Received April 22, 2021, accepted May 7, 2021, date of publication May 13, 2021, date of current version May 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3079508

A Stochastic Gradient Descent Approach for Hybrid mmWave Beamforming With Blockage and CSI-Error Robustness

HIROKI IIMORI[1], (Graduate Student Member, IEEE), GIUSEPPE THADEU FREITAS DE ABREU[1], (Senior Member, IEEE), OMID TAGHIZADEH[2,3], (Member, IEEE), RĂZVAN-ANDREI STOICA[3], (Member, IEEE), TAKANORI HARA[4], (Graduate Student Member, IEEE), AND KOJI ISHIBASHI[4], (Senior Member, IEEE)

[1]Focus Area Mobility, Department of Electrical and Computer Engineering, Jacobs University Bremen, 28759 Bremen, Germany
[2]Network Information Theory Group, Technische Universität Berlin, 10623 Berlin, Germany
[3]5G Wireless Research Group, Lenovo Deutschland GmbH, 70563 Stuttgart, Germany
[4]Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, Tokyo 182-8585, Japan

Corresponding author: Hiroki Iimori (h.iimori@ieee.org)

This work was supported in part by the Ministry of Internal Affairs and Communications in Japan under Grant JPJ000254.

ABSTRACT In this study, we consider the downlink beamforming problem in millimeter wave (mmWave) systems subjected to both path blockages and imperfect channel state information (CSI), and propose a new robust hybrid sum-outage minimizing design as a solution. We first formulate the problem as an empirical risk minimization (ERM) stochastic learning problem, whose solution can be obtained by the alternate iteration of a baseband digital and a radio frequency (RF) analog Riemann manifold-constrained beamforming updates through a mini-batch stochastic gradient descent (MSGD) approach, with gradient minimizing update rules given in closed-form, and learning rates optimized based on the lower-bounds of the corresponding Lipschitz constants. Unlike existing solutions to the path blockage-robust mmWave beamforming problem, wherein out-of-band side information is required or perfect CSI is assumed, our method relies only on the estimates and statistical knowledge of the channel's angles of departure (AoD) and complex gains, which are simultaneously captured in a Bernoulli-Gaussian model and used to generate the training data for the MSGD-based optimizer. Further, unlike preceding fully-digital or fully-connected hybrid contributions, the proposed scheme assumes a virtually-configured partially-connected setup; therefore, it is compatible with coordinated multipoint (CoMP) architectures, which are known to be crucial in terms of exploiting the full potential of mmWave systems. Simulation results confirm the effectiveness of our MSGD-based robust hybrid CoMP mmWave beamformer in mitigating the effects of path blockage and CSI error, demonstrating its superiority to state-of-the-art (SotA) alternatives.

INDEX TERMS Millimeter wave systems, distributed hybrid beamforming, stochastic gradient descent, cooperative multi-point downlink beamforming.

I. INTRODUCTION

The ever-growing demands for data-rate and massive wireless connectivity have driven the fifth generation (5G) standard to incorporate technologies that exploit the mmWave spectrum available in the 24–30 [GHz] bands [1]–[3]. This trend is expected to continue, with 5G New Radio (NR) aiming to support spectrum bands up to 71 [GHz] in Release 17. Higher frequencies in the sub-Terahertz (*i.e.*>100 GHz) bands are planned to be added [4].

MmWave systems have much wider bandwidths than sub-6GHz, and they enable highly directional communications owing to the dense packing of antenna elements, which yields numerically large arrays of small physical dimensions, such that multiple-input multiple-output (MIMO) and beamforming techniques play key roles in mmWave technology [5]. Despite the advantages of higher achievable throughputs (data rates) and improved radio access

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenzhou Tang.

(user capacity), mmWave systems suffer from communications-impairing effects such as increased path loss, higher atmospheric absorption, higher sensitivity to phase noise, and random path blockage [6]–[10]. These phenomena pose challenges to the practical implementation of mmWave, which over the years has attracted the attention of the research community, leading to several important advancements.

Early contributions concentrated on mitigating the path loss issue by exploiting the high directivity of mmWave antenna arrays [11], [12] and therefore aimed at taking advantage of the sparsity of the mmWave channel. To this end, novel sparse signal processing methods were developed, albeit under the assumption of perfect CSI and for a single-user single-carrier setup. These designs were then extended to multi-user and multi-carrier systems by incorporating fully-connected hybrid architectures, where the RF threads are connected to all the phase shifters equipped, *e.g.*, [13]. It was shown that beamforming can take advantage of the homoscedasticity and variance correlations of channel covariance across subcarriers.

Under the argument that accurate RF components at mmWave bands are expensive, cost reduction via the design of partially connected hybrid beamformers such as those in [14], [15] has been motivated. The same argument that lower-cost RF components lead to various imperfections, which in turn can be counterweighted by robust hybrid designs, motivated works such as [16] in which the authors proposed that the alternating maximization of a smooth signal-to-noise-ratio (SNR)-driven optimization problem can be solved via orthogonal matching pursuit (OMP), for the realization of transceivers and amplify-and-forward (AF) relays with CSI uncertainty; in [17], a transmit beamforming scheme was designed via a fractional programming approach and was matched with a minimum mean square error (MMSE) receive beamformer to mitigate hardware and CSI imperfection; and [18] aimed to maximize the sum-rate under imperfect CSI feedback, employing a robust convolutional neural-network approach.

Despite the contributions that helped solidify robust hybrid beamforming as the prevailing approach that ensures the feasibility of mmWave systems, the random path blockage problem inherent to the mmWave channel has not been addressed sufficiently compared to the other issues discussed previously. The fundamental challenge remaining is that mmWave signals are susceptible to random blockage of propagation paths with probabilities ranging between 20% and 60%, which if not counteracted can significantly detract from the high-throughputs that mmWave communication systems can potentially deliver [7], [8], [10].

An early proposal to combat such effects is the CoMP scheme described in [19], in which quality of service (QoS) was maintained despite path blockages by synchronizing transmissions from multiple base stations (BSs) and access points (APs). The CoMP approach is not only attractive in the context of mmWave systems owing to the reduced radio coverage, but it is also shown analytically to offer strong theoretical guarantees to achieve capacity in the presence of a path blockage [20], which, however, is unfortunately not accompanied by procedures for practical implementation.

The latter gap is filled by the current article, which presents an extensive robust stochastic learning approach to minimize outage probability in mmWave CoMP systems subjected to random path blockage, thus effectively maintaining high-throughput service guarantees. To this end, it may be emphasized that the present work is compatible with some existing side-information-aided approaches such as the sub-6 [GHz] side-signaling assumed in [21], [22] or the visual (camera-based) information used in [23]–[25]. In addition, our contributions are also aligned with recent works on robust mmWave beamforming in which path blockage is assumed to be random and dealt-with dynamically, a few examples of which are discussed below.

In [26] a high-speed railway communications environment was considered, where it can be assumed that path blockages, detected during a first probing stage, remain constant during a subsequent transmission stage (*i.e.*, *coherent blockage assumption*). The article then proposed a corresponding two-stage (detect-and-avoid) mmWave hybrid beamforming mechanism based on greedy matching pursuit optimization, aimed at minimizing MMSE and maximizing sum-rate maximization (SRM) under the assumption of perfect CSI.

Besides the coherent blockage assumption, the latter contribution did not consider the benefit of CoMP transmission, which was previously shown in [19], [20] to be crucial for high-performing mmWave systems. In turn, [27] considered a robust CoMP setup under the assumption that blockage affects line-of-sight (LoS) paths in an incoherent manner. Subsequently, a greedy digital beamforming problem was formulated aimed at maximizing the system's minimum sum-rate, which was solved iteratively via convex approximations.

The LoS-only path blockage assumption adopted in [27] can be further extended to match empirical evidence on the behavior of mmWave channels [7], [8], [10]. This generalization has been made by the stochastic gradient descent (SGD) approach employed in [28], where a digital CoMP beamforming scheme was proposed to minimize outage in mmWave systems subjected to Bernoulli-distributed blockages of both LoS and non-line-of-sight (NLoS) paths. This work was later complemented in [29], which extended the digital outage-minimizing (OutMin) contribution of [28] to a fully-connected hybrid design in which the CoMP architecture was replaced by a single BS assisted by the auxiliary support of emerging intelligent reflecting surfaces (IRSs).

Common among the aforementioned references on the design of mmWave beamforming schemes to combat path blockage [26]–[29] are two important challenges that need to be addressed. The first is that these methods are either digital or fully-connected hybrid radio architectures, which can be further extended to the partially-connected hybrid

architecture[1] [12]–[16] owing to their lower costs and scalability advantages. The second is that these recent beamforming schemes, while assuming perfect knowledge of the channel coefficients when considering path blockages, deviate from earlier trends toward solutions that are robust against CSI-errors [16]–[18]. Therefore, it can be said that these first solutions have taken a well-justified scientific license for sacrificing CSI-error robustness and low-cost architectural premises in the design of mmWave beamforming schemes, with the goal of developing insights on combating the path-blockage problem.

Now is the time to address the integration of path-blockage mitigation knowledge with other enabling, building towards feasible and resilient mmWave systems. To this end, we propose a novel SGD-based hybrid-beamforming scheme for mmWave systems that is robust to both CSI imperfection and incoherent random path-blockage under the limited partially-connected CoMP paradigm.

A. CONTRIBUTIONS

The following is a summary of the article's contributions:

- In Subsection II-B, we **incorporate into a Bernoulli-Gaussian probability density function (PDF) the statistical features of both path blockages** [6]–[8], [10], [11] **and CSI errors** [16]–[18], resulting in an integrated stochastic mmWave channel model that enables both challenges to be addressed simultaneously.
- In Subsection II-C, the latter model is employed in the elaboration of the resulting mmWave CoMP downlink receive signal subject to both path blockage and CSI imperfection. Subsequently in Subsection II-D, the corresponding stochastic signal-to-interference-plus-noise ratio (SINR) is formulated, enabling the formulation of **a new stochastic sum-outage-probability minimization problem for the design of robust virtually-configured partially-connected cooperative hybrid beamformers** suitable to mitigate both path blockages and imperfect CSI in mmWave systems.
- In Subsection III-A **the aforementioned formulation is transformed via smooth relaxation into a new ERM stochastic learning problem** for sum-outage minimization with hinge-based soft outage detection, whose solution can be achieved efficiently by the MSGD method, in which gradient minimizing updates of the digital and analog beamforming components are obtained alternately and iteratively.
- In Subsections III-B and III-C, **original closed-form expressions for the SGD-based update rules are obtained**, for the baseband digital and the RF analog Riemann manifold-constrained beamforming sub-problems, respectively.
- In Subection IV, the **designed learning rates to accelerate the convergence of the digital and analog beamforming problems are derived in closed-form**, based on lower-bounds of the corresponding Lipschitz constants, and integrated with the results of the preceding subsection to compose the proposed scheme summarized in Algorithm 1.
- Comparisons of the proposed MSGD-based robust hybrid CoMP mmWave beamformer against SotA methods are offered in Section V, which demonstrate the overall effectiveness of the aforementioned contributions in mitigating the effects of path blockages and CSI errors onto the performance of mmWave systems.

[1]This indicates that the RF beamforming matrix is not dense.

B. NOTATION

The following notations are used throughout the article. The set of complex numbers is denoted by $\mathbb{C}$. Complex-valued matrices and vectors are respectively denoted by bold upper and lower case, respectively, as $\boldsymbol{X}$ and $\boldsymbol{x}$, while scalars are denoted using a normal font as in x. The circularly symmetric complex Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{CN}(\mu, \sigma^2)$. The operators $\Re\{\boldsymbol{X}\}$ and $\Im\{\boldsymbol{X}\}$ denote the real and imaginary parts of $\boldsymbol{X}$, respectively. The ℓ_p-norm is denoted by $\|\mathbf{x}\|_p$, where $p \geq 0$. The transpose, conjugate, and conjugate transpose (Hermitian) of a matrix $\boldsymbol{X}$ are denoted by $\boldsymbol{X}^{\mathrm{T}}$, $\boldsymbol{X}^{*}$ and $\boldsymbol{X}^{\mathrm{H}}$, respectively. The N-sized identity, all-one and all-zero matrices are respectively represented as $\mathbf{I}_N$, $\mathbf{1}_N$ and $\mathbf{0}_N$. The operator $\circ$ and $\otimes$ respectively denotes the Hadamard product and the Kronecker product. The functions vec, unvec, and Retr denote the vectorization, matricization (*i.e.*, the inverse vectorization), and retraction, respectively.

II. CHANNEL AND SYSTEM MODEL

A. MOTIVATION: PATH BLOCKAGE IN mmWave

Consider a CoMP downlink system operating at mmWave frequency bands as illustrated in Figure 1, where multiple synchronized APs simultaneously and cooperatively transmit data symbols to multiple single-antenna users subjected to sudden and random channel path blockages and imperfect small-scale fading coefficient estimates. It has been observed experimentally [7], [19] that mmWave channels are susceptible to path blockages of likelihood between 20% to 60%. Models for such path blockage processes have since been proposed [8], [10] and employed to demonstrate the severe impact of the phenomenon on system performance [20].

It has also been demonstrated that the blockage probability of each channel path component can be estimated and predicted [22]–[24], and thus utilized in the design of robust beamformers with the objective of mitigating performance losses that would result from the aforementioned channel uncertainties [26], [28], [29].

Motivated by the gravity of the aforementioned problem and the relative success of preliminary solutions, we seek to contribute a new and improved robust coordinated hybrid beamforming method to combat path blockage in mmWave systems. To this end, we employ the following system and channel models, consistent with preceding literature.

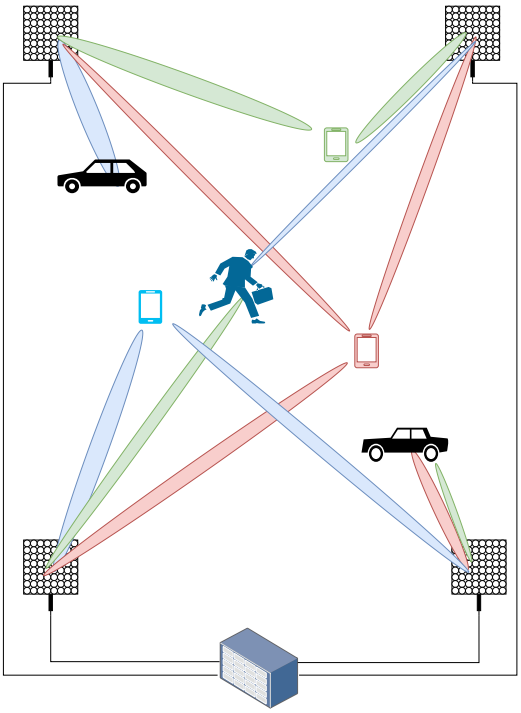


FIGURE 1. Illustration of the downlink system model of a CoMP mmWave communications system with random path blockages. Smart phones represent single-antenna UEs, while cars and a human are regarded as obstacles.

