Robust Energy-Efficient Analog Beamforming under Short Packets and Per-Antenna Power Constraints

Jordi Borras, *Member, IEEE*, Francesc Molina, *Member, IEEE*, Roberto López-Valcarce, *Senior Member, IEEE*, Josep Sala-Álvarez, *Senior Member, IEEE*, and Gregori Vazquez, *Senior Member, IEEE*

***Abstract*—Designing energy-efficient beamformers is a major challenge to the deployment of battery-powered devices transmitting short data packets at millimeter-wave frequencies. In general, directly maximizing energy efficiency will not meet the stringent reliability requirements of certain Internet-of-Things applications. We focus on designing analog beamformers in a device-to-device setting under short packets, a finite encoder set, and imperfect channel knowledge. To guarantee the required error decoding performance, a two-step procedure is proposed such that beamformers are designed to minimize the power consumption subject to reliability constraints under worst-case channel estimation errors, and then the channel encoder maximizing the energy efficiency is selected. The proposed design enjoys a moderate loss under practical low-resolution phase shifters, and exhibits robustness to inaccurate channel estimators, which is crucial to the tradeoff between energy efficiency and error decoding performance.**

***Index Terms*—Analog beamforming, energy efficiency, millimeter-wave communication, short-packet communication, finite-resolution phase shifters.**

I. Introduction

MULTIPLE-INPUT MULTIPLE-OUTPUT (MIMO) at millimeter wave (mmWave) bands is gaining considerable momentum to satisfy the demanding system throughput and to alleviate the spectrum shortage of sub-6 GHz bands [1], [2]. At these frequencies, antenna arrays can be miniaturized and packed into small devices [3], making mmWave communications an attractive solution to enable several sensitive use-cases of Internet of Things (IoT) and Machine-Type Communication (MTC) [4]. In this vein, mmWave MIMO technologies have been considered to achieve low latency and high reliability in some connected vehicle applications, such as autonomous vehicles and pre-crash sensing [5], and to support the implementation of dual-functional radar-communication systems in Internet of Vehicles (IoV) [6]. In the context of industrial IoT, mmWave communications have been identified as an enabler for time-sensitive applications, such as smart manufacturing and augmented reality [7], [8]. However, the implementation of digital MIMO systems at these bands poses severe limitations, due to increased power consumption of the dense antenna arrays needed to overcome the large path losses [9], [10]. This is an important drawback for implementing MIMO processors in battery-powered IoT/MTC devices.

A popular approach to reduce power consumption is to implement fully analog [11]–[13] or hybrid analog-digital beamforming strategies [10], [14]–[16], which permit reducing the number of radio-frequency (RF) chains and power-expensive analog components. In this paper, given the low-complexity hardware requirements of IoT/MTC devices [4], fully analog architectures are adopted as they offer a good performance-complexity tradeoff [17].

A. Motivation and Related Works

Whereas beamformer design in mmWave systems is typically cast as the maximization of attainable rate, energy efficiency (EE)-oriented designs may be preferable in order to account for power consumption. EE is defined as the ratio of information throughput to power consumption [18], and has been addressed in the literature from different points of view, *e.g.,* indirectly, by configuring the RF front-end to reduce the power consumption [19]–[22], or by directly maximizing EE either with fully analog [12] or hybrid analog-digital architectures [23]–[26].

The aforementioned works hinge on two key assumptions: (i) arbitrarily long data packets (which incur an increased system latency); and (ii) an infinite set of channel encoders (which permits selecting any feasible coding rate). These may be questionable in realistic IoT/MTC scenarios [27], [28]. For instance, ultra-reliable low-latency communication (URLLC) stands nowadays as the paradigm that accounts for time delivery constraints [29]: URLLC operates with short packets, wherein channel capacity does not appropriately reflect system performance [30], [31]. Due to their importance in IoT/MTC settings, short-packet communications have garnered substantial interest in recent years, in terms of, *e.g.*, low-complexity reliable channel coding schemes [32], or digital beamforming and resource allocation solutions [33]. Closer to our work, [34] recently proposed a hybrid precoder design maximizing the achievable rate and minimizing the decoding error probability, assuming an infinite encoder set, and [35] improves [34] by incorporating per-user Quality-of-Service (QoS) constraints. However, both [34] and [35] overlooked the interplay between EE and decoding error probability, which

