathbf{h}}_{b,u}[k] = \sum_{d=0}^{D-1} \hat{\mathbf{h}}_{b,u}[d] \mathrm{e}^{-j\frac{2\pi kd}{K}}. \quad (4)$$

During data transmission, objects such as human bodies or vehicles block the estimated path [7], [16]–[20] with probability $p_{b,u}^c$. These blockages are modeled using the random variable $\omega_{b,u}^c \in \{0, 1\}$, which are assumed to follow a Bernoulli distribution, typical in related work on blockage robust CoMP [27]–[30], [34]. The mean of the Bernoulli distribution corresponding to the c-th cluster from the b-th BS toward the u-th UE is blockage probability $p_{b,u}^c$. During data transmission, the actual channel between the b-th BS and u-th UE at the d-th delay tap and the k-th subcarrier can be modeled as

$$\mathbf{h}_{b,u}[d] = \frac{1}{\sqrt{C_{b,u}}} \left[\sum_{c=1}^{C_{b,u}} \omega_{b,u}^c g_{b,u}^c p(dT_\text{s} - \tau_c) \mathbf{a}_{N_\text{t}}(\theta_{b,u}^c, \phi_{b,u}^c) \right], \quad (5)$$

$$\mathbf{h}_{b,u}[k] = \sum_{d=0}^{D-1} \mathbf{h}_{b,u}[d] \mathrm{e}^{-j\frac{2\pi kd}{K}}. \quad (6)$$

Because the sampling duration T_s specified in 5G new radio (NR) is significantly shorter than the continuous blockage time revealed in some studies [17], [18], it is assumed that all the delayed rays of the c-th cluster experience the same blockage. Hence, if $\omega_{b,u}^c$ is zero, the path gain of the c-th cluster from the b-th BS toward the u-th UE disappears completely for any subcarrier. The actual channel in the data transmission phase $\mathbf{h}_{b,u}[k]$ differs from the estimated channel $\hat{\mathbf{h}}_{b,u}[k]$ unless all estimated paths are available $\omega_{b,u}^c = 1, \forall b, u, c$.

C. Received Signal Model

Let $\mathbf{f}_{b,u}[k] \in \mathbb{C}^{N_\text{RF}}$ and $\mathbf{V}_b \in \mathbb{C}^{N_\text{t} \times N_\text{RF}}$ denote the baseband beamforming vector from the b-th BS toward the u-th UE at the k-th subcarrier and the analog beamforming

matrix at the b-th BS, respectively. The received signal at the u-th UE and the k-th subcarrier can be represented as

$$\begin{aligned} y_u[k] &= \sum_{b\in\mathcal{B}} \mathbf{h}_{b,u}^{\mathrm{H}}[k]\mathbf{V}_b \sum_{u\in\mathcal{U}} \mathbf{f}_{b,u}[k]x_u[k] + n_u[k] \\ &= \mathbf{h}_u^{\mathrm{H}}[k]\mathbf{V}\mathbf{f}_u[k]x_u[k] + \sum_{u'\in\mathcal{U}\setminus u} \mathbf{h}_u^{\mathrm{H}}[k]\mathbf{V}\mathbf{f}_{u'}[k]x_{u'}[k] + n_u[k], \end{aligned} \tag{7}$$

where the channel vector $\mathbf{h}_u[k]$, the baseband beamforming vector $\mathbf{f}_u[k]$, and the analog beamforming matrix $\mathbf{V}$ are respectively defined as $\mathbf{h}_u[k] \triangleq \left[\mathbf{h}_{1,u}^{\mathrm{T}}[k], \ldots, \mathbf{h}_{B,1}^{\mathrm{T}}[k]\right]^{\mathrm{T}} \in \mathbb{C}^{BN_{\mathrm{t}}}$, $\mathbf{f}_u[k] \triangleq \left[\mathbf{f}_{1,u}^{\mathrm{T}}[k], \ldots, \mathbf{f}_{B,1}^{\mathrm{T}}[k]\right]^{\mathrm{T}} \in \mathbb{C}^{BN_{\mathrm{RF}}}$, and $\mathbf{V} \triangleq \mathrm{blkdiag}\left(\mathbf{V}_1, \ldots, \mathbf{V}_B\right) \in \mathbb{C}^{BN_{\mathrm{t}}\times BN_{\mathrm{RF}}}$; $x_u[k] \in \mathbb{C}$ are the transmitted data symbols taken from a zero-mean unit-energy constellation, and $n_u[k] \sim \mathcal{CN}(0, \sigma_u^2[k])$ denotes additive white Gaussian noise (AWGN), at the u-th UE and the k-th subcarrier.

D. Proposed Problem Formulation

In view of the system and signal models of the previous subsections, consider the case when the system has only knowledge of blockage probabilities, such that the achievable data rate of each user is probabilistic quantity dependent on the actual (unknown) blockage realization. In this case, actual rate-maximization is impossible, and since we are considering a multi-band system, the most reasonable option to guarantee QoS is to minimize outage probability of the sum of data rates over all subcarriers.

It is known [31], however, that such an OutMin scheme tends to lead to a loss of total sum data rate, in comparing conventional approaches [21], [22]. We therefore propose a beamforming design considering the balance between outage minimization and sum rate-maximization, formulated as the following regularized sum-of-outage-probability minimization problem

$$\underset{\mathbf{V},\ \mathbf{f}[k],\forall k\in\mathcal{K}}{\text{minimize}} \sum_{u\in\mathcal{U}} (1-\mu)\Pr\left\{\sum_{k\in\mathcal{K}} R_u[k] < r_u\right\} + \mu\ell_u, \tag{8a}$$

$$\text{subject to} \sum_{u\in\mathcal{U}}\sum_{k\in\mathcal{K}} \|\mathbf{V}_b\mathbf{f}_{b,u}[k]\|_2^2 \le P_{\max,b}, \forall b\in\mathcal{B}, \tag{8b}$$

$$\mathbf{V}_b \in \mathcal{M}^{N_{\mathrm{t}}N_{\mathrm{RF}}}, \forall b\in\mathcal{B}, \tag{8c}$$

where $\mathbf{f}[k] \triangleq \left[\mathbf{f}_1^{\mathrm{T}}[k], \ldots, \mathbf{f}_U^{\mathrm{T}}[k]\right]^{\mathrm{T}} \in \mathbb{C}^{UBN_{\mathrm{RF}}}$, while $P_{\max,b}$, r_u, $R_u[k]$ denote the maximum transmit power at the b-th BS, the target rate for the u-th UE, and the achievable data rate at the k-th subcarrier of the u-th UE, respectively.

For clarity, the data rate in equation (8a) is given by $R_u[k] = \log_2\{1 + \Gamma(\mathbf{h}_u[k], \mathbf{f}[k], \mathbf{V})\}$ with the signal-to-noise interference ratio (SINR) given by

$$\Gamma(\mathbf{h}_u[k], \mathbf{f}[k], \mathbf{V}) = \frac{\left|\mathbf{h}_u^{\mathrm{H}}[k]\mathbf{V}\mathbf{f}_u[k]\right|^2}{\sum\limits_{u'\in\mathcal{U}\setminus u} \left|\mathbf{h}_u^{\mathrm{H}}[k]\mathbf{V}\mathbf{f}_{u'}[k]\right|^2 + \sigma_u^2[k]}. \tag{9}$$

The set $\mathcal{M}^{N_{\mathrm{t}}N_{\mathrm{RF}}}$ in constraint (8c) is the Riemann circle manifold defined as

$$\mathcal{M}^{N_{\mathrm{t}}N_{\mathrm{RF}}} \triangleq \left\{\mathbf{x}\in\mathbb{C}^{N_{\mathrm{t}}N_{\mathrm{RF}}} \,|\, |x_i|=1, (i=1,\ldots,N_{\mathrm{t}}N_{\mathrm{RF}})\right\}. \tag{10}$$

The second term ℓ_u in the objective function (8a) is a regularizer introduced to control eventual losses in total sum rate resulting from the outage-centric approach, with the scalar $\mu \in \mathbb{R}^+$ denoting, as usual, a hyper-parameter to be determined later. Details of the motivation to employ such regularizer are explained in the next section.

III. Proposed Hybrid Beamforming Design

In this section, we propose methods to solve the problem (8), starting however with a brief comparative review of the approach generally taken in related literature, with the aim of establishing a reference for the future purpose of performance assessment.

A. Conventional Approach: Per-carrier Outage Minimization

In conventional methods [29], [30], rather than attempting to solve the original optimization problem (8), the latter is first divided into K sub-problems with $\mu = 0$, which implies that the multi-band nature of OFDM is not explicitly taken into account. In addition, as described in [31], the mismatch between the estimated and actual CSI (due to post-estimation path blockages), and the lack of an analytical expression of the rate distribution, requires the relaxation of each sub-problem into

$$\underset{\mathbf{V},\mathbf{f}[k]}{\text{minimize}} \sum_{u\in\mathcal{U}} \Pr\left\{R_u[k] < \frac{r_u}{K}\right\}, \tag{11a}$$

$$\text{subject to} \sum_{u\in\mathcal{U}} \|\mathbf{V}_b\mathbf{f}_{b,u}[k]\|_2^2 \le \frac{P_{\max,b}}{K}, \forall b\in\mathcal{B}, \tag{11b}$$

$$\mathbf{V}_b \in \mathcal{M}^{N_{\mathrm{t}}N_{\mathrm{RF}}}, \forall b\in\mathcal{B}. \tag{11c}$$

Notice that the solution of this problem is bound to be sub-optimal in terms of outage probability itself, since the latter objective is not minimized over the ensemble of subcarriers, leading to the allocation of powers that are oblivious to sum-rate distributions, including effects of both frequency selectivity and blockages. Although a beamforming scheme that inherently integrates optimal power allocation based on numerical evaluations of sum-rate distributions over all sub-carriers was proposed in [31], the latter is based on a fully digital architecture ($i.e.$, $N_{\mathrm{RF}} = N_{\mathrm{t}}$ and $\mathbf{V}_b = \mathbf{I}_{N_{\mathrm{t}}}$) and still maintains the focus on outage minimization ($i.e.$, $\mu = 0$), exposing the method to degradation in total system data rate.

In contrast to [31], the proposed formulation in equation (8) enables the suppression in data rate loss. Furthermore, if solved directly in a hybrid fashion, the alternative also results in the optimum power allocation across subcarriers, yielding a cost-effective mechanism to reduce the sum-of-outage probability while maintaining a high total system data rate. This motivates our key contribution, introduced in the next subsection.

B. Proposed Regularization

In principle, the regularizer $\ell_u, \forall u$ can be any outage function that returns a larger value at a lower data rate. In order to evaluate sum-rate distribution properly, however, the regularizer must be stochastic so as to enable the design hybrid beamforming under a stochastic learning framework only.

We therefore consider the regularizer defined by

$$\ell_u \triangleq \Pr\Big\{ \sum_{k\in\mathcal{K}} R_u[k] < \hat{r}_u \Big\}, \tag{12}$$

where $\hat{r}_u \in \mathbb{R}^+$ denotes the predicted data rate at the u-th UE.

This regularizer implicates that the proposed approach improves data rates by minimizing rate loss probability calculated as the difference between achievable $\sum_{k\in\mathcal{K}} R_u[k]$ and predicted data rates $\hat{r}_u$ [2].

In what follows, we show an example of a predicted data rate. Given that minimization of rate loss, defined as the difference between achievable and ideal data rates, is equivalent to SRM, the predicted data rate $\hat{r}_u$ can be determined by hybrid beamformers designed with basis on estimated CSI $\hat{\mathbf{h}}_u[k]$ consisting of unblocked estimated paths only, which can be achieved by solving the following SRM problem

$$\underset{\mathbf{V},\mathbf{f}[k],\forall k\in\mathcal{K}}{\text{maximize}} \sum_{u\in\mathcal{U}}\sum_{k\in\mathcal{K}} \log_2\{1+\Gamma(\hat{\mathbf{h}}_u[k],\mathbf{f}[k],\mathbf{V})\}, \tag{13a}$$

$$\text{subject to} \sum_{u\in\mathcal{U}}\sum_{k\in\mathcal{K}} \|\mathbf{V}_b\mathbf{f}_{b,u}[k]\|_2^2 \le P_{\max,b},\ \forall b\in\mathcal{B}, \tag{13b}$$

$$\mathbf{V}_b \in \mathcal{M}^{N_{\rm t}N_{\rm RF}}, \forall b\in\mathcal{B}, \tag{13c}$$

whose solution can be obtained via the quadratic transform (QT) [35] and any hybrid beamforming designs [13], [15].

Let $\hat{\mathbf{f}}[k], \forall k$ and $\hat{\mathbf{V}}$ denote the optimal solutions for the optimization problem (13). Then, the predicted data rate $\hat{r}_u$ can be calculated as $\hat{r}_u = \sum_{k\in\mathcal{K}} \log_2\{1+\Gamma(\hat{\mathbf{h}}_u[k],\hat{\mathbf{f}}[k],\hat{\mathbf{V}})\}$, which can be taken as the ideal data rate for the u-th UE, such that we may hereafter set $r_u \ll \hat{r}_u$ [3].

