

Pareto Optimal Hybrid Beamforming for Short-Packet Millimeter-Wave Integrated Sensing and Communication

Jitendra Singh, Banda Naveen, *Member, IEEE*, Suraj Srivastava, Aditya K. Jagannatham, *Senior Member, IEEE* and Lajos Hanzo, *Life Fellow, IEEE*

Abstract—Pareto optimal solutions are conceived for radar beamforming error (RBE) and sum rate maximization in short-packet (SP) millimeter-wave (mmWave) integrated sensing and communication (ISAC). Our ultimate goal is to realize ultra-reliable low-latency communication (uRLLC) and real-time sensing capabilities for 6G applications. The ISAC base station (BS) transmits short packets in the downlink (DL) to serve multiple communication users (CUs) and detect multiple radar targets (RTs). We investigate the performance trade-off between the sensing and communication capabilities by optimizing both the radio frequency (RF) and the baseband (BB) transmit precoder (TPC), together with the block lengths. The optimization problem considers the minimum rate requirements of the CUs, the maximum tolerable radar beamforming error (RBE) for the RTs, the unit modulus (UM) elements of the RF TPC, and the finite transmit power as the constraints for SP transmission. The resultant problem is highly non-convex due to the intractable rate expression of the SP regime coupled with the non-convex rate and UM constraints. To solve this problem, we propose an innovative two-layer bisection search (TLBS) algorithm, wherein the RF and BB TPCs are optimized in the inner layer, followed by the block length in the outer layer. Furthermore, a pair of novel methods, namely a bisection search-based majorizer and minimizer (BMM) as well as exact penalty-based manifold optimization (EPMO) are harnessed for optimizing the RF TPC in the inner layer. Subsequently, the BB TPC and the block length are derived via second-order cone programming (SOCP) and mixed integer programming methods, respectively. Finally, our exhaustive simulation results reveal the effect of system parameters for various settings on the RBE-rate region of the SP mmWave ISAC system and demonstrate a significantly enhanced performance compared to the benchmarks.

Index Terms—Ultra-reliable low latency communication, integrated sensing and communication, hybrid beamforming, short packet communication, Pareto boundary.

I. INTRODUCTION

NEXT-generation (NG) wireless networks aim for providing ultra-reliable low-latency connectivity (uRLLC) with

high data rate and precise sensing. This supports demanding applications such as smart grids, industrial automation, autonomous vehicles, and mission-critical communication [1], [2]. Short packet communication (SPC) is a key enabler for realizing uRLLC. However, the intractable expression of the achievable rate corresponding to the finite block length and the decoding error probability requirements in the SPC regime complicates the analysis. As a result, beamforming optimization problems (OPs) in SPC-aided wireless systems become intractable [3], [4].

Recently, integrated sensing and communication (ISAC) in conjunction with millimeter wave (mmWave) technology has gained significant attention due to its excellent ability to provide sensing and communication (SC) capabilities in NG networks [5], [6]. The similarity between the channel characteristics and signal processing tasks encountered both in sensing and communication pave the way for their integration in the existing cellular infrastructure while necessitating only moderate hardware changes [7], [8]. Moreover, the conventional architecture demands a dedicated radio frequency (RF) chain (RFC) for each antenna element, which is a prohibitive requirement. In contrast, the hybrid beamforming (HBF) approach, requiring significantly fewer RFCs, provides a practical and viable solution for the implementation of mmWave multiple-input and multiple-output (MIMO) systems [9], [10]. More specifically, in the HBF scheme, the transmit precoder (TPC) is split into baseband (BB) and RF TPCs, with the RF TPC implemented using digitally controllable phase shifters. A large number of phase shifters enable high-dimensional RF TPCs to counter severe path loss at mmWave bands, while a small number of RFCs support low-dimensional BB TPC for advanced multiplexing and multiuser techniques. Moreover, to study the SC trade-off in ISAC mmWave systems, Pareto optimization-based beamformer design is the ideal method of analyzing the optimal boundary of the SC performance [11]–[13].

However, SPC transmission must be harnessed for supporting uRLLC services in industrial automation, autonomous vehicles, and mission-critical communication, where real-time communication and sensing play a vital role. In a similar fashion, smart cities can leverage uRLLC and ISAC to optimize traffic management systems, ensuring prompt responses to fluctuating traffic conditions and thus enhancing overall urban mobility. Therefore, ISAC requires an additional layer of intelligence for combining the sensing capabilities with uRLLC services via mmWave communication. Thus, the large number of compelling applications gives rise to an increased number of systems requiring uRLLC communications combined with accurate and robust sensing capabilities [14], [15]. Inspired by these trends, we investigate SPC-enabled mmWave

The work of Aditya K. Jagannatham was supported in part by the Qualcomm Innovation Fellowship; in part by the Qualcomm 6G UR Gift; and in part by the Arun Kumar Chair Professorship.

The work of S. Srivastava was supported by IIT Jodhpur's Research Grant No. I/RIG/SUS/20240043.

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council (EPSRC) projects under grant EP/Y037243/1, EP/W016605/1, EP/X01228X/1, EP/Y026721/1, EP/W032635/1, EP/Y037243/1 and EP/X04047X/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028).

J. Singh B. Naveen, and A. K. Jagannatham are with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur, UP 208016, India (e-mail: jitend@iitk.ac.in; naveenb22@iitk.ac.in; adityaj@iitk.ac.in).

S. Srivastava is with the Department of Electrical Engineering, Indian Institute of Technology Jodhpur, Jodhpur, Rajasthan 342030, India (email: surajsri@iitj.ac.in).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

ISAC, which has the potential of significantly improving the overall performance of wireless networks. The optimization of the hybrid beamformer, along with the block length, plays a crucial role in supporting uRLLC services for communication users (CUs) and sensing for the radar targets (RTs) in SPC-enabled mmWave ISAC systems. Specifically, we characterize the trade-off between the sensing and communication tasks via the Pareto optimization of hybrid beamformers and the block length of an SPC-aided mmWave system. To the best of our knowledge, this is the first paper exploring the paradigm of SPC in an mmWave MIMO system, which optimizes the HBF and block length to meet the uRLLC requirements of multiple CUs, while also reliably sensing multiple RTs. The next subsection presents a comprehensive literature survey in the area of SPC-aided mmWave systems.

A. Literature review

The authors of the seminal papers [16]–[25] investigated the trade-off between the performance of SC by exploring the Pareto region of latency-agnostic ISAC-aided systems. Specifically, the authors of [16], [17] conceived cutting-edge techniques for the optimization of the transmit waveform to characterize the trade-off between the SC performance in an ISAC-enabled MIMO system, while considering both shared and separated antenna scenarios. Cao, in the treatise [18], proposed a beamforming strategy for determining the Pareto boundary of ISAC-aided MIMO systems, while considering the sensing- and communication- signal-to-interference-plus-noise ratios (SINRs) as the metrics for the SC trade-off. The authors of [19] developed a revolutionary framework for Pareto optimal beamforming optimization of MIMO ISAC systems with the aim of analyzing the trade-off between the sensing and communication rates. As a further advance, Zou *et al.* [20] explored the Pareto boundary for maximization of the energy efficiency (EE) in ISAC systems. The framework of [20] presented a novel constrained Pareto OP for the maximization of the EE of the CUs, while constraining the sensing-centric EE. To solve this non-convex problem, an iterative successive convex approximation (SCA)-based algorithm is proposed to obtain the approximate Pareto boundary by evaluating a sequence of constrained problems subject to sensing-centric EE thresholds. Furthermore, Yu *et al.*, [21] proposed a majorization and minimization (MM)-based algorithm for optimizing the beamforming in an ISAC-aided wireless system. The authors [23]–[25] proposed various beamforming frameworks for ISAC systems, which are robust to imperfect knowledge of the channel state information (CSI)

It is important to note that the fully-digital beamforming schemes discussed in [16]–[25] are inefficient for ISAC-aided mmWave MIMO systems, since they require an excessive number of RFCs. To this end, the related expositions [26]–[33] conceived HBF designs for ISAC-aided mmWave MIMO systems, which significantly reduces the number of RFCs and yet performs close to the fully-digital schemes. However, the main challenge in designing the HBF lies in the practical constraints of the RF TPC, particularly the need to implement the RF TPC

using unit modulus (UM) phase shifters [26]. To this end, the authors of [27] proposed an innovative orthogonal matching pursuit (OMP)-based algorithm for optimizing the BB and RF TPCs for the mmWave MIMO systems. Along similar lines, Yu *et al.* [28] proposed the HBF schemes for the ISAC-aided mmWave MIMO systems that employed the Riemannian conjugate gradient (RCG) method to handle the UM constraint for RF TPC. The simulation results therein [28] demonstrated that proposed HBF schemes having fewer RFCs approached the performance of the optimal FDB scheme. As a further advance, the authors of [29] consider a partially connected hybrid architecture for ISAC-aided mmWave MIMO systems, where they focused on the dual objectives of minimizing the Cramér-Rao bound (CRB) for the estimation of the direction of arrival (DOA) and maximizing the SINR of the received radar echos. Furthermore, Liu *et al.* [30] present innovative techniques for minimizing the weighted radar beamforming error (RBE) and communication beamforming error (CBE) to optimize both the RF and BB TPCs of the HBF scheme. Furthermore, the terms RBE and CBE are defined as the error between the optimal radar-only and transmit beamformer as well as between the communication-only beamformers and the transmit beamformer, respectively. Moreover, Xiao *et al.* [31] proposed multi-beam sweeping schemes for beamforming optimization in mmWave MIMO ISAC systems, which rely only on the RF TPC. Although the proposed schemes [31] are cost-effective, the performance degrades upon increasing the number of RTs and CUs in the system. To handle these issues, our recent work [32] proposed the HBF scheme for multi-beam framework, which maximizes the sensing beam-pattern gain of the RTs while considering the minimum SINR requirement of the CUs. In this work [32], we adopted the block coordinate descent (BCD) method for decoupling the optimization variables and employed the second-order cone programming (SOCP) and RCG algorithms for optimizing the BB and RF TPCs, respectively. Furthermore, the authors of [33] proposed the hybrid transceiver design for mmWave MIMO full-duplex ISAC systems, wherein they mitigate the self-interference caused in the mono-static radars via the signal processing in the RF domain.

The transmission models underlying a large fraction of the literature surveyed above, namely [16]–[33] are based on communication with infinite block lengths (IBLs), hence they rely on the conventional Shannon capacity formula. While their analyses are ingenious and immensely useful in their specific settings, these models are agnostic of the stringent reliability and latency requirements of uRLLC applications. More specifically, the classical Shannon capacity formula that considers IBL transmission, is inapplicable in the SPC regime due to the finite block length and non-zero code word error probability specifications. Thus, following the innovative rate expression provided by [3], the authors of [34]–[37] proposed inspiring beamforming designs for SPC-enabled systems. Specifically, Huang *et al.* [34] investigated the rate region of the SPC-enabled multiple-input and single-output (MISO) interference channel. The authors therein [34] optimized the beamformer

weights and block length to maximize the sum rate of the system, considering the resource allocation and block length as constraints. He *et al.* [35] derived novel beamforming techniques for the SPC-enabled multi-CU MISO downlink, where they addressed the optimization of multiple objectives including the weighted sum rate, EE, and CU fairness, while also considering the minimum CU-rate and transmission power constraints. On the other hand, the authors of [36] designed efficient resource allocation and beamforming algorithms for SPC-enabled multi-CU MISO systems that achieve transmit power minimization, while constraining the decoding-error probability of SPC transmission. Furthermore, the authors of [37] presented state-of-the-art beamformer designs for an SPC-enabled MIMO system using alternating optimization and fractional programming. Their innovative beamforming techniques maximize the achievable data rate for a given transmit power budget.

Following the above discussion of ISAC and SPC in separate contexts, we now move our focus to the literature that explored their integration. The authors of the path-breaking works [38]–[42] proposed a transmit beamforming paradigm for SPC-enabled ISAC systems, wherein the base station (BS) performs detection of an RT and provides uRLLC services for the CUs. The pioneering research in [38] presents a framework for transmit power minimization, while meeting the critical radar sensing and uRLLC latency requirements. The authors of [38] proposed a creative quadratic transform-based fractional programming approach in conjunction with an interior point method to solve the pertinent OP. To explore the trade-off between uRLLC data transmission and target localization in SPC-aided ISAC systems, Zabini *et al.* derived a novel beamforming scheme for minimizing the CRB of RT localization under constraints on the block error probability for the CUs in [39]. Along similar lines, in the avant-grade investigation of [42], the researchers succeeded in developing a joint beamforming and scheduling scheme for an SPC-enabled ISAC system that meets the stringent uRLLC requirements of the CUs.

The authors of [40] explored the SC trade-off for an SPC-enabled ISAC system with single input and single output (SISO) configuration, wherein they characterized the Pareto boundary of the rate and error metrics for the communication and sensing, respectively. As a further advance, Huang *et al.* [41] investigated SPC-aided ISAC for mobile edge computing, wherein multiple ISAC devices adopt SPC for offloading their sensed radar data to an edge server.

Moreover, considering the strong possibility that next-generation networks may harness mmWave technology in view of the imminent spectrum crunch in the sub-6 GHz band, it is crucial to explore the SC trade-off in SPC-enabled mmWave ISAC systems. Exploring the Pareto boundary of such systems holds the key for SPC. Clearly, there is a paucity of SPC-enabled mmWave ISAC system studies in the open literature. Explicitly, in the complex face of challenges, such as the complex rate expression of SPC transmission, integration of sensing and communication tasks, coupled with the hybrid

design of the TPC, the associated beamforming optimization is not well documented. Inspired by this knowledge-gap, we conceive a novel HBF scheme to achieve Pareto optimal SC trade-offs in SPC-enabled mmWave ISAC systems. Furthermore, a systematic contrasting of our novel contributions to the existing literature is shown in Table I, and the main contributions of this paper are enumerated next.

B. Contributions of this work

- 1) Explicitly, this is the first paper to investigate SPC in the context of mmWave ISAC systems where an ISAC BS transmits SPC-encoded signals to multiple CUs and detects several RTs, determining the RBE-rate region to reveal the trade-off between sensing and communication. To this end, we formulate the OP for the Pareto boundary of the RBE-rate region, incorporating constraints for minimum CU rates, maximum tolerable RBE for RTs, and additional limitations from hybrid MIMO architecture, transmit power, and SPC block length, resulting in a highly non-convex problem.
- 2) To solve the above problem, we propose an iterative two-layer bisection search (TLBS) algorithm, where the inner layer minimizes the RBE of the system by optimizing the BB and RF TPCs for a fixed sum rate, and subsequently, the outer layer optimizes the block length and updates the achievable rate via the bisection search method.
- 3) More specifically, in the inner layer, we reformulate the rate expression to transform the intractable QoS constraint into a tractable SINR constraint and, subsequently, adopt the BCD principle for iteratively optimizing the BB and RF TPCs. For the optimization of the RF TPC, we formulate the quadratically constrained quadratic program (QCQP) OP with an extra UM constraint. To solve this, we propose two novel methods: bisection-based majorization and minimization (BMM) as well as exact penalty-based manifold optimization (EPMO). Following this, the SOCP method is employed to optimize the BB TPC. Furthermore, we update the block length and sum rate via mixed integer programming and a bisection search in the outer layer.
- 4) Finally, we evaluate the performance of the proposed scheme by evaluating the Pareto boundaries for SC trade-off, achievable sum rates, and beam patterns for a variety of settings and compare them to the benchmarks for verifying the effectiveness of our proposed algorithm.