B. BERNOULLI-GAUSSIAN mmWave PATH BLOCKAGE CHANNEL MODEL

It is assumed that APs are each equipped with a uniform planar array (UPA) of N_t antennas – and thus individually capable of 3D beamforming [30] – and are also connected through a wired fronthaul link to a common central processing unit (CPU) such that the ensemble of CPU-coordinated APs can perform cooperative robust beamforming.

It is assumed that the mmWave channel contains a random number, $K_{b,u}$, of clusters [5], modeled as a lower-bounded Poisson random variable, $K_{b,u} \sim \max(1, \text{Possion}(\lambda))$, with the intensity parameter λ, as proposed in [6]. It is also assumed that CSI is acquired and tracked continuously, exploiting the reciprocity between the uplink and downlink channels of standard time division duplex (TDD) systems. Consequently, channel estimates can be modeled as

$$\hat{\boldsymbol{h}}_{b,u} = \sqrt{\frac{1}{K_{b,u}}} \sum_{k=1}^{K_{b,u}} g_{b,u}^{k} \boldsymbol{a}_{N_t}(\theta_{b,u}^{k}, \phi_{b,u}^{k}), \tag{1}$$

where $b \in \mathcal{B} \triangleq \{1, 2, \cdots, B\}$ and $u \in \mathcal{U} \triangleq \{1, 2, \cdots, U\}$ denote the AP and UE indices, respectively, with B and U denoting the total number of APs and UEs; $\theta_{b,u}^{k}$ and $\phi_{b,u}^{k}$ are the elevation and azimuth AoD of the m-th cluster from the b-th AP towards the u-th downlink user, respectively; $\boldsymbol{a}_T(\theta_{b,u}^{k}, \phi_{b,u}^{k})$ represents the array response vector; and $g_{b,u}^{k}$ is the associated channel gain modeled as $g_{b,u}^{k} \sim \mathcal{CN}\left(0, \sigma_{b,u,k}^{2}\right)$, with $\sigma_{b,u,k}^{2} = 10^{-\text{PL}_{b,u}^{k}/10}$ and $\text{PL}_{b,u}^{k} = \alpha + 10\beta \log_{10}\left(d_{b,u}\right) + \xi$ [dB], in which $d_{b,u}$ is the distance (in meters) between the b-th AP and the u-th user, with the parameters α, β, and ξ as given in [6, Table 1].[2]

Assuming that the UPAs equipping the APs have a regular square shape with $\sqrt{N_t}$ antenna elements on each axis, the array response vector $\boldsymbol{a}_T(\theta_{b,u}^{k}, \phi_{b,u}^{k})$ can be written as

$$\begin{aligned}&\boldsymbol{a}_{N_t}(\theta_{b,u}^{k}, \phi_{b,u}^{k}) \\ &\quad = \mathbf{c}_{\sqrt{N_t}}\left(\tfrac{1}{2}\sin(\theta_{b,u}^{k})\cos(\phi_{b,u}^{k})\right) \otimes \mathbf{c}_{\sqrt{N_t}}\left(\tfrac{1}{2}\cos(\theta_{b,u}^{k})\right),\end{aligned} \tag{2}$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{c}_N$ represents the array response vector of a uniform linear array (ULA) with N antenna elements, which can be expressed as

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

Despite the knowledge of $\hat{\boldsymbol{h}}_{b,u}$, during actual downlink, the system might be subjected to partial blockage if and when any or some of the LoS and NLoS clusters become temporarily blocked, such that the actual channel between the b-th AP and the u-th user can be modeled as

$$\boldsymbol{h}_{b,u} = \sqrt{\frac{1}{K_{b,u}}} \sum_{k=1}^{K_{b,u}} \omega_{b,u}^{k}(g_{b,u}^{k})\boldsymbol{a}_{N_t}(\theta_{b,u}^{k}, \phi_{b,u}^{k}), \tag{4}$$

with $\omega_{b,u}^{k} \sim f(p_{b,u}^{k}; g_{b,u}^{k}, \zeta_{b,u,k}^{2})$, where $f(p_{b,u}^{k}; g_{b,u}^{k}, \zeta_{b,u,k}^{2})$ denotes the Bernoulli-Gaussian distribution given by

$$\begin{aligned}f(p_{b,u}^{k}; g_{b,u}^{k}, \zeta_{b,u,k}^{2}) &= p_{b,u}^{k}\delta(\omega_{b,u}^{k}) \\ &\quad + (1 - p_{b,u}^{k})\mathcal{CN}(g_{b,u}^{k}; \zeta_{b,u,k}^{2}),\end{aligned} \tag{5}$$

with $p_{b,u}^{k}$, $\delta(\cdot)$, and $\zeta_{b,u,k}^{2}$ denoting the corresponding blockage probability, Dirac delta function, and channel gain estimation uncertainty variance, respectively.

It should be noted that the random variable model $\omega_{b,u}^{k}$ adopted here is a generalization of that utilized in most related literature, including our own previous work [28]. In particular, while the blockage model typically used in the mmWave literature is simple Bernoulli, the model in equation (5) is Bernoulli-Gaussian [34], with the first term modeling the mean random blockage that occurs with probability $p_{b,u}^{k}$, while the second term captures variations of the small-scale fading coefficients, which occur due to a summation of effects including CSI imperfection, channel aging, and partial blockage. Setting $\zeta_{b,u,k}^{2} = 0$, the Bernoulli-Gaussian model of equation (5) reduces to the usual Bernoulli blockage model considered in the literature [27]–[29], and the path of the m-th cluster from the b-th AP to the u-th user gets completely blocked when $\omega_{b,u}^{k} = 0$.

[2] Although we consider a frequency-independent outdoor urban mmWave channel model, one can consider other mmWave channel models such as indoor scenarios [31], rural macrocell scenarios [32], and frequency-dependent wideband scenarios [33]. In any case, since the proposed algorithm presented later does not assume any particular channel model, the proposed algorithm can be easily extended to such different scenarios, which is however out of scope of this article.

C. PARTIALLY CONNECTED HYBRID COOPERATIVE DOWNLINK SIGNAL MODEL

Under the assumption that the number of RF chains at each AP is limited to $N_{\mathrm{RF}} \ll N_t$, and considering the channel model detailed above, let $\boldsymbol{f}_{b,u} \in \mathbb{C}^{N_{\mathrm{RF}}\times 1}$ denote the transmit digital baseband beamforming vector from the b-th AP towards the u-th user, and $\boldsymbol{V}_b \in \mathbb{C}^{N_t\times N_{\mathrm{RF}}}$ be the transmit analog beamforming matrix employed by the b-th AP when transmitting to all users, subject to the unit modulus constraint (*i.e.*, $|[\boldsymbol{V}_b]_{i,j}| = 1$) [12].

Then, introducing the aggregate digital baseband beamforming matrix $\boldsymbol{F}_b \triangleq [\boldsymbol{f}_{b,1}, \cdots, \boldsymbol{f}_{b,u}]$ employed by the b-th AP, the received signal y_u at the u-th user can be written as

$$\begin{aligned} y_u &= \sum_{b\in\mathcal{B}} \boldsymbol{h}_{b,u}^{\mathrm{H}} \boldsymbol{V}_b \boldsymbol{F}_b \boldsymbol{x} + n_u \\ &= \sum_{b\in\mathcal{B}} \boldsymbol{h}_{b,u}^{\mathrm{H}} \boldsymbol{V}_b \boldsymbol{f}_{b,u} x_u + \sum_{u'\in\mathcal{U}\setminus u} \sum_{b\in\mathcal{B}} \boldsymbol{h}_{b,u}^{\mathrm{H}} \boldsymbol{f}_{b,u'} x_{u'} + n_u \\ &= \underbrace{\boldsymbol{h}_u^{\mathrm{H}} \boldsymbol{V} \boldsymbol{f}_u x_u}_{\text{intended signal}} + \underbrace{\sum_{u'\in\mathcal{U}\setminus u} \boldsymbol{h}_u^{\mathrm{H}} \boldsymbol{V} \boldsymbol{f}_{u'} x_{u'}}_{\text{interuser interference}} + n_u, \end{aligned} \tag{6}$$

where $\boldsymbol{x} \triangleq [x_1, \cdots, x_U]^{\mathrm{T}}$ is the symbol vector transmitted cooperatively from all APs to all users, with x_u denoting a symbol targeting at the u-th user; n_u denotes independent and identically distributed (i.i.d.) circularly symmetric zero-mean additive white Gaussian noise (AWGN) at the u-th user,[3] *i.e.* $n_u \sim \mathcal{CN}(0, \xi_u^2)$; and finally the cooperative analog beamforming matrix, $\boldsymbol{V} \triangleq \mathrm{blkdiag}(\boldsymbol{V}_1, \cdots, \boldsymbol{V}_B) \in \mathbb{C}^{BN_t\times BN_{\mathrm{RF}}}$, the aggregate channel vector, $\boldsymbol{h}_u \triangleq \left[\boldsymbol{h}_{1,u}^{\mathrm{T}}, \cdots, \boldsymbol{h}_{B,u}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{BN_t\times 1}$, and digital baseband beamformer, $\boldsymbol{f}_u \triangleq \left[\boldsymbol{f}_{1,u}^{\mathrm{T}}, \cdots, \boldsymbol{f}_{B,u}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{BN_{\mathrm{RF}}\times 1}$ from all APs to the u-th user are implicitly defined and introduced in the last equation, for notational simplicity.

We emphasize that the block-diagonal structure of the analog beamforming matrix $\boldsymbol{V}$ implies that the cooperative downlink transmission scheme yielding the received signal described by equation (6) is a virtually-configured partially-connected hybrid beamforming architecture, unlike the fully-connected counterpart considered in related works.

D. STOCHASTIC FORMULATION OF ROBUST COOPERATIVE BEAMFORMING PROBLEM

The task of performing mmWave robust CoMP downlink hybrid beamforming subject to random path blockage can be formulated as a constrained optimization problem without resorting to the standard deterministic robust optimization framework. To elaborate further, our strategy is to build resilience against random blockages by minimizing the stochastic sum-outage-probability subjected to this phenomenon, which can be formulated as

$$\underset{\boldsymbol{f},\boldsymbol{V}}{\text{minimize}} \sum_{u\in\mathcal{U}} \Pr\left\{\Gamma_u(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u) < \gamma_u\right\} \tag{7a}$$

$$\text{subject to } \|\boldsymbol{V}_b\boldsymbol{F}_b\|_{\mathrm{F}}^2 \le P_{\max,b}, \quad \forall b, \tag{7b}$$

$$\boldsymbol{V}_b \in \mathcal{M}_{cc}^{N_{\mathrm{RF}}N_t}, \quad \forall b, \tag{7c}$$

where γ_u denotes the target SINR for the u-th user, and the effective SINR Γ_u can be given by

$$\Gamma_u(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u) = \frac{|\boldsymbol{h}_u^{\mathrm{H}}\boldsymbol{V}\boldsymbol{f}_u|^2}{\sum_{u'\in\mathcal{U}\setminus u}|\boldsymbol{h}_u^{\mathrm{H}}\boldsymbol{V}\boldsymbol{f}_{u'}|^2 + \xi_u^2}, \tag{8}$$

with $\boldsymbol{f} \triangleq \left[\boldsymbol{f}_1^{\mathrm{T}}, \cdots, \boldsymbol{f}_U^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{BUN_{\mathrm{RF}}\times 1}$.

In equation (7), $\mathcal{M}_{cc}^{N_{\mathrm{RF}}N_t}$ denotes an N_{RF}-by-N_t sub-Riemannian circle manifold in $\mathbb{C}^{N_{\mathrm{RF}}\times N_t}$, defined as

$$\mathcal{M}_{cc}^{N_{\mathrm{RF}}N_t} \triangleq \left\{\boldsymbol{z}\in\mathbb{C}^{N_{\mathrm{RF}}\times N_t} \,\middle|\, |z_i| = 1,\, i = \{1,\cdots,N_{\mathrm{RF}}\times N_t\}\right\}. \tag{9}$$

It should be noted that the problem formulated in equation (7) differs fundamentally from related robust beamforming methods. The objective of the problem formulated in equation (7) is to enable reliable communication even under harmful random blockages and CSI errors, by minimizing sum-outage, which is more practical from a system point-of-view. In addition, unlike deterministic formulations, which are often based on a max-min sum rate maximization framework followed by convex approximations with complexity that grows rapidly, equation (7) is stochastic and has only as many constraints as the number of APs owing to the power constraints on top of the unit modular constraint for analog beamforming.

Therefore, it can be said that the proposed approach has a core advantage of being inherently practical, both from mathematical and application viewpoints.

III. PROPOSED METHOD PART 1: UPDATE RULES

A. REFORMULATION AS TRAINED EMPIRICAL RISK MINIMIZATION PROBLEM

In this section, we determine the update rules for the digital and analog beamformers, $\boldsymbol{f}$ and $\boldsymbol{V}$, required to solve equation (7). To that end, we employ the Gauss-Seidel-type block-coordinate stochastic gradient technique [35] in conjunction with manifold optimization to decouple the variables $\boldsymbol{f}$ and $\boldsymbol{V}$, so that the intractability imposed jointly by the stochastic objective and the non-convex constraint can be mitigated, providing an iterative alternate solution for the problem formulated in equation (7).