C. Proposed Approach: Empirical Risk Minimization

In order to evaluate the data rate distribution numerically, we introduce the indicator function $\mathbb{1}(\cdot)$ with $z_u \in \mathbb{R}^+$, denoting the scalar representing the data rate, defined as [31]

$$\mathbb{1}\Big(z_u \mid \sum_{k\in\mathcal{K}} R_u[k]\Big) = \begin{cases} 0 \text{ if } \sum_{k\in\mathcal{K}} R_u[k] \ge z_u \\ 1 \text{ otherwise} \end{cases}, \tag{14}$$

which substituted into problem (8) yields

$$\underset{\mathbf{V},\mathbf{f}[k],\forall k\in\mathcal{K}}{\text{minimize}} \sum_{u\in\mathcal{U}} \mathbb{E}_{\omega^c_{b,u}}\Big[t_{\mathbb{1}}\Big(r_u,\hat{r}_u \mid \sum\nolimits_{k\in\mathcal{K}} R_u[k]\Big)\Big], \tag{15a}$$

$$\text{subject to} \sum_{u\in\mathcal{U}}\sum_{k\in\mathcal{K}} \|\mathbf{V}_b\mathbf{f}_{b,u}[k]\|_2^2 \le P_{\max,b},\forall b\in\mathcal{B}, \tag{15b}$$

$$\mathbf{V}_b \in \mathcal{M}^{N_{\rm t}N_{\rm RF}}, \forall b\in\mathcal{B}, \tag{15c}$$

where $\mathbb{E}_{\omega^c_{b,u}}[\cdot]$ denotes the expectation of channel realization caused by path blockages, and the function $t_{\mathbb{1}}$ is defined as

$$t_{\mathbb{1}}(r_u,\hat{r}_u \mid z_u) \triangleq (1-\mu)\mathbb{1}(r_u \mid z_u) + \mu\mathbb{1}(\hat{r}_u \mid z_u). \tag{16}$$

[2]Notice that the convex combination of the regularizer and outage probability is tighter than the weighted rate approach $r_u \to (1-\mu)r_u + \mu\hat{r}_u$ for both standard outage minimization ($\mu=0$) and rate maximization ($\mu=1$).

[3]Although the predicted data rate is determined based on the optimization problem (13) to focus on balancing outage minimization and rate maximization rather than hardware limitations in this paper, the total system data rate can be improved by any predicted data rate greater than the target rate.

Since the indicator function in (14) is not smooth, the gradient of the objective (15a) can not be directly evaluated. To circumvent this problem, we introduce the following generalized smooth-hinge surrogate function $\nu(\cdot)$, defined as [31]

$$\nu\Big(z_u \mid \sum_{k\in\mathcal{K}} R_u[k]\Big) = \begin{cases} 0 \ \text{ if } 1-\sum\limits_{k\in\mathcal{K}} \frac{R_u[k]}{z_u} < 0 \\ 1-\sum\limits_{k\in\mathcal{K}} \frac{R_u[k]}{z_u} \ \text{otherwise} \end{cases}. \tag{17}$$

Thanks to the above, the following cost function $t_\nu(\cdot)$ can be defined as

$$t_\nu(r_u,\hat{r}_u \mid z_u) \triangleq (1-\mu)\nu(r_u \mid z_u) + \mu\nu(\hat{r}_u \mid z_u), \tag{18}$$

which introduced into objective (15a) enables the original optimization problem (8) to be rewritten as the following ERM problem

$$\underset{\mathbf{V},\mathbf{f}[k],\forall k\in\mathcal{K}}{\text{minimize}} \sum_{u\in\mathcal{U}} \mathbb{E}_{\omega^c_{b,u}}\Big[t_\nu\Big(r_u,\hat{r}_u \mid \sum\nolimits_{k\in\mathcal{K}} R_u[k]\Big)\Big], \tag{19a}$$

$$\text{subject to} \sum_{u\in\mathcal{U}}\sum_{k\in\mathcal{K}} \|\mathbf{V}_b\mathbf{f}_{b,u}[k]\|_2^2 \le P_{\max,b},\forall b\in\mathcal{B}, \tag{19b}$$

$$\mathbf{V}_b \in \mathcal{M}^{N_{\rm t}N_{\rm RF}}, \forall b\in\mathcal{B}, \tag{19c}$$

that can be solved efficiently, for instance by stochastic approximation approaches [36].

D. Proposed Solver: Block Stochastic Gradient Descent

If information on the path gain coefficients $g^c_{b,u}$, the AoDs $\theta^c_{b,u}$ and $\phi^c_{b,u}$, the time delays τ_c, and the blockage probabilities $p^c_{b,u}$ are available, channels with various blockage patterns can be generated as training data for stochastic optimization. In turn, referring on the mini-batch approach [36], the ERM problem in equation (19) can be reduced to an equivalent problem by replacing the expectation value with an empirical mean calculated over the training the data $\mathbf{h}^m_u[k]$, namely

$$\underset{\mathbf{V},\mathbf{f}[k],\forall k\in\mathcal{K}}{\text{minimize}} \frac{1}{M}\sum_{m=1}^{M}\sum_{u\in\mathcal{U}} t_\nu\Big(r_u,\hat{r}_u \mid \sum\nolimits_{k\in\mathcal{K}} R^m_u[k]\Big), \tag{20a}$$

$$\text{subject to} \sum_{u\in\mathcal{U}}\sum_{k\in\mathcal{K}} \|\mathbf{V}_b\mathbf{f}_{b,u}[k]\|_2^2 \le P_{\max,b},\forall b\in\mathcal{B}, \tag{20b}$$

$$\mathbf{V}_b \in \mathcal{M}^{N_{\rm t}N_{\rm RF}}, \forall b\in\mathcal{B}, \tag{20c}$$

where $R^m_u[k] \triangleq \log_2\{1+\Gamma(\mathbf{h}^m_u[k],\mathbf{f}[k],\mathbf{V})\}$, and $M \in \mathbb{N}^+$ denotes the mini-batch size.

Although baseband beamforming vectors for all subcarriers $\mathbf{f}[k]$, $\forall k\in\mathcal{K}$ and analog beamforming matrices $\mathbf{V}_b, \forall b\in\mathcal{B}$ should be updated to minimize the empirical risk, it is difficult to calculate the gradient of the objective function for all variables. Fortunately, such nonconvex multivariate ERM problems with manifold constraints can be solved via MSGD, using either an alternate update approach [29] or BMSGD algorithms [30].

Since, however, the contributions in [29] and [30] do not incorporate optimal power allocation over the subcarriers, nor do they account for the summation inside the hinge function, we introduce in the sequel a new, purpose-built BMSGD approach to update such variables efficiently and thus solve problem (20).

1) Baseband Beamforming Design: First, the gradient of the objective function (20a) is calculated for each beamforming vector $\mathbf{f}[k']$, $k' \in \mathcal{K}$, with all the remaining beamforming vectors $\mathbf{f}[k]$, $k \in \mathcal{K}\backslash k'$ and analog beamforming matrix $\mathbf{V}$ fixed, using

$$\nabla_{k'} t_\nu \Big(r_u, \hat{r}_u \mid \sum_{k\in\mathcal{K}} R_u^m[k] \Big) = \begin{cases} 0 \;\; \text{if } \hat{r}_u \le \sum\limits_{k\in\mathcal{K}} R_u^m[k] \\ -\beta_u \frac{\nabla_{\mathbf{f}} \Gamma_u^m[k']}{1+\Gamma_u^m[k']} \;\; \text{if } r_u \le \sum\limits_{k\in\mathcal{K}} R_u^m[k] \\ -\Lambda_u \frac{\nabla_{\mathbf{f}} \Gamma_u^m[k']}{1+\Gamma_u^m[k']} \;\; \text{otherwise,} \end{cases} \tag{21}$$

where $\beta_u \triangleq \frac{1-\mu}{\hat{r}_u \log_e 2}$, $\Lambda_u \triangleq \frac{1}{\log_e 2} \cdot (\frac{\mu}{r_u} + \frac{1-\mu}{\hat{r}_u})$, $\Gamma_u^m[k] \triangleq \Gamma(\mathbf{h}_u^m[k], \mathbf{f}[k], \bar{\mathbf{V}})$, and $\nabla_{\mathbf{f}} \Gamma_u^m[k']$ denotes the gradient of the SINR for the baseband beamforming vector $\mathbf{f}[k']$, given by

$$\nabla_{\mathbf{f}} \Gamma_u^m[k'] = \frac{\mathbf{A}_u^m[k']\mathbf{f}[k']}{\mathbf{f}^{\mathrm{H}}[k']\bar{\mathbf{A}}_u^m[k']\mathbf{f}[k'] + \sigma_u^2[k']} - \frac{\mathbf{f}^{\mathrm{H}}[k']\mathbf{A}_u^m[k']\mathbf{f}[k']}{\left(\mathbf{f}^{\mathrm{H}}[k']\bar{\mathbf{A}}_u^m[k']\mathbf{f}[k'] + \sigma_u^2[k']\right)^2} \bar{\mathbf{A}}_u^m[k']\mathbf{f}[k'], \tag{22}$$

with

$$\mathbf{A}_u^m[k] \triangleq \mathrm{diag}\left(\mathbf{e}_u\right) \otimes \mathbf{V}^{\mathrm{H}} \mathbf{h}_u^m[k]\, \mathbf{h}_u^{m\mathrm{H}}[k] \mathbf{V}, \tag{23}$$

$$\bar{\mathbf{A}}_u^m[k] \triangleq \mathrm{diag}\left(\bar{\mathbf{e}}_u\right) \otimes \mathbf{V}^{\mathrm{H}} \mathbf{h}_u^m[k]\, \mathbf{h}_u^{m\mathrm{H}}[k] \mathbf{V}, \tag{24}$$

where $\mathbf{e}_u \in \{0,1\}^U$ denotes the vector of length U with only the u-th element equal to 1, and $\bar{\mathbf{e}}_u \in \{0,1\}^U$ denotes its complement, such that $\mathbf{e}_u + \bar{\mathbf{e}}_u = \mathbf{1}_U$.

With possession of the gradient in equation (21), the corresponding baseband beamforming vector $\mathbf{f}[k']$ is updated as

$$\mathbf{f}^{(i)}[k'] = \mathbf{f}^{(i-1)}[k'] - \frac{\alpha_i^f[k']}{M} \sum_{m=1}^{M} \sum_{u\in\mathcal{U}} \nabla_{k'} t_\nu \Big(r_u, \hat{r}_u \mid \sum_{k\in\mathcal{K}} R_u^m[k] \Big), \tag{25}$$

where $\alpha_i^f[k] \in \mathbb{R}^+$ and $\mathbf{f}^{(i)}[k] \in \mathbb{C}^{UBN_{\mathrm{RF}}}$ denote the learning rate, and the baseband beamforming vector at the i-th iteration and the k-th subcarrier, respectively.

On the matter of baseband beamforming, similarly to the fully digital approach of [31], after updating one vector $\mathbf{f}[k'], k' \in \mathcal{K}$, the vectors for all subcarriers $\mathbf{f}[k], \forall k \in \mathcal{K}$ are projected onto the feasible region by normalizing to satisfy power constraints. The baseband beamforming for each subcarrier is sequentially updated following these operations with the same training dataset $\mathbf{h}_u^m[k]$, with $m = \{1, \ldots, M\}, \forall k, u$, in order to combine the beamforming design with power allocation over subcarriers based on the stochastic learning of blockage effects.

2) Analog Beamforming Design: After updating all baseband beamforming vectors, the analog beamforming matrix is updated using the same training dataset. The received signal is rewritten using the relation $\mathrm{vec}(\mathbf{AXB}) = (\mathbf{B}^{\mathrm{T}} \otimes \mathbf{A})\mathrm{vec}(\mathbf{X})$ so as to enable the calculation of the gradient of the objective function (20a) for the analog beamforming, which is given by

$$y_u[k] = \left\{ (x_u[k]\mathbf{f}_u[k])^{\mathrm{T}} \otimes \mathbf{h}_u^{\mathrm{H}}[k] \right\} \mathrm{vec}(\mathbf{V}) + \sum_{u'\in\mathcal{U}\backslash u} \left\{ (x_{u'}[k]\mathbf{f}_{u'}[k])^{\mathrm{T}} \otimes \mathbf{h}_u^{\mathrm{H}}[k] \right\} \mathrm{vec}(\mathbf{V}) + n_u[k]. \tag{26}$$