Work supported by grants PID2022-136512OB-C21/C22 funded by MCIN/AEI/10.13039/501100011033 and by ERDF "A way of making Europe" (EU), and by Margarita Salas Fellowship.

J. Borras and R. López-Valcarce are with the atlanTTic Research Center, Universidade de Vigo, 36310 Vigo, Spain. E-mail: jordi.borras@ieee.org, valcarce@gts.uvigo.es.

F. Molina was with the Department of Signal Theory and Communications, Technical University of Catalonia, 08034 Barcelona, Spain, and is now with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain. E-mail: francesc.molina@upf.edu.

J. Sala-Álvarez and G. Vazquez are with the Department of Signal Theory and Communications, Technical University of Catalonia, 08034 Barcelona, Spain. E-mail: josep.sala@upc.edu, gregori.vazquez@upc.edu.

is crucial at mmWave frequencies, since severe path losses reduce the operating Signal-to-Noise Ratio (SNR) and can compromise the reliability.

It is important to realize that in URLLC applications there is a threefold tradeoff between EE, reliability, and latency [28]. If analog beamformers are obtained with the sole goal of maximizing EE, and given that information throughput is a function of the Packet Error Rate (PER), the corresponding design may yield impractical PER values. An illustrative example is shown in Fig. 1, where the maximum EE design can only guarantee a PER of roughly 10^{-1}, significantly below typical reliability requirements for URLLC scenarios [36] which, according to the 3GPP Release 17 [37], range between $1-10^{-3}$ and $1-10^{-5}$; and with some applications demanding a reliability as stringent as $1-10^{-9}$ [38].

B. Contributions

We address the EE-based analog beamforming design for a point-to-point setting in the *finite-blocklength* regime, assuming that the transmitter is equipped with a *finite set* of channel encoders. In contrast to conventional constant-modulus analog beamforming (CMAB) approaches, we assume the transmitter is not only equipped with switches and phase shifters (PSs) but also with per-antenna variable-gain amplifiers (VGAs), similarly to, *e.g.,* [39], [40]. For the receiver node, a low-complexity architecture consisting of a phased array is adopted. To satisfy the demanding reliability requirements and circumvent the limitations illustrated in Fig. 1, transmit and receive beamformers are sought to minimize power consumption, subject to a maximum PER requirement; then, EE is maximized through a code allocation policy. It must be noted that, for a finite set of channel encoders, minimizing power consumption at a fixed coding rate and operating PER is approximately equivalent to maximizing EE when the PER requirement is sufficiently stringent, as in URLLC scenarios.

Some preliminary results relative to the present problem were presented in our previous work [42], in which we considered the analog beamformer design for a single channel encoder, in terms of minimum power consumption under perfect channel state information (CSI). However, the applicability of the design from [42] is limited by its computational complexity and its lack of robustness to CSI errors. Channel estimation in mmWave bands is a cumbersome task [43], emphasizing the importance of robust designs [44], [45]. In analog beamforming, robustness to CSI uncertainty is even more critical as channel estimators have larger errors [46] given the necessity of adopting beamspace angle-of-arrival (AoA) estimators, whose precision depend on the quantization bits of finite-resolution PSs and the size of predefined beamspace codebooks [47], [48].

Our main contributions can be summarized as follows:

- The design of energy-efficient analog transmit and receive beamformers is cast as a power consumption minimization subject to SNR constraints. To account for CSI errors, we adopt the worst-case SNR model in terms of mutual information [49], *i.e.*, CSI uncertainty contributes as an extra noise term degrading the SNR.