C. Organization and Notation

The rest of this paper is organized as follows. Section II presents the system model of our SPC-aided mmWave ISAC system, where we discuss the performance metric of interest and formulate the design problem. Next, we introduce the proposed algorithms to jointly optimize the RF and BB TPCs, together with the block lengths in Section III. Section IV presents numerical results to characterize the performance of our proposed approach and its effectiveness. Finally, Section V concludes the paper.

Table I: Contrasting our novel contributions to the literature of short-packet mmWave ISAC systems

	[18]	[19]	[21]	[28]	[29]	[30]	[32]	[34]	[35]	[37]	[38]	[39]	[40]	[41]	Proposed
ISAC	✓	✓	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓
mmWave MIMO				✓	✓	✓	✓								✓
Short packet communication (SPC)								✓	✓	✓	✓	✓	✓	✓	✓
Hybrid beamforming				✓	✓	✓	✓								✓
Multi-users		✓	✓	✓	✓		✓	✓	✓	✓	✓			✓	✓
Multi-targets		✓		✓		✓								✓	✓
Pareto boundary	✓	✓						✓					✓		✓
Sum rate		✓	✓	✓		✓	✓	✓	✓	✓			✓		✓
Radar beamforming error				✓		✓									✓
SOC					✓		✓	✓							✓
Mixed integer program								✓							✓
SPC-aided mmWave ISAC															✓
Layer-based bisection search															✓
Bisection-based MM method															✓
Penalty-based manifold optimization															✓

Notations: We use the following notations throughout the paper: \mathbf{A} , \mathbf{a} , and a represent a matrix, a vector, and a scalar quantities, respectively. The (i, j) th element, and Hermitian of matrix \mathbf{A} are denoted by $\mathbf{A}(i, j)$, and \mathbf{A}^H , respectively. The trace, Frobenius norm and vectorization of a matrix \mathbf{A} are represented as $\text{tr}(\cdot)$, $\|\mathbf{A}\|_F$ and $\text{vec}(\cdot)$. The expectation operator is represented as $\mathbf{E}\{\cdot\}$; the real part of a quantity is denoted by $\text{Re}(\cdot)$. \mathbf{I}_M denotes an $M \times M$ identity matrix; the symmetric complex Gaussian distribution of mean μ and covariance matrix σ^2 is represented as $\mathcal{CN}(\mu, \sigma^2)$. The operators \odot and \otimes denote the Hadamard product and Kronecker product respectively, while $\mathcal{W}(\cdot, \cdot; \cdot)$ represents the generalized Lambert function. The elements in $\arg[\mathbf{z}]$ are the phases of the input complex vectors. Furthermore $\lfloor a \rfloor$ denotes the largest integer less than or equal to a and $\lceil a \rceil$ denotes the smallest integer greater than or equal to a .

II. SYSTEM MODEL AND PROBLEM FORMULATION OF SPC-AIDED MMWAVE ISAC SYSTEMS

As shown in Fig. 1a, we consider an SPC-enabled ISAC downlink operating in the mmWave band, where an ISAC BS communicates with M uRLLC CUs and detects N_{tar} RTs, simultaneously. The ISAC BS relies on a fully-connected hybrid MIMO architecture [32], [43] having N_t transmit antennas and $N_{\text{RF}} \leq N_t$ RFCs, where each RF chain is connected to each antenna via phase shifters, as shown in Fig. 1b. In addition, we assume the phase synchronization of RFCs via following [44] Moreover, each CU is equipped with a single antenna. Therefore, to support simultaneous service for M single antenna CUs while detecting N_{tar} RTs through SPC transmission at the angles of interest, one has to follow the condition $M \leq N_{\text{RF}} \ll N_t$.

A. mmWave channel model

This paper employs the popular Saleh-Valenzuela channel model [18]–[22] to capture the geometric properties of the mmWave channel, which includes complex-valued path losses, angles-of-arrival (AoAs), and angles-of-departure (AoDs) arising due to the paths scattered by a finite number of dominant clusters. Mathematically, the frequency-flat mmWave MISO channel $\mathbf{h}_m^H \in \mathbb{C}^{N_t \times 1}$ between the ISAC BS and the m th CU

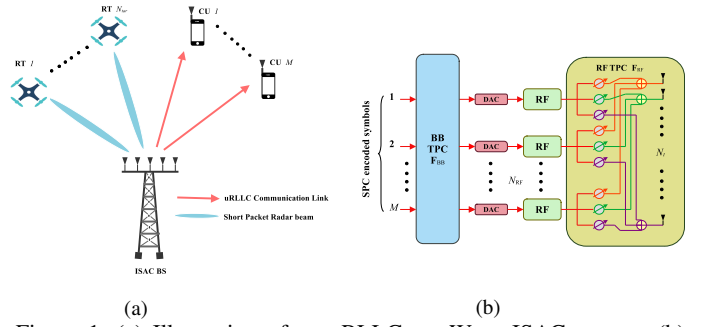


Figure 1: (a) Illustration of an uRLLC mmWave ISAC system. (b) Block diagram of hybrid beamforming architecture at the ISAC BS. is expressed as

$$\mathbf{h}_m^H = \sqrt{\frac{N_t}{N_{\text{clu}} N_{\text{ray}}}} \sum_{i=1}^{N_{\text{clu}}} \sum_{j=1}^{N_{\text{ray}}} \alpha_{i,j}^m \mathbf{a}_{\text{BS}}^H(\phi_{i,j}^m), \quad (1)$$

where N_{clu} and N_{ray} are the number of scattering clusters and scattered rays per cluster, respectively. The quantity $\alpha_{i,j}^m$ in (1) is the multipath channel gain that is distributed as $\mathcal{CN}(0, 10^{-0.1 PL(d_m)})$, $\forall i = \{1, \dots, N_{\text{clu}}\}$, and $j = \{1, \dots, N_{\text{ray}}\}$ where $PL(d_m)$ is the path loss in dB that depends on the distance d_m associated with the corresponding link. Moreover, we consider that the ISAC BS is employed with a uniform linear array (ULA), owing to the array response vector $\mathbf{a}_{\text{BS}}(\phi_{i,j}^m) \in \mathbb{C}^{N_t \times 1}$ as [32]

$$\mathbf{a}_{\text{BS}}(\phi_{i,j}^m) = \frac{1}{\sqrt{N_t}} \left[1, e^{j \frac{2\pi d}{\lambda} \sin(\phi_{i,j}^m)}, \dots, e^{j(N_t-1) \frac{2\pi d}{\lambda} \sin(\phi_{i,j}^m)} \right]^T, \quad (2)$$

where $\phi_{i,j}^m$ denotes the AoD, λ is the carrier wavelength, and d is the spacing between adjacent antennas, which is set as $d = \lambda/2$. Based on this channel model, let us describe the signal-, radar- and communication models in the subsequent subsections.

B. Signal model

Let us consider that the data of each CU is encoded with an individual encoder of finite block length β_m , $m = 1, \dots, M$ at the ISAC BS. Thus, the information bits of the uRLLC CUs are transmitted via packets. In a similar fashion, each CU decodes its data independently, considering a non-zero decoding error probability of ϵ_m , $m = 1, \dots, M$. Furthermore,

the encoded symbols of all the CUs at the ISAC BS are initially processed by the BB TPC $\mathbf{F}_{\text{BB}} = [\mathbf{f}_{\text{BB},1}, \dots, \mathbf{f}_{\text{BB},M}] \in \mathbb{C}^{N_{\text{RF}} \times M}$, followed by the RF TPC $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_t \times N_{\text{RF}}}$. Thus, the downlink transmitted signal $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ from the ISAC BS is given by [28], [29]

$$\mathbf{x} = \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s} = \mathbf{F}_{\text{RF}} \sum_{m=1}^M \mathbf{f}_{\text{BB},m} s_m, \quad (3)$$

where $\mathbf{s} = [s_1, s_2, \dots, s_M]^T \in \mathbb{C}^{M \times 1}$ is the encoded signal, which serves as a combined signal for the radar detection as well as downlink communication transmission [28]. Moreover, the encoded symbols are assumed to be statistically independent and identically distributed (i.i.d), which satisfy $\mathbb{E}\{\mathbf{s}\} = \mathbf{0}$ and $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_M$. Consequently, the covariance matrix $\mathbf{C}_x \in \mathbb{C}^{N_t \times N_t}$ of the transmitted signal \mathbf{x} is given by [28]

$$\mathbf{C}_x = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H. \quad (4)$$

C. Radar model

This paper considers the mono-static MIMO radar used at the ISAC BS for target detection, where the same antenna arrays are used for transmitting and receiving radar signals. However, this leads to a signal leakage problem in such an architecture, which can be efficiently mitigated via RF and baseband cancellers, as discussed in [45]. Therefore, we assume that the noise term includes the residual self-interference, and the echo signal $\mathbf{y}_{\text{rad}} \in \mathbb{C}^{N_t \times 1}$ received at the ISAC BS, which can be written as [28]

$$\begin{aligned} \mathbf{y}_{\text{rad}} = & \sum_{t=1}^{N_{\text{tar}}} \zeta_t^{\text{tar}} \mathbf{a}_{\text{BS}}(\theta_t^{\text{tar}}) \mathbf{a}_{\text{BS}}^H(\theta_t^{\text{tar}}) \mathbf{x} \\ & + \sum_{c=1}^{N_{\text{ct}}} \zeta_c^{\text{ct}} \mathbf{a}_{\text{BS}}(\theta_c^{\text{ct}}) \mathbf{a}_{\text{BS}}^H(\theta_c^{\text{ct}}) \mathbf{x} + \mathbf{n}_{\text{rad}}, \end{aligned} \quad (5)$$

where the first and second terms in (5) are the desired target signal and the echo signal due to clutter, respectively, and $\mathbf{n}_{\text{rad}} \in \mathbb{C}^{N_t \times 1}$ is the noise encountered in the radar sensing environment. The quantities ζ_t^{tar} and ζ_c^{ct} are the complex-valued path loss reflection coefficients of the RTs and clutters located at angles θ_t^{tar} and at θ_c^{ct} , respectively [28], [29]. Note that here we ignore the echo signal reflected by the CUs, which is in line with the seminal literature [28]–[32]. Furthermore, based on the received signal (5), one can estimate the angles of the RTs by employing various schemes, such as the multiple signal classification (MUSIC) algorithm [5] and the affine-precoded superimposed pilot (AP-SIP) [46]. For the target-tracking mode of the radar system, we assume perfect estimation of the target angle θ [7], [9], [28]–[32]. Consequently, the transmit beampattern gains $G(\theta)$ can be expressed as

$$G(\theta) = \mathbf{a}_{\text{BS}}^H(\theta) \mathbf{C}_x \mathbf{a}_{\text{BS}}(\theta), \quad (6)$$

where $G(\theta)$ is the spatial beam pattern, which has to be synthesized for the target sensing environment. Observe that designing $G(\theta)$ is equivalent to designing the covariance matrix \mathbf{C}_x . Hence, one has to design \mathbf{F}_{RF} and \mathbf{F}_{BB} , that meet

the sensing requirements of the RTs. Therefore, to optimize the radar sensing performance, we design \mathbf{C}_x to approach the ideal desired radar covariance matrix $\mathbf{C}_d = \mathbf{F}_r \mathbf{F}_r^H$, where $\mathbf{F}_r \in \mathbb{C}^{N_t \times N_{\text{tar}}}$ is the ideal radar beamformer used for the RTs [30], which is given by

$$\mathbf{F}_r = [\mathbf{a}_{\text{BS}}(\theta_1^{\text{tar}}), \mathbf{a}_{\text{BS}}(\theta_2^{\text{tar}}), \dots, \mathbf{a}_{\text{BS}}(\theta_{N_{\text{tar}}}^{\text{tar}})]. \quad (7)$$

To evaluate the performance of the sensing, we consider the radar beamforming error (RBE) of the RTs denoted by \mathcal{E} [28], [30], which is given by

$$\mathcal{E}(\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}, \mathbf{U}) = \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} - \mathbf{F}_r \mathbf{U}\|_F^2, \quad (8)$$

where $\mathbf{U} \in \mathbb{C}^{N_{\text{tr}} \times M}$ is an auxiliary unitary matrix obeying $\mathbf{U}\mathbf{U}^H = \mathbf{I}_{N_{\text{tar}}}$. Specifically, the RBE metric \mathcal{E} quantifies the error between the optimal radar beamformer and the transmit beamformer. Therefore, based on the potential directions of the desired RTs, one can maximize the transmit beamforming gain towards the RTs by minimizing the RBE \mathcal{E} .