The analog beamforming vector $\mathrm{vec}(\mathbf{V})$, which has a sparse structure representing CoMP transmission, can be decomposed into the non-sparse vector $\mathbf{v} \triangleq \mathrm{vec}(\tilde{\mathbf{V}}) \in \mathbb{C}^{BN_{\mathrm{t}}N_{\mathrm{RF}}}$ and the matrix $\mathbf{W} \in \{0,1\}^{B^2 N_{\mathrm{t}} N_{\mathrm{RF}} \times BN_{\mathrm{t}} N_{\mathrm{RF}}}$, where $\tilde{\mathbf{V}} \triangleq [\mathbf{V}_1, \ldots, \mathbf{V}_B] \in \mathbb{C}^{BN_{\mathrm{t}} \times BN_{\mathrm{RF}}}$ satisfying $\mathrm{vec}(\mathbf{V}) = \mathbf{W}\mathbf{v}$. The Euclidean gradient of the objective function (20a) for the analog beamforming vector $\mathbf{v}$ can be calculated as

$$\nabla_v t_\nu \Big(r_u, \hat{r}_u \mid \sum_{k\in\mathcal{K}} R_u^m[k] \Big) = \begin{cases} 0 \qquad \text{if } \hat{r}_u \le \sum\limits_{k\in\mathcal{K}} R_u^m[k] \\ -\beta_u \sum\limits_{k\in\mathcal{K}} \frac{\nabla_{\mathbf{v}} \Gamma_u^m[k]}{1+\Gamma_u^m[k]} \;\; \text{if } r_u \le \sum\limits_{k\in\mathcal{K}} R_u^m[k] \\ -\Lambda_u \sum\limits_{k\in\mathcal{K}} \frac{\nabla_{\mathbf{v}} \Gamma_u^m[k]}{1+\Gamma_u^m[k]} \;\; \text{otherwise} \end{cases} \tag{27}$$

where $\nabla_{\mathbf{v}} \Gamma_u[k]$ denotes the gradient of the SINR for the analog beamforming vector, given by

$$\nabla_{\mathbf{v}} \Gamma_u[k] = \frac{\mathbf{B}_u^m[k]\mathbf{v}}{\mathbf{v}^{\mathrm{H}}\bar{\mathbf{B}}_u^m[k]\mathbf{v} + \sigma_u^2[k]} - \frac{\mathbf{v}^{\mathrm{H}}\mathbf{B}_u^m[k]\mathbf{v} \cdot \bar{\mathbf{B}}_u^m[k]\mathbf{v}}{(\mathbf{v}^{\mathrm{H}}\bar{\mathbf{B}}_u^m[k]\mathbf{v} + \sigma_u^2[k])^2}, \tag{28}$$

with

$$\mathbf{B}_u^m[k] \triangleq \mathbf{W}^{\mathrm{H}} \left(\mathbf{f}_u^*[k]\mathbf{f}_u^{\mathrm{T}}[k] \otimes \mathbf{h}_u^m[k]\mathbf{h}_u^{m\mathrm{H}}[k] \right) \mathbf{W}, \tag{29}$$

$$\bar{\mathbf{B}}_u^m[k] \triangleq \mathbf{W}^{\mathrm{H}} \Bigg(\sum_{u'\in\mathcal{U}\backslash u} \mathbf{f}_{u'}^*[k]\mathbf{f}_{u'}^{\mathrm{T}}[k] \otimes \mathbf{h}_u^m[k]\mathbf{h}_u^{m\mathrm{H}}[k] \Bigg) \mathbf{W}. \tag{30}$$

Unlike the previous baseband beamforming process, here updates based on Euclidean gradients is insufficient, since manifold optimization [37] must be performed to satisfy the unit modular constraints (20c) defined by the Riemann circle manifold in (10). In this case, the tangent space at a given $\mathbf{x} \in \mathcal{M}^{N_{\mathrm{t}}N_{\mathrm{RF}}}$ can be defined as in [29], while the Riemann gradient can be expressed as an orthogonal projection of the Euclidean gradient

$$\begin{aligned} &\nabla_v^{\mathrm{Rie}} t_\nu \Big(r_u, \hat{r}_u \mid \sum_{k\in\mathcal{K}} R_u^m[k] \Big) \\ &= \nabla_v t_\nu \Big(r_u, \hat{r}_u \mid \sum_{k\in\mathcal{K}} R_u^m[k] \Big) \\ &\quad - \Re \Big\{ \nabla_v t_\nu \Big(r_u, \hat{r}_u \mid \sum_{k\in\mathcal{K}} R_u^m[k] \Big) \circ \mathbf{v}^* \Big\} \circ \mathbf{v}, \end{aligned} \tag{31}$$

such that the analog beamforming vector can be updated as

$$\mathbf{v}^{(i)} = \mathrm{Retr} \Bigg[\mathbf{v}^{(i-1)} - \frac{\alpha_i^v}{M} \sum_{m=1}^{M} \sum_{u\in\mathcal{U}} \nabla_v^{\mathrm{Rie}} t_\nu \Big(r_u, \hat{r}_u \mid \sum_{k\in\mathcal{K}} R_u^m[k] \Big) \Bigg], \tag{32}$$

where $\alpha_i^v \in \mathbb{R}^+$ and $\mathbf{v}^{(i)} \in \mathbb{C}^{BN_{\mathrm{t}}N_{\mathrm{RF}}}$ denote the learning rate and the analog beamforming vector at the i-th iteration, respectively, and $\mathrm{Retr}[\cdot]$ is a retraction operation to satisfy the unit modular constraints in equation (20c).

We summarize the proposed OutMin hybrid beamforming design in Algorithm 1, where $\mathbf{f}^{\mathrm{ini}}[k]$ and $\mathbf{v}^{\mathrm{ini}}$ denote the initial baseband and analog beamforming for stochastic learning, respectively.

Algorithm 1 BMSGD-based hybrid beamforming design

Input: $g_{b,u}^c$, $\phi_{b,u}^c$, $\theta_{b,u}^c$, τ_c, $T_{\rm s}$, $p_{b,u}^c$, $\mathbf{f}^{\rm ini}[k]$, $\mathbf{v}^{\rm ini}$, $\forall b,u,c,k$

Output: $\mathbf{f}^{\rm opt}[k]$, $\forall k$, $\mathbf{v}^{\rm opt}$

1: Initializing $\mathbf{v}^{(0)}=\mathbf{v}^{\rm ini}$
2: Initializing $\mathbf{f}^{(0)}[k]=\mathbf{f}^{\rm ini}[k]$, $\forall k$
3: **for** $i=1,2,\ldots,I_{\rm BMSGD}$ **do**
4: Generate training data $\mathbf{h}_u^m[k], m=1,...,M_{\rm mini}, \forall u,k$
5: Determine update order randomly $\mathcal{K}'=\{k_1,\ldots,k_K\}$
6: **for** $k'=k_1,k_2,\ldots,k_K$ **do**
7: Update $\mathbf{f}^{(i)}[k']$ following equation (25)
8: Projection of $\mathbf{f}^{(i)}[k], \forall k$ onto the feasible region
9: **end for** on k'
10: Update the vector $\mathbf{v}^{(i)}$ following equation (32)
11: Retraction $\mathbf{v}^{(i)}=\mathrm{Retr}(\mathbf{v}^{(i)})$
12: Projection of $\mathbf{f}^{(i)}[k], \forall k$ onto the feasible region
13: Go to line 14 if convergence before $i=I_{\rm BMSGD}$
14: **end for** on i
15: $\mathbf{f}^{\rm opt}[k]=\mathbf{f}^{(i)}[k]$, $\forall k$
16: $\mathbf{v}^{\rm opt}=\mathbf{v}^{(i)}$

3) Initial Beamforming Design: It is well-known [38] that the convergence behavior of stochastic learning methods can be significantly affected by the initial point used to kick-start the optimization process. It is therefore worthwhile to briefly discuss a suitable initialization alternative, especially in the case of hybrid designs where the unit modular constraint imposed onto the analog beamforming component makes the design of initializers more challenging.

In order to satisfy such constraints, we first select analog beamforming vector among the column vectors of the DFT matrix $\mathbf{D}\triangleq[\mathbf{d}_1,\cdots,\mathbf{d}_{N_{\rm t}}]\in\mathbb{C}^{N_{\rm t}\times N_{\rm t}}$, and then align them with the estimated channels via the inner products $\|\hat{\mathbf{H}}_b^{\rm H}[k]\mathbf{d}_i\|_2$ between the channel matrix $\hat{\mathbf{H}}_b[k]\triangleq\big[\hat{\mathbf{h}}_{b,1}[k],\ldots,\hat{\mathbf{h}}_{b,U}[k]\big]\in\mathbb{C}^{N_{\rm t}\times U}$ and each i-th column $\mathbf{d}_i\in\mathbb{C}^{N_{\rm t}}$ of the DFT matrix.

Denoting the indices of the columns of the DFT matrix by $\mathcal{D}$, and the indices of vectors assigned to the analog beamformer by $\mathcal{V}$, we can concisely describe the $j\in\{1,2,\ldots,N_{\rm RF}\}$-th column of the initial analog beamformer at the b-th BS as

$$[\mathbf{V}_b^{\rm ini}]_j=\underset{\mathbf{d}_i,i\in\mathcal{D}\setminus\mathcal{V}}{\rm argmax}\sum_{k\in\mathcal{K}}\|\hat{\mathbf{H}}_b^{\rm H}[k]\mathbf{d}_i\|_2. \tag{33}$$

Finally, the initial baseband beamformer is then obtained based on the maximum ratio transmission (MRT) criterion to maximize signal-to-noise ratio (SNR), which contributes to both outage minimization and rate maximization, namely

$$\mathbf{f}_u^{\rm ini}[k]=(\hat{\mathbf{h}}_u^{\rm H}[k]\mathbf{V}^{\rm ini})^*, \tag{34}$$

where $\mathbf{V}^{\rm ini}\triangleq\mathrm{blkdiag}(\mathbf{V}_1^{\rm ini},\ldots,\mathbf{V}_B^{\rm ini})$ and $\mathbf{f}_u^{\rm ini}[k]$ is normalized to satisfy the power constraints, considering equal power allocation over the subcarriers.

E. Learning Rates for Convergence Guarantee

It is well known that sufficient conditions for the convergence of stochastic learning algorithms with shrinking learning rates α_i are [39]

$$\sum_{i=1}^{\infty}\alpha_i\to\infty,\ \ \sum_{i=1}^{\infty}\alpha_i^2<\infty. \tag{35}$$

In turn, the Lipschitz criterion [38] ensures the convergence of the BMSGD algorithm by adjusting the learning rate via $\alpha_i=\rho/(\sqrt{i}\cdot L^\star)$, where $\rho\in\mathbb{R}$ and $L^\star$ denote the scaling coefficient and lower bound of the Lipschitz constant, respectively. In what follows we therefore derive a lower bound on the Lipschitz constants $L_f^\star$ and $L_v^\star$, to be respectively used in baseband and analog beamforming, so as to obtain the learning rates $\alpha_i^f[k],\forall k$ and α_i^v, that ensure the convergence of Algorithm 1.

Let $\nabla^2 f$ be the Hessian of a generic objective function, then, as discussed in [29], it can be shown using the Taylor theorem that the lower bound on the Lipschitz constant is given by

$$L^\star=\sum_{u\in\mathcal{U}}\lambda_{\max}(\nabla^2 f), \tag{36}$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue.

It follows that the largest eigenvalues of the Hessians $\nabla^2_{k'}t_{\nu_u}$ and $\nabla^2_v t_{\nu_u}$ of the objective function (20a) with respect to the baseband and the analog beamformers for the u-th UE, respectively, satisfy the following inequalities:

$$\lambda_{\max}(\nabla^2_{k'}t_{\nu_u})\le 3\Lambda_u\frac{(BN_{\rm t}N_{\rm RF})^2}{\sigma_u^4[k']}\|\hat{\mathbf{h}}_u[k']\|_2^4\sum_{b\in\mathcal{B}}P_{\max,b}, \tag{37a}$$

$$\lambda_{\max}(\nabla^2_v t_{\nu_u})\le 5\Lambda_u BN_{\rm t}N_{\rm RF}\Big(\sum_{b\in\mathcal{B}}P_{\max,b}\Big)^2\sum_{k\in\mathcal{K}}\frac{\|\hat{\mathbf{h}}_u[k]\|_2^4}{\sigma_u^4[k]}, \tag{37b}$$

such that the corresponding learning rates that ensure convergence, are given by

$$\alpha_i^f[k']=\rho\bigg(\sqrt{i}3(BN_{\rm t}N_{\rm RF})^2\sum_{b\in\mathcal{B}}P_{\max,b}\sum_{u\in\mathcal{U}}\Lambda_u\frac{\|\hat{\mathbf{h}}_u[k']\|_2^4}{\sigma_u^4[k']}\bigg)^{-1}, \tag{38a}$$

$$\alpha_i^v=\rho\bigg(\sqrt{i}5BN_{\rm t}N_{\rm RF}\Big(\sum_{b\in\mathcal{B}}P_{\max,b}\Big)^2\sum_{u\in\mathcal{U}}\Lambda_u\sum_{k\in\mathcal{K}}\frac{\|\hat{\mathbf{h}}_u[k]\|_2^4}{\sigma_u^4[k]}\bigg)^{-1}. \tag{38b}$$

Proof: See Appendix.

F. Learning Rates for Performance Improvement

Although the learning rates in (38) guarantee convergence, they may be quite small depending on the tightness of the inequalities in (37), which may either cause excessive delays or, when combined with the shrinking criterion may lead to a premature termination prior to sufficient learning, resulting in sub-optimal local solutions. To mitigate this problem, we propose in the sequel alternative (heuristic) learning rates which are later shown via simulations to lead to lower outage probabilities.

The key idea is to tune the learning rates considering that the gradient at the first iteration is steeper for larger target rates. Therefore, in order to avoid undesirable updates caused by exceedingly steep gradients, the proposed learning rates consist of the inverse function of the target rate, while subsequently maintaining the shrinking strategy [39] to satisfy

the convergence criteria (35). Mathematically, the alternative learning rates are given by

$$\alpha_i^f[k'] = \alpha_i^v = \Big(f_i(i)\cdot f_t\Big(\sum_{u\in\mathcal{U}}\frac{(1-\lambda)r_u+\lambda\hat{r}_u}{U}\Big)\Big)^{-1} \quad (39)$$

where $f_i \triangleq \sqrt[3]{(\cdot)}$ and $f_t \triangleq \sqrt{(\cdot)}$ are defined from the results of the hyperparameter tuning.

It will be shown in Section IV-A that the learning rates obtained from equation (39) lead to a convergence behavior similar to that obtained via hyperparameter optimization approaches such as Optuna [40], which is the current SotA.