To this end, owing to the ambiguity of the objective function, we first introduce the following indicator function $\mathbb{1}_{\Gamma_u(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u)<\gamma_u}$ given by

$$\mathbb{1}_{\Gamma_u(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u)<\gamma_u} = \begin{cases} 1 & \text{if } \Gamma_u(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u) < \gamma_u \\ 0 & \text{otherwise} \end{cases}, \tag{10}$$

[3] It should be noted that this method will be proposed later for different types of noise. For the sake of simplicity, the common AWGN model is assumed here.

such that equation (7) can be written as a type of expectation minimization problem, namely

$$\underset{\boldsymbol{f},\boldsymbol{V}}{\text{minimize}} \sum_{u\in\mathcal{U}} \mathbb{E}_{\omega_{b,u}^{k}}\left[\mathbb{1}_{\Gamma_u(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u)<\gamma_u}\right] \tag{11a}$$

$$\text{subject to } \|\boldsymbol{V}_b\boldsymbol{F}_b\|_{\mathrm{F}}^2 \le P_{\max,b}, \quad \forall\, b, \tag{11b}$$

$$\boldsymbol{V}_b \in \mathcal{M}_{cc}^{N_{\mathrm{RF}}N_t}, \quad \forall\, b. \tag{11c}$$

Under the assumption that the channel gains and AoDs are obtained from the preceding channel estimation process, while also considering that some of those paths might be under random blockage during the actual data transmission, one may notice that possible combinations of blockage patterns due to $\omega_{b,u}^{k}$ can be randomly generated and utilized as a training dataset for problem (11). This is based on the established fact that the path blockage probabilities of mmWave channels can be estimated, as demonstrated in [22]–[24].

Considering the above, we define $\boldsymbol{h}_u^m$ as the m-th training data batch for the channel $\boldsymbol{h}_u$ of the u-th user, such that the ERM problem formulation of equation (11) can be further rewritten in terms of the summation, *i.e.*

$$\underset{\boldsymbol{f},\boldsymbol{V}}{\text{minimize}} \frac{1}{M}\sum_{m=1}^{M}\sum_{u\in\mathcal{U}} \mathbb{1}_{\Gamma_u(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u^m)<\gamma_u} \tag{12a}$$

$$\text{subject to } \|\boldsymbol{V}_b\boldsymbol{F}_b\|_{\mathrm{F}}^2 \le P_{\max,b} \quad \forall\, b, \tag{12b}$$

$$\boldsymbol{V}_b \in \mathcal{M}_{cc}^{N_{\mathrm{RF}}N_t} \quad \forall\, b, \tag{12c}$$

where M denotes the size of the training dataset.

To address the intractable non-smoothness of the ERM problem given in equation (12), we further introduce a smooth surrogate function $\nu_u\left(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u^m\right)$ so that a gradient expression of (12a) can be efficiently computed. To this end, we exploit the fact that equation (12) can be seen as a classification problem; therefore, the indicator function (12a) can be replaced by the hinge surrogate function [36], [37]

$$\nu_u\left(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u^m\right) = \begin{cases} 0 & \text{if } 1-\dfrac{\Gamma_u\left(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u^m\right)}{\gamma_u} \\ & < 0, \\ 1-\dfrac{\Gamma_u\left(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u^m\right)}{\gamma_u} & \text{otherwise}, \end{cases} \tag{13}$$

which in turn enables equation (12) to be rewritten as

$$\underset{\boldsymbol{f},\boldsymbol{V}}{\text{minimize}} \frac{1}{M}\sum_{m=1}^{M}\sum_{u\in\mathcal{U}} \nu_u\left(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u^m\right) \tag{14a}$$

$$\text{subject to } \|\boldsymbol{V}_b\boldsymbol{F}_b\|_{\mathrm{F}}^2 \le P_{\max,b} \quad \forall\, b, \tag{14b}$$

$$\boldsymbol{V}_b \in \mathcal{M}_{cc}^{N_{\mathrm{RF}}N_t} \quad \forall\, b. \tag{14c}$$

The latter formulation of the proposed robust cooperative hybrid mmWave beamforming problem can be recognized as a type of differentiable non-convex stochastic optimization problem with manifold constraints, whose solution can be found via a block-coordinate descent algorithm such as that proposed in [35]. To this end, however, the gradients of the objective (14a) with respect to the digital baseband matrix $\boldsymbol{f}$ and the analog RF beamforming matrix $\boldsymbol{V}$ must be derived, which is the subject of the next two subsections.

Following a standard MSGD framework [38], this task will be pursued under a hybrid and alternate approach in which the gradients of the objective function (14a) with respect to the digital baseband component $\boldsymbol{f}$ with $\boldsymbol{V}$ constant, and with respect to the analog RF component $\boldsymbol{V}$ with $\boldsymbol{f}$ constant will be considered. In addition, we consider a minimization algorithm for variance reduction in terms of gradient direction at each iteration [39] by means of mini-batches.

B. DIGITAL BASEBAND COMPONENT: GRADIENT WITH RESPECT TO $\boldsymbol{f}$ WITH $\boldsymbol{V}$ CONSTANT

For notation simplicity, we hereafter define the mini-batch size employed at the i-th algorithmic iteration as M_i, with $1 \ll M_i \ll M$.

For a fixed $\boldsymbol{V}$, and over a mini batch of size M_i, the optimization problem (14) reduces to the subproblem

$$\underset{\boldsymbol{f}}{\text{minimize}} \frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{u\in\mathcal{U}} \nu_u\left(\boldsymbol{f}|\boldsymbol{V},\boldsymbol{h}_u^m\right) \tag{15a}$$

$$\text{subject to } \|\boldsymbol{V}_b\boldsymbol{F}_b\|_{\mathrm{F}}^2 \le P_{\max,b}\ \forall\, b. \tag{15b}$$

Since the entries of the digital baseband beamforming vector $\boldsymbol{f}$ can be any complex value that satisfies the power constraint (15b), the gradient of $\nu_u\left(\boldsymbol{f}|\boldsymbol{V},\boldsymbol{h}_u^m\right)$ with respect to $\boldsymbol{f}$ can be computed by taking its derivative for a fixed $\boldsymbol{V}$, which yields

$$\nabla\nu_u\left(\boldsymbol{f}|\boldsymbol{V},\boldsymbol{h}_u^m\right) = \begin{cases} 0 & \text{if } 1-\dfrac{\Gamma_u\left(\boldsymbol{f}|\boldsymbol{V},\boldsymbol{h}_u^m\right)}{\gamma_u} \\ & < 0 \\ -\dfrac{\nabla\Gamma_u\left(\boldsymbol{f}|\boldsymbol{V},\boldsymbol{h}_u^m\right)}{\gamma_u} & \text{otherwise}, \end{cases} \tag{16}$$

where the gradient of the SINR expression Γ_u can be computed by introducing its alternative formula for the stacked baseband beamforming vector $\boldsymbol{f}$

$$\begin{aligned}\nabla\Gamma_u(\boldsymbol{f}|\boldsymbol{V},\boldsymbol{h}_u^m) &= \nabla\frac{\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Phi}_f^m\boldsymbol{f}}{\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m\boldsymbol{f}+\xi_u^2}, \\ &= \frac{\boldsymbol{\Phi}_f^m\boldsymbol{f}}{\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m\boldsymbol{f}+\xi_u^2} - \frac{\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Phi}_f^m\boldsymbol{f}}{(\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m\boldsymbol{f}+\xi_u^2)^2}\boldsymbol{\Psi}_f^m\boldsymbol{f}, \end{aligned} \tag{17}$$

where the second equality follows from the Wirtinger derivative and the auxiliary matrices $\boldsymbol{\Phi}_f^m$ and $\boldsymbol{\Psi}_f^m$ are respectively defined as

$$\boldsymbol{\Phi}_f^m \triangleq \mathrm{diag}\left(\boldsymbol{e}_u\right) \otimes \boldsymbol{V}^{\mathrm{H}}\boldsymbol{h}_u^m\boldsymbol{h}_u^{m\mathrm{H}}\boldsymbol{V}, \tag{18a}$$

$$\boldsymbol{\Psi}_f^m \triangleq \mathrm{diag}\left(\bar{\boldsymbol{e}}_u\right) \otimes \boldsymbol{V}^{\mathrm{H}}\boldsymbol{h}_u^m\boldsymbol{h}_u^{m\mathrm{H}}\boldsymbol{V}, \tag{18b}$$

where $\boldsymbol{e}_u \in \{0,1\}^{U\times 1}$ denotes the u-th vector of the standard orthonormal basis of dimension U and $\bar{\boldsymbol{e}}_u$ denotes the logical negation of $\boldsymbol{e}_u$ (*i.e.*, $\bar{\boldsymbol{e}}_u = \boldsymbol{1} - \boldsymbol{e}_u$).

From the above, the update of $\boldsymbol{f}$ for a fixed analog RF beamforming matrix $\boldsymbol{V}$ can be written as

$$\boldsymbol{f}^{(i)} = \mathcal{P}_{\|\boldsymbol{V}_b \boldsymbol{F}_b\|_{\mathrm{F}}^2 \le P_{\max,b}} \left(\boldsymbol{f}^{(i-1)} - \frac{\alpha_i^f}{M_i} \sum_{m=1}^{M_i} \sum_{u \in \mathcal{U}} \nabla \nu_u \left(\boldsymbol{f} | \boldsymbol{V}, \boldsymbol{h}_u^m \right) \right), \tag{19}$$

where $\boldsymbol{f}^{(i-1)}$ indicates the solution obtained at the $(i-1)$-th iteration, α_i^f is the corresponding step-size to be tuned later, and $\mathcal{P}_{\|\boldsymbol{V}_b \boldsymbol{F}_b\|_{\mathrm{F}}^2 \le P_{\max,b}}(\cdot)$ denotes the projection onto the feasible convex set defined by the power constraint per-AP given by inequality (15b).

C. ANALOG RF COMPONENT: GRADIENT WITH RESPECT TO V WITH f CONSTANT

Analogous to the above, for a fixed $\boldsymbol{f}$, and over a mini batch of size M_i, the optimization problem (14) with respect to $\boldsymbol{V}$ reduces to the subproblem

$$\underset{\boldsymbol{V}}{\text{minimize}} \ \frac{1}{M_i} \sum_{m=1}^{M_i} \sum_{u \in \mathcal{U}} \nu_u \left(\boldsymbol{V} | \boldsymbol{f}, \boldsymbol{h}_u^m \right) \tag{20a}$$

$$\text{subject to } \boldsymbol{V}_b \in \mathcal{M}_{cc}^{N_{\mathrm{RF}} N_t} \quad \forall\, b. \tag{20b}$$

Unlike digital baseband beamforming $\boldsymbol{f}$, however, calculating the gradient direction with respect to $\boldsymbol{V}$ is challenging. This is because of the unit modular constraint defined by the complex circle manifold $\mathcal{M}_{cc}^{N_{\mathrm{RF}} N_t}$; thus, manifold optimization techniques [40] must be employed. To that end, we first define the tangent space at a given $\boldsymbol{z} \in \mathcal{M}_{cc}^{N_{\mathrm{RF}} N_t}$, which can be written as

$$\mathcal{T}_{\boldsymbol{z}} \mathcal{M}_{cc}^{N_{\mathrm{RF}} N_t} \triangleq \{ \boldsymbol{\tau} \in \mathbb{C}^{N_{\mathrm{RF}} N_t} | \Re\{ \boldsymbol{\tau} \circ \boldsymbol{z}^* \} = \boldsymbol{0} \}, \tag{21}$$

which contains all tangent vectors to $\mathcal{M}_{cc}^{N_{\mathrm{RF}} N_t}$ at a certain point $\boldsymbol{z}$.

Since the complex circle manifold $\mathcal{M}_{cc}^{N_{\mathrm{RF}} N_t}$ is a smooth Riemannian manifold, a positive-definite inner product can be defined on the tangent space given by equation (21), such that the Riemannian gradient $\nabla^{\mathcal{M}} \nu_u \left(\boldsymbol{V} | \boldsymbol{f}_u, \boldsymbol{h}_u^m \right)$ at $\boldsymbol{z}$ on $\mathcal{M}_{cc}^{N_{\mathrm{RF}} N_t}$ can be expressed as the orthogonal projection of the Euclidean gradient $\nabla \nu_u \left(\boldsymbol{V} | \boldsymbol{f}_u, \boldsymbol{h}_u^m \right)$ onto its tangent space. In other words, we may define

$$\begin{aligned} &\nabla^{\mathcal{M}} \nu_u \left(\boldsymbol{V} | \boldsymbol{f}, \boldsymbol{h}_u^m \right) \\ &\quad \triangleq \mathcal{P}_{\mathcal{T}_{\boldsymbol{z}} \mathcal{M}_{cc}^{N_{\mathrm{RF}} N_t}} \left(\frac{\sum_{m=1}^{M_i} \nabla \nu_u \left(\boldsymbol{V} | \boldsymbol{f}, \boldsymbol{h}_u^m \right)}{M_i} \right) \\ &\quad = \frac{1}{M_i} \sum_{m=1}^{M_i} \nabla \nu_u \left(\boldsymbol{V} | \boldsymbol{f}, \boldsymbol{h}_u^m \right) \\ &\qquad - \Re \left\{ \frac{1}{M_i} \sum_{m=1}^{M_i} \nabla \nu_u \left(\boldsymbol{V} | \boldsymbol{f}, \boldsymbol{h}_u^m \right) \circ \boldsymbol{V}^* \right\} \circ \boldsymbol{V} \end{aligned} \tag{22}$$

To facilitate the derivation of the Euclidean gradient $\nabla \nu_u \left(\boldsymbol{V} | \boldsymbol{f}_u, \boldsymbol{h}_u^m \right)$ with respect to $\boldsymbol{V}$, we reformulate the received signal and the SINR expressions so as to form a tractable optimization variable. In particular, we rewrite the received signal originally given in equation (6) as

$$\begin{aligned} y_u &= \boldsymbol{h}_u^{\mathrm{H}} \boldsymbol{V} \boldsymbol{f}_u x_u + \sum_{u' \in \mathcal{U} \setminus u} \boldsymbol{h}_u^{\mathrm{H}} \boldsymbol{V} \boldsymbol{f}_{u'} x_{u'} + n_u \\ &= ((x_u \boldsymbol{f}_u)^{\mathrm{T}} \otimes \boldsymbol{h}_u^{\mathrm{H}}) \mathrm{vec}(\boldsymbol{V}) \qquad (23a) \\ &\quad + \sum_{u' \in \mathcal{U} \setminus u} ((x_{u'} \boldsymbol{f}_{u'})^{\mathrm{T}} \otimes \boldsymbol{h}_u^{\mathrm{H}}) \overbrace{\mathrm{vec}(\boldsymbol{V})}^{\triangleq \boldsymbol{v}} + n_u \qquad (23b) \end{aligned}$$

where $\mathrm{vec}(\cdot)$ denotes the vectorized (column stacked) representation of a matrix, such that the implicitly defined vectorized representation $\boldsymbol{v} \triangleq \mathrm{vec}(\boldsymbol{V})$ is a sparse vector owing to the partially connected structure of $\boldsymbol{V}$.