D. Communication model

Based on the mmWave MISO channel (1) and the transmission signal model (3), the received signal y_m at the m th CU can be written as [28], [29]

$$y_m = \mathbf{h}_m^H \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s} + n_m \quad (9a)$$

$$= \mathbf{h}_m^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},m} s_m + \sum_{n=1, n \neq m}^M \mathbf{h}_m^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},n} s_n + n_m, \quad (9b)$$

where n_m is the i.i.d. complex additive white Gaussian noise (AWGN) having the distribution $n_m \sim \mathcal{CN}(0, N_o)$. It is assumed that the channel state information (CSI) is perfectly known at all CUs [7]–[10], [28]–[30], noting that both the accurate channel estimation and the robust TPC design relying on partial CSI constitute rather different problems in ISAC systems [23]–[25]. Given the strict length-limitation, these have to be set aside for future research. Then, the widely adopted metric in wireless communication, namely, SINR γ_m of the m th CU, is evaluated as [8]

$$\gamma_m = \frac{|\mathbf{h}_m^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},m}|^2}{\sum_{n=1, n \neq m}^M |\mathbf{h}_m^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},n}|^2 + N_o}. \quad (10)$$

Following Shannon's capacity formula, the maximum achievable transmission rate S_m of this CU, in nats/s/Hz/channel, is given by

$$S_m = \ln(1 + \gamma_m), \forall m. \quad (11)$$

However, the conventional Shannon capacity relation holds true only for IBL transmission, wherein the error probability tends to zero. Since this paper considers practical SPC, the achievable rate given by (11) is not a realistic model. In this context, thanks to the results in [3], [4], [35], [36], the achievable rate R_m of the m th uRLLC CU owing to the SPC

transmission with a finite block length of β_m and transmission error probability ϵ_m is given by

$$R_m = \ln(1 + \gamma_m) - \sqrt{\frac{V_m}{\beta_m}} Q^{-1}(\epsilon_m), \forall m, \quad (12)$$

where $Q^{-1}(\cdot)$ is the inverse of the Gaussian Q-function and V_m^{-1} represents the channel dispersion of the uRLLC CU m , which is given by

$$V_m(\gamma_m) = 1 - \frac{1}{(1 + \gamma_m)^2}. \quad (13)$$

Consequently, the achievable sum rate of the CUs is given by

$$\mathcal{R} = \sum_{m=1}^M R_m. \quad (14)$$

E. Problem Formulation

This paper aims for jointly optimizing the Pareto optimal RF TPC \mathbf{F}_{RF} , BB TPC \mathbf{F}_{BB} and the block lengths $\{\beta_m\}_{m=1}^M$ to characterize the RBE-rate region of an SPC-enabled ISAC mmWave system. Specifically, the RBE-rate region of the system under consideration is defined as the collection of all the feasible twin tuples $(\mathcal{E}, \mathcal{R})$ that can be simultaneously achieved, where \mathcal{E} and \mathcal{R} are sensing and communication metrics, respectively. Therefore, we are interested in evaluating the Pareto front [11], [12] constituted by all optimal twin tuples $(\mathcal{E}, \mathcal{R})$ in the boundary of the RBE-rate region. Typically, the Pareto front consists of $(\mathcal{E}, \mathcal{R})$ pairs at which it is impossible to simultaneously improve the communication and sensing performance without a compromise between them. More specifically, for a given SPC-aided mmWave ISAC system, any $(\mathcal{E}, \mathcal{R})$ twin tuple located on the Pareto boundary of the rate-RBE region is formulated as

$$\mathcal{P}_0 : \max_{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}, \mathbf{U}, \{\beta_m\}_{m=1}^M} \mathcal{R}, \quad (15a)$$

$$\text{s. t. } R_m \geq \eta_m \mathcal{R}, \forall m, \quad (15b)$$

$$\mathcal{E} \leq \mathcal{E}_{\max}, \quad (15c)$$

$$\mathbf{U}\mathbf{U}^H = \mathbf{I}_{N_{\text{tar}}}, \quad (15d)$$

$$|\mathbf{F}_{\text{RF}}(i, j)| = 1, \forall i, j, \quad (15e)$$

$$\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F^2 \leq P_{\max} \quad (15f)$$

$$\sum_{m=1}^M \beta_m \leq N, \quad (15g)$$

$$\beta_m \in \mathbb{Z}^+, \forall m, \quad (15h)$$

where (15b) denotes the QoS constraint for the CUs with $\eta_m \in (0, 1)$ denoting the ratio between the achievable rate of the m th CU and the sum rate \mathcal{R} , such that $\sum_{m=1}^M \eta_m = 1$. Furthermore, the constraint (15c) is the RBE required for the RTs, while (15e) is the UM constraint due to the phase shifters of the hybrid architecture, and (15f) is the maximum transmit power budget constraint. Moreover, the last two constraints (15g) and (15h) are due to the SPC regime, where

¹It gauges the variability of the channel relative to a deterministic bit pipe with the same capacity.

all the uRLLC CU symbols are transmitted within a maximum of N symbols and the block length β_m must be a non-negative integer. Since the calculation of RBE depends on \mathbf{F}_{r} , the problem formulated is also applicable for tracking the mobile RTs, where the sensing beam is formed towards the estimated/predicted direction to track the movement of the RTs [22]. By minimizing the RBE for a specific direction, one can improve the sensing capabilities in this desired sensing direction.

The complete Pareto boundaries of the achievable RBE-rate region of the SPC-enabled mmWave ISAC system above can be characterized by solving the OP \mathcal{P}_0 , which is however, challenging due to the non-convex constraints (15b) and (15e) and tightly coupled variables in the objective function and the constraints. Moreover, the rate expression defined by (12) is more decimate than the traditional Shannon formula, which exacerbates the challenge. Therefore, the next section proposes a novel TLBS method that overcomes this obstacle by intelligently exploiting a bisection search.

III. TWO LAYER BISECTION SEARCH FOR JOINT OPTIMIZATION

In the proposed TLBS method, the inner layer evaluates $\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}$ and \mathbf{U} to minimize the RBE $\mathcal{E} = \|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} - \mathbf{F}_{\text{r}}\mathbf{U}\|_F^2$ for the SPC parameters $\beta_m, \forall m$, whereas the outer layer updates the achievable rate \mathcal{R} by employing the bisection search. Specifically, in the inner layer, we obtain $\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}$ and \mathbf{U} for feasible values of \mathcal{R} and subsequently update the sum rate \mathcal{R} and $\{\beta_m\}_{m=1}^M$ in the outer layer. For any given sum rate $\mathcal{R} \geq 0$ and β_m , the equivalent feasible problem of \mathcal{P}_0 in the inner layer is the minimization of the RBE, which is given by

$$\begin{aligned} \mathcal{P}_1 : \min_{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}, \mathbf{U}} \mathcal{E}(\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}, \mathbf{U}) &= \|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} - \mathbf{F}_{\text{r}}\mathbf{U}\|_F^2 \\ \text{s. t. } & \text{(15b), (15d), (15e), and (15f).} \end{aligned} \quad (16)$$

To obtain the optimal solution $\{\mathbf{F}_{\text{RF}}^*, \mathbf{F}_{\text{BB}}^*, \mathbf{U}^*\}$ in the inner layer, we solve the above OP (16). Subsequently, if the OP (16) is feasible for the given $\{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}, \mathbf{U}\}$, we perform a bisection search by solving a sequence of feasibility problems corresponding to the problem \mathcal{P}_0 to update the optimal value of the rate \mathcal{R}^* and the block length $\{\beta_m^*\}_{m=1}^M$ in the outer layer. Moreover, upon denoting the optimal solutions of \mathcal{P}_1 as $\mathbf{F}_{\text{RF}}^*, \mathbf{F}_{\text{BB}}^*$ and \mathbf{U}^* , it is evident that the OP \mathcal{P}_1 is feasible if $\|\mathbf{F}_{\text{RF}}^*\mathbf{F}_{\text{BB}}^* - \mathbf{F}_{\text{r}}\mathbf{U}^*\|_F^2 \leq \mathcal{E}_{\max}$, and $\|\mathbf{F}_{\text{RF}}^*\mathbf{F}_{\text{BB}}^*\|_F^2 \leq P_{\max}$, otherwise, it is considered infeasible.

A. Inner layer: BCD algorithm for solving \mathcal{P}_1

Because of the finite block length β_m and the transmission error probability ϵ_m in the rate expression, the constraint (15b) is highly non-convex. Therefore, to solve \mathcal{P}_1 , we first transform the non-convex constraint (15b) into a tractable form by using the following proposition:

Proposition 1. Given $\beta_m \geq 0$ and $\epsilon_m \geq 0, \forall m$, the constraint $R_m \geq \eta_m \mathcal{R}$ is equivalent to $\gamma_m \geq \Gamma_m, \forall m$, where Γ_m is given

by

$$\Gamma_m = e^{\eta_m \mathcal{R} + \frac{\kappa_m}{2}} - 1, \quad (17)$$

and κ_m is the generalized Lambert \mathcal{W} function, which is given by [34]

$$\kappa_m = \mathcal{W} \left(\frac{2Q^{-1}(\epsilon_m)}{\sqrt{\beta_m}}, \frac{-2Q^{-1}(\epsilon_m)}{\sqrt{\beta_m}}; -4\delta_m^2 \left(\frac{Q^{-1}(\epsilon_m)}{\sqrt{\beta_m}} \right)^2 \right), \quad (18)$$

where $\delta_m = e^{-\eta_m \mathcal{R}}$.

Proof. Please refer to Appendix A of [47] for the detailed proof. \square

Upon employing the above proposition, the OP \mathcal{P}_1 can be recast as follows

$$\min_{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}, \mathbf{U}} \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} - \mathbf{F}_r \mathbf{U}\|_F^2 \quad (19a)$$

$$\text{s. t. } \gamma_m \geq \Gamma_m, \forall m, \quad (19b)$$

$$(15d), (15e), \text{ and } (15f). \quad (19c)$$

Since the optimization variables \mathbf{F}_{RF} , \mathbf{F}_{BB} and \mathbf{U} are coupled in both the objective function and the constraints of (19), we adopt the BCD method to decouple the optimization variables \mathbf{F}_{RF} , \mathbf{F}_{BB} , and \mathbf{U} in (19). It is worth noting that the BCD method renders it easier to break down the intricate problem into distinct sub-problems, each of which is focused on maximizing a particular block of variables [32].

1) *Sub-problem for \mathbf{F}_{RF} :* For fixed \mathbf{F}_{BB} and \mathbf{U} , the equivalent OP for the design of \mathbf{F}_{RF} is given by

$$\min_{\mathbf{F}_{\text{RF}}} \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} - \mathbf{F}_r \mathbf{U}\|_F^2, \quad (20a)$$

$$\text{s. t. } \gamma_m \geq \Gamma_m, \forall m, \quad (15e), \text{ and } (15f). \quad (20b)$$

The OP (20) above is still non-convex due to the non-convex SINR constraints (19b) and UM constraint (15e). To handle this issue, let us rewrite the SINR γ_m of the m th CU as follows

$$\gamma_m = \frac{\text{tr}(\mathbf{F}_{\text{RF}}^H \mathbf{H}_m \mathbf{F}_{\text{RF}} \mathbf{B}_m)}{\sum_{n=1, n \neq m}^M \text{tr}(\mathbf{F}_{\text{RF}}^H \mathbf{H}_m \mathbf{F}_{\text{RF}} \mathbf{B}_n) + N_o}, \quad (21)$$

where the quantities $\mathbf{H}_m \in \mathbb{C}^{N_t \times N_t}$ and $\mathbf{B}_m \in \mathbb{C}^{N_{\text{RF}} \times N_{\text{RF}}}$ are given by $\mathbf{H}_m = \mathbf{h}_m \mathbf{h}_m^H$ and $\mathbf{B}_m = \mathbf{f}_{\text{BB}, m} \mathbf{f}_{\text{BB}, m}^H$, respectively. Consequently, one can reformulate the SINR constraint (19b) as follows

$$\frac{\text{tr}(\mathbf{F}_{\text{RF}}^H \mathbf{H}_m \mathbf{F}_{\text{RF}} \mathbf{B}_m)}{\sum_{n=1, n \neq m}^M \text{tr}(\mathbf{F}_{\text{RF}}^H \mathbf{H}_m \mathbf{F}_{\text{RF}} \mathbf{B}_n) + N_o} \geq \Gamma_m, \forall m, \quad (22a)$$

$$\Rightarrow \sum_{n=1, n \neq m}^M \text{tr}(\mathbf{F}_{\text{RF}}^H \mathbf{H}_m \mathbf{F}_{\text{RF}} \mathbf{B}_n) - \frac{1}{\Gamma_m} \text{tr}(\mathbf{F}_{\text{RF}}^H \mathbf{H}_m \mathbf{F}_{\text{RF}} \mathbf{B}_m) + N_o \leq 0, \forall m. \quad (22b)$$

Upon employing the $\text{vec}(\cdot)$ operation to (22b), one can rewrite it as follows

$$\sum_{n=1, n \neq m}^M \text{vec}(\mathbf{F}_{\text{RF}})^H (\mathbf{B}_n^T \otimes \mathbf{H}_m) \text{vec}(\mathbf{F}_{\text{RF}}) - \frac{1}{\Gamma_m} \text{vec}(\mathbf{F}_{\text{RF}})^H (\mathbf{B}_m^T \otimes \mathbf{H}_m) \text{vec}(\mathbf{F}_{\text{RF}}) + N_o \leq 0. \quad (23)$$

Furthermore, we define $\mathbf{d} = \text{vec}(\mathbf{F}_{\text{RF}}) \in \mathbb{C}^{N_t N_{\text{RF}} \times 1}$, where $\mathbf{d}(l) = 1, \forall l$ and $\Upsilon_{n,m} = \mathbf{B}_n^T \otimes \mathbf{H}_m \in \mathbb{C}^{N_t N_{\text{RF}} \times N_t N_{\text{RF}}}$. Then (23) is equivalently written as

$$g_m(\mathbf{d}) \triangleq \mathbf{d}^H \Delta_m \mathbf{d} + N_o \leq 0, \quad (24)$$

where $\Delta_m = \left(\sum_{n=1, n \neq m}^M \Upsilon_{n,m} - \frac{1}{\Gamma_m} \Upsilon_{m,m} \right) \in \mathbb{C}^{N_t N_{\text{RF}} \times N_t N_{\text{RF}}}$.

In a similar fashion, let us define $\mathbf{T} = \mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H \in \mathbb{C}^{N_{\text{RF}} \times N_{\text{RF}}}$. Consequently, we express the total power constraint (15f) as

$$\|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = \text{tr}(\mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}) \quad (25a)$$

$$= \text{tr}(\mathbf{F}_{\text{RF}}^H \mathbf{I}_{N_t} \mathbf{F}_{\text{RF}} \mathbf{T}). \quad (25b)$$

Subsequently, by employing $\mathbf{d} = \text{vec}(\mathbf{F}_{\text{RF}})$ in the above equation, (15f) can be redefined as

$$\omega_p(\mathbf{d}) \triangleq \mathbf{d}^H \Omega_p \mathbf{d} - P_{\max} \leq 0, \quad (26)$$

where $\Omega_p = (\mathbf{T}^T \otimes \mathbf{I}_{N_t}) \in \mathbb{C}^{N_t N_{\text{RF}} \times N_t N_{\text{RF}}}$. In addition, we express the objective function RBE of (20) using the $\text{vec}(\cdot)$ operation as follows

$$\|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} - \mathbf{F}_r \mathbf{U}\|_F^2 = \|\text{vec}(\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}) - \text{vec}(\mathbf{F}_r \mathbf{U})\|_2^2 \quad (27a)$$

$$= \|(\mathbf{F}_{\text{BB}}^T \otimes \mathbf{I}_{N_t}) \mathbf{d} - \text{vec}(\mathbf{F}_r \mathbf{U})\|_2^2. \quad (27b)$$

Subsequently, the RBE can be expressed in terms of \mathbf{d} as follows

$$\omega_r(\mathbf{d}) = \|\Omega_r \mathbf{d} - \mathbf{f}_r\|_2^2, \quad (28)$$

where the quantities $\mathbf{f}_r \in \mathbb{C}^{N_t M \times 1}$ and $\Omega_r \in \mathbb{C}^{N_t M \times N_t N_{\text{RF}}}$ are defined as $\mathbf{f}_r = \text{vec}(\mathbf{F}_r \mathbf{U})$ and $\Omega_r = (\mathbf{F}_{\text{BB}}^T \otimes \mathbf{I}_{N_t})$, respectively. As a further advance, (28) is translated into a quadratic expression as follows

$$\omega_r(\mathbf{d}) \triangleq \|\Omega_r \mathbf{d} - \mathbf{f}_r\|_2^2 \quad (29a)$$

$$= (\Omega_r \mathbf{d} - \mathbf{f}_r)^H (\Omega_r \mathbf{d} - \mathbf{f}_r) \quad (29b)$$

$$= \mathbf{d}^H \Xi_r \mathbf{d} - 2\text{Re}(\mathbf{a}_r^H \mathbf{d}) + e_r, \quad (29c)$$

where $\Xi_r = \Omega_r^H \Omega_r \in \mathbb{C}^{N_t N_{\text{RF}} \times N_t N_{\text{RF}}}$, $\mathbf{a}_r = \Omega_r^H \mathbf{f}_r \in \mathbb{C}^{N_t N_{\text{RF}} \times 1}$ and $e_r = \mathbf{f}_r^H \mathbf{f}_r$. Therefore, following the above mathematical manipulations spanning from (21) to (29), the OP (20) can be recast as follows

$$\min_{\mathbf{d}} \omega_r(\mathbf{d}), \quad (30a)$$

$$\text{s.t. } g_m(\mathbf{d}) \leq 0, \forall m, \quad (30b)$$

$$\omega_p(\mathbf{d}) \leq 0, \quad (30c)$$

$$|\mathbf{d}(l)| = 1, \forall l. \quad (30d)$$

The above problem (30) is a quadratically constrained quadratic program (QCQP) with an extra UM constraint, which is a highly non-convex problem. The semi-definite relaxation with randomization (SDR) framework can also be used to solve (30), but it may not guarantee a feasible solution with a fixed randomization number and involves high computational cost due to the non-convex UM constraint [10]. To handle these issues, we propose low-complex two innovative methods, namely the bisection-based majorization-minimization (BMM) method and the exact penalty manifold optimization (EPMO), which are discussed next.