G. Computational Complexity

The most expensive step of the proposed algorithm is the computation of the SINR expression for gradients and sum rates $\sum_{k\in\mathcal{K}} R_u[k], \forall u$. It is clear from equations (22) and (28) that the complexity orders of the gradient calculations for the baseband and analog beamformers are $\mathcal{O}(\max\{KUB^2N_{\mathrm{RF}}^2, KBN_{\mathrm{t}}N_{\mathrm{RF}}\})$ and $\mathcal{O}(KB^2N_{\mathrm{t}}^2N_{\mathrm{RF}}^2)$, respectively, considering the block diagonal structure of matrix $\mathbf{V}$, the sparsity of matrix $\mathbf{W}$, and all the subcarriers.

The proposed algorithm must also recalculate the sum rate over subcarriers $\sum_{k\in\mathcal{K}} R_u[k]$ for each beamformer update, resulting in K times sum rate calculation at each iteration ($i = 1,\ldots,I_{\mathrm{BMSGD}}$). In the first update of the baseband beamforming vector $\mathbf{f}^{(i)}[k_1]$ in Algorithm 1, the complexity order of the sum rate calculation is the same as that of gradients. In subsequent update phases, the SINR expression, consisting of the beamformer updated via equation (25) in the previous phase, requires the complexity order $\mathcal{O}(\max\{UB^2N_{\mathrm{RF}}^2, BN_{\mathrm{t}}N_{\mathrm{RF}}\})$. For the remaining $K-1$ subcarriers, calculations of the SINR require only scalar multiplications (*i.e.*, normalizations), whose complexity order is $\mathcal{O}((K-1)BU)$. Therefore, the complexity order of sum rate calculations is $\mathcal{O}(\max\{K(K-1)BU, 2KUB^2N_{\mathrm{RF}}^2, 2KBN_{\mathrm{t}}N_{\mathrm{RF}}\})$ in total at each iteration.

From the above, the complexity order of the proposed algorithm is $\mathcal{O}(\max\{K(K-1)BU, 2KUB^2N_{\mathrm{RF}}^2, KB^2N_{\mathrm{t}}^2N_{\mathrm{RF}}^2\})$ owing to the SINR expression, which is common for both data rates and gradient expressions[4]. Considering mmWave systems, which usually employ large or massive antenna arrays, such that the inequalities $N_{\mathrm{t}}^2 > 2U$ and $\frac{B}{U}N_{\mathrm{t}}^2N_{\mathrm{RF}}^2 - 1 > K$ typically hold, the computational complexity of the proposed hybrid beamforming design is $\mathcal{O}(KB^2N_{\mathrm{t}}^2N_{\mathrm{RF}}^2)$, which is the same order as the conventional design of [29], on a per-subcarrier basis.

IV. Performance Assessment

In this section, we evaluate the proposed OutMin hybrid beamforming scheme, contrasting its performance with those of the comparable hybrid alternating minimization (AltMin) approach of [15], as well as with SotA fully digital beamforming techniques such as the MRT, the minimum mean square error (MMSE) [21], [22], and the conventional OutMin methods of [28], [31].

[4]Notice that the complexity order to calculate the learning rates in (38) is $\mathcal{O}(UKBN_{\mathrm{t}})$ owing to norm calculations for the channels $\hat{\mathbf{h}}_u[k], \forall u,k$, and the training data generation step requires additions only. Therefore, the complexity of the proposed algorithm is determined by gradient calculations.

In order to serve as a reference lower bound on the achievable outage probability, comparisons with an ideal CoMP-SRM transmission scheme with perfect knowledge of the actual CSI $\mathbf{h}_{b,u}[k], \forall b,u,k$ and their instantaneous blockages is also offered.

In our computer simulations we consider a square cell with a width of 100 [m] and $B = 4$ BSs located at the corners, which cooperatively serve $U = 2$ single-antenna UEs randomly located within the cell. Each BS is equipped with an UPA with $N_{\mathrm{t}}^{\mathrm{h}} = 4$ horizontal and $N_{\mathrm{t}}^{\mathrm{v}} = 4$ vertical antenna elements, but only $N_{\mathrm{RF}} = 2$ RF chains in the case of hybrid methods [5].

The maximum transmit power per BS is set to $P_{\max,b} = 30$ [dBm], and a total of $K = 36$ subcarriers with $W = 240$ [kHz] subcarrier spacing operating at the 28 [GHz] band is assumed, which leads to a sampling period of $T_{\mathrm{s}} = 1/(WK) = 0.115$ [μs] [6]. Pulse shaping is performed with a root-raised cosine roll-off filter with the roll-off rate of 0.8, and the equivalent pulse response $p(dT_s - \tau_c)$ is calculated using the raised cosine roll-off filter. It is also assumed that the time delay of each cluster τ_c follows a uniform distribution in the interval $[0, DT_{\mathrm{s}}]$ as [14], where $D = K/4 = 9$.

The blockage probability $p_{b,u}^c$ of each path follows an independent equivalence uniform distribution in the interval $[0.2, 0.6]$ [16], [18], and the AWGN variance at the u-th UE and the k-th subcarriers is given by

$$10\log_{10}(\sigma_u^2[k]) = 10\log_{10}(1000\kappa T) + 10\log_{10}(W) + \mathrm{NF} \quad (40)$$

where κ denotes the Boltzmann constant, $T = 293.15$ [K] denotes the physical temperature, and $\mathrm{NF} = 5$ [dB] is the noise figure.

The MRT and MMSE beamformers are designed based on estimated CSI and equal power allocation, which is optimal under unpredictable CSI errors [41]–[43], to keep fair comparisons. Then, these beamformers are computed by normalizing to the power constraint the expressions

$$\mathbf{F}_b[k] = \hat{\mathbf{H}}^*[k], \quad (41)$$

$$\mathbf{F}_b[k] = \Big(\hat{\mathbf{H}}[k]\hat{\mathbf{H}}^{\mathrm{H}}[k] + U\frac{K\sigma_u^2[k]}{P_{\max,b}}\mathbf{I}_{BN_{\mathrm{t}}}\Big)^{-1}\hat{\mathbf{H}}[k], \quad (42)$$

where the matrices are respectively defined as $\mathbf{F}_b[k] \triangleq [\mathbf{f}_{b,1}[k],\ldots,\mathbf{f}_{b,U}[k]] \in \mathbb{C}^{BN_{\mathrm{t}}\times U}$ and $\hat{\mathbf{H}}[k] \triangleq \big[\hat{\mathbf{h}}_1[k],\ldots,\hat{\mathbf{h}}_U[k]\big] \in \mathbb{C}^{BN_{\mathrm{t}}\times U}$.

The hybrid AltMin beamformer [15] is obtained via the solution of following the Frobenius minimization problem

$$\underset{\mathbf{V},\mathbf{F}_b[k],\forall b,k}{\text{minimize}} \sum_{b\in\mathcal{B}}\sum_{k\in\mathcal{K}} \|\mathbf{F}_b^{\mathrm{opt}}[k] - \mathbf{V}_b\mathbf{F}_b[k]\|_{\mathrm{F}}^2, \quad (43a)$$

$$\text{subject to} \sum_{k\in\mathcal{K}} \|\mathbf{V}_b\mathbf{F}_b[k]\|_2^2 \le P_{\max,b}, \forall b\in\mathcal{B}, \quad (43b)$$

$$\mathbf{V}_b \in \mathcal{M}^{N_{\mathrm{t}}N_{\mathrm{RF}}}, \forall b\in\mathcal{B}, \quad (43c)$$

where $\mathbf{F}_b^{\mathrm{opt}}[k] \in \mathbb{C}^{BN_{\mathrm{t}}\times U}$ denotes a pre-calculated fully digital beamformer, which for the sake of a fair comparison is

[5]For the fully-digital methods, obviously $N_{\mathrm{RF}} = N_{\mathrm{t}} = 16$.

[6]Although we consider only 36 subcarriers to reduce the computational complexity to a level compatible with laptop computers, the sampling period is still much shorter than typical blocking times, which are on the order of milliseconds [17], [18].

here given by the proposed baseband beamformer of Section III-D1, with $N_{\mathrm{RF}} = N_{\mathrm{t}}$ and $\mathbf{V}_b = \mathbf{I}_{N_{\mathrm{t}}}, \forall b$.

In turn, the fully digital MSGD-based OutMin beamformer, is obtained from the optimization problem (11) with $N_{\mathrm{t}} = N_{\mathrm{RF}}$ and $\mathbf{V}_b = \mathbf{I}_{N_{\mathrm{t}}}, \forall b$. Notice that this beamforming achieves the upper bound on the performance of the conventional hybrid beamforming proposed in [29].

As for the fully digital BMSGD-based OutMin approach [31], the beamformers are designed with basis on the optimization problem (8) with $\mu = 0$, $N_{\mathrm{t}} = N_{\mathrm{RF}}$, and $\mathbf{V}_b = \mathbf{I}_{N_{\mathrm{t}}}, \forall b$. We remark that in all fully digital beamforming designs based on stochastic learning, the initial beamformer is obtained via the MMSE approach summarize by equation (42), and the mini-batch size is set to $M = 16$ in all stochastic learning approaches.

In the ideal CoMP-SRM scheme, fully digital beamforming is designed via the optimization problem (13) with $N_{\mathrm{t}} = N_{\mathrm{RF}}$, $\mathbf{V}_b = \mathbf{I}_{N_{\mathrm{t}}}, \forall b$, and actual CSI $\mathbf{h}_u[k]$, as described in equation (6), including realizations of blockage patterns $\omega^c_{b,u}, \forall b, u, c$. Note that CoMP transmissions in [27] can achieve the same outage probabilities and total system data rate as the ideal scheme if worst-case optimization is solved with fully digital architecture $N_{\mathrm{RF}} = N_{\mathrm{t}}$ and realizations of blockage patterns $\omega^c_{b,u}, \forall b, u, c$.

A. Convergence Behavior

Let us start by assessing the convergence of the proposed approach, capturing in particular the effects of the beamformer initialization methods described in Section III-D3, the learning rate mechanisms of Section III-E, and the hyperparameter μ. To this end, we first compare in Fig. 3 the convergence behavior of the proposed OutMin beamforming design algorithm at the target rate of $r_u = 20$ [Mbps], $\forall u$, and the specific hyperparameter $\mu = 0.1$, with the various initialization and learning rate adaptation schemes.

The horizontal and vertical axes of this figure correspond to the number of iterations (*i.e.*, training-dataset generations) and the objective function in equation (19a), respectively. The values of the objective function with learning rates that ensure the convergence in equations (38a) and (38b) are denoted by lines without markers. Conversely, convergence behaviors determined by tuned learning rates are denoted by lines with markers. Curves with dotted line show the convergence behavior with randomly initialized analog beamforming and the initial baseband beamforming following equation (34).

The results confirms that Algorithm 1 converges to local optimal points regardless of the initialization method employed under the learning rates satisfying the convergence criterion. The effectiveness of the initialization method described at Section III-D3 is also corroborated.

The gap between lines with and without markers shows the gain due to learning rate tuning. In particular, it is found that if learning rates are tuned via Optuna, the best convergence behavior is obtained, which confirms both the necessity of the hyperparameter tuning, and the effectiveness of the proposed method. For instance, the tuning results for the proposed approach with initialization via the DFT codebook are $\alpha^f_i[k] = 0.0228$, $\alpha^v_i = 0.9837$, $\forall k, i$.

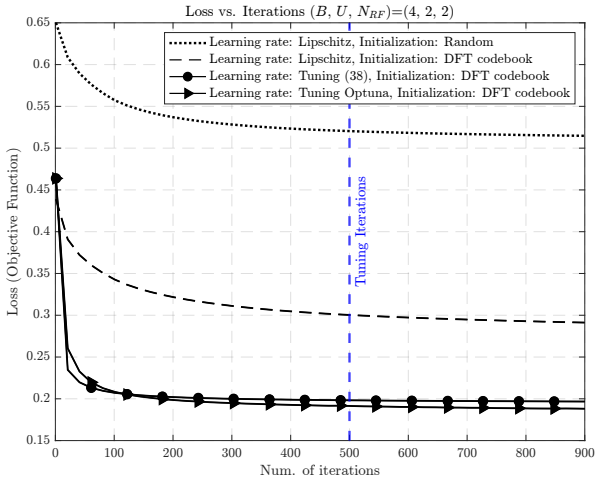
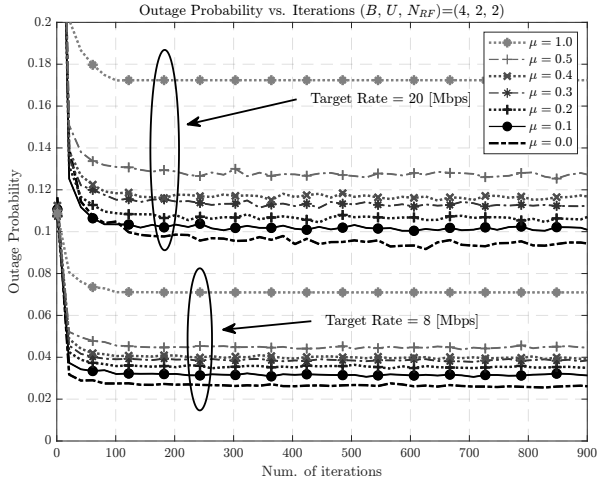
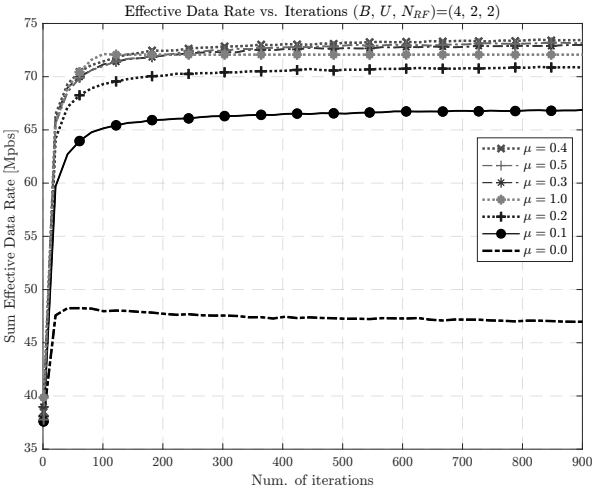


Fig. 3. Convergence behavior with different initial beamforming designs and learning rates at the target rate of 20 [Mbps].