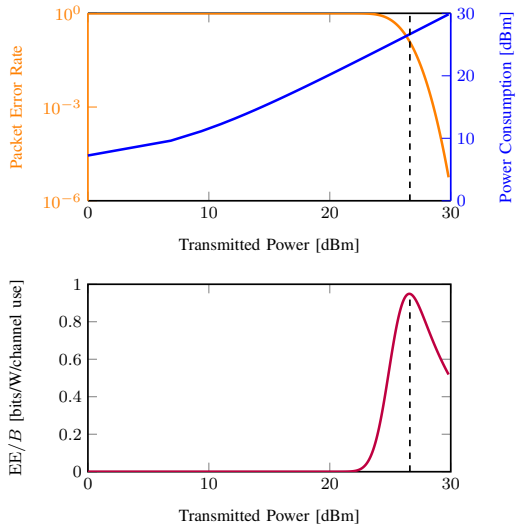


Fig. 1: Maximum-EE analog beamforming performance. Top: PER (as per [30]) and power consumption (as per [41]) vs. transmit power. Bottom: EE normalized by system bandwidth vs. transmit power. Setup: transmitter equipped with a 16-element fully analog array with unquantized phase shifters, per-antenna variable gain control, and a single channel encoder with rate $R = 1/2$ [see (1)]; single-antenna receiver. Packet length: 100 symbols. All channel paths have the same gain, such that the only design parameter is the power allocation, and the SNR depends only on transmit power.

- We prove that to attain the reliability constraint, some of the transmit antennas can be turned off; then, transmit power across the remaining active antennas is proportional to effective channel gains, except for the strongest channels which are limited by the per-antenna power constraint. To circumvent the non-convexity in determining the active antennas, we present an efficient cyclic approach, whose convergence is guaranteed and numerically illustrated to occur in a few iterations.
- For finite-resolution PSs, exhaustive search over all possible configurations is avoided by replacing the objective function by an appropriate lower bound which can be readily maximized.
- At the simulation level, we evaluate the performance of the proposed designs and their robustness to CSI acquisition errors. We discuss the advantages of optimizing the power and code allocation to improve the EE performance and the importance of appropriately setting the size of the uncertainty region to balance the EE-PER tradeoff even in the presence of slight CSI estimation errors.

C. Organization and Notation

The system model and problem statement are presented in Sec. II, whereas Sec. III describes the proposed analog beamformer design. Simulation results are presented in Sec. IV, and conclusions are drawn in Sec. V.

Notation: Boldface lowercase (resp., uppercase) symbols denote vectors (resp., matrices). Sets are denoted by calligraphic uppercase. $\boldsymbol{A}^T$ and $\boldsymbol{A}^H$ denote the transpose and the transpose conjugate of matrix $\boldsymbol{A}$, respectively. $\boldsymbol{I}_K$ is the $K \times K$ identity matrix. $\|\cdot\|_{\mathrm{F}}$ and $\|\cdot\|_p$ stand for the Frobenius and L_p norms, $\odot$ stands for the Schur-Hadamard (element-wise) product, $\mathcal{Q}[\theta; \mathcal{S}]$ returns the element of $\mathcal{S}$ closest to

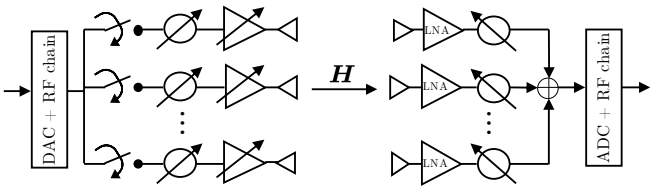


Fig. 2: Analog beamforming system model adopted in this work. The transmit beamformer is composed of switches, phase shifters, and VGAs; the receive beamformer comprises fixed-gain low-noise amplifiers and phase shifters.