BMM optimization: Let us assume $\mathbf{d}^{(\kappa)}$ to be the feasible point for the problem (30) that is found from the $(\kappa - 1)$ th iteration. Following the inequalities (72), (73) of Appendix B of [47], the respective majorizer functions for (30a), (30b) and (30c) are given by (31), (32) and (33), respectively, where $D = N_t N_{\text{RF}}$ and $c_r = \text{tr}(\Xi_r)$, $c_m = \text{tr}\left(\sum_{n=1, n \neq m}^M \Upsilon_{n,m}\right)$, and $c_p = \text{tr}(\Omega_p)$. Thus, to generate the next feasible point $\mathbf{d}^{(\kappa+1)}$, we solve the following MM OP in the κ th iteration

$$\min_{\mathbf{d}} \omega_r^{(\kappa)}(\mathbf{d}) \quad (34a)$$

$$\text{s.t. } g_m^{(\kappa)}(\mathbf{d}) \leq 0, \forall m, \quad (34b)$$

$$\omega_p^{(\kappa)}(\mathbf{d}) \leq 0, \quad (34c)$$

$$|\mathbf{d}(l)| = 1, \forall l. \quad (34d)$$

To solve the OP (34), we adopt the Lagrange dual optimization by applying the Karush-Kuhn-Tucker (KKT) framework [10]. To this end, the Lagrange function associated with the problem (34) is given by

$$\mathcal{L}^{(\kappa)}(\mathbf{d}, \boldsymbol{\lambda}^{(\kappa)}, \vartheta^{(\kappa)}) = \omega_r^{(\kappa)}(\mathbf{d}) + \sum_{m=1}^M \lambda_m^{(\kappa)} g_m^{(\kappa)}(\mathbf{d}) + \vartheta^{(\kappa)} \omega_p^{(\kappa)}(\mathbf{d}), \quad (35)$$

where $\boldsymbol{\lambda}^{(\kappa)} = [\lambda_1^{(\kappa)}, \dots, \lambda_M^{(\kappa)}]^T \in \mathbb{C}^{M \times 1}$. Moreover, $\lambda_m^{(\kappa)} \geq 0$ and $\vartheta^{(\kappa)} \geq 0$ denote the Lagrange dual multiplier associated with $g_m^{(\kappa)}$ and $\omega_p^{(\kappa)}$, respectively, in the κ th iteration. Subsequently, the Lagrange dual function $\mathcal{D}_{\mathbf{d}}^{(\kappa)}(\boldsymbol{\lambda}^{(\kappa)}, \vartheta^{(\kappa)})$ over the variable \mathbf{d} is expressed as

$$\mathcal{D}_{\mathbf{d}}^{(\kappa)}(\boldsymbol{\lambda}^{(\kappa)}, \vartheta^{(\kappa)}) = \min_{|\mathbf{d}(l)|=1, \forall l \in \mathcal{L}} \mathcal{L}^{(\kappa)}(\mathbf{d}, \boldsymbol{\lambda}^{(\kappa)}, \vartheta^{(\kappa)}). \quad (36)$$

Since the Lagrange function $\mathcal{L}^{(\kappa)}(\mathbf{d}, \boldsymbol{\lambda}^{(\kappa)}, \vartheta^{(\kappa)})$ is linear with respect to the variable \mathbf{d} , the primal optimal point of (36) can be written as a function of the dual multipliers as follows

$$\mathbf{d}^{(\kappa)}(\boldsymbol{\lambda}^{(\kappa)}, \vartheta^{(\kappa)}) = \exp \left(j \arg \left[(c_r \mathbf{I}_D - \Xi_r) \mathbf{d}^{(\kappa)} + \mathbf{a}_r \right. \right. \\ \left. \left. + \sum_{m=1}^M [(c_m \mathbf{I}_D - \Delta_m) \mathbf{d}^{(\kappa)}] \lambda_m^{(\kappa)} + [(c_p \mathbf{I}_D - \Omega_p) \mathbf{d}^{(\kappa)}] \vartheta^{(\kappa)} \right] \right). \quad (37)$$

Furthermore, the optimal solution for the Lagrange Dual problem (36) is obtained as follows

$$\{\boldsymbol{\lambda}^{(\kappa+1)}, \vartheta^{(\kappa+1)}\} \\ = \arg \sup_{\{\lambda_m^{(\kappa)} \geq 0\}_{m=1}^M, \vartheta^{(\kappa)} \geq 0} \mathcal{D}_{\mathbf{d}}^{(\kappa)}(\boldsymbol{\lambda}^{(\kappa)}, \vartheta^{(\kappa)})|_{\mathbf{d}=\mathbf{d}^{(\kappa)}(\boldsymbol{\lambda}^{(\kappa)}, \vartheta^{(\kappa)})}. \quad (38)$$

Due to the strong duality between the primal problem (34) and its dual (38), the primal optimal point $\mathbf{d}^{(\kappa+1)}$ can be obtained by employing the KKT conditions as follows

$$\mathbf{d}^{(\kappa+1)}(\boldsymbol{\lambda}^{(\kappa)}, \vartheta^{(\kappa)}) = \exp \left(j \arg \left[(c_r \mathbf{I}_D - \Xi_r) \mathbf{d}^{(\kappa)} + \mathbf{a}_r \right. \right. \\ \left. \left. + \sum_{m=1}^M [(c_m \mathbf{I}_D - \Delta_m) \mathbf{d}^{(\kappa)}] \lambda_m^{(\kappa)} + [(c_p \mathbf{I}_D - \Omega_p) \mathbf{d}^{(\kappa)}] \vartheta^{(\kappa)} \right] \right), \quad (39a)$$

$$0 \leq \lambda_m^{(\kappa)} \leq \infty, \quad g_m^{(\kappa)}(\mathbf{d}^{(\kappa)}(\boldsymbol{\lambda}^{(\kappa)})) \leq 0, \quad \forall m \quad (39b)$$

$$0 \leq \vartheta^{(\kappa)} \leq \infty, \quad \omega_p^{(\kappa)}(\mathbf{d}^{(\kappa)}(\vartheta^{(\kappa)})) \leq 0, \quad (39c)$$

$$\lambda_m^{(\kappa)} g_m^{(\kappa)}(\mathbf{d}^{(\kappa)}(\boldsymbol{\lambda}^{(\kappa)})) = 0 \quad \forall m, \quad (39d)$$

$$\vartheta^{(\kappa)} \omega_p^{(\kappa)}(\mathbf{d}^{(\kappa)}(\vartheta^{(\kappa)})) = 0, \quad (39e)$$

where (39b) and (39c) are the dual and primal feasibility for $g_m^{(\kappa)}$ and $\omega_p^{(\kappa)}$, respectively. Furthermore, (39d) and (39e) indicate the complementary slackness of the functions $g_m^{(\kappa)}$ and $\omega_p^{(\kappa)}$, respectively. Therefore, one can compute the next feasible point of the dual multipliers $\boldsymbol{\lambda}^{(\kappa+1)}$ and $\vartheta^{(\kappa+1)}$ by solving (38) with the aid of the KKT conditions (39) for the given feasible primal point $\mathbf{d}^{(\kappa)}$ and dual multipliers $\boldsymbol{\lambda}^{(\kappa)}, \vartheta^{(\kappa)}$ in the κ th iteration. We adopt the coordinate ascent technique to compute the next feasible point of the dual multipliers $\boldsymbol{\lambda}^{(\kappa+1)}$ and $\vartheta^{(\kappa+1)}$. In addition, for the given $\{\lambda_n^{(\kappa)} \geq 0\}_{n=1, n \neq m}^M$ and $\vartheta^{(\kappa)}$, if $g_m^{(\kappa)}(\mathbf{d}^{(\kappa)}(\boldsymbol{\lambda}^{(\kappa)}))|_{\lambda_m^{(\kappa)}=0} \leq 0$, we set $\lambda_m^{(\kappa+1)} = 0$ by following (39d). Otherwise, there exists a non-zero $g_m^{(\kappa)}(\mathbf{d}^{(\kappa)}(\boldsymbol{\lambda}^{(\kappa)}))|_{\lambda_m^{(\kappa+1)} \approx 0} \approx 0$. Similarly, for a given $\vartheta^{(\kappa)} \geq 0$, if $\vartheta^{(\kappa)} \omega_p^{(\kappa)}(\mathbf{d}^{(\kappa)}(\vartheta^{(\kappa)}))|_{\vartheta^{(\kappa)}=0} \leq 0$, then we set $\vartheta^{(\kappa+1)} = 0$ by following (39e). Otherwise, there exists a non-zero $\vartheta^{(\kappa)} \omega_p^{(\kappa)}(\mathbf{d}^{(\kappa)}(\vartheta^{(\kappa)}))|_{\vartheta^{(\kappa+1)} \approx 0} \approx 0$. To obtain such non-zero dual multipliers within a limited number of iterations, we employ the bisection method. Finally, if $\{\lambda_m^{(\kappa)}\}_{m=1}^M$ and $\vartheta^{(\kappa)}$ satisfy all the constraints, then the next feasible point of the primal OP is found as

$$\mathbf{d}^{(\kappa+1)} = \exp \left(\arg \left[(c_r \mathbf{I}_D - \Xi_r) \mathbf{d}^{(\kappa)} + \mathbf{a}_r \right] \right). \quad (40)$$

Algorithm 1 summarizes the computational procedure of the BMM method harnessed for solving the problem (30). Note that the next feasible point for the dual multipliers $\{\lambda_m^{(\kappa+1)}, \vartheta^{(\kappa+1)}\}$ of the problem (38) is obtained by optimizing a single dual multiplier at a time, while keeping the other dual multipliers fixed, until the Lagrange function (35) converges at the corresponding primal optimal point $\mathbf{d}^{(\kappa+1)}$.

$$\omega_r(\mathbf{d}) \leq \omega_r^{(\kappa)}(\mathbf{d}) \triangleq 2\text{Re} \left(\mathbf{d}^H [(\Xi_r - c_r \mathbf{I}_D) \mathbf{d}^{(\kappa)} - \mathbf{a}_r] \right) - (\mathbf{d}^{(\kappa)})^H \Xi_r \mathbf{d}^{(\kappa)} + 2Dc_r + E_r, \quad (31)$$

$$g_m(\mathbf{d}) \leq g_m^{(\kappa)}(\mathbf{d}) \triangleq 2\text{Re} \left(\mathbf{d}^H (\Delta_m - c_m \mathbf{I}_D) \mathbf{d}^{(\kappa)} \right) - (\mathbf{d}^{(\kappa)})^H \Delta_m \mathbf{d}^{(\kappa)} + 2Dc_m + N_o, \forall m, \quad (32)$$

$$\omega_p(\mathbf{d}) \leq \omega_p^{(\kappa)}(\mathbf{d}) \triangleq 2\text{Re} \left(\mathbf{d}^H (\Omega_p - c_p \mathbf{I}_D) \mathbf{d}^{(\kappa)} \right) - (\mathbf{d}^{(\kappa)})^H \Omega_p \mathbf{d}^{(\kappa)} + 2Dc_p - P_{\max}, \quad (33)$$

Algorithm 1 BMM algorithm for solving (30)

Input: Feasible \mathbf{F}_{RF} , \mathbf{F}_{BB} , and stopping parameters τ_1 , τ_2

Output: Optimal RF TPC \mathbf{F}_{RF}^*

```

1: Initialize:  $\kappa = 0$ ,  $\nu = 0$ ,  $\mathbf{d}^{(\kappa)} = \text{vec}(\mathbf{F}_{\text{RF}})$ , and dual
   multipliers  $\{\lambda_m^{(\kappa)}\}_{m=1}^M = 0$  and  $\vartheta^{(\kappa)} = 0$ 
2: repeat
3:   for  $m = 1$  to  $K$ 
4:     if  $g_m^{(\kappa)}(\mathbf{d}^{(\kappa)}(\lambda^{(\kappa)}))|_{\lambda_m^{(\kappa)}=0} \leq 0$ , then  $\lambda_m = 0$ 
5:     else
6:        $\lambda_m^L = 0$  and  $\lambda_m^U = 1$ 
7:       if  $g_m^{(\kappa)}(\mathbf{d}^{(\kappa)}(\lambda^{(\kappa)}))|_{\lambda_m^{(\kappa)}=1} \leq 0$ , then  $\lambda_m^U = 1$ 
8:       else
9:         while  $g_m^{(\kappa)}(\mathbf{d}^{(\kappa)}(\lambda^{(\kappa)}))|_{\lambda_m^{(\kappa)}=\lambda_m^U} \geq 0$  do
10:           $\lambda_m^U = 2\lambda_m^U$ 
11:        end while
12:         $\lambda_m^L = \lambda_m^U/2$ 
13:      end if
14:       $\lambda_m = (\lambda_m^L + \lambda_m^U)/2$ 
15:      while  $|g_m^{(\kappa)}(\mathbf{d}^{(\kappa)}(\lambda^{(\kappa)}))|_{\lambda_m^{(\kappa)}=\lambda_m} + \frac{\tau_1}{2}| \geq \frac{\tau_1}{2}$  do
16:        if  $g_m^{(\kappa)}(\mathbf{d}^{(\kappa)}(\lambda^{(\kappa)}))|_{\lambda_m^{(\kappa)}=\lambda_m} \geq 0$ , set  $\lambda_m^L =$ 
            $\lambda_m$ 
17:        else  $\lambda_m^U = \lambda_m$ 
18:        end if
19:      end while
20:    end if
21:  end for
22:  follow steps from (4) to (20) to get  $\vartheta$  using function
    $\omega_p^{(\kappa)}(\mathbf{d}^{(\kappa)}(\lambda^{(\kappa)}))$  for given  $\{\lambda_m\}_{m=1}^M$ 
23:  set  $\nu \leftarrow \nu + 1$ ,  $\lambda^{(\kappa)}[\nu] = [\lambda_1, \dots, \lambda_M]^T$  and  $\vartheta^{(\kappa)}[\nu] = \vartheta$ ,
24:  compute  $L[\nu] = \mathcal{L}^{(\kappa)}(\mathbf{d}, \lambda^{(\kappa)}, \vartheta^{(\kappa)})|_{\mathbf{d}=\mathbf{d}^{(\kappa)}(\lambda^{(\kappa)}[\nu], \vartheta^{(\kappa)}[\nu])}$ 
25: until  $|L[\nu] - L[\nu - 1]| / L[\nu] \leq \tau_1$ 
26: set  $\kappa \leftarrow \kappa + 1$  and  $\mathbf{d}^{(\kappa+1)} = \mathbf{d}^{(\kappa)}(\lambda^{(\kappa)}[\nu], \vartheta^{(\kappa)}[\nu])$ 
27: evaluate  $\omega_r(\mathbf{d})$  using (29) and set  $\omega_r^{(\kappa)} = \omega_r(\mathbf{d}^{(\kappa+1)})$ 
28: if  $|(\omega_r^{(\kappa)} - \omega_r^{(\kappa-1)}) / \omega_r^{(\kappa)}| \leq \tau_2$ , then  $\mathbf{d}^* = \mathbf{d}^{(\kappa+1)}$  stop
29: else go to step 2
30: end if
31: return:  $\mathbf{F}_{\text{RF}}^* = \text{reshape}(\mathbf{d}^*)$  to  $N_t \times N_{\text{RF}}$  matrix.