(a) Outage probability



(b) Effective data rate at the target rate of 20 [Mbps]

Fig. 4. Comparison of convergence behavior with different hyperparameters.

Building onto the findings of Fig. 3, in what follows, the performance of the proposed approach with initialization via the DFT codebook and learning rates as in equation (39) is further assessed, considering the balance between the performance and complexity. In particular, the impact of the

hyperparameter μ on the convergence behavior of both the outage probability and the data rate achieved by Algorithm 1 are evaluated in Figs. 4(a) and 4(b), respectively.

For outage probability, Fig. 4(a) confirms that the proposed algorithm converges to local optimum points within only a few iterations, regardless of the hyperparameter values, although the level of outage actually achieved is found to be lower for lower values of μ, as can be trivially expected from the objective function in equation (8a).

In turn, the vertical axis in Fig. 4(b) corresponds to the total effective data rate for all users $\sum_{u\in\mathcal{U}} R_{u,\text{eff}}$, with the effective rate for each user defined as

$$R_{u,\text{eff}} \triangleq \begin{cases} 0 & \text{if } \sum\limits_{k\in\mathcal{K}} R_u[k] < r_u, \\ \sum\limits_{k\in\mathcal{K}} R_u[k] & \text{otherwise.} \end{cases} \tag{44}$$

It can be seen from the results of Fig. 4(b) that, complementary to the findings in Fig. 4(a), a higher effective aggregate rate is achieved with larger values of μ in the range where $0 \leq \mu \leq 0.4$, again as expected from the formulation of the problem (8), such that altogether, the results of Fig. 4 demonstrate that the proposed regularized formulation is suitable for hybrid beamforming design to improve both outage probability and data rate.

To elaborate, we emphasize that the results of the proposed method with $\mu = 0$ are equivalent to SotA OutMin approaches such as those in [28], [31], which achieve the lowest outage probabilities at the expense of decreased data rates. In contrast, the proposed regularized OutMin approach achieves low outage probabilities, comparable to those of previous methods [28], [31], while avoiding data rate losses.

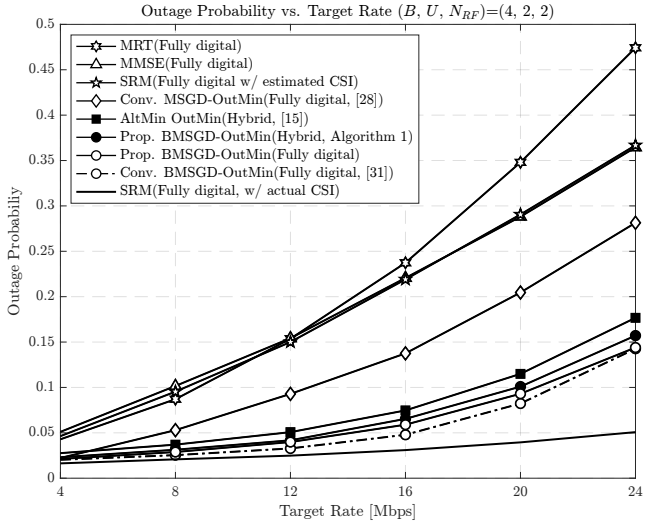
In the following subsection, the proposed approach with $\mu = 0.1$ is evaluated in more detail, considering different values of target rates.

B. Outage and Rate Performance

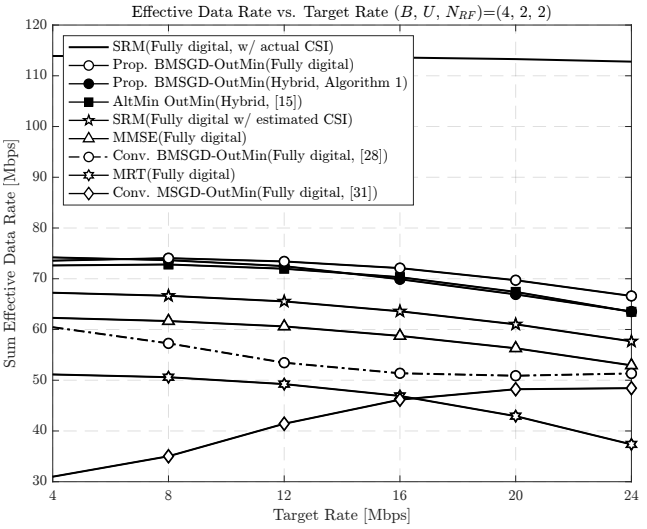
The improvement achieved with the proposed method in terms of both outage probability and effective data rate is assessed under various system conditions in Fig. 5. To that end, we first compare the outage probabilities of the proposed and SotA schemes directly in Fig. 5(a), with the performances of fully digital and hybrid beamforming methods depicted in white and black markers, respectively. The results confirm that all BMSGD-based OutMin approaches achieve lower outage probabilities than SotA alternatives based on other techniques. In particular, it is found that despite relying only on statistical path blockage information, the BMSGD-based OutMin methods come closest to the performance of the ideal CoMP-SRM with full and instantaneous knowledge of path blockages, achieving nearly the same result for low target rates.

As a highlight, the wide gap between the performance of the proposed approach and the fully-digital MSGD-OutMin [28], [29] approach demonstrates the effectiveness of the joint beamforming and power allocation approach here introduced.

It can also be clearly observed that methods such as the MRT, MMSE, and SRM with estimated CSI, which ignore blockages altogether, are found to result in poor outage performances compared to the proposed scheme and to all BMSGD-based OutMin methods in general.



(a) Outage probability as a function of target rates

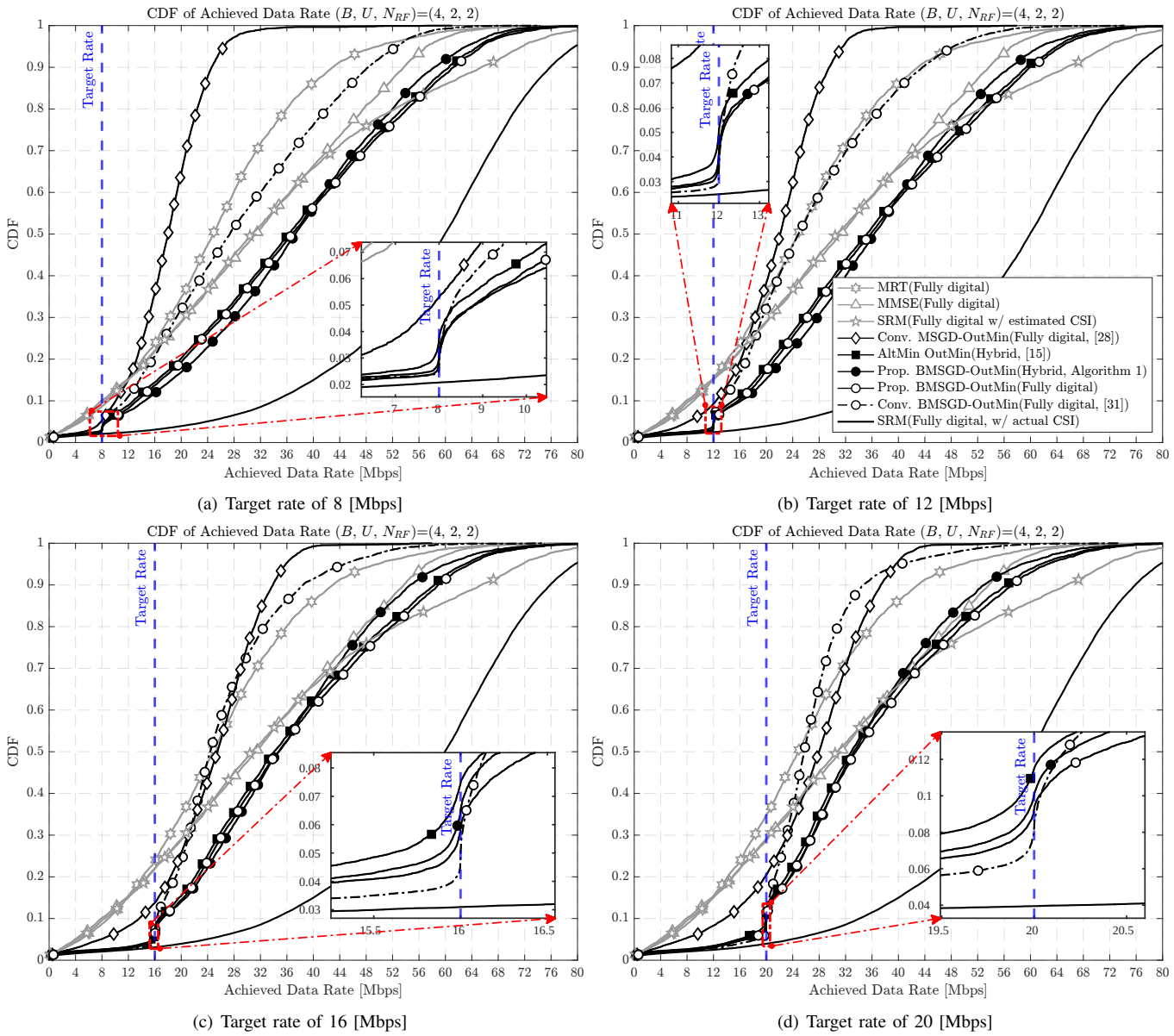


(b) Effective data rate as a function of target rates

Fig. 5. Outage probability and effective data rate comparison with SotA.

In order to further elucidate the advantage of the proposed scheme over other BMSGD-based methods [31], we compare in Fig. 5(b) the sum effective data rates achieved by the various techniques. These results indicate that, unlike the OutMin approach without regularization, the proposed approach outperforms the MMSE and SRM methods without blockage information in terms of both outage probability and total system data rates.

In addition, the narrower performance gap between the fully digital and hybrid variations of the proposed method than the that of the AltMin approach [15] illustrates the small loss in performance paid for the significant reduction in the number of RF chains achieved by the technique here contributed.



(a) Target rate of 8 [Mbps]

(b) Target rate of 12 [Mbps]

(c) Target rate of 16 [Mbps]

(d) Target rate of 20 [Mbps]

Fig. 6. cumulative distribution functions (CDFs) of achievable data rates with target rates.

C. Data Rate Statistics

Given that outage probability and sum rate are opposite performance metrics – in the sense that a high sum rate at the expense of a high outage is as undesirable as a low outage at the cost of a low sum rate – we conclude our performance assessment by evaluating the CDFs of the data rates achieved by the various methods compared, as shown in Fig. 6, for four distinct target rates.

It is found that the rate CDFs for the beamformers that are oblivious to target rates, such as the MRT, the MMSE, and the SRM with estimated CSI, are similar in all figures, in the sense that they do not exhibit a reduction in the likelihood of rates below the target. For this reason, the lines corresponding to the latter three methods are shown in grey, not to disturb the visibility of the better-performing and more recent alternative schemes. In contrast, the curves for all OutMin schemes have a "knee shape" in the vicinity of the target rate, clearly showing a reduction in that region of their respective CDFs, which of BMSGD-based schemes remarkably come close to the curve corresponding to the ideal CoMP-SRM, the only method assumed to have full and perfect knowledge of instantaneous blockages.

The CDFs in the whole region, including the non-outage region, show that the proposed hybrid approach achieve higher data rates than the fully digital SotA technique [31] with comparable outage probabilities, despite the significant savings in the number of RF chains, which is reduced from $N_{\mathrm{RF}} = N_{\mathrm{t}} = 16$ to $N_{\mathrm{RF}} = 2$. The results also confirm that the OutMin approaches achieve higher data rates than the conventional SRM approach with probabilities of about 80% and 70% at target rates of 8 [Mbps] and 20 [Mbps],

respectively, thanks to the proposed regularization[7].

V. Conclusion

We proposed new schemes to jointly optimize hybrid beamformers and the per-carrier allocation of transmit powers, with aim at mitigating the effect of random path blockages in CoMP systems using OFDM in mmWave channels. Our designs are based on a newly formulated sum-of-outage minimization problem with manifold constraints, per-user sum data rates, and a regularizer corresponding to the ideal data rate. In order to enable efficient solution, the latter problem is transformed into an ERM problem, solved via a stochastic learning method here introduced, which requires only knowledge of path blockage probabilities. To further improve the convergence behavior of the proposed technique, beamforming initialization and learning rate adaptation schemes are also contributed. Numerical results confirm that under realistic conditions, the proposed approach outperforms SotA methods in terms of convergence, outage probability, total system data rate, and requirements in number of RF chains.

A possible future work might aim to reduce the complexity of the blockage-robust hybrid beamforming design, aiming at enabling practical mmWave systems employing a larger antenna array and more UEs. In addition, since the stochastic approach may require a large amount of training data, owing to mini-batch size tuning and imperfect knowledge of channel gains and angles, the design of deterministic approach with comparable performance remains as a possible target for future work.

Acknowledgments

A preliminary version of parts of this work was presented in part at WCNC 2023 [44].