It will prove convenient, therefore, to further define the auxiliary matrix $\tilde{\boldsymbol{V}} \triangleq \left[\boldsymbol{V}_1, \cdots, \boldsymbol{V}_B \right]$ such that its vectorized representation $\tilde{\boldsymbol{v}} \triangleq \mathrm{vec}(\tilde{\boldsymbol{V}})$ is a dense vector, and we may write

$$y_u = ((x_u \boldsymbol{f}_u)^{\mathrm{T}} \otimes \boldsymbol{h}_u^{\mathrm{H}}) \boldsymbol{W} \tilde{\boldsymbol{v}} + \sum_{u' \in \mathcal{U} \setminus u} ((x_{u'} \boldsymbol{f}_{u'})^{\mathrm{T}} \otimes \boldsymbol{h}_u^{\mathrm{H}}) \boldsymbol{W} \tilde{\boldsymbol{v}} + n_u, \tag{23c}$$

where $\boldsymbol{W}$ denotes the transform matrix mapping the dense vector $\tilde{\boldsymbol{v}}$ onto its sparse representation $\boldsymbol{v}$.

In light of equation (23), the SINR originally expressed as in equation (8) can be rewritten for a given channel $\boldsymbol{h}_u^m$ as

$$\Gamma_u \left(\tilde{\boldsymbol{v}} | \boldsymbol{f}, \boldsymbol{h}_u^m \right) = \frac{\tilde{\boldsymbol{v}}^{\mathrm{H}} \boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m \tilde{\boldsymbol{v}}}{\tilde{\boldsymbol{v}}^{\mathrm{H}} \boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m \tilde{\boldsymbol{v}} + \xi_u^2}, \tag{24}$$

where

$$\boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m \triangleq \boldsymbol{W}^{\mathrm{H}} \left(\boldsymbol{f}_u^{\mathrm{T}} \otimes \boldsymbol{h}_u^{m\mathrm{H}} \right)^{\mathrm{H}} \left(\boldsymbol{f}_u^{\mathrm{T}} \otimes \boldsymbol{h}_u^{m\mathrm{H}} \right) \boldsymbol{W}, \tag{25a}$$

$$\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m \triangleq \boldsymbol{W}^{\mathrm{H}} \sum_{u' \in \mathcal{U} \setminus u} \left(\boldsymbol{f}_{u'}^{\mathrm{T}} \otimes \boldsymbol{h}_u^{m\mathrm{H}} \right)^{\mathrm{H}} \left(\boldsymbol{f}_{u'}^{\mathrm{T}} \otimes \boldsymbol{h}_u^{m\mathrm{H}} \right) \boldsymbol{W}. \tag{25b}$$

Given the above, the Euclidean gradient $\nabla \nu_u \left(\tilde{\boldsymbol{v}} | \boldsymbol{f}, \boldsymbol{h}_u^m \right)$ can be represented as

$$\nabla \nu_u \left(\tilde{\boldsymbol{v}} | \boldsymbol{f}, \boldsymbol{h}_u^m \right) = \begin{cases} 0 & \text{if } 1 - \dfrac{\Gamma_u \left(\tilde{\boldsymbol{v}} | \boldsymbol{f}, \boldsymbol{h}_u^m \right)}{\gamma_u} < 0 \\ -\dfrac{\nabla \Gamma_u \left(\tilde{\boldsymbol{v}} | \boldsymbol{f}, \boldsymbol{h}_u^m \right)}{\gamma_u} & \text{otherwise,} \end{cases} \tag{26}$$

where, based on the SINR reformulation given in equation (24), the Euclidean gradient of the SINR can be expressed similarly to equation (17), *i.e.*

$$\nabla \Gamma_u \left(\tilde{\boldsymbol{v}} | \boldsymbol{f}, \boldsymbol{h}_u^m \right) = \frac{\boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m \tilde{\boldsymbol{v}}}{\tilde{\boldsymbol{v}}^{\mathrm{H}} \boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m \tilde{\boldsymbol{v}} + \xi_u^2} - \frac{\tilde{\boldsymbol{v}}^{\mathrm{H}} \boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m \tilde{\boldsymbol{v}}}{\left(\tilde{\boldsymbol{v}}^{\mathrm{H}} \boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m \tilde{\boldsymbol{v}} + \xi_u^2 \right)^2} \boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m \tilde{\boldsymbol{v}}. \tag{27}$$

In light of the above, the Riemannian gradient can be finally rewritten as

$$\begin{aligned} \nabla^{\mathcal{M}} \nu_u \left(\tilde{\boldsymbol{v}} | \boldsymbol{f}, \boldsymbol{h}_u^m \right) &= \frac{1}{M_i} \sum_{m=1}^{M_i} \nabla \nu_u \left(\tilde{\boldsymbol{v}} | \boldsymbol{f}, \boldsymbol{h}_u^m \right) \\ &\quad - \Re \left\{ \frac{1}{M_i} \sum_{m=1}^{M_i} \nabla \nu_u \left(\tilde{\boldsymbol{v}} | \boldsymbol{f}, \boldsymbol{h}_u^m \right) \circ \tilde{\boldsymbol{v}}^* \right\} \circ \tilde{\boldsymbol{v}}, \end{aligned} \tag{28}$$

which can be recognized as a variation of equation (22) over the vectorized dense representation $\tilde{\boldsymbol{v}}$ of $\boldsymbol{V}$.

Based on the Riemannian gradient of equation (28), the updated vector $\tilde{\boldsymbol{v}}$ obtained over the tangent space and mapped onto the complex circle manifold $\mathcal{M}_{cc}^{N_{\mathrm{RF}}N_t}$ can be written as

$$\tilde{\boldsymbol{v}}^{(i)} = \mathrm{Retr}\Big(\tilde{\boldsymbol{v}}^{(i-1)} - \alpha_i^v \sum_{u\in\mathcal{U}} \nabla^{\mathcal{M}} \nu_u\big(\tilde{\boldsymbol{v}}^{(i-1)}|\boldsymbol{f}, \boldsymbol{h}_u^m\big)\Big), \tag{29a}$$

where $\mathrm{Retr}(\cdot)$ denotes the retraction operator that rescales a given vector to element-wise unit modular entries [40] and α_i^v is a step size to be given later.

After obtaining $\tilde{\boldsymbol{v}}^{(i)}$, the update of $\boldsymbol{V}$ for a fixed digital baseband beamforming matrix $\boldsymbol{f}$ can finally be obtained by unvectorizing its associated sparse version, *i.e.*

$$\boldsymbol{V}^{(i)} = \mathrm{unvec}(\boldsymbol{W}\tilde{\boldsymbol{v}}^{(i)}). \tag{29b}$$

IV. PROPOSED METHOD PART 2: LEARNING RATES

A. CRITERIA FOR CONVERGENCE GUARANTEE

Having derived gradient updates corresponding to both the digital baseband beamforming matrix $\boldsymbol{f}$ and the virtually configured partially connected analog RF beamforming matrix $\boldsymbol{V}$ constrained by the unit modular manifold $\mathcal{M}_{cc}$, we proceed to propose a new blockage-robust hybrid beamforming algorithm via the block-coordinate stochastic gradient descent (BSGD) framework for CoMP systems operating at mmWave bands. To that end, we first need to determine the learning rates α_i^f and α_i^v employed in equations (19) and (29a), respectively.

There are two well-known (but not necessary) convergence criteria for SGD algorithms, namely, the *shrinkage criterion* [41], which entails a shrinking learning rate $\alpha_1 > \cdots > \alpha_i > \cdots$ satisfying $\sum_{i=1}^{\infty}\alpha_i \to \infty$ and $\sum_{i=1}^{\infty}\alpha_i^2 < \infty$; and the *Lipschitz criterion* [35], according to which convergence is ensured by setting $\alpha_i \in (0, 1/L]$, $\forall i$, where L is the Lipschitz constant.

Although it is highly challenging to determine the exact Lipschitz constant for the problem at hand, it was shown in [35] that setting $\alpha_i = \rho/(\sqrt{i}\cdot L^*)$, where L^* is a lower bound of the Lipschitz constant L and ρ is a scaling coefficient, has led to convergence. With that in mind, one can leverage the well-known Taylor theorem to obtain a lower bound of the Lipschitz constant. In particular, upon momentarily altering our notation and using ν_u to denote the u-th term of either of the surrogate objective functions in equations (15a) and (20a), the Taylor Theorem yields $L_u\cdot\mathbf{I} \succeq \nabla^2\nu_u \Longrightarrow L_u \geq \lambda_{\max}(\nabla^2\nu_u)$, where $\lambda_{\max}(\cdot)$ is the largest eigenvalue (spectral norm) of a given matrix, such that the lower bound on the Lipschitz constant is given by

$$L^* = \sum_{u=1}^{U} \lambda_{\max}(\nabla^2\nu_u). \tag{30}$$

In light of the above, we offer below the derivation of the Lipschitz constant lower bounds L_f^* and L_v^* corresponding to the stochastic-gradient-based solutions of the subproblems described by equations (15) and (20), respectively.

B. DIGITAL BASEBAND COMPONENT: LIPSCHITZ BOUND FOR f WITH V CONSTANT

The Hessian of $\nu_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right)$ is given by

$$\begin{aligned} &H_f(\nu_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right)) \\ &\triangleq \nabla^2\nu_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right) \\ &= \begin{cases} \boldsymbol{0} & \text{if } 1 - \dfrac{\Gamma_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right)}{\gamma_u} < 0 \\ \dfrac{-H_f\left(\Gamma_u(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m)\right)}{\gamma_u} & \text{otherwise,} \end{cases} \end{aligned} \tag{31}$$

where $H_f(\Gamma_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right))$ is the Hessian of the SINR with respect to $\boldsymbol{f}$, which in turn is given by

$$\begin{aligned} &H_f\left(\Gamma_u(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m)\right) \\ &\triangleq \nabla^2\Gamma_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right) \\ &= \begin{bmatrix} \dfrac{\partial^2\Gamma_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right)}{\partial\boldsymbol{f}^*\partial\boldsymbol{f}^{\mathrm{T}}} & \dfrac{\partial^2\Gamma_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right)}{\partial\boldsymbol{f}^*\partial\boldsymbol{f}^{\mathrm{H}}} \\ \dfrac{\partial^2\Gamma_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right)}{\partial\boldsymbol{f}\partial\boldsymbol{f}^{\mathrm{T}}} & \dfrac{\partial^2\Gamma_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right)}{\partial\boldsymbol{f}\partial\boldsymbol{f}^{\mathrm{H}}} \end{bmatrix} \\ &= \overbrace{\frac{1}{\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m\boldsymbol{f} + \xi_u^2}\begin{bmatrix} \boldsymbol{\Phi}_f^m & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Phi}_f^{m\mathrm{T}} \end{bmatrix}}^{\triangleq\boldsymbol{q}_1} \\ &\quad - \overbrace{\frac{\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Phi}_f^m\boldsymbol{f}}{(\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m\boldsymbol{f} + \xi_u^2)^2}\begin{bmatrix} \boldsymbol{\Psi}_f^m & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi}_f^{m\mathrm{T}} \end{bmatrix}}^{\triangleq\boldsymbol{q}_2} \\ &\quad - \overbrace{\frac{2\begin{bmatrix} \boldsymbol{\Phi}_f^m\boldsymbol{f}\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m & \boldsymbol{\Phi}_f^m\boldsymbol{f}\boldsymbol{f}^{\mathrm{T}}\boldsymbol{\Psi}_f^{m\mathrm{T}} \\ \boldsymbol{\Phi}_f^{m\mathrm{T}}\boldsymbol{f}^*\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m & \boldsymbol{\Phi}_f^{m\mathrm{T}}\boldsymbol{f}^*\boldsymbol{f}^{\mathrm{T}}\boldsymbol{\Psi}_f^{m\mathrm{T}} \end{bmatrix}}{(\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m\boldsymbol{f} + \xi_u^2)^2}}^{\triangleq\boldsymbol{q}_3} \\ &\quad + \underbrace{\frac{2\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Phi}_f^m\boldsymbol{f}}{(\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m\boldsymbol{f} + \xi_u^2)^3}\begin{bmatrix} \boldsymbol{\Psi}_f^m\boldsymbol{f}\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m & \boldsymbol{\Psi}_f^m\boldsymbol{f}\boldsymbol{f}^{\mathrm{T}}\boldsymbol{\Psi}_f^{m\mathrm{T}} \\ \boldsymbol{\Psi}_f^{m\mathrm{T}}\boldsymbol{f}^*\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m & \boldsymbol{\Psi}_f^{m\mathrm{T}}\boldsymbol{f}^*\boldsymbol{f}^{\mathrm{T}}\boldsymbol{\Psi}_f^{m\mathrm{T}} \end{bmatrix}}_{\triangleq\boldsymbol{q}_4}. \end{aligned} \tag{32}$$

Given the relation between $\nabla^2\nu_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right)$ and $H_f(\Gamma_u\left(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m\right))$, as per equation (31), it follows from equations (30) and (32) that, for a given user, we have

$$\begin{aligned} L_{f_u}^* &= \frac{1}{\gamma_u}\lambda_{\max}(-H_f(\Gamma_u(\boldsymbol{f}|\boldsymbol{V}, \boldsymbol{h}_u^m))) \\ &\leq \frac{1}{\gamma_u}\big(\lambda_{\max}(-\boldsymbol{q}_1) + \lambda_{\max}(\boldsymbol{q}_2) + \lambda_{\max}(\boldsymbol{q}_3) + \lambda_{\max}(-\boldsymbol{q}_4)\big), \end{aligned} \tag{33}$$

where the latter inequality follows straightforwardly from the triangular inequality.