```

Moreover, observe that the following conditions hold

$$\omega_r(\mathbf{d}^{(\kappa+1)}) \leq \omega_r^{(\kappa)}(\mathbf{d}^{(\kappa+1)}) < \omega_r^{(\kappa)}(\mathbf{d}^{(\kappa)}) = \omega_r(\mathbf{d}^{(\kappa)}), \quad (41)$$

which shows that $\mathbf{d}^{(\kappa+1)}$ is an improved feasible point over $\mathbf{d}^{(\kappa)}$ for problem (30). Thus, Algorithm 1 generates a sequence of upgraded feasible points by iteratively solving the problem

(30) until it converges to its locally optimal solution. It is worth noting that the solution given by Algorithm 1 is an optimal solution of the sub-problem (34), since strong duality holds between the problems (38) and (34), and its optimal solution is obtained from the limit solution of the bisection method. In the theoretical analysis, the optimal solution is obtained by setting the termination threshold to $\tau_2 = 0$ in Algorithm 1 for zero numerical error.

EPMO method: The BMM algorithm requires a bisection search for $M+1$ dual multipliers, which renders it highly complex. Therefore, we propose a low-complexity exact penalty-based manifold optimization (EPMO) method to solve (30). To relax the SINR constraint (30b) and TPC constraint (30c), we add them into the objective function as a penalty term and subsequently, the problem is solved by employing manifold optimization. To this end, let us redefine the SINR constraint (30b) as $\psi_m(\mathbf{d}) \triangleq (\max(0, g_m(\mathbf{d})))^2, \forall m$ and (30c) as $\chi_p(\mathbf{d}) \triangleq (\max(0, \omega_p(\mathbf{d})))^2$. Consequently, the equivalent OP for (30) can be expressed as

$$\min_{\mathbf{d}} f(\mathbf{d}) = \omega_r(\mathbf{d}) + \mu \left(\sum_{m=1}^M \psi_m(\mathbf{d}) + \chi_p(\mathbf{d}) \right) \quad (42)$$

s.t. (30d),

where $\mu > 1$ is a penalty factor. Specifically, μ is obtained by adopting sequential optimization, wherein the penalty parameter μ is increased successively, followed by solving the problem (42) until the solutions eventually converge to that of the original problem (30). Observe that the constraint (30d) represents a complex circle Riemannian manifold $\mathcal{M} = \{\mathbf{d} \in \mathbb{C}^{N_t N_{\text{RF}} \times 1} : |\mathbf{d}(l)| = 1, \forall 1 \leq l \leq N_t N_{\text{RF}}\}$. Therefore, (42) can be solved by using the manifold optimization method. Specifically, for $\mu > 1$, we adopt the Riemannian conjugate gradient (RCG) optimization method to find a near-optimal solution. Note that the RCG algorithm relies on computing the Riemannian gradient to obtain the steepest direction in the decreasing objective function. However, computing the Riemannian gradient differs from obtaining the traditional gradient in the Euclidean space. Toward this, let us evaluate the Euclidean gradient of the objective function $f(\mathbf{d})$ as follows

$$\nabla f(\mathbf{d}) = 2\Xi_r \mathbf{d} - 2\mathbf{a}_r + \mu \left(\sum_{m=1}^M \xi_m + \xi_p \right), \quad (43)$$

where ξ_m and ξ_p are given by

$$\xi_m = \begin{cases} 4g_m(\mathbf{d})\Delta_m\mathbf{d}, & \text{if } g_m(\mathbf{d}) \geq 0, \\ 0, & \text{otherwise,} \end{cases} \quad (44a)$$

$$\xi_p = \begin{cases} 4\omega_p(\mathbf{d})\Omega_p\mathbf{d}, & \text{if } \omega_p(\mathbf{d}) \geq 0 \\ 0, & \text{otherwise.} \end{cases} \quad (44b)$$

Furthermore, computing the Riemannian gradient involves the tangent space, which comprises the vectors that are tangential to any smooth curves on the manifold \mathcal{M} . In addition, the tangent space at a point \mathbf{d} on the complex circle manifold \mathcal{M} is defined as

$$T_{\mathbf{d}}\mathcal{M} = \{\mathbf{z} \in \mathbb{C}^{N_t N_{RF} \times 1} | \text{Re}(\mathbf{z} \odot \mathbf{d}^*) = \mathbf{0}_{N_t N_{RF} \times 1}\}. \quad (45)$$

Thereby, the Riemannian gradient $\nabla_{\mathcal{M}}f(\mathbf{d})$ can be obtained by projecting $\nabla f(\mathbf{d})$ onto the tangent space of the manifold \mathcal{M} using a projection operator, which is given by

$$\begin{aligned} \nabla_{\mathcal{M}}f(\mathbf{d}) &= \text{Proj}_{\mathbf{d}}\nabla f(\mathbf{d}) \\ &= \nabla f(\mathbf{d}) - \text{Re}\{\nabla f(\mathbf{d}) \odot \mathbf{d}^*\} \odot \mathbf{d}. \end{aligned} \quad (46)$$

Employing the Riemannian gradient, one can follow the same steps as that of the Euclidean space for optimization. Thus, the steepest search direction in the $(\kappa+1)$ th iteration is given by

$$\zeta^{(\kappa+1)} = -\nabla_{\mathcal{M}}f(\mathbf{d}^{(\kappa+1)}) + \rho T_{\mathbf{d}^{(\kappa)} \mapsto \mathbf{d}^{(\kappa+1)}}(\zeta^{(\kappa)}), \quad (47)$$

where $\zeta^{(\kappa)}$ is the search direction at $\mathbf{d}^{(\kappa)}$, ρ is Polak-Ribière's conjugate parameter [28] and $T_{\mathbf{d}^{(\kappa)} \mapsto \mathbf{d}^{(\kappa+1)}}(\zeta^{(\kappa)})$ is the transport operation used to map the search direction from its original tangent space to the current tangent space. The transport operation is expressed as

$$\begin{aligned} \mathcal{T}_{\mathbf{d}^{(\kappa)}}\mathcal{M} &\rightarrow \mathcal{T}_{\mathbf{d}^{(\kappa+1)}}\mathcal{M} : \\ \mathcal{T}_{\mathbf{d}^{(\kappa)} \mapsto \mathbf{d}^{(\kappa+1)}}(\zeta^{(\kappa)}) &= \\ \zeta^{(\kappa)} - \text{Re}\left\{\zeta^{(\kappa)} \odot \left(\mathbf{d}^{(\kappa+1)}\right)^*\right\} \odot \mathbf{d}^{(\kappa+1)}. \end{aligned} \quad (48)$$

Moreover, in the Euclidean gradient, the next point is computed as $\mathbf{d}^{(\kappa+1)} = \mathbf{d}^{(\kappa)} + \delta^{(\kappa)}\zeta^{(\kappa)}$ with $\delta^{(\kappa)}$ as step size, which lies on the tangent space $T_{\mathbf{d}}\mathcal{M}$. Therefore, to project the point to the manifold \mathcal{M} , we perform retraction mapping [16], which is given by

$$\text{Retr}_{\mathbf{d}} : \mathcal{T}_{\mathbf{d}}\mathcal{M} \rightarrow \mathcal{M} :$$

$$\mathbf{d}^{(\kappa+1)} = \left[\frac{(\mathbf{d}^{(\kappa)} + \delta^{(\kappa)}\zeta^{(\kappa)})_1}{|(\mathbf{d}^{(\kappa)} + \delta^{(\kappa)}\zeta^{(\kappa)})_1|}, \dots, \frac{(\mathbf{d}^{(\kappa)} + \delta^{(\kappa)}\zeta^{(\kappa)})_D}{|(\mathbf{d}^{(\kappa)} + \delta^{(\kappa)}\zeta^{(\kappa)})_D|} \right]^T, \quad (49)$$

where $\delta^{(\kappa)}$ is the step size at the κ th iteration, which is obtained by Armijo's backtracking line search algorithm [16]. Furthermore, Algorithm 2 summarizes the complete procedure of solving (30) using the EPMO method, which involves updating the penalty parameter μ until the measures of violating the constraints (30b) and (30c) satisfy the condition

$$\omega_r(\mathbf{d}_1)_{\mu_1} \geq \omega_r(\mathbf{d}_2)_{\mu_2}, \quad \forall m, \quad (50)$$

Algorithm 2 EPMO algorithm for solving (30)

Input: \mathbf{F}_{BB} and thresholds $\tau_3 > 0, \tau_4 > 0, 0 < c < 1$

Output: Optimal RF TPC \mathbf{F}_{RF}^*

```

1: Initialize:  $\mathbf{F}_{\text{RF}}, \kappa = 0, \mathbf{d}^{(\kappa)} = \text{vec}(\mathbf{F}_{\text{RF}}), \zeta^{(\kappa)} = -\nabla_{\mathcal{M}}f(\mathbf{d}^{(\kappa)})$ 
2: while  $(\|\nabla_{\mathcal{M}}f(\mathbf{d}^{(\kappa)})\|^2 \geq \tau_3)$  do
3:   find Armijo backtracking line search step size  $\delta^{(\kappa)}$ 
4:   obtain the next point  $\mathbf{d}^{(\kappa+1)}$  using the retraction (49)
5:   compute the Riemannian gradient  $\nabla_{\mathcal{M}}f(\mathbf{d}^{(\kappa+1)})$  using (46).
6:   evaluate the transport  $T_{\mathbf{d}^{(\kappa)} \mapsto \mathbf{d}^{(\kappa+1)}}(\zeta^{(\kappa)})$  using (48)
7:   determine the steepest direction  $\zeta^{(\kappa+1)}$  according to (47)
8:   set  $\kappa \leftarrow \kappa + 1$ 
9: end while
10: if  $\left(\sum_{m=1}^M \psi_m(\mathbf{d}^{(\kappa)}) + \chi_p(\mathbf{d}^{(\kappa)})\right) \leq \tau_4$ ,
11:   return  $\mathbf{d}^* = \mathbf{d}^{(\kappa)}$  stop
12: else
13:   update  $\mu = \frac{\mu}{c}$  and go to step 2
14: end if
15: return:  $\mathbf{F}_{\text{RF}}^* = \text{reshape}(\mathbf{d}^*)$  to  $N_t \times N_{\text{RF}}$  matrix

```

where \mathbf{d}_1 and \mathbf{d}_2 are the optimal solutions of problem (42) for given $\mu_1 < \mu_2$, respectively. Hence, the penalty factor μ is initialized to be a small number to find a good starting point, then gradually increased to tighten the penalty. Specifically,

$$\mu = \frac{\mu}{c}, \quad 0 < c < 1, \quad (51)$$

where c is a constant scaling parameter. A larger c may lead to a more precise solution with a longer running time.

2) *Sub-problem for \mathbf{F}_{BB} :* For the given \mathbf{F}_{RF} and \mathbf{U} , the resultant OP for \mathbf{F}_{BB} is given by

$$\min_{\mathbf{F}_{\text{BB}}} \|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} - \mathbf{F}_r\mathbf{U}\|_F^2 \quad \text{s. t.} \quad (19b) \text{ and } (15f). \quad (52)$$

To solve the problem (52), we reformulate the non-convex SINR constraint as a second-order cone (SOC) constraint by introducing a common phase shift for $\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},m}$. Thus, the equivalent second-order cone programming (SOCP) problem constructed for (52) is given by

$$\begin{aligned} \min_{\mathbf{F}_{\text{BB}}} \quad & \|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} - \mathbf{F}_r\mathbf{U}\|_F^2 \\ \text{s. t.} \quad & \left\| \frac{\mathbf{A}^H \mathbf{e}}{\sqrt{N_o}} \right\|_2 \leq \sqrt{1 + \frac{1}{\Gamma_m} \text{Re}(t_{m,n})}, \\ & \|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F \leq \sqrt{P_{\text{max}}}, \end{aligned} \quad (53)$$

where $t_{m,n} = \mathbf{h}_m^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},n}$, $\mathbf{A}(m,n) = t_{m,n}$, and $\mathbf{e} \in \mathbb{C}^{M \times 1}$ is the vector with one in its m th position and zero elsewhere. The OP (53) above is a widely known SOCP convex OP [32], which can be efficiently solved using a standard convex optimization tool package.