Appendix

Then, the Hessian of the objective function t_{ν_u} for the baseband and analog beamformers for the u-th UE are given by

$$\nabla^2_{k'} t_{\nu_u} = \begin{bmatrix} \frac{\partial^2 t_{\nu_u}}{\partial \mathbf{f}^*[k'] \partial \mathbf{f}^{\mathrm{T}}[k']} & \frac{\partial^2 t_{\nu_u}}{\partial \mathbf{f}^*[k'] \partial \mathbf{f}^{\mathrm{H}}[k']} \\ \frac{\partial^2 t_{\nu_u}}{\partial \mathbf{f}[k'] \partial \mathbf{f}^{\mathrm{T}}[k']} & \frac{\partial^2 t_{\nu_u}}{\partial \mathbf{f}[k'] \partial \mathbf{f}^{\mathrm{H}}[k']} \end{bmatrix} \tag{45a}$$

$$= -S_u \begin{bmatrix} \frac{\partial}{\partial \mathbf{f}^{\mathrm{T}}[k']} \left(\frac{\nabla_{\mathbf{f}^*} \Gamma_u[k']}{1+\Gamma_u[k']} \right) & \frac{\partial}{\partial \mathbf{f}^{\mathrm{H}}[k']} \left(\frac{\nabla_{\mathbf{f}^*} \Gamma_u[k']}{1+\Gamma_u[k']} \right) \\ \frac{\partial}{\partial \mathbf{f}^{\mathrm{T}}[k]} \left(\frac{\nabla_{\mathbf{f}} \Gamma_u[k']}{1+\Gamma_u[k']} \right) & \frac{\partial}{\partial \mathbf{f}^{\mathrm{H}}[k]} \left(\frac{\nabla_{\mathbf{f}} \Gamma_u[k']}{1+\Gamma_u[k']} \right) \end{bmatrix},$$

$$\nabla^2_{v} t_{\nu_u} = \begin{bmatrix} \frac{\partial^2 t_{\nu_u}}{\partial \mathbf{v}^* \partial \mathbf{v}^{\mathrm{T}}} & \frac{\partial^2 t_{\nu_u}}{\partial \mathbf{v}^* \partial \mathbf{v}^{\mathrm{H}}} \\ \frac{\partial^2 t_{\nu_u}}{\partial \mathbf{v} \partial \mathbf{v}^{\mathrm{T}}} & \frac{\partial^2 t_{\nu_u}}{\partial \mathbf{v} \partial \mathbf{v}^{\mathrm{H}}} \end{bmatrix} \tag{45b}$$

$$= -S_u \sum_{k \in \mathcal{K}} \begin{bmatrix} \frac{\partial}{\partial \mathbf{v}^{\mathrm{T}}} \left(\frac{\nabla_{v^*} \Gamma_u[k]}{1+\Gamma_u[k]} \right) & \frac{\partial}{\partial \mathbf{v}^{\mathrm{H}}} \left(\frac{\nabla_{v^*} \Gamma_u[k]}{1+\Gamma_u[k]} \right) \\ \frac{\partial}{\partial \mathbf{v}^{\mathrm{T}}} \left(\frac{\nabla_{v} \Gamma_u[k]}{1+\Gamma_u[k]} \right) & \frac{\partial}{\partial \mathbf{v}^{\mathrm{H}}} \left(\frac{\nabla_{v} \Gamma_u[k]}{1+\Gamma_u[k]} \right) \end{bmatrix},$$

where $\nabla_{\mathbf{f}^*}$ and ∇_{v^*} denote conjugate gradients for the baseband and analog beamforming, respectively, and the scalar S_u takes the value β_u or Λ_u.

[7]The performance behavior is determined by the spectrum efficiency rather than the specific target rate with the given bandwidth. Therefore, considering the balance between the simulation time and reasonable rate, we chosen these rates.

To calculate the largest eigenvalues for Hessian of these, we consider a scalar function $\Gamma \in \mathbb{R}$ in the form

$$\Gamma = \frac{\mathbf{x}^{\mathrm{H}} \mathbf{H} \mathbf{x}}{\mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}} \mathbf{x} + \sigma^2}, \tag{46}$$

where $\mathbf{x} \in \mathbb{C}^N$ and $\sigma^2 \in \mathbb{R}$ are given vector and scalar, respectively, while $\mathbf{H} \in \mathbb{C}^{N \times N}$ and $\bar{\mathbf{H}} \in \mathbb{C}^{N \times N}$ are positive semi-definite matrices, respectively.

Then, the partial derivative and conjugate partial derivative of the scalar function $\Gamma \in \mathbb{R}$ with respect to the vector $\mathbf{x}$ are respectively given by

$$\nabla_x \Gamma \triangleq \frac{\partial \Gamma}{\partial \mathbf{x}} = \left(\frac{\mathbf{H}^{\mathrm{T}} \mathbf{x}^*}{\mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}} \mathbf{x} + \sigma^2} - \frac{\mathbf{x}^{\mathrm{H}} \mathbf{H} \mathbf{x}}{(\mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}} \mathbf{x} + \sigma^2)^2} \bar{\mathbf{H}}^{\mathrm{T}} \mathbf{x}^* \right), \tag{47}$$

$$\nabla_{x^*} \Gamma \triangleq -\frac{\partial \Gamma}{\partial \mathbf{x}^*} = \left(\frac{\mathbf{H} \mathbf{x}}{\mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}} \mathbf{x} + \sigma^2} - \frac{\mathbf{x}^{\mathrm{H}} \mathbf{H} \mathbf{x}}{(\mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}} \mathbf{x} + \sigma^2)^2} \bar{\mathbf{H}} \mathbf{x} \right), \tag{48}$$

Hence, the Hessian of the function $-\log_2(1+\Gamma)$ with respect to the vector $\mathbf{x}$ is calculated as

$$H_\Gamma = -\beta \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}^{\mathrm{T}}} \left(\frac{\nabla_{x^*} \Gamma}{1+\Gamma} \right) & \frac{\partial}{\partial \mathbf{x}^{\mathrm{H}}} \left(\frac{\nabla_{x^*} \Gamma}{1+\Gamma_u} \right) \\ \frac{\partial}{\partial \mathbf{x}^{\mathrm{T}}} \left(\frac{\nabla_{x} \Gamma}{1+\Gamma} \right) & \frac{\partial}{\partial \mathbf{x}^{\mathrm{H}}} \left(\frac{\nabla_{x} \Gamma}{1+\Gamma} \right) \end{bmatrix}, \tag{49}$$

with

$$\frac{\nabla_{x^*} \Gamma}{1+\Gamma} = \frac{\mathbf{H} \mathbf{x} - \frac{\mathbf{x}^{\mathrm{H}} \mathbf{H} \mathbf{x}}{\mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}} \mathbf{x} + \sigma^2} \bar{\mathbf{H}} \mathbf{x}}{\mathbf{x}^{\mathrm{H}} (\mathbf{H} + \bar{\mathbf{H}}) \mathbf{x} + \sigma^2}, \tag{50a}$$

$$\frac{\nabla_{x} \Gamma}{1+\Gamma} = \frac{\mathbf{H}^{\mathrm{T}} \mathbf{x}^* - \frac{\mathbf{x}^{\mathrm{H}} \mathbf{H} \mathbf{x}}{\mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}} \mathbf{x} + \sigma^2} \bar{\mathbf{H}}^{\mathrm{T}} \mathbf{x}^*}{\mathbf{x}^{\mathrm{H}} (\mathbf{H} + \bar{\mathbf{H}}) \mathbf{x} + \sigma^2}, \tag{50b}$$

where $\beta \triangleq 1/\log_2 \in \mathbb{R}^+$.

Define the matrix $\mathbf{\Phi} \triangleq \mathbf{H} + \bar{\mathbf{H}}$. Then, the partial derivatives of the vector in equation (50a) with respect to the vectors $\mathbf{x}^{\mathrm{T}}$ and $\mathbf{x}^{\mathrm{H}}$ are given by

$$\frac{\partial}{\partial \mathbf{x}^{\mathrm{T}}} \left(\frac{\nabla_{x^*} \Gamma}{1+\Gamma} \right) = \frac{\mathbf{H}}{\xi} - \frac{\mathbf{\Phi} \mathbf{x} \mathbf{x}^{\mathrm{H}} \mathbf{H}}{\xi^2} - \frac{\mathbf{H} \mathbf{x} \mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}}}{\xi \cdot \bar{\xi}} + \frac{\mathbf{x}^{\mathrm{H}} \mathbf{H} \mathbf{x}}{\xi \cdot \bar{\xi}^2} \bar{\mathbf{H}} \mathbf{x} \mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}} - \frac{\mathbf{x}^{\mathrm{H}} \mathbf{H} \mathbf{x}}{\xi \cdot \bar{\xi}} \bar{\mathbf{H}} + \frac{\mathbf{\Phi} \mathbf{x} \mathbf{x}^{\mathrm{H}} \mathbf{H} \mathbf{x} \mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}}}{\xi^2 \cdot \bar{\xi}}, \tag{51a}$$

$$\frac{\partial}{\partial \mathbf{x}^{\mathrm{H}}} \left(\frac{\nabla_{x^*} \Gamma}{1+\Gamma} \right) = -\frac{\mathbf{\Phi} \mathbf{x} \mathbf{x}^{\mathrm{T}} \mathbf{H}^{\mathrm{T}}}{\xi^2} - \frac{\mathbf{H} \mathbf{x} \mathbf{x}^{\mathrm{T}} \mathbf{H}^{\mathrm{T}}}{\xi \cdot \bar{\xi}} + \frac{\mathbf{x}^{\mathrm{H}} \mathbf{H} \mathbf{x}}{\xi \cdot \bar{\xi}^2} \bar{\mathbf{H}} \mathbf{x} \mathbf{x}^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} + \frac{\mathbf{\Phi} \mathbf{x} \mathbf{x}^{\mathrm{H}} \mathbf{H} \mathbf{x} \mathbf{x}^{\mathrm{T}} \mathbf{H}^{\mathrm{T}}}{\xi^2 \cdot \bar{\xi}}, \tag{51b}$$

where $\xi \in \mathbb{R}$ and $\bar{\xi} \in \mathbb{R}$ are defined as $\xi \triangleq \mathbf{x}^{\mathrm{H}} \mathbf{\Phi} \mathbf{x} + \sigma^2$ and $\bar{\xi} \triangleq \mathbf{x}^{\mathrm{H}} \bar{\mathbf{H}} \mathbf{x} + \sigma^2$, respectively.

Similarly, the partial derivatives of the vector in equation (50b) are given by

$$\frac{\partial}{\partial \mathbf{x}^{\mathrm{T}}}\Big(\frac{\nabla_x \Gamma}{1+\Gamma}\Big) = -\frac{\mathbf{\Phi}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{H}}\mathbf{H}}{\xi^2} - \frac{\mathbf{H}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{H}}\bar{\mathbf{H}}}{\xi\cdot\bar{\xi}} + \frac{\mathbf{x}^{\mathrm{H}}\mathbf{H}\mathbf{x}}{\xi\cdot\bar{\xi}^2}\bar{\mathbf{H}}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{H}}\bar{\mathbf{H}} + \frac{\mathbf{\Phi}^{\mathrm{T}}\mathbf{v}^*\mathbf{v}^{\mathrm{H}}\mathbf{H}\mathbf{v}\mathbf{v}^{\mathrm{H}}\bar{\mathbf{H}}}{\xi^2\cdot\bar{\xi}}, \tag{52a}$$

$$\frac{\partial}{\partial \mathbf{x}^{\mathrm{H}}}\Big(\frac{\nabla_x \Gamma}{1+\Gamma}\Big) = \frac{\mathbf{H}^{\mathrm{T}}}{\xi} - \frac{(\mathbf{H}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{T}}\bar{\mathbf{H}}^{\mathrm{T}} + \mathbf{x}^{\mathrm{H}}\mathbf{H}\mathbf{x}\cdot\bar{\mathbf{H}}^{\mathrm{T}})}{\xi\cdot\bar{\xi}} + \frac{\mathbf{x}^{\mathrm{H}}\mathbf{H}\mathbf{x}}{\xi\cdot\bar{\xi}^2}\bar{\mathbf{H}}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{T}}\bar{\mathbf{H}}^{\mathrm{T}} + \frac{\mathbf{\Phi}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{H}}\mathbf{H}\mathbf{x}\mathbf{x}^{\mathrm{T}}\bar{\mathbf{H}}^{\mathrm{T}}}{\xi^2\cdot\bar{\xi}} - \frac{\mathbf{\Phi}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}}{\xi^2}. \tag{52b}$$