Note that $\boldsymbol{q}_1$ and $\boldsymbol{q}_4$ are both rank-1 positive semi-definite matrices, which implies that $\lambda_{\max}(-\boldsymbol{q}_1) = 0$ and $\lambda_{\max}(-\boldsymbol{q}_4) = 0$. In addition, for $\boldsymbol{q}_3$, we have

$$\begin{aligned}\lambda_{\max}(\boldsymbol{q}_3) &= \frac{2(\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Phi}_f^m\boldsymbol{\Psi}_f^m\boldsymbol{f} + \boldsymbol{f}^{\mathrm{T}}\boldsymbol{\Phi}_f^{m^{\mathrm{T}}}\boldsymbol{\Psi}_f^{m^{\mathrm{T}}}\boldsymbol{f}^*)}{(\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m\boldsymbol{f}+\xi_u^2)^2} \\ &= \frac{4\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Phi}_f^m\boldsymbol{\Psi}_f^m\boldsymbol{f}}{(\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m\boldsymbol{f}+\xi_u^2)^2} = 0, \end{aligned} \tag{34}$$

where we employed the fact that $\boldsymbol{\Phi}_f^m\boldsymbol{\Psi}_f^m = \boldsymbol{0}$.

Using these results in equation (33), we obtain

$$L_{f_u}^* \leq \frac{1}{\gamma_u}\lambda_{\max}(\boldsymbol{q}_2), \tag{35}$$

with

$$\begin{aligned}\lambda_{\max}(\boldsymbol{q}_2) &= \frac{\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Phi}_f^m\boldsymbol{f}}{(\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Psi}_f^m\boldsymbol{f}+\xi_u^2)^2}\lambda_{\max}(\boldsymbol{\Psi}_f^m) \\ &\leq \frac{\boldsymbol{f}^{\mathrm{H}}\boldsymbol{\Phi}_f^m\boldsymbol{f}}{\xi_u^4}\lambda_{\max}(\boldsymbol{\Psi}_f^m) \\ &= \frac{\mathrm{Tr}\left(\boldsymbol{V}\boldsymbol{f}_u\boldsymbol{f}_u^{\mathrm{H}}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{h}_u^m\boldsymbol{h}_u^{m\mathrm{H}}\right)\|\boldsymbol{h}_u^{m\mathrm{H}}\boldsymbol{V}\|_2^2}{\xi_u^4} \\ &\leq BN_tN_{\mathrm{RF}}\frac{\|\hat{\boldsymbol{h}}_u\|_2^4}{\xi_u^4}\sum_{b=1}^{B}P_{\max,b}, \end{aligned} \tag{36}$$

where the identity $\lambda_{\max}\left(\left[\begin{smallmatrix}\boldsymbol{A} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}^{\mathrm{T}}\end{smallmatrix}\right]\right) = \lambda_{\max}(\boldsymbol{A})$ and the bound $\mathrm{Tr}\,(\boldsymbol{AB}) \leq \lambda_{\max}(\boldsymbol{A})\mathrm{Tr}\,(\boldsymbol{B})$ combined with the trivial result $\mathrm{Tr}\left(\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}\right) = BN_tN_{\mathrm{RF}}$ were used in the first equation and the last inequality, respectively.

Combining equations (30), (35), and (36), we finally have

$$L_f^* \leq BN_tN_{\mathrm{RF}}\sum_{b=1}^{B}P_{\max,b}\cdot\sum_{u=1}^{U}\frac{\|\hat{\boldsymbol{h}}_u\|_2^4}{\gamma_u\xi_u^4}, \tag{37}$$

from which it follows that the learning rates α_i^f to be employed at each iteration of equation (19) in order to solve problem (15) and obtain the optimal digital baseband beamformer $\boldsymbol{f}$ is given by

$$\alpha_i^f = \rho\left(\sqrt{i}BN_tN_{\mathrm{RF}}\sum_{b=1}^{B}P_{\max,b}\cdot\sum_{u=1}^{U}\frac{\|\hat{\boldsymbol{h}}_u\|_2^4}{\gamma_u\xi_u^4}\right)^{-1}. \tag{38}$$

C. ANANLOG RF COMPONENT: LIPSCHITZ BOUND FOR V WITH f CONSTANT

Following steps similar to the above, the Hessian matrix with respect to $\tilde{\boldsymbol{v}}$ is given by

$$\begin{aligned}&H_v\left(\nu_u(\tilde{\boldsymbol{v}}|\boldsymbol{f},\boldsymbol{h}_u^m)\right) \\ &\triangleq \nabla^2\Gamma_u(\tilde{\boldsymbol{v}}|\boldsymbol{f},\boldsymbol{h}_u^m) \\ &= \begin{bmatrix}\dfrac{\partial^2\Gamma_u\left(\tilde{\boldsymbol{v}}|\boldsymbol{f},\boldsymbol{h}_u^m\right)}{\partial\tilde{\boldsymbol{v}}^*\partial\tilde{\boldsymbol{v}}^{\mathrm{T}}} & \dfrac{\partial^2\Gamma_u\left(\tilde{\boldsymbol{v}}|\boldsymbol{f},\boldsymbol{h}_u^m\right)}{\partial\tilde{\boldsymbol{v}}^*\partial\tilde{\boldsymbol{v}}^{\mathrm{H}}} \\ \dfrac{\partial^2\Gamma_u\left(\tilde{\boldsymbol{v}}|\boldsymbol{f},\boldsymbol{h}_u^m\right)}{\partial\tilde{\boldsymbol{v}}\partial\tilde{\boldsymbol{v}}^{\mathrm{T}}} & \dfrac{\partial^2\Gamma_u\left(\tilde{\boldsymbol{v}}|\boldsymbol{f},\boldsymbol{h}_u^m\right)}{\partial\tilde{\boldsymbol{v}}\partial\tilde{\boldsymbol{v}}^{\mathrm{H}}}\end{bmatrix} \end{aligned}$$

$$\begin{aligned}&= \overbrace{\frac{1}{\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}+\xi_u^2}\begin{bmatrix}\boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^{m^{\mathrm{T}}}\end{bmatrix}}^{\triangleq\boldsymbol{\ell}_1} \\ &\quad - \overbrace{\frac{\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}}{(\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}+\xi_u^2)^2}\begin{bmatrix}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^{m^{\mathrm{T}}}\end{bmatrix}}^{\triangleq\boldsymbol{\ell}_2} \\ &\quad - \overbrace{\frac{2\begin{bmatrix}\boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m & \boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^{\mathrm{T}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^{m^{\mathrm{T}}} \\ \boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^{m^{\mathrm{T}}}\tilde{\boldsymbol{v}}^*\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m & \boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^{m^{\mathrm{T}}}\tilde{\boldsymbol{v}}^*\tilde{\boldsymbol{v}}^{\mathrm{T}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^{m^{\mathrm{T}}}\end{bmatrix}}{(\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}+\xi_u^2)^2}}^{\triangleq\boldsymbol{\ell}_3} \\ &\quad + \underbrace{\frac{2\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}}{(\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}+\xi_u^2)^3}\begin{bmatrix}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m & \boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^{\mathrm{T}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^{m^{\mathrm{T}}} \\ \boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^{m^{\mathrm{T}}}\tilde{\boldsymbol{v}}^*\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m & \boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^{m^{\mathrm{T}}}\tilde{\boldsymbol{v}}^*\tilde{\boldsymbol{v}}^{\mathrm{T}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^{m^{\mathrm{T}}}\end{bmatrix}}_{\triangleq\boldsymbol{\ell}_4}, \end{aligned} \tag{39}$$

with $\lambda_{\max}(-\boldsymbol{\ell}_1) = \lambda_{\max}(-\boldsymbol{\ell}_4) = 0$, such that

$$L_{v_u}^* = \frac{1}{\gamma_u}\lambda_{\max}(-H_v(\Gamma_u(\tilde{\boldsymbol{v}}|\boldsymbol{f},\boldsymbol{h}_u^m)) \leq \frac{\lambda_{\max}(\boldsymbol{\ell}_2)+\lambda_{\max}(\boldsymbol{\ell}_3)}{\gamma_u}, \tag{40}$$

where

$$\lambda_{\max}(\boldsymbol{\ell}_2) = \frac{\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}}{(\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}+\xi_u^2)^2}\lambda_{\max}(\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m), \tag{41a}$$

$$\lambda_{\max}(\boldsymbol{\ell}_3) = \frac{4\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}}{(\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}+\xi_u^2)^2}. \tag{41b}$$

For the sake of future convenience, we introduce the following equalities:

$$\begin{aligned}&\boldsymbol{W}^{\mathrm{H}}\left(\boldsymbol{f}_u^{\mathrm{T}}\otimes\boldsymbol{h}_u^{m\mathrm{H}}\right)^{\mathrm{H}}\left(\boldsymbol{f}_u^{\mathrm{T}}\otimes\boldsymbol{h}_u^{m\mathrm{H}}\right)\boldsymbol{W} \\ &\quad = \boldsymbol{W}^{\mathrm{H}}\left(\boldsymbol{f}_u^*\boldsymbol{f}_u^{\mathrm{T}}\otimes\boldsymbol{h}_u^m\boldsymbol{h}_u^{m\mathrm{H}}\right)\boldsymbol{W} \end{aligned} \tag{42a}$$

$$\begin{aligned}&\boldsymbol{W}^{\mathrm{H}}\sum_{u'\in\mathcal{U}\setminus u}\left(\boldsymbol{f}_{u'}^{\mathrm{T}}\otimes\boldsymbol{h}_u^{m\mathrm{H}}\right)^{\mathrm{H}}\left(\boldsymbol{f}_{u'}^{\mathrm{T}}\otimes\boldsymbol{h}_u^{m\mathrm{H}}\right)\boldsymbol{W} \\ &\quad = \boldsymbol{W}^{\mathrm{H}}\sum_{u'\in\mathcal{U}\setminus u}\left(\boldsymbol{f}_u^*\boldsymbol{f}_{u'}^{\mathrm{T}}\otimes\boldsymbol{h}_u^m\boldsymbol{h}_u^{m\mathrm{H}}\right)\boldsymbol{W}, \end{aligned} \tag{42b}$$

where we utilized $(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D}) = (\boldsymbol{AC}\otimes\boldsymbol{BD})$.

Then, we obtain

$$\begin{aligned}\lambda_{\max}(\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m) &= \mathrm{Tr}\left(\boldsymbol{W}^{\mathrm{H}}\left(\boldsymbol{f}_u^*\boldsymbol{f}_u^{\mathrm{T}}\otimes\boldsymbol{h}_u^m\boldsymbol{h}_u^{m\mathrm{H}}\right)\boldsymbol{W}\right), \\ &\leq \sum_{b=1}^{B}P_{\max,b}\|\hat{\boldsymbol{h}}_u\|_2^2, \end{aligned} \tag{43}$$

and

$$\frac{\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}}{(\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}+\xi_u^2)^2} \leq \frac{BN_tN_{\mathrm{RF}}\sum_{b=1}^{B}P_{\max,b}}{\xi_u^4}\|\hat{\boldsymbol{h}}_u\|_2^2, \tag{44a}$$

$$\frac{4\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Phi}_{\tilde{\boldsymbol{v}}}^m\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}}{(\tilde{\boldsymbol{v}}^{\mathrm{H}}\boldsymbol{\Psi}_{\tilde{\boldsymbol{v}}}^m\tilde{\boldsymbol{v}}+\xi_u^2)^2} \leq \frac{4BN_tN_{\mathrm{RF}}(\sum_{b=1}^{B}P_{\max,b})^2}{\xi_u^4}\|\hat{\boldsymbol{h}}_u\|_2^4, \tag{44b}$$

Algorithm 1 Proposed Outage Minimum MSGD Cooperative MmWave Blockage-Robust Hybrid Beamformer

Parameters: Number of iterations I_{tot} and mini-batch size(s) M_i.
Input: Channel gain and angle estimates $g_{b,u}^k$, $\theta_{b,u}^k$ and $\phi_{b,u}^k$; corresponding uncertainty variance $\zeta_{b,u,k}^2$; received signals y_u; maximum transmit power $P_{\max,b}$ and target SINR $\gamma_u \; \forall \, u, b, k$
Output: Digital and analog beamformers $\boldsymbol{V}$ and $\boldsymbol{f}$.

1: Initialize $\boldsymbol{V}^{(0)}$ by phase matching to $\hat{\boldsymbol{h}}_{b,u}$.
2: Initialize $\boldsymbol{f}^{(0)}$ by phase matching to $\hat{\boldsymbol{h}}_{b,u}$.
3: **for** $i = 1, 2, \cdots I_{\text{tot}}$
4: Generate mini batch of M_i samples of channel vectors $\boldsymbol{h}_u^m \triangleq \left[\boldsymbol{h}_{1,u}^{m\text{T}}, \cdots, \boldsymbol{h}_{B,u}^{m\text{T}}\right]^{\text{T}}$, with $m = 1, \cdots, M_i$ and where each $\boldsymbol{h}_{b,u}^m$ is in likeness of equation (4).
5: Update $\boldsymbol{f}^{(i)}$ via equation (19), with analog beamformer $\boldsymbol{V}^{(i-1)}$ and learning rate α_i^f as in equation (38).
6: Update $\boldsymbol{V}^{(i)}$ via equations (29a) and (29b) with digital beamformer $\boldsymbol{f}^{(i)}$ and learning rate α_i^v as in equation (46).
7: **end for**
8: Retain last $\boldsymbol{V} = \boldsymbol{V}^{(i)}$ as the final analog beamformer.
9: Obtain final digital beamformer $\boldsymbol{f}$ via equation (19) with the last beamformer $\boldsymbol{V}$ and learning rate α_i^f as in equation (38).

which, when combined with equations (40) and (41) and ultimately substituted into equation (30), yield

$$L_v^* \leq 5BN_tN_{\text{RF}}\left(\sum_{b=1}^{B} P_{\max,b}\right)^2 \cdot \sum_u \frac{\|\hat{\boldsymbol{h}}_u\|_2^4}{\gamma_u \xi_u^4}. \tag{45}$$

Consequently, the learning rates α_i^v to be employed at each iteration of equation (29a) in order to solve problem (20) and obtain the optimal digital baseband beamformer $\boldsymbol{V}$ according to equation (29b) are given by

$$\alpha_i^v = \rho\left(\sqrt{i}5BN_tN_{\text{RF}}\left(\sum_{b=1}^{B} P_{\max,b}\right)^2 \cdot \sum_u \frac{\|\hat{\boldsymbol{h}}_u\|_2^4}{\gamma_u \xi_u^4}\right)^{-1}. \tag{46}$$

D. BLOCK-COORDINATE STOCHASTIC GRADIENT

Combining the results of Subsections IV-B, IV-C, III-B, and III-C, a complete scheme to solve the stochastic formulation of the mmWave cooperative hybrid blockage-robust beamforming design described originally in equation (7) is achieved, which is summarized in Algorithm 1.