3) *Sub-Problem for \mathbf{U} :* For the given \mathbf{F}_{RF} and \mathbf{F}_{BB} , the OP constructed for \mathbf{U} is given by

$$\min_{\mathbf{U}} \|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} - \mathbf{F}_r\mathbf{U}\|_F^2 \quad \text{s. t.} \quad (15d). \quad (54)$$

Problem (54) is the orthogonal Procrustes problem (OPP) [30], which is the least-squares problem associated with a non-convex UM [30]. Interestingly, its optimal solution can be obtained via the singular value decomposition (SVD), which is given by

$$\mathbf{U} = \tilde{\mathbf{U}} \mathbf{I}_{N_{tar} \times M} \tilde{\mathbf{V}}^H, \quad (55)$$

where $\mathbf{I}_{N_{tar} \times M}$ is constructed by concatenating the $[N_{tar} \times N_{tar}]$ identity matrix and the $[N_{tar} \times (M - N_{tar})]$ zero matrix, while the matrices $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are derived from the following equation

$$\text{SVD}(\mathbf{F}_r^H \mathbf{F}_{RF} \mathbf{F}_{BB}) = \tilde{\mathbf{U}} \tilde{\Sigma} \tilde{\mathbf{V}}^H. \quad (56)$$

B. Outer layer: update $\{\beta_m\}_{m=1}^M$ and \mathcal{R}

For the fixed RBE, $\mathcal{E} = \|\mathbf{F}_{RF} \mathbf{F}_{BB} - \mathbf{F}_r \mathbf{U}\|_F^2$, the next step is to optimize the block length $\{\beta_m\}_{m=1}^M$ in the outer layer. The equivalent sub-problem is constructed for addressing the blocklength $\{\beta_m\}_{m=1}^M$ as follows

$$\begin{aligned} & \max_{\{\beta_m\}_{m=1}^M} \mathcal{R} \\ & \text{s. t.} \quad (15b), (15g), \text{ and } (15h). \end{aligned} \quad (57)$$

Proposition 2. *To find a point on the Pareto boundary for the given RBE \mathcal{E} , the block length constraint must be met with equality, i.e.,*

$$\sum_{m=1}^M \beta_m = N. \quad (58)$$

Proof. This is proved by considering two CUs in the system. For the given BB and RF TPC, the corresponding rates R_1 and R_2 of the two CUs for the system under consideration are given, for the block lengths of β_1 and β_2 , as well as for the decoding error probabilities of ϵ_1 and ϵ_2 , respectively, as:

$$R_1 = \ln(1 + \gamma_1) - \sqrt{\frac{V_1}{\beta_1}} Q^{-1}(\epsilon_1) \quad (59a)$$

$$R_2 = \ln(1 + \gamma_2) - \sqrt{\frac{V_2}{\beta_2}} Q^{-1}(\epsilon_2), \quad (59b)$$

where $\beta_1 + \beta_2 = N$. Let us assume that there exists a positive value \hat{N} such that $\hat{N} < N$. Based on this assumption, the corresponding point on the RBE-rate region should adhere to $\beta_1 + \beta_2 = \hat{N}$. Let us fix the block length of CU 1 to β_1 . Hence, the block length of CU 2 is given by $\hat{\beta}_2 = \hat{N} - \beta_1$. As a result, the achievable rate of CU 2 \hat{R}_2 for the block length of $\hat{\beta}_2$ is modified as follows

$$\hat{R}_2 = \ln(1 + \gamma_2) - \sqrt{\frac{V_2}{\hat{\beta}_2}} Q^{-1}(\epsilon_2). \quad (60)$$

Note that the achievable rate of the SPC regime monotonically increases with block length [3]. For the given $\hat{N} < N$, it follows that $\hat{\beta}_2 = \hat{N} - \beta_1 < N - \beta_1$. Hence, $\hat{R}_2 < R_2$, which reduces the achievable sum rate of the system. Therefore, for any achievable sum rate point on the Pareto boundary of the RBE-rate region, \hat{N} must be equal to N . This holds for more than two CUs. \square

Algorithm 3 Two layer bisection search (TLBS) algorithm for solving (15)

Input: \mathbf{F}_r , $\{\eta_m\}_{m=1}^M$, $\{\epsilon_m\}_{m=1}^M$, $\mathcal{R}_L = 0$, \mathcal{R}_U , \mathcal{E}_{\max} , P_{\max} , N , and thresholds $\tau_5 > 0$, $\tau_6 > 0$

- 1: **initialize:** \mathbf{F}_{RF} , \mathbf{F}_{BB} , \mathbf{U} , $\{\beta_m\}_{m=1}^M$, and $\mathcal{E}(\mathbf{U}, \mathbf{F}_{BB}, \mathbf{F}_{RF})$
- 2: **repeat**
- 3: $\mathcal{R} = (\mathcal{R}_L + \mathcal{R}_U)/2$
- 4: evaluate $\Gamma_m, \forall m$ using (17)
- 5: **repeat**
- 6: set $\kappa = 0$, $\mathcal{E}^{(\kappa)} = \infty$
- 7: given $\mathbf{F}_{BB}^{(\kappa)}$ and $\mathbf{U}^{(\kappa)}$, obtain $\mathbf{F}_{RF}^{(\kappa+1)}$ by solving (20)
- 8: given $\mathbf{F}_{RF}^{(\kappa+1)}$ and $\mathbf{U}^{(\kappa)}$, evaluate $\mathbf{F}_{BB}^{(\kappa+1)}$ by solving (53)
- 9: given $\mathbf{F}_{RF}^{(\kappa+1)}$ and $\mathbf{F}_{BB}^{(\kappa+1)}$, calculate $\mathbf{U}^{(\kappa+1)}$ by solving (54)
- 10: compute $\mathcal{E}^{(\kappa+1)} = \mathcal{E}(\mathbf{U}^{(\kappa+1)}, \mathbf{F}_{BB}^{(\kappa+1)}, \mathbf{F}_{RF}^{(\kappa+1)})$
- 11: set $\kappa \leftarrow \kappa + 1$
- 12: **until** $|(\mathcal{E}^{(\kappa)} - \mathcal{E}^{(\kappa-1)})/\mathcal{E}^{(\kappa)}| \leq \tau_5$
- 13: **if** $\mathcal{E}^{(\kappa)} \leq \mathcal{E}_{\max}$ and $\|\mathbf{F}_{RF}^{(\kappa)} \mathbf{F}_{BB}^{(\kappa)}\|_F^2 \leq P_{\max}$
- 14: obtain $\{\beta_m\}_{m=1}^M$ using (62)
- 15: **if** (62) is feasible
- 16: set $\mathcal{R}_L = \mathcal{R}$.
- 17: **else** set $\mathcal{R}_U = \mathcal{R}$.
- 18: **end if**
- 19: **else** set $\mathcal{R}_U = \mathcal{R}$.
- 20: **end if**
- 21: **until** $\mathcal{R}_U - \mathcal{R}_L \leq \tau_6$
- 22: **output:** \mathbf{F}_{RF} , \mathbf{F}_{BB} , $\{\beta_m\}_{m=1}^M$ and RBE-rate tuple $(\mathcal{E}, \mathcal{R})$

Moreover, for any target rate $\mathcal{R} \geq 0$, the constraint (15b) can be modified as follows

$$\beta_m \geq \left(\frac{\sqrt{V_m} Q^{-1}(\epsilon_m)}{\ln(1 + \gamma_m) - \eta_m \mathcal{R}} \right)^2. \quad (61)$$

Consequently, following Proposition 2 and (61), the modified block length optimization is given by

$$\text{Find : } [\beta_1, \dots, \beta_M] \text{ s. t. } (61), (58), \text{ and } (15h), \quad (62)$$

where (62) is a mixed integer program for fixed \mathbf{F}_{RF} , \mathbf{F}_{BB} and \mathbf{U} [26]. In order to solve (62), we first relax $\beta_m, \forall m$ into a continuous value, solve the resulting problem for $\tilde{\beta}_m$, and finally round each optimal continuous value $\tilde{\beta}_m^*$ to the nearest integer. To determine the best integer quantization block length and efficiently control the quantization error, we employ the following criterion for $\forall m$:

$$\beta_m^* = \begin{cases} \lfloor \tilde{\beta}_m^* \rfloor, & \text{if } \tilde{\beta}_m^* - \lfloor \tilde{\beta}_m^* \rfloor \leq \varsigma, \\ \lceil \tilde{\beta}_m^* \rceil, & \text{otherwise,} \end{cases} \quad (63)$$

where $\varsigma \in [0, 1]$ is chosen so that $\sum_{m=1}^M \beta_m^* = N$ is satisfied.

Finally, for a fixed BB TPC \mathbf{F}_{BB} , RF TPC \mathbf{F}_{RF} , and block length $\{\beta_m\}_{m=1}^M$, we update the achievable sum rate \mathcal{R} via the bisection search method [34].

C. Summary

This subsection describes the overall algorithm. Since the proposed algorithm relies on the BCD algorithm, we iteratively optimize the RF TPC \mathbf{F}_{RF} , BB TPC \mathbf{F}_{BB} and the block length $\beta_m, \forall m$, until convergence is achieved. Now let us discuss the initialization of $\mathbf{F}_{\text{RF}}^{(0)}, \mathbf{F}_{\text{BB}}^{(0)}, \mathbf{U}^{(0)}$ and $\beta_m^{(0)}, \forall m$, for the proposed TLBS algorithm.

1) $\mathbf{F}_{\text{RF}}^{(0)}$: we initialize the each element of $\mathbf{F}_{\text{RF}}^{(0)}$ as $e^{j\nu_{i,j}}$ with $\nu_{i,j}$ denoting random phases chosen from $[-\pi, \pi], \forall i \in N_t, \forall j \in N_{\text{RF}}$.

2) $\mathbf{U}^{(0)}$: recalling (55), the auxiliary matrix $\mathbf{U}^{(0)}$ is initialized as $\mathbf{U}^{(0)} = \tilde{\mathbf{U}}^{(0)} \mathbf{I}_{N_{\text{tar}} \times M} (\tilde{\mathbf{V}}^{(0)})^H$, where we generate unitary matrices $\tilde{\mathbf{U}}^{(0)}$ and $\tilde{\mathbf{V}}^{(0)}$ via the SVD of a random matrix.

3) $\mathbf{F}_{\text{BB}}^{(0)}$: for a given \mathbf{F}_r , $\mathbf{F}_{\text{BB}}^{(0)}$ is initialized by employing the least squares method [8], [27] as $\mathbf{F}_{\text{BB}}^{(0)} = \sqrt{P_{\text{max}}} \left((\mathbf{F}_{\text{RF}}^{(0)})^H \mathbf{F}_{\text{RF}}^{(0)} \right)^{-1} (\mathbf{F}_{\text{RF}}^{(0)})^H \mathbf{F}_r \mathbf{U}^{(0)}$, which satisfies the power constraint at the ISAC BS.

4) $\beta_m^{(0)}$: for the given total block length N and number of CUs M , we set $\beta_m^{(0)}$ as $\beta_m^{(0)} = \frac{N}{M}, \forall m$. Furthermore, the complete procedure of the proposed TLBS-based joint optimization is summarized in Algorithm 3.

IV. SIMULATION RESULTS

In this section, we present simulation results that characterize the Pareto boundaries of the RBE-rate region across various scenarios, along with the beam pattern, for an SPC-enabled mmWave ISAC system. The ISAC BS is assumed to be equipped with a ULA consisting of N_t transmit antennas and N_{RF} RFCs. Each CU and RT is assumed to be within a range of 100 meters from the ISAC BS, with the path loss for the mmWave channel modeled as $PL(d_m)$, given by [8], [9], [13], [32]

$$PL(d_m) [\text{dB}] = \varepsilon + 10\varphi \log_{10}(d_m) + \varpi, \quad (64)$$

where ϖ follows a distribution $\mathcal{CN}(0, \sigma_\varpi^2)$ with $\sigma_\varpi = 5.8 \text{ dB}$, $\varepsilon = 61.4$ and $\varphi = 2$ [32]. For the propagation environment, we set $N_{\text{clu}} = 5$ and $N_{\text{ray}} = 10$ per cluster, with an angular spread of 10 degrees [8]. Furthermore, the AoDs $\phi_{i,j}, \forall i, j$ are generated from a Laplacian distribution and uniformly distributed within the range $[-90^\circ, 90^\circ]$ [8]. We consider a scenario with two RTs and two CUs, and the system operates at a frequency of 28 GHz with a bandwidth of 251.1886 MHz. Consequently, the noise variance N_0 at each CU is set to $N_0 = -174 + 10 \log_{10} B = -90 \text{ dBm}$ [32]. Unless stated otherwise, the key simulation parameters are those listed in Table II. Moreover, all the simulation results are averaged over 100 channel realizations.

A. Convergence behavior of the proposed algorithms

In this subsection, we characterize the convergence behavior of the proposed TLBS-based joint optimization Algorithm 3 to solve (16). Observe that the inner layer of the proposed

Table II: Simulation Parameters and Corresponding Values

Parameter	Value
Maximum allowable power budget P_{max}	30 dBm
Number of transmit antennas, N_t	128
Number of RF chains, N_{RF}	{4, 6}
Maximum allowable block length, N	{128, 256}
Decoding error probability of each CU, ϵ	$\{10^{-5}, 10^{-6}\}$
Number of uRLLC CUs	2
Number of RTs	2
Radar beam pattern error, \mathcal{E}	{0.15, 0.45}
Noise power, N_0	-90 dBm
Target rate ratio of each CU, $\{\eta_1, \eta_2\}$	(0.5, 0.5)

algorithm employs the BCD method for iteratively updating $\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}$ and \mathbf{U} to minimize the RBE \mathcal{E} , in the $(\kappa + 1)$ th iteration, we have

$$\begin{aligned} \mathcal{E}(\mathbf{F}_{\text{RF}}^{(\kappa+1)}, \mathbf{F}_{\text{BB}}^{(\kappa+1)}, \mathbf{U}^{(\kappa+1)}) &\leq \mathcal{E}(\mathbf{F}_{\text{RF}}^{(\kappa+1)}, \mathbf{F}_{\text{BB}}^{(\kappa+1)}, \mathbf{U}^{(\kappa)}) \\ &\leq \mathcal{E}(\mathbf{F}_{\text{RF}}^{(\kappa+1)}, \mathbf{F}_{\text{BB}}^{(\kappa)}, \mathbf{U}^{(\kappa)}) \leq \mathcal{E}(\mathbf{F}_{\text{RF}}^{(\kappa)}, \mathbf{F}_{\text{BB}}^{(\kappa)}, \mathbf{U}^{(\kappa)}), \end{aligned} \quad (65)$$

where the RF TPC $\mathbf{F}_{\text{RF}}^{(\kappa)}$ is optimized via the BMM and EP MO methods. Moreover, the BMM method returns progressively improved feasibility points with lower values of the objective function, while the EP MO method determines descent directions within the feasible region of the complex circle Riemannian manifold to achieve the same goal. Furthermore, for the BB and RF TPCs designed, the outer layer subsequently optimizes the block length $\{\beta_m\}_{m=1}^M$ via solving (62) and updates the achievable sum rate via the bisection search method, until convergence is achieved.