In turn, Hessian H_Γ in equation (49) can be written as

$$\begin{aligned} H_\Gamma = &-\underbrace{\frac{\beta}{\xi}\begin{bmatrix}\mathbf{H} & \mathbf{0}\\ \mathbf{0} & \mathbf{H}^{\mathrm{T}}\end{bmatrix}}_{\triangleq\mathbf{Q}_1} + \underbrace{\frac{\beta}{\xi^2}\begin{bmatrix}\mathbf{\Phi}\mathbf{x}\mathbf{x}^{\mathrm{H}}\mathbf{H} & \mathbf{\Phi}\mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{H}}\mathbf{H} & \mathbf{\Phi}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\end{bmatrix}}_{\triangleq\mathbf{Q}_2}\\ &+\underbrace{\frac{\beta}{\xi\cdot\bar{\xi}}\begin{bmatrix}\mathbf{H}\mathbf{x}\mathbf{x}^{\mathrm{H}}\bar{\mathbf{H}} & \mathbf{H}\mathbf{x}\mathbf{x}^{\mathrm{T}}\bar{\mathbf{H}}^{\mathrm{T}}\\ \mathbf{H}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{H}}\bar{\mathbf{H}} & \mathbf{H}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{T}}\bar{\mathbf{H}}^{\mathrm{T}}\end{bmatrix}}_{\triangleq\mathbf{Q}_3}\\ &-\underbrace{\frac{\beta\mathbf{x}^{\mathrm{H}}\mathbf{H}\mathbf{x}}{\xi\cdot\bar{\xi}^2}\begin{bmatrix}\bar{\mathbf{H}}\mathbf{x}\mathbf{x}^{\mathrm{H}}\bar{\mathbf{H}} & \bar{\mathbf{H}}\mathbf{x}\mathbf{x}^{\mathrm{T}}\bar{\mathbf{H}}^{\mathrm{T}}\\ \bar{\mathbf{H}}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{H}}\bar{\mathbf{H}} & \bar{\mathbf{H}}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{T}}\bar{\mathbf{H}}^{\mathrm{T}}\end{bmatrix}}_{\triangleq\mathbf{Q}_4}\\ &+\underbrace{\frac{\beta\mathbf{x}^{\mathrm{H}}\mathbf{H}\mathbf{x}}{\xi\cdot\bar{\xi}}\begin{bmatrix}\bar{\mathbf{H}} & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{H}}^{\mathrm{T}}\end{bmatrix}}_{\triangleq\mathbf{Q}_5}\\ &-\underbrace{\frac{\beta\mathbf{x}^{\mathrm{H}}\mathbf{H}\mathbf{x}}{\xi^2\cdot\bar{\xi}}\begin{bmatrix}\mathbf{\Phi}\mathbf{x}\mathbf{x}^{\mathrm{H}}\bar{\mathbf{H}} & \mathbf{\Phi}\mathbf{x}\mathbf{x}^{\mathrm{T}}\bar{\mathbf{H}}\\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{H}}\bar{\mathbf{H}} & \mathbf{\Phi}^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{\mathrm{T}}\bar{\mathbf{H}}^{\mathrm{T}}\end{bmatrix}}_{\triangleq\mathbf{Q}_6}, \end{aligned} \tag{53}$$

From the triangle inequality, the largest eigenvalue of the Hessian H_Γ satisfies

$$\begin{aligned}\lambda_{\max}(H_\Gamma) \le &\lambda_{\max}(-\mathbf{Q}_1) + \lambda_{\max}(\mathbf{Q}_2) + \lambda_{\max}(\mathbf{Q}_3)\\ &+ \lambda_{\max}(-\mathbf{Q}_4) + \lambda_{\max}(\mathbf{Q}_5) + \lambda_{\max}(-\mathbf{Q}_6),\end{aligned} \tag{54}$$

where the matrices $\mathbf{Q}_1$ through $\mathbf{Q}_6$ are rank 1 and positive semi-definite.

Next, notice that the largest eigenvalues of the matrices $-\mathbf{Q}_1$, $-\mathbf{Q}_4$, and $-\mathbf{Q}_6$ are zero, while the largest eigenvalues of the remaining matrices are given by

$$\lambda_{\max}(\mathbf{Q}_2) = \beta\frac{2\Re\big\{\mathrm{Tr}(\mathbf{\Phi}\mathbf{x}\mathbf{x}^{\mathrm{H}}\mathbf{H})\big\}}{\xi^2}, \tag{55a}$$

$$\lambda_{\max}(\mathbf{Q}_3) = \beta\frac{2\Re\big\{\mathrm{Tr}(\mathbf{H}\mathbf{x}\mathbf{x}^{\mathrm{H}}\bar{\mathbf{H}})\big)\big\}}{\xi\cdot\bar{\xi}}, \tag{55b}$$

$$\lambda_{\max}(\mathbf{Q}_5) = \beta\frac{\mathbf{x}^{\mathrm{H}}\mathbf{H}\mathbf{x}}{\xi^2\cdot\bar{\xi}}\lambda_{\max}(\bar{\mathbf{H}}), \tag{55c}$$

where the relation

$$\lambda_{\max}\left(\begin{bmatrix}\mathbf{X} & \mathbf{0}\\ \mathbf{0} & \mathbf{X}^{\mathrm{T}}\end{bmatrix}\right) = \lambda_{\max}(\mathbf{X}), \tag{56}$$

was used in (55c).

In baseband beamforming design, it is considered that $\mathbf{x} = \mathbf{f}[k']$, $\mathbf{H} = \mathbf{A}_u^m[k']$, $\beta = S_u$, and $\bar{\mathbf{H}} = \bar{\mathbf{A}}_u^m[k']$, such that the largest eigenvalue of the matrix $\mathbf{Q}_3$ becomes

$$\begin{aligned}\lambda_{\max}(\mathbf{Q}_3) &= S_u\frac{2\Re\big\{\mathrm{Tr}(\mathbf{A}_u^m[k']\mathbf{f}[k']\mathbf{f}^{\mathrm{H}}[k']\bar{\mathbf{A}}_u^m[k'])\big\}}{\xi^m[k']\cdot\bar{\xi}^m[k']}\\ &= S_u\frac{2\Re\big\{\mathrm{Tr}(\bar{\mathbf{A}}_u^m[k']\mathbf{A}_u^m[k']\mathbf{f}[k']\mathbf{f}^{\mathrm{H}}[k'])\big\}}{\xi^m[k']\cdot\bar{\xi}^m[k']} = 0,\end{aligned} \tag{57}$$

where $\xi^m[k'] = (\mathbf{f}^{\mathrm{H}}[k']\mathbf{\Phi}_u^m[k']\mathbf{f}[k'] + \sigma_u^2[k'])^2$, $\bar{\xi}^m = (\mathbf{f}^{\mathrm{H}}[k']\bar{\mathbf{A}}_u^m[k']\mathbf{f}[k'] + \sigma_u^2[k'])^2$, $\mathbf{\Phi}_u^m[k'] = \bar{\mathbf{A}}_u^m[k'] + \mathbf{A}_u^m[k']$, and the identity $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$ and $\bar{\mathbf{A}}_u^m[k]\mathbf{A}_u^m[k] = \mathbf{0}, \forall k$ were used.

In turn, the expression of the largest eigenvalue of the matrix $\mathbf{Q}_2$ can be simplified as follows

$$\begin{aligned}\lambda_{\max}(\mathbf{Q}_2) &= S_u\frac{2\Re\left\{\mathrm{Tr}(\mathbf{\Phi}_u^m[k']\mathbf{f}[k']\mathbf{f}^{\mathrm{H}}[k']\mathbf{A}_u^m[k'])\right\}}{(\mathbf{f}^{\mathrm{H}}[k']\mathbf{\Phi}_u^m[k']\mathbf{f}[k'] + \sigma_u^2[k'])^2}\\ &= S_u\frac{2\mathrm{Tr}(\mathbf{A}_u^m[k']\mathbf{A}_u^m[k']\mathbf{f}[k']\mathbf{f}^{\mathrm{H}}[k'])}{(\mathbf{f}^{\mathrm{H}}[k']\mathbf{\Phi}_u^m[k']\mathbf{f}[k'] + \sigma_u^2[k'])^2}\\ &\le \Lambda_u\frac{2\mathrm{Tr}(\mathbf{A}_u^m[k']\mathbf{A}_u^m[k'])}{\sigma_u^4[k']}\lambda_{\max}(\mathbf{f}[k']\mathbf{f}^{\mathrm{H}}[k'])\\ &\le \Lambda_u\frac{2\mathrm{Tr}(\mathbf{V}\mathbf{V}^{\mathrm{H}}\mathbf{V}\mathbf{V}^{\mathrm{H}})}{\sigma_u^4[k']}\|\mathbf{h}_u^m[k']\|_2^4\|\mathbf{f}\,[k']\,\|_2^2\\ &\le \Lambda_u\frac{2\left(BN_{\mathrm{t}}N_{\mathrm{RF}}\right)^2}{\sigma_u^4\,[k']}\|\hat{\mathbf{h}}_u[k']\|_2^4\sum_{b\in\mathcal{B}}P_{\max,b},\end{aligned} \tag{58}$$

where $\mathrm{Tr}\left(\mathbf{AB}\right) \le \lambda_{\max}\left(\mathbf{A}\right)\mathrm{Tr}\left(\mathbf{B}\right)$, $\beta_u < \Lambda_u$ and $\mathrm{Tr}(\mathbf{V}\mathbf{V}^{\mathrm{H}}) = BN_{\mathrm{t}}N_{\mathrm{RF}}$ were used.

The final inequality in (58) follows from two facts. First, channels $\mathbf{h}_u^m[k']$ affected by path blockages have smaller l_2-norms than their corresponding estimated channels $\hat{\mathbf{h}}_u[k']$. Secondly, under power constraints, the square value of the l_2-norm of the baseband beamformer $\|\mathbf{f}[k]\|_2^2$ is less than the total maximum transmit power under power constraints.

Similarly, the largest eigenvalue of the matrix $\mathbf{Q}_5$ becomes

$$\begin{aligned}\lambda_{\max}(\mathbf{Q}_5) &\le \Lambda_u\frac{\mathbf{f}^{\mathrm{H}}[k']\mathbf{A}_u^m[k']\mathbf{f}[k']}{\sigma_u^4[k']}\lambda_{\max}\left(\bar{\mathbf{A}}_u^m[k']\right)\\ &= \Lambda_u\frac{\mathrm{Tr}\left(\mathbf{f}_u[k']\mathbf{f}_u^{\mathrm{H}}[k']\mathbf{A}_u^m[k']\right)}{\sigma_u^4[k']}\|\mathbf{V}^{\mathrm{H}}\mathbf{h}_u^m[k']\|_2^2\\ &\le \Lambda_u\frac{\mathrm{Tr}\left(\mathbf{V}^{\mathrm{H}}\mathbf{V}\right)}{\sigma_u^4[k']}\|\mathbf{V}^{\mathrm{H}}\|_2^2\|\mathbf{h}_u^m[k]\|_2^4\sum_{b\in\mathcal{B}}P_{\max,b}\\ &\le \Lambda_u\frac{\left(BN_{\mathrm{t}}N_{\mathrm{RF}}\right)^2}{\sigma_u^4\,[k']}\|\hat{\mathbf{h}}_u[k]\|_2^4\sum_{b\in\mathcal{B}}P_{\max,b}.\end{aligned} \tag{59}$$

The largest eigenvalue of the Hessian then satisfies

$$\begin{aligned}\lambda_{\max}(\nabla^2\nu_u) &\le \lambda_{\max}(\mathbf{Q}_2) + \lambda_{\max}(\mathbf{Q}_5)\\ &= 3\Lambda_u\frac{\left(BN_{\mathrm{t}}N_{\mathrm{RF}}\right)^2}{\sigma_u^4[k']}\|\hat{\mathbf{h}}_u[k']\|_2^4\sum_{b\in\mathcal{B}}P_{\max,b},\end{aligned} \tag{60}$$

which is the same as in equation (37a).

$$\mathrm{Tr}(\mathbf{B}_u^m[k]\mathbf{B}_u^m[k]\mathbf{v}\mathbf{v}^{\mathrm{H}}) = \mathrm{Tr}\Big(\mathbf{W}^{\mathrm{H}}(\mathbf{f}_u^*[k]\mathbf{f}_u^{\mathrm{T}}[k]\otimes\mathbf{h}_u[k]\mathbf{h}_u^{\mathrm{H}}[k])\mathbf{W}\mathbf{W}^{\mathrm{H}}(\mathbf{f}_u^*[k]\mathbf{f}_u^{\mathrm{T}}[k]\otimes\mathbf{h}_u[k]\mathbf{h}_u^{\mathrm{H}}[k])\mathbf{W}\mathbf{v}\mathbf{v}^{\mathrm{H}}\Big)$$
$$\leq \mathrm{Tr}\left(\mathbf{W}\mathbf{v}\mathbf{v}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}\right)\lambda_{\max}^2(\mathbf{f}_u^*[k]\mathbf{f}_u^{\mathrm{T}}[k])\lambda_{\max}^2\left(\mathbf{h}_u\left[k\right]\mathbf{h}_u^{\mathrm{H}}[k]\right) \leq BN_{\mathrm{t}}N_{\mathrm{RF}}\Big(\sum_{b\in\mathcal{B}}P_{\max,b}\Big)^2\|\hat{\mathbf{h}}_u[k]\|_2^4, \quad (62a)$$

$$\mathrm{Tr}(\bar{\mathbf{B}}_u^m[k]\mathbf{B}_u^m[k]\mathbf{v}\mathbf{v}^{\mathrm{H}}) = \mathrm{Tr}\Big(\mathbf{W}^{\mathrm{H}}\sum_{u'\in\mathcal{U}\setminus u}(\mathbf{f}_{u'}^*[k]\mathbf{f}_{u'}^{\mathrm{T}}[k]\otimes\mathbf{h}_u[k]\mathbf{h}_u^{\mathrm{H}}[k])\mathbf{W}\mathbf{B}_u^m[k]\mathbf{v}\mathbf{v}^{\mathrm{H}}\Big)BN_{\mathrm{t}}N_{\mathrm{RF}}\Big(\sum_{b\in\mathcal{B}}P_{\max,b}\Big)^2\|\hat{\mathbf{h}}_u[k]\|_2^4, \quad (62b)$$

$$\mathrm{Tr}(\bar{\mathbf{B}}_u^m[k]) = \mathrm{Tr}\Big(\mathbf{W}^{\mathrm{H}}\sum_{u'\in\mathcal{U}\setminus u}(\mathbf{f}_{u'}^*[k]\mathbf{f}_{u'}^{\mathrm{T}}[k]\otimes\mathbf{h}_u[k]\mathbf{h}_u^{\mathrm{H}}[k])\mathbf{W}\Big) \leq \Big(\sum_{b\in\mathcal{B}}P_{\max,b}\Big)^2\|\hat{\mathbf{h}}_u[k]\|_2^4, \quad (62c)$$