Note that in order to control the variance of the gradient direction, we adopt an MSGD-type variance reduction technique, which updates the solution by taking the sample mean over a small fraction of the training dataset at each algorithmic iteration. With regards to generating the training dataset, we remark that the CSI error level $\zeta_{b,u,k}^2$ is typically several orders of magnitude smaller than its corresponding mean quantity $g_{b,u}^k$, which can also be systematically incorporated into the robust beamforming design in Algorithm 1.

In summary, the proposed algorithm exhibits robustness not only against random blockages but also against CSI errors, which to the best of our knowledge is a novel contribution that has not been presented before. Finally, the efficacy of the method is elucidated in the next section.

V. PERFORMANCE ASSESSMENT

In this section, we evaluate the sum-outage-minimizing hybrid robust mmWave beamforming method proposed above, comparing its performance against three fully digital schemes, namely, the OutMin technique of [28], the SRM method [42], and the classic maximum ratio transmission (MRT).

Since fully digital approaches assume that transmit beams are optimized without restrictions (employing a $BN_t \times U$ fully digital beamforming matrix to serve U users), the performances of fully digital methods serve as bounds on those of their hybrid counterparts.

In our simulations, a setup similar to that employed in the measurement campaign carried out for the 28 [GHz] band reported in [43] is considered, unless mentioned otherwise. In particular, a mmWave square microcell model, with sides 100 meters wide and with $B = 4$ APs each located at a corner of the square, is assumed, while $U = 2$ UEs are randomly placed within the square being served by the system. We set the heights of the APs and UEs to 15 [m] and 1.6 [m], respectively, and it is assumed that each AP is equipped with $N_t = 16$ transmit antennas but only $N_{\text{RF}} = 2$ digital RF chains, with the maximum transmit power constrained to $P_{\max,b} = 30$ [dBm].

The mmWave channel propagation model proposed in [6, Table 1] is utilized to characterize the path loss and the number of clusters, and the AWGN variance ξ_u^2 of each UE is set such that

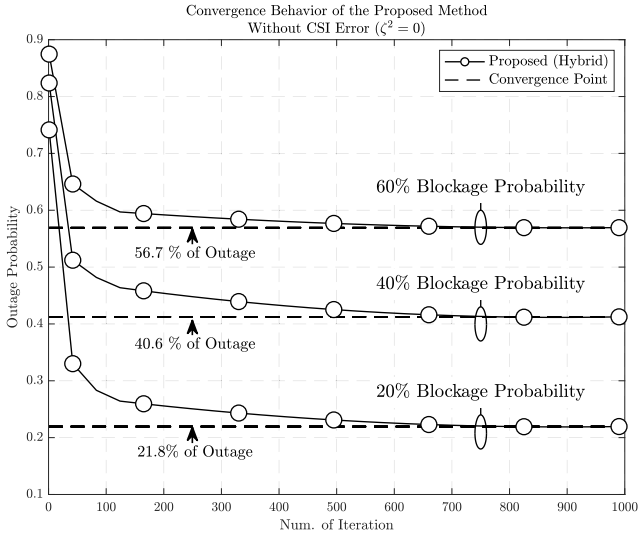
$$10\log_{10}\left(\xi_u^2\right) = 10\log_{10}\left(1000\kappa T\right) + 10\log_{10}\left(W\right) + \text{NF}, \tag{47}$$

where $\kappa \approx 1.38 \times 10^{-23}$ is the Boltzmann constant, $T = 293.15$ [K] is the physical temperature at each user location in Kelvins (*i.e.*, 20 degrees Celsius), $W = 100$ [MHz] is the subcarrier bandwidth, and NF $= 5$ [dB] is the standard noise figure corresponding to each UE, ultimately yielding $\xi_u^2 \approx -89$ [dBm].

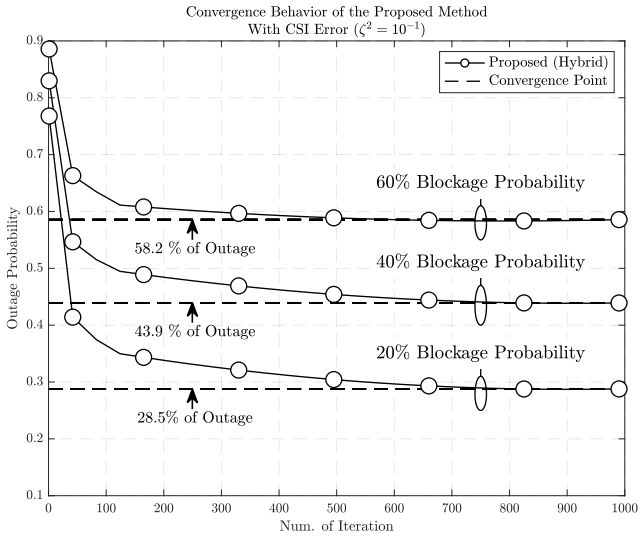
Finally, we set the target data rate to be either 3 or 5 [bits/s/Hz] depending on the user requirements, where γ_u is the desired SINR as defined in equation (7), while the mini-batch size $M_i = 16$.

A. CONVERGENCE BEHAVIOR

We start our assessment by evaluating the convergence behavior of the proposed hybrid beamforming method in terms of outage probability, which is the objective cost



(a) Without CSI Error



(b) With CSI Error

FIGURE 2. Convergence behavior of the proposed method for different channel conditions.

function of interest in this study. For the sake of simplicity and illustration, we assume that the blockage probabilities ($p^k_{b,u}\ \forall\ m, b, u$) are identical, regardless of the superscripts and subscripts, *i.e.*, $p^k_{b,u} = p\ \forall\ m, b, u$, and p is assumed to vary from 20% to 60% as suggested by the findings in [8], [10]. With that in mind, three blockage scenarios are investigated in the simulations, namely, 20%, 40%, and 60%, which are referred to as moderate, severe, and critical blockage conditions, respectively. In turn, the CSI uncertainty $\zeta^2_{b,u,k}$ is considered to be proportional to the corresponding average small-scale fading coefficient, that is, $\zeta^2_{b,u,k} = \zeta^2 \cdot g^k_{b,u}$, where ζ^2 is a parameter that controls the CSI accuracy level.

The first set of results is presented in Figure 2, which depicts the convergence of the proposed iterative hybrid beamforming design for blockage-robust CoMP mmWave systems, as a function of the number of algorithmic iterations, for the three distinct blockage scenarios described above. It is found that regardless of the severity of the scenario in terms of blockage probability, the proposed method quickly learns the crucial blockage patterns and converges to a stationary point, not only under the ideal assumption of perfect CSI typically adopted in related literature, as shown in Figure 2a, but also under the more realistic assumption that fluctuation of the small-scale fading coefficients takes place (*i.e.*, imperfect CSI), as shown in Figure 2b. In fact, in both cases (perfect or imperfect CSI), and under all blockage scenarios (moderate, severe, and critical), the number of iterations required for the proposed beamformer to converge is approximately the same, which demonstrates the remarkable robustness of the method.

Finally, it is also observed (interestingly but as expected) that the outage performance of the proposed method is dominated by probabilities of random path blockages themselves, rather than by the CSI error variances. This fact will be elucidated further in the next subsection.

B. STATISTICAL ANALYSIS OF THROUGHPUT

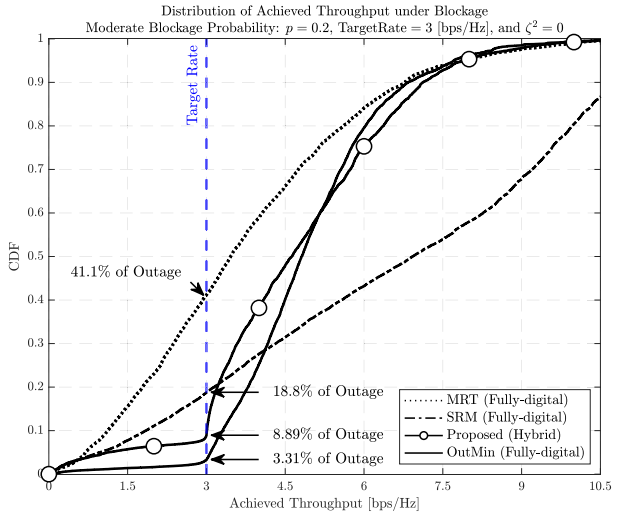
Next, we numerically analyze, via Monte Carlo simulations, the statistical behavior of the four beamforming methods considered herein—the proposed design, MRT, OutMin method of [28], and SRM method—in terms of their achievable throughput and outage probability under different channel conditions.

We again emphasize that the fully digital versions of the MRT, OutMin, and SRM schemes are simulated such that the corresponding performances should be considered performance upper-bounds for their hybrid counterparts. Finally, to demonstrate that the advantage of the proposed method extends beyond average performance, all results are offered in the form of throughput cumulative density functions (CDFs).

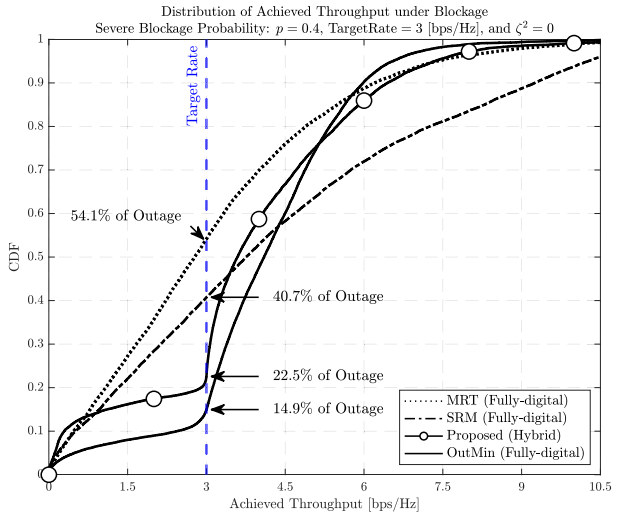
In Figures 3 and 4, the CDFs of the achievable throughput obtained by the four beamforming methods, with target rates of 3 and 5 [bps/Hz], respectively, are shown for different blockage scenarios and under perfect CSI, while Figures 5 and 6 display equivalent results obtained under CSI uncertainty. Since the bandwidth of the system is $W = 100$ [MHz], these target rates imply that in the cases corresponding to Figures 3 and 5, the system is set to ideally serve each user with at least 300 [Mbps], while in Figures 4 and 6, the system aims to serve each user with at least 500 [Mbps], respectively.

In all figures, lines without markers are used for the SotA methods, while a solid line with a white circular marker indicates the proposed method. For readability, the target throughput is highlighted in the figures by a blue vertical line annotated with the text "Target Rate". In addition, the outage probability achieved by each beamforming method — as defined in equation (48) — is marked by an arrow also annotated with the corresponding numerical value.

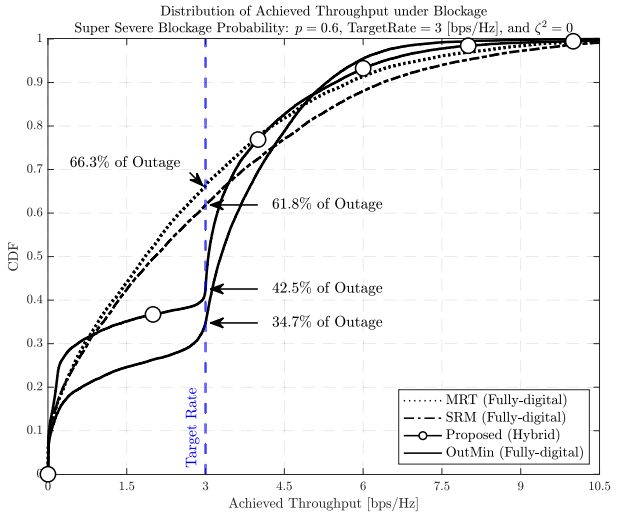
$$\Pr\left\{\log_2(1+\Gamma_u(\boldsymbol{f},\boldsymbol{V}|\boldsymbol{h}_u)) < \log_2(1+\gamma_u)\right\}. \quad (48)$$



(a) 20% of Blockage Probability

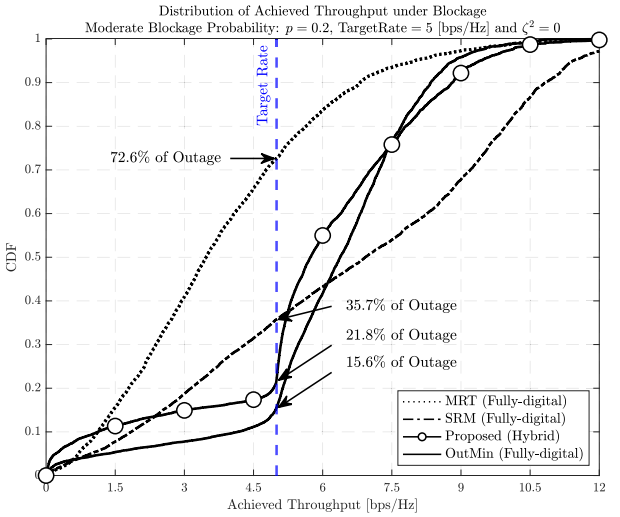


(b) 40% of Blockage Probability

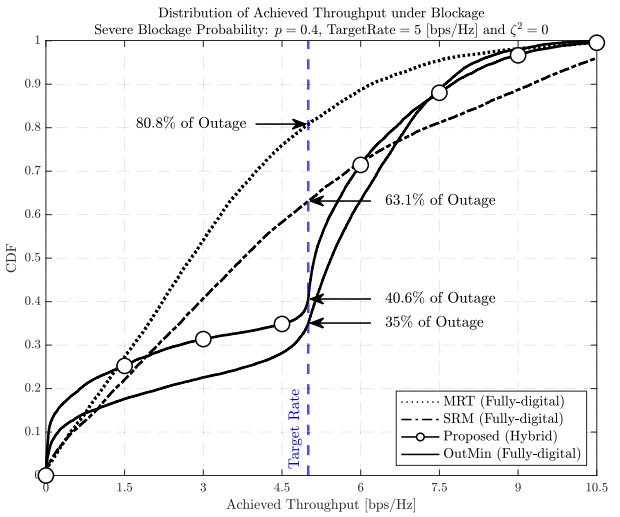


(c) 60% of Blockage Probability

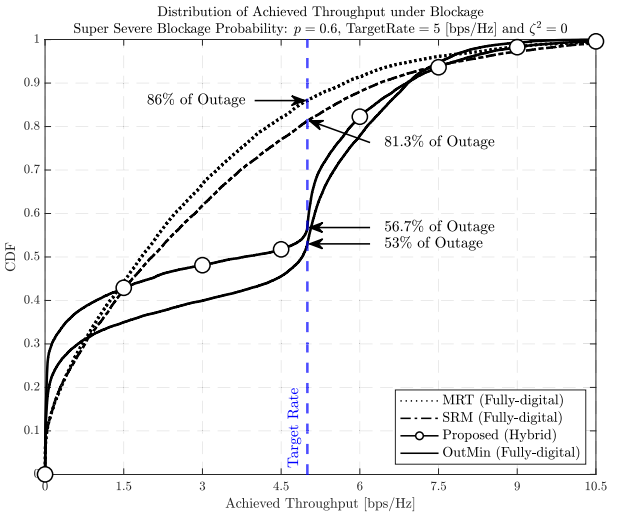
FIGURE 3. CDF of achieved data rates for different blockage probabilities with target rate of 3 [bps/Hz] and perfect CSI.