Furthermore, Fig. 2a and Fig. 2b present the convergence of RBE \mathcal{E} in the inner layer of the proposed algorithms with respect to the block length N , the number of RFCs, N_{RF} , respectively, for a fixed sum rate of $\mathcal{R} = 10 \text{ bps/Hz}$. We compare the convergence performance of the proposed TLBS-BMM and TLBS-EP MO. One can observe from both figures that the RBE of the TLBS-BMM and TLBS-EP MO methods decreases monotonically, verifying the convergence of Algorithm 3 in the inner layer. Moreover, the RBE of the TLBS-EP MO method is much lower than that of the TLBS-BMM method for a fixed sum rate \mathcal{R} . This is due to the fact that the TLBS-BMM involves the approximation of the majorizer functions, which increases the RBE. However, the TLBS-EP MO method does not require such approximations, which reduces the gap between the optimal radar beamformer and the HBF designed. Furthermore, the RBE corresponding to $N = 256$ is lower than that of $N = 128$ since, upon increasing the block length, the SPC rate approaches the Shannon capacity, and hence for the given target rate, an increased quantum of transmit power is available towards the RTs. In a similar fashion, the RBE of $N_{\text{RF}} = 6$ is much lower than that of $N_{\text{RF}} = 4$, which is due to the fact that increasing N_{RF} improves the approximation of HBF for the ideal radar beamformer. Fig. 2c shows the convergence of the proposed algorithm in the outer layer comprising a binary search approach, for $\mathcal{R} = \{8, 10\} \text{ bps/Hz}$. Observe that the proposed algorithm converges within 10 iterations for both the BMM and EP MO techniques, which evidences the

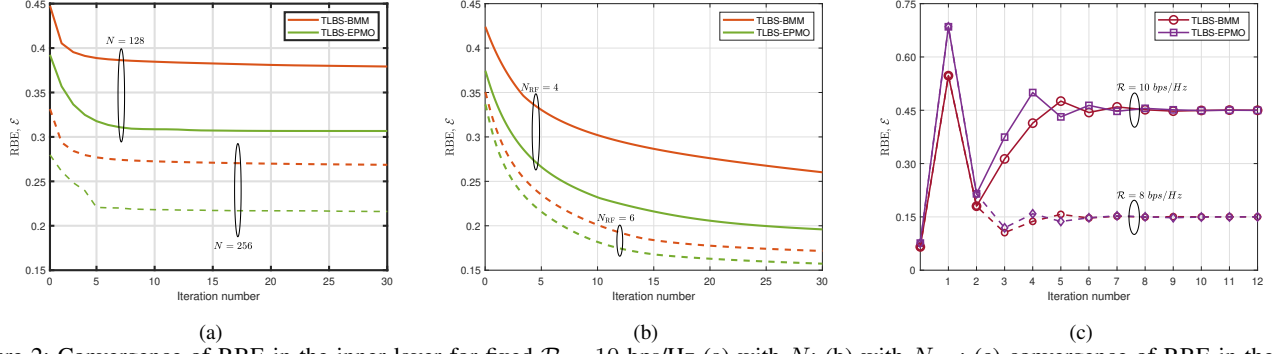


Figure 2: Convergence of RBE in the inner layer for fixed $\mathcal{R} = 10$ bps/Hz (a) with N ; (b) with N_{RF} ; (c) convergence of RBE in the outer layer with $N = 128$ and $N_{\text{RF}} = 4$.

convergence of the proposed TLBS algorithm. Moreover, the RBE is higher for $\mathcal{R} = 10$ bps/Hz than $\mathcal{R} = 8$ bps/Hz, which is due to the fact that a large sum rate requirement for the CUs reduces the power radiated towards the RTs.

Furthermore, to demonstrate the efficiency of the proposed algorithms and to glean interesting design insights, we compare the proposed method to the following schemes.

- *Scheme 1* (Optimal IBL-FDB): For this scheme, IBL is employed at the ISAC BS, which follows the Shannon capacity (SC) (11). Furthermore, FDB is used for designing the beamformer.
- *Scheme 2* (TLBS-FDB): This scheme corresponds to SPC transmission along with the FDB scheme to design the beamformer, where the TLBS algorithm is employed for optimizing the FDB and block length.
- *Scheme 3* (TLBS-OMP): In this scheme, we employ the orthogonal matching pursuit (OMP) [27] in the inner layer of the TLBS algorithm to optimize the RF and BB TPCs.

We compare the performance by evaluating the RBE-rate region and the sum rate versus several important parameters, which are discussed in the subsequent subsections.

B. Pareto boundary of the RBE-rate region

In this subsection, we investigate the behavior of the Pareto boundary of the RBE-rate region in SPC-enabled mmWave MIMO ISAC systems by varying some important parameters.

1) *Pareto boundary of RBE-rate region for different block lengths N* : In Fig. 3a, we plot the Pareto boundary of the RBE-rate region for block lengths of $N = 128$ and 256 at a fixed decoding error probability² of $\epsilon = 10^{-5}$ when the number of RFCs is $N_{\text{RF}} = 4$. As seen from the figure, the Pareto boundary of the RBE-rate region increases with N , since a larger N results in a higher sum rate, which reveals the impact of the block length on the system due to the SPC transmission. Moreover, the IBL-FDB scheme serves as the global upper bound for the RBE-rate region due to the resultant gain of the IBL transmission coupled with the FDB scheme. Meanwhile, TLBS-FDB acts as the local upper bound for the

proposed schemes in the SPC regime for both $N = 128$ and $N = 256$ due to the FDB scheme. Furthermore, the TLBS-EPMO scheme yields improved performance over the TLBS-BMM scheme and it is close to the locally optimal curve of the TLBS-FDB for both $N = 128$ and 256, which shows the efficacy of the RCG approach in the context of the EPMO technique. Moreover, both the proposed TLBS-EPMO and TLBS-BMM schemes are clearly superior to the TLBS-OMP method, which shows the effectiveness of the EPMO and BMM methods in optimizing the RF TPC.

2) *Pareto boundary of the RBE-rate region for different decoding error probabilities ϵ* : Fig. 3b investigates the impact of decoding error probability on the Pareto boundary of the RBE-rate region. As seen from the figure, the gap in the RBE-rate region increases as the decoding error probability decreases from $\epsilon = 10^{-5}$ to 10^{-6} dB since a reduction in the decoding error probability of the SPC regime results in an increase in the achievable rate. Thus, upon decreasing the decoding error probability, the power available for the RTs increases for a given sum rate, leading to an RBE reduction. Furthermore, the Pareto boundary of the proposed schemes is superior to that of the TLBS-OMP method for both $\epsilon = 10^{-5}$ and 10^{-6} dB, which shows the efficacy of the MM and RCG steps employed in the BMM and EPMO algorithms, respectively.

3) *Pareto boundary of the RBE-rate region for different RFCs N_{RF}* : Fig. 3c reveals the Pareto boundary of the RBE-rate region for $N_{\text{RF}} = \{4, 6\}$ along with $N = 128$ and $\epsilon = 10^{-5}$. It can be observed from the figure that the Pareto boundary of the RBE-rate region expands upon increasing the values of N_{RF} . This can be explained by the fact that the error between the ideal radar beamformer and the HBF designed decreases upon increasing N_{RF} , which therefore results in a reduced RBE. Consequently, more power is available for the CUs for a given RBE, which in turn leads to an increase in the sum rate. Furthermore, the Pareto boundary of the proposed TLBS-BMM and TLBS-EPMO methods approaches that of the locally optimal TLBS-FDB for both $N_{\text{RF}} = 4$ and 6. This shows that our proposed methods in the SPC regime achieve optimal performance with fewer RFCs. Hence, the proposed schemes save power and cost by employing the HBF scheme, while still achieving a performance that is close to that of

²Note that ϵ represents the decoding error probability due to SPC, whereas \mathcal{E} is the RBE.

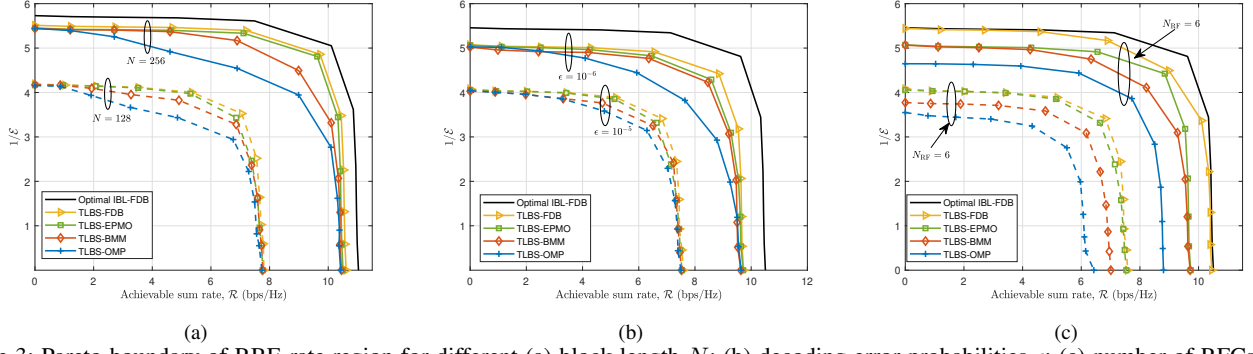


Figure 3: Pareto boundary of RBE-rate region for different (a) block length N ; (b) decoding error probabilities ϵ ; (c) number of RFCs N_{RF} .

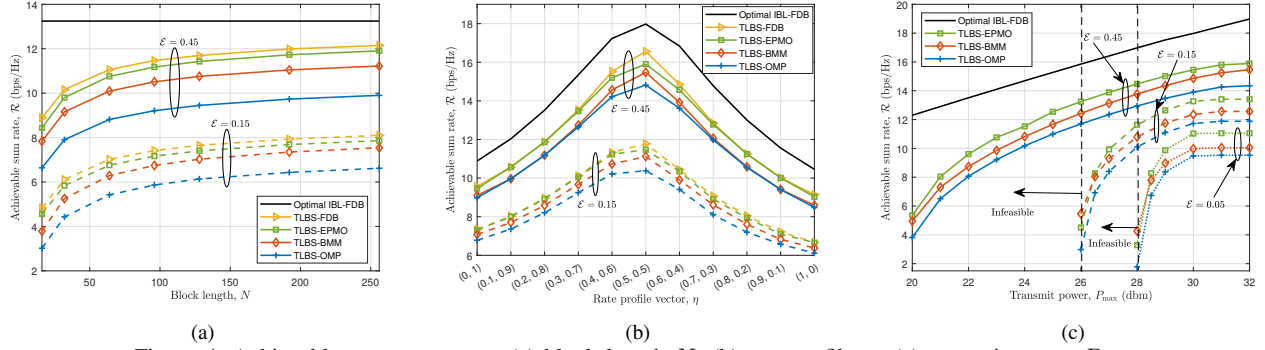


Figure 4: Achievable sum rate versus (a) block length N ; (b) rate profile η ; (c) transmit power P_{max} .

the optimal FDB scheme. Moreover, the Pareto boundary of the proposed TLBS-EPMO scheme is very close to that of the globally optimal IBL-FDB technique. Therefore, one can approach the Shannon capacity of the SPC-enabled mmWave MIMO ISAC system at a fixed block length and decoding error probability by increasing the number of RFCs N_{RF} in the TLBS-EPMO approach. It is worth noting that the radar sensing capabilities and SINR of the system are proportional to $\frac{1}{\epsilon}$ and \mathcal{R} , respectively. Therefore, one can also corroborate the Pareto boundary of the RBE-rate region by underlining the radar sensing and SINR trade-off.

C. Achievable sum rate of the SPC-enabled mmWave MIMO ISAC systems

1) *Achievable sum rate versus block length N* : In Fig. 4a, we plot the achievable sum rate versus the block length for different RBEs $\mathcal{E} = \{0.15, 0.45\}$. It can be seen from the figure that the IBL-FDB scheme is independent of the block length and acts as the global optimum. Moreover, the achievable sum rate increases upon increasing the block length N due to the influence of the block length on the rate expression. Furthermore, the sum rates of the proposed TLBS-EPMO and TLBS-BMM schemes approach that of the locally optimal TLBS-FDB, are seen to be improved over the TLBS-OMP scheme for increasing N , which shows the efficacy of the proposed designs. In addition, one can observe from the figure that reducing the RBE from $\mathcal{E} = 0.45$ to 0.15 results in a decrease in the achievable sum rate. This is due to the fact that reducing the RBE results in an increased focus on the RTs, leading to a reduced sum rate, as expected. Moreover, the

TLBS-EPMO scheme has a performance edge over its TLBS-BMM counterpart for both $\mathcal{E} = 0.15$ and $\mathcal{E} = 0.45$, which is due to the RCG step involved in the TLBS-EPMO approach conceived for the design of the RF beamformer.

2) *Achievable sum rate versus the rate profile vector η* : Fig. 4b investigates the impact of the rate profile vector η^3 on the achievable sum rate of the system for the RBEs of $\mathcal{E} = \{0.15, 0.45\}$. As discussed, the elements of η denote the target ratio of the m th CU rate and to the sum rate of the system and satisfy the constraint $\sum_{m=1}^M \eta_m = 1$ as associated with $\eta_m \in (0, 1)$. Therefore, in pair of the uRLLC CUs, we set η as $\eta = [\eta, 1 - \eta]$ and vary η from 0 to 1 with increments of 0.1. As seen from the figure, the achievable sum rate of the system is quasi-concave in nature with respect to the rate profile vector. Therefore, an optimal value of the rate profile vector exists at which the achievable rate is maximum.

3) *Achievable sum rate versus transmit power P_{max}* : We plot the achievable sum rate versus the transmit power in Fig. 4c for the fixed RBEs of $\mathcal{E} = \{0.05, 0.15, 0.45\}$. For a fixed RBE, the transmit power is a feasibility parameter for the TLBS algorithm. Therefore, as seen from the figure, for the RBE values of $\mathcal{E} = 0.05$ and $\mathcal{E} = 0.15$, the TLBS algorithm is infeasible for $P_t < 28$ dBm and $P_t < 26$ dBm, respectively. However, a large RBE of $\mathcal{E} = 0.45$ is always feasible for the TLBS algorithm. This trend is due to the fact that a small RBE leads to focusing a large fraction of the available power for the RTs. Hence, the power radiated towards the CUs is low, which

³In case of two CUs, rate profile vector η is given by $\eta = [\eta_1, \eta_2]$, where η_1 and η_2 represents the target rate ratio of CU 1 and 2, respectively, with $\eta_1 + \eta_2 = 1$.

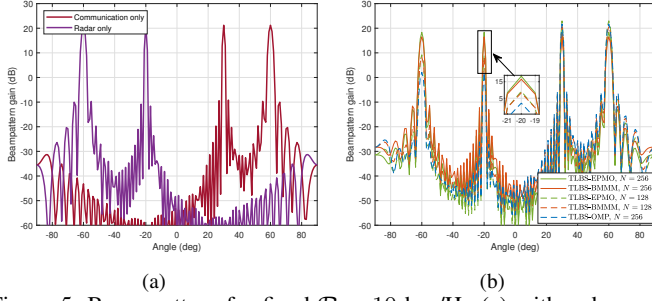


Figure 5: Beam pattern for fixed $\mathcal{R} = 10$ bps/Hz (a) with radar and communication only; (b) with HBF.

renders the problem infeasible due to the inability of achieving their QoS requirement. Moreover, the achievable sum rate of the system increases with the transmit power, and the proposed schemes yield an improved sum rate in comparison to the TLBS-OMP method.