In analog beamforming, it is considered that $\mathbf{x}=\mathbf{v}$, $\mathbf{H}=\mathbf{B}_u^m[k]$, and $\bar{\mathbf{H}}=\bar{\mathbf{B}}_u^m[k]$. Then, the largest eigenvalues of the matrices $\mathbf{Q}_2$, $\mathbf{Q}_3$, and $\mathbf{Q}_5$ are given by

$$\lambda_{\max}(\mathbf{Q}_2) \leq \Lambda_u\sum_{k\in\mathcal{K}}\frac{2\Re\mathrm{Tr}(\mathbf{B}_u^m[k]\mathbf{B}_u^m[k]\mathbf{v}\mathbf{v}^{\mathrm{H}})}{\sigma_u^4[k]}, \quad (61a)$$

$$\lambda_{\max}(\mathbf{Q}_3) \leq \Lambda_u\sum_{k\in\mathcal{K}}\frac{2\Re\mathrm{Tr}(\bar{\mathbf{B}}_u^m[k]\mathbf{B}_u^m[k]\mathbf{v}\mathbf{v}^{\mathrm{H}})}{\sigma_u^4[k]}, \quad (61b)$$

$$\lambda_{\max}(\mathbf{Q}_5) \leq \Lambda_u\sum_{k\in\mathcal{K}}\frac{\mathbf{v}^{\mathrm{H}}\mathbf{B}_u^m[k]\mathbf{v}}{\sigma_u^4[k]}\lambda_{\max}\left(\bar{\mathbf{B}}_u^m[k]\right). \quad (61c)$$

Using the bounds on the trace operator for the rank 1 positive semi-defined matrices given by equation (62), on the top of this page, we obtain

$$\lambda_{\max}(\mathbf{Q}_2) \leq 2\Lambda_u BN_{\mathrm{t}}N_{\mathrm{RF}}\Big(\sum_{b\in\mathcal{B}}P_{\max,b}\Big)^2\sum_{k\in\mathcal{K}}\frac{\|\hat{\mathbf{h}}_u[k]\|_2^4}{\sigma_u^4[k]}, \quad (63a)$$

$$\lambda_{\max}(\mathbf{Q}_3) \leq 2\Lambda_u BN_{\mathrm{t}}N_{\mathrm{RF}}\Big(\sum_{b\in\mathcal{B}}P_{\max,b}\Big)^2\sum_{k\in\mathcal{K}}\frac{\|\hat{\mathbf{h}}_u[k]\|_2^4}{\sigma_u^4[k]}, \quad (63b)$$

$$\lambda_{\max}(\mathbf{Q}_5) \leq \Lambda_u BN_{\mathrm{t}}N_{\mathrm{RF}}\Big(\sum_{b\in\mathcal{B}}P_{\max,b}\Big)^2\sum_{k\in\mathcal{K}}\frac{\|\hat{\mathbf{h}}_u[k]\|_2^4}{\sigma_u^4[k]}, \quad (63c)$$

such that the largest eigenvalue of the Hessian satisfies

$$\begin{aligned}&\lambda_{\max}(\nabla^2\nu_u)\\ &\leq \lambda_{\max}(\mathbf{Q}_2)+\lambda_{\max}(\mathbf{Q}_3)+\lambda_{\max}(\mathbf{Q}_5)\\ &= 5\Lambda_u BN_{\mathrm{t}}N_{\mathrm{RF}}\Big(\sum_{b\in\mathcal{B}}P_{\max,b}\Big)^2\sum_{u\in\mathcal{U}}\sum_{k\in\mathcal{K}}\frac{\|\hat{\mathbf{h}}_u[k']\|_2^4}{\sigma_u^4[k']}, \quad (64)\end{aligned}$$

which is the same as in equation (37b).

REFERENCES

[1] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, Mar. 2013.

[2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[3] T. S. Rappaport *et al.*, "Overview of millimeter wave communications for fifth-generation 5G wireless networks—with a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.

[4] Y. C. *et al.*, "Sub-THz spectrum as enabler for 6G wireless communications up to 1 Tbit/s," in *6G Wireless Summit*, Levi Lapland, Finland, Mar. 2019.

[5] W. Hong *et al.*, "The role of millimeter-wave technologies in 5G/6G wireless communications," *IEEE J. Microw.*, vol. 1, no. 1, pp. 101–122, Jan. 2021.

[6] K. Ishibashi *et al.*, "User-centric design of millimeter wave communications for beyond 5G and 6G," *IEICE Trans. Commun.*, vol. E105-B, Oct. 2022.

[7] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5070–5083, Sep. 2014.

[8] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, Jun. 2019.

[9] M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "A comparison of MIMO techniques in downlink millimeter wave cellular networks with hybrid beamforming," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1952–1967, May 2016.

[10] T. Kebede, Y. Wondie, J. Steinbrunn, H. B. Kassa, and K. T. Kornegay, "Precoding and beamforming techniques in mmWave-massive MIMO: Performance assessment," *IEEE Access*, vol. 10, pp. 16 365–16 387, Feb. 2022.

[11] Y. Ahn, J. Kim, S. Kim, K. Shim, J. Kim, S. Kim, and B. Shim, "Towards intelligent millimeter and terahertz communication for 6G: Computer vision-aided beamforming," *IEEE Wireless Communications*, pp. 1–18, Sept. 2022.

[12] C. Fang, B. Makki, J. Li, and T. Svensson, "Hybrid precoding in cooperative millimeter wave networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5373–5388, Aug. 2021.

[13] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[14] Venugopal *et al.*, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sept. 2017.

[15] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[16] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep 2015.

[17] 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. (GLOBECOM)*, Dec. 2017, pp. 1–7.

[18] V. Raghavan *et al.*, "Statistical blockage modeling and robustness of beamforming in millimeter-wave systems," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 3010–3024, Mar. 2019.

[19] S. Ju, O. Kanhere, Y. Xing, and T. S. Rappaport, "A millimeter-wave channel simulator NYUSIM with spatial consistency and human blockage," in *Proc IEEE Global Commun. Conf. (GLOBECOM)*, Feb. 2019, pp. 1–6.

[20] L. Ahumada Fierro, E. C. n. Maggi, A. Anglés Vazquez, and D. Schkolnik, "Empirical results for human-induced shadowing events for indoor 60 GHz wireless links," *IEEE Access*, vol. 8, pp. 44 522–44 533, 2020.

[21] 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, July 2019.

[22] B. Maham and T. Svensson, "Performance analysis of millimeter wave CoMP networks under blockage," in *Proc 93rd Veh. Technol. Conf. (VTC2021-Spring)*, Apr. 2021, pp. 1–5.

[23] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5504–5518, Sept. 2020.

[24] T. Nishio *et al.*, "Proactive received power prediction using machine learning and depth images for mmWave networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2413–2427, Aug. 2019.

[25] S. Wu, M. Alrabeiah, C. Chakrabarti, and A. Alkhateeb, "Blockage prediction using wireless signatures: Deep learning enables real-world demonstration," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 776–796, Mar. 2022.

[26] G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-aided 6G wireless communications: Blockage prediction and proactive handoff," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10 193–10 208, Aug. 2021.

[27] D. Kumar, J. Kaleva, and A. Tölli, "Blockage-aware reliable mmWave access via coordinated multi-point connectivity," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4238–4252, Feb. 2021.

[28] H. Iimori *et al.*, "Stochastic learning robust beamforming for millimeter-wave systems with path blockage," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1557–1561, May 2020.

[29] ——, "A stochastic gradient descent approach for hybrid mmWave beamforming with blockage and CSI-error robustness," *IEEE Access*, vol. 9, pp. 74 471–74 487, May 2021.

[30] G. Zhou *et al.*, "Stochastic learning-based robust beamforming design for RIS-aided millimeter-wave systems in the presence of random blockages," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 1057–1061, Jan. 2021.

[31] S. Uchimura, H. Iimori, G. T. F. de Abreu, and K. Ishibashi, "Outage-minimization coordinated multi-point for millimeter-wave OFDM with random blockages," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 8783–8796, Jul. 2023.

[32] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[33] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc Asilomar Conf. Sig. Syst. Comput.*, Nov. 2014, pp. 273–277.

[34] D. Kumar, S. K. Joshi, and A. Tölli, "Latency-aware highly-reliable mmWave systems via multi-point connectivity," *IEEE Access*, vol. 10, pp. 32 822–32 835, Mar. 2022.

[35] K. Shen and W. Yu, "Fractional programming for communication systems—part i: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, Mar. 2018.

[36] R. Johnson and T. Zhang, "Accelerating stochastic gradient descent using predictive variance reduction," in *Advances in Neural Information Processing Systems*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds., vol. 26. Curran Associates, Inc., 2013.

[37] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press, 2008.

[38] Y. Xu and W. Yin, "Block stochastic gradient iteration for convex and nonconvex optimization," *SIAM J. Optim.*, vol. 25, no. 3, pp. 1686–1716, 2015.

[39] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400 – 407, 1951. [Online]. Available: https://doi.org/10.1214/aoms/1177729586

[40] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc 25rd ACM SIGKDD Int. Conf. Knowl. Discovery and Data Mining*, 2019.

[41] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[42] B. A. Bash, D. Goeckel, and D. Towsley, "Asymptotic optimality of equal power allocation for linear estimation of wss random processes," *IEEE Wireless Commun. Lett.*, vol. 2, no. 3, pp. 247–250, Jun. 2013.

[43] L. Sun, T. Xu, S. Yan, J. Hu, X. Yu, and F. Shu, "On resource allocation in covert wireless communication with channel estimation," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6456–6469, Oct. 2020.

[44] S. Uchimura, G. T. Freitas de Abreu, and K. Ishibashi, "Hybrid beamforming for outage-minimization in frequency selective millimeter-wave channels," in *Proc IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2023, pp. 1–6.

Sota Uchimura (Graduate Student Member, IEEE) received the B.E. and M.E. degree in engineering from The University of Electro-Communications, Tokyo, Japan in 2021 and 2023, respectively. His current research interests include millimeter-wave communications and signal processing. He was the recipient of the IEICE RCS Active Research Award in 2022 and the IEICE SITA Young Researcher Award in 2022.

Giuseppe Thadeu Freitas de Abreu (Senior Member, IEEE) received the B.Eng. degree in electrical engineering and the specialization Latu Sensu degree in telecommunications engineering from the Universidade Federal da Bahia (UFBa), Salvador, Bahia, Brazil in 1996 and 1997, respectively, and the M.Eng. and D.Eng. degrees in physics, electrical, and computer engineering from Yokohama National University, Japan, in March 2001 and March 2004, respectively. He was a postdoctoral fellow and later an adjunct professor (docent) in statistical signal processing and communications theory at the Department of Electrical and Information Engineering, University of Oulu, Finland from 2004 to 2006 and from 2006 to 2011, respectively. Since 2011, he has been a professor of electrical engineering at Jacobs University, Bremen, Germany. From April 2015 to August 2018, he simultaneously held a full professorship at the Department of Computer and Electrical Engineering, Ritsumeikan University, Japan. His research interests include communications and signal processing, including communications theory, estimation theory, statistical modeling, wireless localization, cognitive radio, wireless security, MIMO systems, ultrawideband and millimeter wave communications, full-duplex and cognitive radio, compressive sensing, energy harvesting networks, random networks, connected vehicles networks, and many other topics. He received the Uenohara Award at Tokyo University in 2000 for his master ' s thesis. He has been a co-recipient of the Best Paper Award at several international conferences. He was awarded prestigious JSPS, Heiwa Nakajima, and NICT Fellowships in 2010, 2013, and 2015, respectively. He served as an associate editor for the IEEE Transactions on Wireless Communications from 2009 to 2014 and the IEEE Transactions on Communications from 2014 to 2017; and as an executive editor for IEEE Transactions on Wireless Communications from 2017 to 2021. He is currently serving as an editor to the IEEE Signal Processing Letters and the IEEE Communications Letters.

Koji Ishibashi (Senior Member, IEEE) received the B.E. and M.E. degrees in engineering from The University of Electro-Communications, Tokyo, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in engineering from Yokohama National University, Yokohama, Japan, in 2007. From 2007 to 2012, he was an Assistant Professor at the Department of Electrical and Electronic Engineering, Shizuoka University, Hamamatsu, Japan. Since April 2012, he has been with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, Tokyo, Japan where he is currently a Professor. From 2010 to 2012, he was a Visiting Scholar at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA. He is a senior member of IEICE and IEEE. He is a recipient of Takayanagi Research Encouragement Award in 2009 and KDDI Foundation Award in 2023. He was certified as an Exemplary Reviewer of IEEE Communications Wireless Letters in 2015 and awarded by the Telecommunication Technology Committee (TTC) for his devotion to standardization activities in 2020. He in the past served as an Associate Editor for IEICE Transactions on Communications, an Associate Editor for IEEE Journal on Selected Areas in Communications (JSAC) and a Guest Editor for IEEE Open Journal of the Communications Society. His current research interests are robust beamforming, grant-free access, energy-harvesting, compressed sensing, coding, cell-free network, and MIMO technologies.