(a) 20% of Blockage Probability

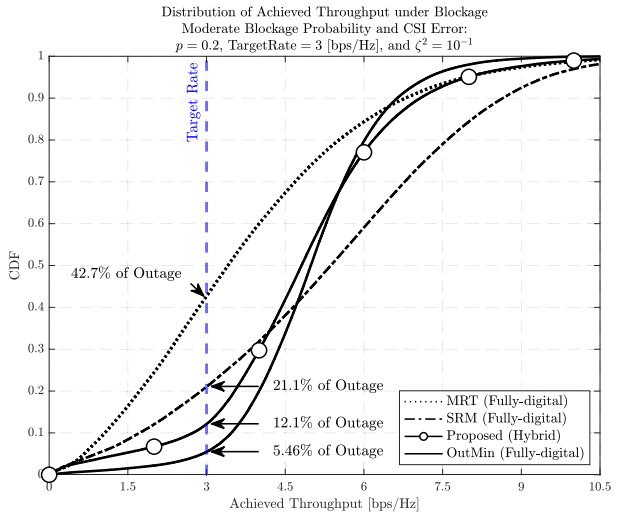


(b) 40% of Blockage Probability

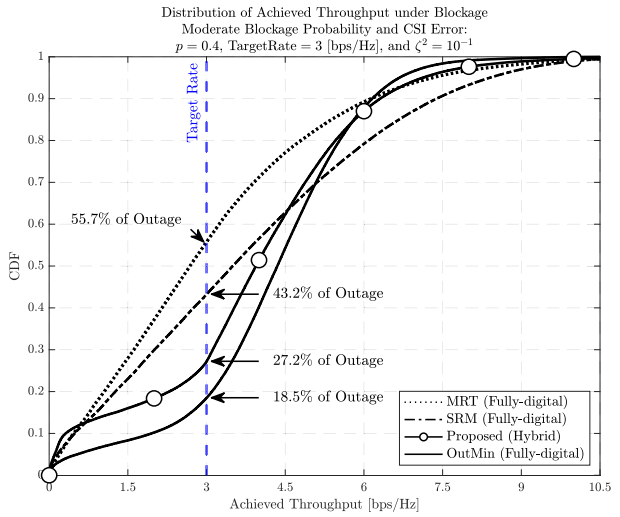


(c) 60% of Blockage Probability

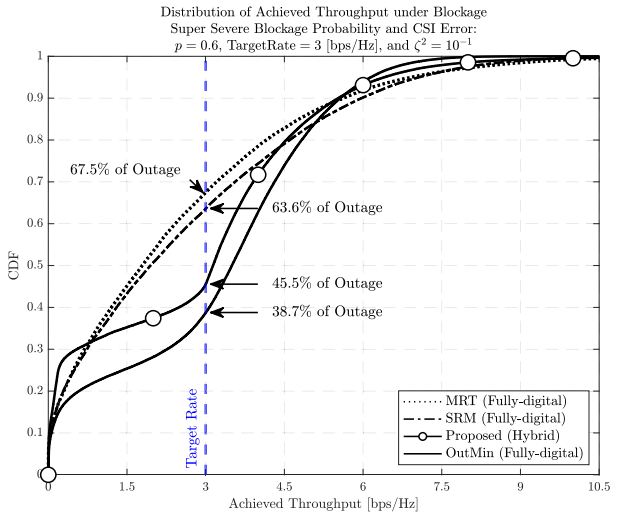
FIGURE 4. CDF of achieved data rates for different blockage probabilities with target rate of 5 [bps/Hz] and perfect CSI.



(a) 20% of Blockage Probability

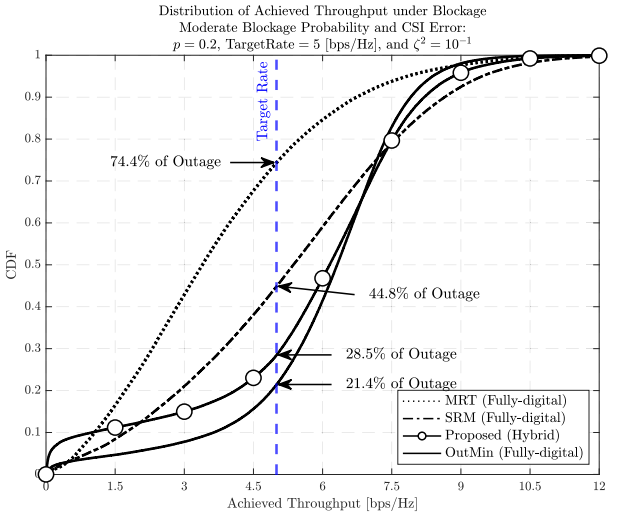


(b) 40% of Blockage Probability

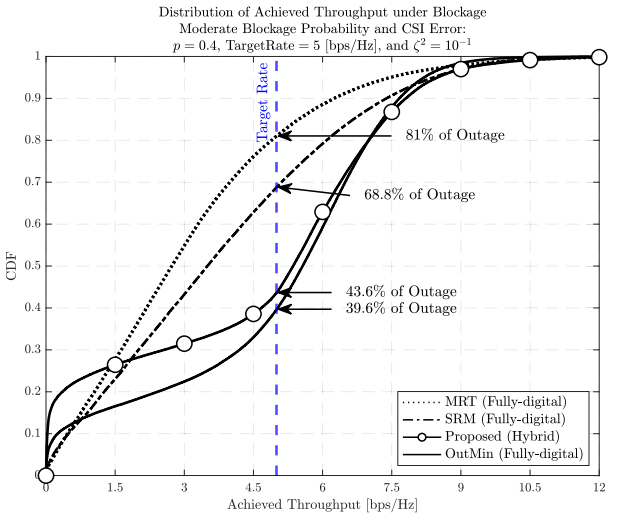


(c) 60% of Blockage Probability

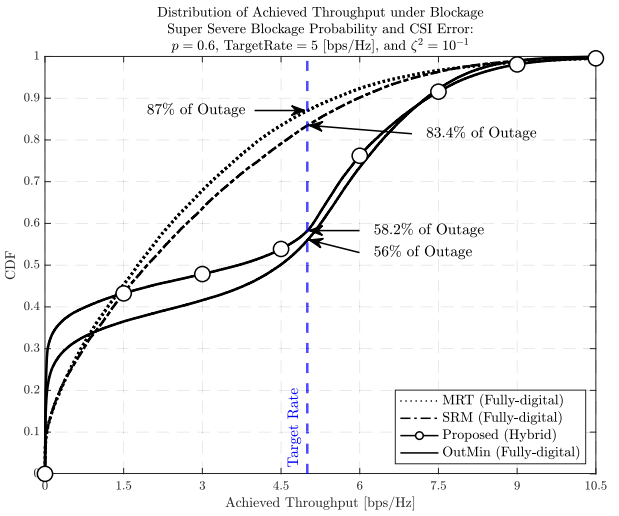
FIGURE 5. CDF of achieved data rates for different blockage probabilities with target rate of 3 [bps/Hz] and CSI errors.



(a) 20% of Blockage Probability



(b) 40% of Blockage Probability



(c) 60% of Blockage Probability

FIGURE 6. CDF of achieved data rates for different blockage probabilities with target rate of 5 [bps/Hz] and CSI errors.

In all the figures, the proposed hybrid method achieves a lower outage probability than the fully digital MRT and SRM beamformers, regardless of CSI quality and blockage conditions. Furthermore, the new technique is found to consistently approach the performance of the also-fully-digital OutMin method. Taking into account that the proposed hybrid scheme makes use of only $N_{\mathrm{RF}} = 2$ RF chains per AP, as opposed to fully digital methods which require all $N_{\mathrm{RF}} = N_t = 16$ RF chains per AP, the results demonstrate the remarkable effectiveness of the proposed hybrid method in combatting path blockage with significant potential to reduce hardware costs (by alleviating RF chain requirements) without compromising outage performance.

As for the impact of CSI errors, a comparison of Figures 3 and 4 against Figures 5 and 6 indicates that the overall impact of CSI errors is to retract some sharpness from the outage minimization approaches in enforcing the prescribed target rate. Even under such an effect, however, both the relative gain over the MRT and SRM beamformers, as well as the proximity in performance of the proposed scheme to the ideal OutMin method, remain mostly unchanged.

To cite a few examples, both under perfect and imperfect CSI, the proposed hybrid CoMP beamformer, under 20% path blockage probability and a target throughput of 3 [bps/Hz], achieves an outage reduction of approximately 30% and 10% over the MRT and SRM beamformers, respectively (see Figures 3a and 5b). Similarly, at 60% blockage probability and a target rate of 5 [bps/Hz], the new hybrid scheme outperforms the MRT and SRM methods by approximately 30% and 25% in terms of the reduction of outage probability, respectively, both under perfect and imperfect CSI conditions (see Figures 4c and 6c).

Thus, the results demonstrate that the proposed SGD approach for cooperative hybrid beamforming aimed at minimizing outage in mmWave systems subjected to path blockage and CSI imperfections is highly effective.

An interesting and final conclusion that can be drawn from the comparison of the results shown in Figures 3 through 6 is that among the considered methods, the MRT, OutMin, and new hybrid techniques are in fact all somewhat robust to CSI imperfections, whereas the SRM schemes seems to be the most sensitive to the quality of channel estimates. This is somewhat unsurprising, since the SRM method is not designed to minimize outage, but rather to maximize the achievable rate, such that one could argue that the aforenoted comparison is "unfair" to that particular approach. Scrutinizing this conclusion is a worthy exercise, nevertheless, as it provides insight on the conceptual question of whether rate maximization is the most suitable figure of merit in the optimization of mmWave systems, which in practice needs to be carefully chosen depending on the system objectives and requirements, as implied by Figures 3 through 6.

To that end, we compare in Figure 7 the effective throughput (*i.e.*, the average throughput at and above the target rate)

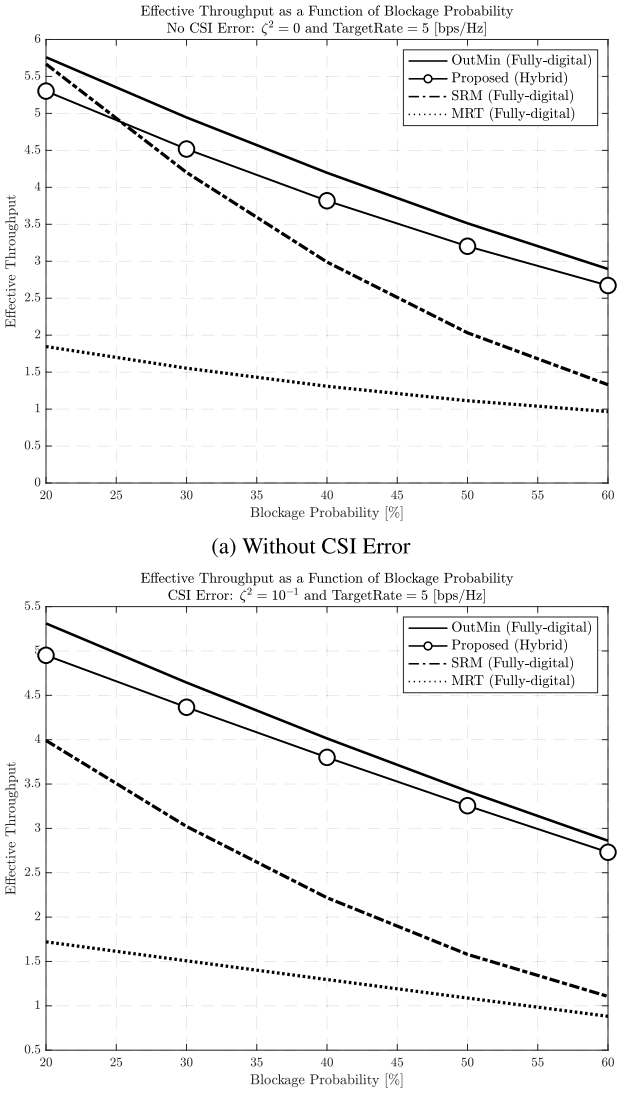


FIGURE 7. Effective throughput as a function of the blockage probabilities, with and without CSI uncertainty.

achieved by each of the considered beamforming schemes, as a function of the path blockage probability and for both evaluated channel estimation conditions, namely, perfect CSI ($\zeta = 0$) and imperfect CSI with $\zeta = 10^{-1}$. Following the related literature, the effective throughput is defined as $\mathbb{E}[\log_2(1+x)]$, where $x = \Gamma_u(\boldsymbol{f}, \boldsymbol{V}|\boldsymbol{h}_u)$ if $\Gamma_u(\boldsymbol{f}, \boldsymbol{V}|\boldsymbol{h}_u) \geq \gamma_u$ and $x = 0$ otherwise.