D. Beampattern of the SPC-enabled mmWave MIMO ISAC system

For this scenario, we consider the RTs and uRLLC CUs to be located at $[-60^\circ, -20^\circ]$ and $[30^\circ, 60^\circ]$, respectively. Therefore, the desired beampattern is given by

$$G_d(\theta_i) = \begin{cases} 1, & \theta_i \in (\bar{\theta}_i \pm \sigma_\theta), i = 1, 2, \\ 0, & \text{otherwise,} \end{cases} \quad (66)$$

where $\bar{\theta}$ is the direction of the target and σ_θ denotes a constant angular spread of σ_θ , which is assumed to be $\frac{1}{\sqrt{2}}$. Fig. 5a shows the ideal beam pattern of the radar- and communication-only beamformer. As seen from the figure, the main lobes of the beam pattern are directed towards the location of the RTs and the communicating CUs. Furthermore, Fig. 5b plots the beam pattern of the proposed HBF schemes and compares it to the baseline for block lengths of $N = 128$ and 256 for a fixed RBE of $\mathcal{E} = 0.15$ and sum rate of $\mathcal{R} = 10$ bps/Hz. As seen from the figure, the main lobes of the HBF beampattern are directed toward the RTs and the CUs. Moreover, the beamforming gain of the system toward the RTs is higher for $N = 256$ than $N = 128$. This is due to the fact that a large N increases the sum rate. As a result of this, higher power is available for the target at a given sum rate and RBE. Additionally, the gain of the proposed schemes is higher than that of the TLBS-OMP method, which once again vindicates the efficacy of the EP MO and BMM algorithms conceived.

E. Complexity analysis

We now evaluate the overall computational complexity of the proposed TLBS algorithm. In the inner layer, the complexity of computing the RF TPC \mathbf{F}_{RF} via the BMM and EP MO methods is $\mathcal{O}(\mathcal{I}_b K N_t^2 N_{RF}^2)$ and $\mathcal{O}(\mathcal{I}_e N_t^2 N_{RF}^2)$, respectively, where \mathcal{I}_b and \mathcal{I}_e are the number of iterations required to update \mathbf{d} in the BMM and EP MO methods. Furthermore, the complexities involved in obtaining the BB TPC \mathbf{F}_{BB} , the auxiliary matrix \mathbf{U} and block length $\{\beta_m\}_{m=1}^M$ are given as $\mathcal{O}(N_{RF}^{3.5} M^{3.5})$, $\mathcal{O}(N_t N_{tar} N_{RF})$ and $\mathcal{O}(\mathcal{I}_{BL} M^2)$, respectively, where \mathcal{I}_{BL} is the number of

Table III: Computational Complexity Comparison

Schemes	Computational Complexity
IBL-FDB	$\mathcal{O}(N_t^4)$
TLBS-FDB	$\mathcal{O}(N_t^4)$
TLBS-BMM	$\mathcal{O}(K N_t^2 N_{RF}^2)$
TLBS-EP MO	$\mathcal{O}(N_t^2 N_{RF}^2)$
TLBS-OMP	$\mathcal{O}(N_t^3 N_{RF})$

bisection iterations. Therefore, the overall complexity of the TLBS algorithm harnessing the BMM and EP MO methods, namely TLBS-BMM and TLBS-EP MO, is given by $\mathcal{O}[I_{out} I_{in} (\mathcal{I}_b K N_t^2 N_{RF}^2 + N_{RF}^{3.5} M^{3.5} + N_t N_{tar} N_{RF})] + \mathcal{O}(\mathcal{I}_{BL} M^2)$ and $\mathcal{O}[I_{out} I_{in} (\mathcal{I}_e N_t^2 N_{RF}^2 + N_{RF}^{3.5} M^{3.5} + N_t N_{tar} N_{RF})] + \mathcal{O}(\mathcal{I}_{BL} M^2)$, respectively, where I_{out} and I_{in} denote the number of iterations required in the outer and inner layers. Observe that a significant fraction of the computational complexities of the proposed algorithms arise from the optimization of the RF TPC. For completeness, we compare the dominant complexities at each iteration for various schemes by omitting the lower order term in Table III. Since both the IBL-FDB and TLBS-FDB schemes employ the fully digital TPCs, which require N_t number of RF chains, they have higher complexities. It is worth noting that the major complexity of optimizing the RF TPC via the OMP scheme originates from calculating the inner product between the codebook matrix and the residual matrix [27]. Considering the optimal size of the codebook, the overall complexity of the TLBS-OMP scheme is given by $\mathcal{O}(N_t^3 N_{RF})$ [27]. In summary, the complexity of the proposed schemes is proportional to N_t^2 , which is relatively low compared to the benchmark scheme. This demonstrates the feasibility of the proposed schemes for real-time implementation in practical SPC-aided mmWave ISAC systems.

V. CONCLUSION

Pareto-optimal joint HBF and block length designs were conceived by considering the SPC transmission in the mmWave ISAC systems to meet the uRLLC requirements of the CUs, while also accomplishing sensing of the RTs. To this end, a Pareto-optimal framework was developed for characterizing the RBE-rate region of the model considered via the optimization of the RF and BB TPCs, as well as the block lengths. A novel TLBS algorithm was proposed for HBF design that comprises two layers. The inner layer computed the RF and BB TPCs minimizing the RBE of the RTs for a fixed sum rate of the CUs. Subsequently, the outer layer achieved block length optimization and evaluated the sum rate achievable for the given RBE. As a further advance, a pair of algorithms were proposed to design the RF TPC for the given system, namely, the BMM and EP MO schemes, which are based on the MM and RCG principles, respectively. Finally, simulation results were presented for characterizing the complete RBE-rate region and the sum rate of the system achievable for various parameter settings. The results evidence the fact that, through careful design, the mmWave ISAC system relying on SPC achieves the performance of an ideal

IBL-aided mmWave ISAC system, despite using substantially fewer RFCs and a finite block length. Thus, the proposed design is cost- and power-efficient, while supporting uRLLC services in 6G ISAC mmWave systems. Furthermore, future research should focus on the extension of this work to more practical scenarios with imperfect channel state information (CSI) that may degrade the communication and sensing performance, necessitating the development of robust beamforming techniques that account for channel uncertainties.

REFERENCES

- [1] X. Gao, W. Yi, Y. Liu, and L. Hanzo, "Multi-objective optimization of URLLC-based metaverse services," *IEEE Trans. Commun.*, vol. 71, no. 11, pp. 6745–6761, 2023.
- [2] G. J. Sutton *et al.*, "Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives," *IEEE Commun. Surv. Tutor.*, vol. 21, no. 3, pp. 2488–2524, 2019.
- [3] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
- [4] T. Li *et al.*, "Robust beamforming design with finite blocklength for URLLC," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2604–2608, 2022.
- [5] F. Liu, C. Masouros *et al.*, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, 2020.
- [6] Z. Wei *et al.*, "Integrated sensing and communication signals toward 5G-A and 6G: A survey," *IEEE Internet Things J.*, vol. 10, no. 13, pp. 11 068–11 092, 2023.
- [7] C. Qi, W. Ci, J. Zhang, and X. You, "Hybrid beamforming for millimeter wave MIMO integrated sensing and communications," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1136–1140, 2022.
- [8] J. Singh *et al.*, "Joint hybrid transceiver and reflection matrix design for RIS-aided mmWave MIMO cognitive radio systems," *IEEE Trans. Cognitive Commun. and Networking*, pp. 1–1, 2024.
- [9] J. Gong *et al.*, "Hybrid beamforming for millimeter wave integrated sensing and communications," in *ICC 2023 - IEEE Int. Conf. Commun.*, 2023, pp. 2958–2963.
- [10] X. He and J. Wang, "QCQP with extra constant modulus constraints: Theory and application to SINR constrained mmWave hybrid beamforming," *IEEE Trans. Signal Process.*, vol. 70, pp. 5237–5250, 2022.
- [11] P. Gao, L. Lian, and J. Yu, "Cooperative ISAC with direct localization and rate-splitting multiple access communication: A pareto optimization framework," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1496–1515, 2023.
- [12] Y. Xiong *et al.*, "On the fundamental tradeoff of integrated sensing and communications under gaussian channels," *IEEE Trans. Inf. Theory*, vol. 69, no. 9, pp. 5723–5751, 2023.
- [13] J. Singh, S. Srivastava, and A. K. Jagannatham, "Energy-efficient hybrid beamforming for integrated sensing and communication enabled mmWave MIMO systems," in *2024 IEEE 99th Veh. Technol. Conf (VTC2024-Spring)*, 2024, pp. 1–6.
- [14] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.
- [15] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE communications magazine*, vol. 57, no. 8, pp. 84–90, 2019.
- [16] F. Liu *et al.*, "Toward dual-functional radar-communication systems: Optimal waveform design," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4264–4279, 2018.
- [17] F. Liu, C. Masouros *et al.*, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2755–2770, 2018.
- [18] P. Cao, "Pareto optimal analog beamforming design for integrated MIMO radar and communication," *IEEE Signal Processing Letters*, 2022.
- [19] C. Ouyang, Y. Liu, and H. Yang, "MIMO-ISAC: Performance analysis and rate region characterization," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 669–673, 2023.
- [20] J. Zou, S. Sun, C. Masouros, Y. Cui, Y. Liu, and D. W. K. Ng, "Energy-efficient beamforming design for integrated sensing and communications systems," *arXiv preprint arXiv:2307.04002*, 2023.
- [21] Z. Yu *et al.*, "Active RIS-aided ISAC systems: Beamforming design and performance analysis," *IEEE Trans. Commun.*, vol. 72, no. 3, pp. 1578–1595, 2024.
- [22] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, "Cramér-rao bound optimization for joint radar-communication beamforming," *IEEE Trans. Signal Process.*, vol. 70, pp. 240–253, 2022.
- [23] J. Choi, J. Park, N. Lee, and A. Alkhateeb, "Joint and robust beamforming framework for integrated sensing and communication systems," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2024.
- [24] Y. Zhang, W. Ni, W. Tang, Y. C. Eldar, and D. Niyato, "Robust transceiver design for ISAC with imperfect CSI," in *Global Commun. Conf. (GLOBECOM)*, 2023, pp. 1320–1325.
- [25] J. Dai *et al.*, "Joint radar-communication beamforming considering both transceiver hardware impairments and imperfect CSI," *IEEE Wireless Commun. Lett.*, vol. 13, no. 7, pp. 1898–1902, 2024.
- [26] Y. Wang *et al.*, "Channel estimation for hybrid massive MIMO systems with adaptive-resolution ADCs," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 2131–2146, 2022.
- [27] O. E. Ayach *et al.*, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.
- [28] X. Yu *et al.*, "Hybrid beamforming in mmWave massive MIMO for IoV with dual-functional radar communication," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 9017–9030, 2023.
- [29] X. Wang *et al.*, "Partially-connected hybrid beamforming design for integrated sensing and communication systems," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6648–6660, 2022.
- [30] F. Liu and C. Masouros, "Hybrid beamforming with sub-arrayed MIMO radar: Enabling joint sensing and communication at mmwave band," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. IEEE, 2019, pp. 7770–7774.
- [31] Z. Xiao *et al.*, "Simultaneous multi-beam sweeping for mmWave Massive MIMO integrated sensing and communication," *IEEE Trans. Veh. Technol.*, vol. 73, no. 6, pp. 8141–8152, 2024.
- [32] J. Singh *et al.*, "Multi-beam object-localization for millimeter-wave ISAC-aided connected autonomous vehicles," *IEEE Trans. Veh. Technol.*, pp. 1–6, 2024.
- [33] M. Bayraktar *et al.*, "Hybrid precoding and combining for mmWave full-duplex joint radar and communication systems under self-interference," in *ICC 2024 - IEEE Int. Conf. Commun.*, 2024, pp. 915–920.
- [34] W. Huang, S. Xiao, L. Wu, C. Kai, S. He, and C. Li, "Achievable rate region for urllc interference channel with finite blocklength transmission," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 8857–8868, 2023.
- [35] S. He *et al.*, "Beamforming design for multiuser URLLC with finite blocklength transmission," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8096–8109, 2021.
- [36] B. Zhu *et al.*, "Resource allocation and beamforming design for downlink MISO-URLLC systems," in *in Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*. IEEE, 2022, pp. 1–5.
- [37] C. Ye *et al.*, "Beamforming optimization in MU-MISO URLLC systems," in *in Proc. Int. Conf. Commun. Technol. (ICCT)*. IEEE, 2022, pp. 520–525.
- [38] C. Ding *et al.*, "Joint precoding for MIMO radar and URLLC in ISAC systems," in *Proceedings of the 1st ACM MobiCom Workshop on Integrated Sensing and Communications Systems*, 2022, pp. 12–18.
- [39] F. Zabini *et al.*, "Joint sensing and communications in finite block-length regime," in *Global Commun. Conf. (GLOBECOM)*. IEEE, 2022, pp. 5595–5600.
- [40] X. Shen, N. Zhao, and Y. Shen, "On the performance tradeoff of an ISAC system with finite blocklength," in *ICC 2023 - IEEE Int. Conf. Commun.*, 2023, pp. 4628–4633.
- [41] N. Huang *et al.*, "Mobile edge computing aided integrated sensing and communication with short-packet transmissions," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7759–7774, 2024.
- [42] X. Zhao and Y.-J. A. Zhang, "Joint beamforming and scheduling for integrated sensing and communication systems in URLLC," in *Global Commun. Conf. (GLOBECOM)*. IEEE, 2022, pp. 3611–3616.
- [43] J. Singh *et al.*, "Hybrid transceiver design and optimal power allocation for the cognitive mmWave multiuser MIMO downlink relying on limited feedback," *IEEE Open J. Veh. Technol.*, vol. 4, pp. 241–256, 2023.

- [44] M. Salarpour *et al.*, “Synchronization-phase alignment of all-digital phase-locked loop chips for a 60-GHz MIMO transmitter and evaluation of phase noise effects,” *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 3187–3199, 2019.
- [45] C. Baquero Barneto *et al.*, “Full-duplex OFDM radar with LTE and 5G NR waveforms: Challenges, solutions, and measurements,” *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 4042–4054, 2019.
- [46] A. Gupta *et al.*, “An affine precoded superimposed pilot-based mmWave MIMO-OFDM ISAC system,” *IEEE Open J. Commun. Soc.*, vol. 5, pp. 1504–1524, 2024.
- [47] J. Singh *et al.*, “Pareto optimal hybrid beamforming for short-packet millimeter-wave integrated sensing and communication,” 2024. [Online]. Available: <https://arxiv.org/abs/2406.01945>