It is found, once again, that the proposed hybrid beamforming scheme performs only slightly worse than its fully digital couterpart [28], outperforming the other SotA methods, regardless of the CSI conditions and blockage probabilities. The only exception is the case of perfect CSI for blockage probabilities below 25%, where it is found that the SRM scheme outperforms the proposed method. It must be emphasized, however, that: a) the SRM scheme is fully digital, employing a total of $N_t \times B = 16 \times 4 = 64$ RF chains to serve $U = 2$ users, in contrast to the proposed hybrid scheme,

which requires only $N_{\mathrm{RF}} \times B = 2 \times 4 = 8$ RF chains; and b) that the SRM scheme is designed to maximize effective throughput, paying to that end the price of sacrificing overall QoS by allowing higher outage probabilities (see *e.g.*, Figure 4a where it is shown that the SRM beamformer with perfect CSI under 20% blockage probability leads to outages above 35%).

By contrast, when subjected to CSI errors, even that eventual localized "advantage" of the SRM approach is lost. In fact, as shown in Figure 7b, the SRM approach proves the most sensitive to both path blockages and CSI errors, exhibiting a significantly higher degree of degradation in performance than all other methods.

In summary, the results demonstrate collectively that the proposed method offers a competitive approach to mitigate the path blockage challenge in mmWave systems; notably, it also possesses robustness against CSI imperfections and exhibits little sacrifice in performance, compared to the fully digital OutMin approach, for the significant potential reduction in hardware cost due to its hybrid architecture.

VI. CONCLUSION

Aiming at the minimization of sum-outage in mmWave systems subjected to both path blockages and imperfect CSI, we proposed a new hybrid CoMP beamforming algorithm.

The proposed design is based on an empirical risk minimization stochastic learning problem formulation, solved through the alternate iteration of baseband digital and RF analog Riemann-manifold-constrained beamforming subproblems via a mini-batch stochastic gradient descent approach, with gradient-minimizing update rules given in closed form; further, learning rates are designed via lower bounds of the corresponding Lipschitz constants.

In contrast with previously proposed solutions, in which perfect CSI is assumed [26]–[29], our method relies on conventional initial estimates of the channels' angles of departure and complex gains, and on statistical knowledge of the corresponding blockages [6]–[8], [10], [11] and estimation error probabilities [16]–[18]. The channel uncertainties, including path blockages and imperfect CSI, are captured jointly in a newly introduced Bernoulli-Gaussian model, which is used to generate the training data for the MSGD-based optimizer, altogether resulting in a stochastic learning beamforming solution that is robust to both types of impairments.

Unlike preceding contributions, which employ either digital or fully connected hybrid beamforming architectures, the proposed scheme employs a virtually configured partially connected hybrid beamforming architecture, having therefore the advantage of being suitable for CoMP systems. This advantage is aligned with the existing theoretical findings [19], [20] that illustrate that the CoMP approach can achieve the full potential (*i.e.*, capacity-approaching rates and low outages) of mmWave systems in the presence of path blockages.

Simulation results were offered to confirm the effectiveness of our MSGD-based robust hybrid CoMP mmWave beamformer in mitigating the effects of path blockages and CSI errors.

Lastly, we point out that since mmWave CoMP systems might be required to serve extremely high data rates, wideband technology needs to be considered in order to meet such requirements. When considering wideband transmissions, one of the main bottlenecks is known to be beam squint, which may degrade the communication performance. Given the above, a wideband extension taking into account path blockages, CSI errors, and beam squint is left for a future work.

REFERENCES

[1] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "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.

[2] A. N. Uwaechia and N. M. Mahyuddin, "A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: Feasibility and challenges," *IEEE Access*, vol. 8, pp. 62367–62414, 2020.

[3] R.-A. Stoica, H. Iimori, G. T. F. de Abreu, and K. Ishibashi, "Frame theory and fractional programming for sparse recovery-based mmWave channel estimation," *IEEE Access*, vol. 7, pp. 150757–150774, 2019.

[4] A. Dogra, R. K. Jha, and S. Jain, "A survey on beyond 5G network with the advent of 6G: Architecture and emerging technologies," *IEEE Access*, vol. 9, pp. 67512–67547, 2021.

[5] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[6] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[7] 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)*, Singapore, Dec. 2017, pp. 1–7.

[8] 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)*, Honolulu, HI, USA, Dec. 2019, pp. 1–6.

[9] H. Iimori, G. T. F. De Abreu, T. Hara, K. Ishibashi, R.-A. Stoica, D. Gonzalez G, and O. Gonsa, "Robust symbol detection in large-scale overloaded NOMA systems," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 512–533, 2021.

[10] V. Raghavan, L. Akhoondzadeh-Asl, V. Podshivalov, J. Hulten, M. A. Tassoudji, O. H. Koymen, A. Sampath, and J. Li, "Statistical blockage modeling and robustness of beamforming in millimeter-wave systems," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 3010–3024, Jul. 2019.

[11] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[12] 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.

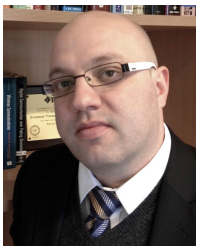
[13] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.

[14] R. Magueta, D. Castanheira, A. Silva, R. Dinis, and A. Gameiro, "Hybrid multi-user equalizer for massive MIMO millimeter-wave dynamic subconnected architecture," *IEEE Access*, vol. 7, pp. 79017–79029, 2019.

[15] A. M. Elbir and K. V. Mishra, "Joint antenna selection and hybrid beamformer design using unquantized and quantized deep learning networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1677–1688, Mar. 2020.

[16] Z. Luo, H. Liu, Y. Li, H. Wang, and L. Zhang, "Robust hybrid transceiver design for AF relaying in millimeter wave systems under imperfect CSI," *IEEE Access*, vol. 6, pp. 29739–29746, 2018.
[17] H. Iimori, G. T. F. de Abreu, and G. C. Alexandropoulos, "MIMO beamforming schemes for hybrid SIC FD radios with imperfect hardware and CSI," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4816–4830, Oct. 2019.
[18] A. M. Elbir and A. K. Papazafeiropoulos, "Hybrid precoding for multiuser millimeter wave massive MIMO systems: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 552–563, Jan. 2020.
[19] G. R. MacCartney and T. S. Rappaport, "Millimeter-wave base station diversity for 5G coordinated multipoint (CoMP) applications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3395–3410, Jul. 2019.
[20] B. Maham and P. Popovski, "Capacity analysis of coordinated multipoint reception for mmWave uplink with blockages," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16299–16303, Dec. 2020.
[21] A. Ali, N. González-Prelcic, and R. W. Heath, Jr., "Millimeter wave beam-selection using out-of-band spatial information," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1038–1952, Feb. 2018.
[22] 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, Sep. 2020.
[23] T. Nishio, H. Okamoto, K. Nakashima, Y. Koda, K. Yamamoto, M. Morikura, Y. Asai, and R. Miyatake, "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, Nov. 2019.
[24] Y. Koda, J. Park, M. Bennis, K. Yamamoto, T. Nishio, M. Morikura, and K. Nakashima, "Communication-efficient multimodal split learning for mmWave received power prediction," *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1284–1288, Jun. 2020.
[25] S. Mihara, S. Ito, T. Murakami, and H. Shinbo, "Positioning for user equipment of a mmWave system using RSSI and stereo camera images," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Nanjing, China, Mar. 2021, pp. 1–7.
[26] M. Gao, B. Ai, Y. Niu, W. Wu, P. Yang, F. Lyu, and X. Shen, "Efficient hybrid beamforming with anti-blockage design for high-speed railway communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 9643–9655, Sep. 2020.
[27] D. Kumar, J. Kaleva, and A. Tolli, "Blockage-aware reliable mmWave access via coordinated multi-point connectivity," *IEEE Trans. Wireless Commun.*, early access, Feb. 11, 2021, doi: 10.1109/TWC.2021.3057227.
[28] H. Iimori, G. T. F. de Abreu, O. Taghizadeh, R.-A. Stoica, T. Hara, and K. Ishibashi, "Stochastic learning robust beamforming for millimeter-wave systems with path blockage," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1557–1561, Sep. 2020.
[29] G. Zhou, C. Pan, H. Ren, K. Wang, M. Elkashlan, and M. D. Renzo, "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.
[30] Q. Sultan, M. S. Khan, and Y. S. Cho, "Fast 3D beamforming technique for millimeter-wave cellular systems with uniform planar arrays," *IEEE Access*, vol. 8, pp. 123469–123482, 2020.
[31] S. Sun, G. R. MacCartney, and T. S. Rappaport, "Millimeter-wave distance-dependent large-scale propagation measurements and path loss models for outdoor and indoor 5G systems," in *Proc. 10th Eur. Conf. Antennas Propag. (EuCAP)*, Davos, Switzerland, Apr. 2016, pp. 1–5.
[32] G. R. MacCartney and T. S. Rappaport, "Rural macrocell path loss models for millimeter wave wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1663–1677, Jul. 2017.
[33] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, "Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 5893–5908, Dec. 2019.
[34] T. Hara, H. Iimori, and K. Ishibashi, "Hyperparameter-free receiver for grant-free NOMA systems with MIMO-OFDM," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 810–814, Apr. 2021.
[35] Y. Xu and W. Yin, "Block stochastic gradient iteration for convex and nonconvex optimization," *SIAM J. Optim.*, vol. 25, no. 3, pp. 1686–1716, Jan. 2015.
[36] H. Masnadi-Shirazi and N. Vasconcelos, "On the design of loss functions for classification: Theory, robustness to outliers, and savageboost," in *Proc. Int. Conf. Neural Info. Process. (NIPS)*, 2008, pp. 1049–1056.
[37] P. L. Bartlett and M. H. Wegkamp, "Classification with a reject option using a hinge loss," *J. Mach. Learn. Res.*, vol. 9, pp. 1823–1840, Jun. 2008.
[38] E. Bottou, "Stochastic gradient learning in neural networks," in *Proc. Neuro-Nimes*, 1991, pp. 1–12.
[39] R. Johnson and T. Zhang, "Accelerating stochastic gradient descent using predictive variance reduction," in *Proc. Int. Conf. Neural Info. Process. (NIPS)*, 2013, pp. 315–323.
[40] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[41] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, Sep. 1951.
[42] O. Taghizadeh, V. Radhakrishnan, A. C. Cirik, R. Mathar, and L. Lampe, "Hardware impairments aware transceiver design for bidirectional full-duplex MIMO OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7450–7464, Aug. 2018.
[43] C. U. Bas, R. Wang, S. Sangodoyin, S. Hur, K. Whang, J. Park, J. Zhang, and A. F. Molisch, "28 GHz microcell measurement campaign for residential environment," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.

HIROKI IIMORI (Graduate Student Member, IEEE) received the B.Eng. degree in electrical and electronic engineering and the M.Eng. degree (Hons.) in advanced electrical, electronic, and computer systems from Ritsumeikan University, Kyoto, Japan, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Jacobs University Bremen, Germany. He was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. He was a recipient of the YKK Doctoral Fellowship awarded by the Yoshida Scholarship Foundation, Japan. His research interests include optimization theory, wireless communications, and signal processing.

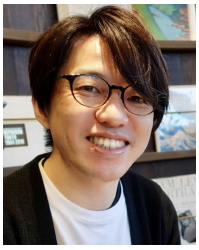


GIUSEPPE THADEU FREITAS DE ABREU (Senior Member, IEEE) received the B.Eng. degree in electrical engineering and the Latu Sensu degree in telecommunications engineering from the Universidade Federal da Bahia (UFBa), Salvador, Brazil, in 1996 and 1997, respectively, and the M.Eng. and D.Eng. degrees in physics, electrical, and computer engineering from Yokohama National University, Japan, in March 2001 and March 2004, respectively. From 2004 to 2006, he was a Postdoctoral Fellow with the Department of Electrical and Information Engineering, University of Oulu, Finland, where he was an Adjunct Professor (Docent) on statistical signal processing and communications theory, from 2006 to 2011. Since 2011, he has been a Professor of electrical engineering with Jacobs University Bremen, Germany. From April 2015 to August 2018, he held a full professorship position at the Department of Computer and Electrical Engineering, Ritsumeikan University, Japan. His research interest includes communications and signal processing, including communications theory, estimation theory, statistical modeling, wireless localization, cognitive radio, wireless security, MIMO systems, ultra-wideband and millimeter wave communications, full-duplex and cognitive radio, compressive sensing, energy harvesting networks, random networks, and connected vehicles networks. In 2000, he received the Uenohara Award by Tokyo University for his master's thesis work. He was a co-recipient of best paper awards at several international conferences. He was also awarded the prestigious JSPS Fellowship, in 2010; the Heiwa Nakajima Fellowship, in 2013; and the NICT Fellowship, in 2015. He served as an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, from 2009 to 2014; and IEEE TRANSACTIONS ON COMMUNICATIONS, from 2014 to 2017. He currently serves as an Executive Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

OMID TAGHIZADEH (Member, IEEE) received the M.Sc. degree in communications and signal processing from the Ilmenau University of Technology, Ilmenau, Germany, and the Ph.D. degree from RWTH Aachen University, Aachen, Germany. He is currently a Research Associate with the Network Information Theory Group, Technical University of Berlin, and with the 5G Wireless Research Group, Lenovo Deutschland GmbH. His research interests include full-duplex wireless systems, MIMO communications, machine learning, optimization, and resource allocation in wireless networks.

RĂZVAN-ANDREI STOICA (Member, IEEE) received the B.Sc. degree in electrical and computer engineering, the M.Sc. degree in communications, systems, and electronics, and the Ph.D. degree *(summa cum laude)* in electrical engineering from Jacobs University Bremen, Germany, in 2014, 2016, and 2019, respectively. He is currently with the 5G Wireless Group, Lenovo Research. His research interests include communications systems, signal processing, statistical learning, and convex optimization.

TAKANORI HARA (Graduate Student Member, IEEE) received the B.E. and M.E. degrees in engineering from The University of Electro-Communications, Tokyo, Japan, in 2017 and 2019, respectively, where he is currently pursuing the Ph.D. degree. His current research interests include communication theory, channel coding, grant-free access, and information theory.



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 with the Department of Electrical and Electronic Engineering, Shizuoka University, Hamamatsu, Japan. From 2010 to 2012, he was a Visiting Scholar with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. Since April 2012, he has been with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, where he is currently a Professor. His current research interests include grant-free access, cell-free architecture, millimeter-wave communications, energy harvesting communications, wireless power transfer, channel codes, signal processing, and information theory.