θ (modulo 2π), and $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a circular complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

II. System Model and Problem Formulation

A. System Model

We consider a point-to-point communication link in the mmWave band, with a single data stream being transmitted using short packets. The transmitter is equipped with p encoders $\mathcal{R} \doteq \{\mathfrak{R}_1, \ldots, \mathfrak{R}_p\}$, each of them producing packets of the same blocklength n. We consider that encoder $\mathfrak{R}_q \in \mathcal{R}$ comprises both the channel code and the constellation mapper; it maps m_q information bits to n symbols with the coding rate

$$R_q = \frac{m_q}{n} \quad \text{(info. bits per symbol)}. \tag{1}$$

To enable single-stream transmission in a power-efficient and hardware-saving manner, transmitter and receiver incorporate a single RF chain each, and are equipped with fully analog beamforming architectures with L_t and L_r antennas, respectively. Each transmit antenna is preceded by an on-off switch, a PS, and a VGA with maximum output power P. Each of the L_r receive antennas is connected to a PS via a low-noise amplifier (LNA), with no individual control on amplitude gains. The considered models are shown in Fig. 2. The transmit and receive beamformers are respectively denoted by $\boldsymbol{f} \in \mathbb{C}^{L_t}$ and $\boldsymbol{w} \in \mathbb{C}^{L_r}$. The analog implementation imposes constraints on their entries, *i.e.*, $f_i \in \mathcal{F}$ for $1 \leq i \leq L_t$, and $w_j \in \mathcal{W}$ for $1 \leq j \leq L_r$; the feasible sets are given by

$$\mathcal{F} \doteq \left\{ re^{j\phi} \,|\, \phi \in \mathcal{S}_f \,,\, r \in [0,\, \sqrt{P}] \right\}, \tag{2}$$

$$\mathcal{W} \doteq \left\{ e^{j\psi} \,|\, \psi \in \mathcal{S}_w \right\}. \tag{3}$$

Note that (2) incorporates the associated per-antenna power constraint, and that $\boldsymbol{w}^H\boldsymbol{w} = L_r$ for all $\boldsymbol{w} \in \mathcal{W}^{L_r}$. The phase sets $\mathcal{S}_f$, $\mathcal{S}_w$ contain the feasible phase values and have cardinalities $|\mathcal{S}_f| = 2^{b_t}$ and $|\mathcal{S}_w| = 2^{b_r}$, where b_t, b_r are the PS bit-resolutions at the transmitter and receiver, respectively. We assume that VGA resolution is sufficiently high, so that they can deliver any output power level in $[0, P]$.

The mmWave channel is assumed frequency-flat and characterized by matrix $\boldsymbol{H} \in \mathbb{C}^{L_r \times L_t}$. Using pilot signals, each node obtains a channel estimate $\widehat{\boldsymbol{H}}$. The true channel matrix reads as $\boldsymbol{H} = \widehat{\boldsymbol{H}} + \boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ stands for the channel estimation error. For the sake of generality, the only assumption we make on $\boldsymbol{\Delta}$ is that it belongs to the bounded isotropic uncertainty set (cf. [50], [51]) given by

$$\mathcal{E} \doteq \left\{ \boldsymbol{\Delta} \mid \|\boldsymbol{\Delta}\|_{\mathrm{F}}^2 \leq \delta^2 \right\}, \tag{4}$$

where the *uncertainty level* $\delta^2 > 0$ defines the radius of $\mathcal{E}$. The chosen value of δ^2 is a designer's choice that mainly depends on the adopted channel estimator and the second-order statistics of the estimation error (see, *e.g.*, [52]).

Let x be the transmitted symbol with zero mean and variance σ_x^2. After combining and frequency downconversion, the received signal is

$$\begin{aligned} y &= \boldsymbol{w}^H\boldsymbol{H}\boldsymbol{f}x + \boldsymbol{w}^H\boldsymbol{n} \\ &= \boldsymbol{w}^H\widehat{\boldsymbol{H}}\boldsymbol{f}x + \boldsymbol{w}^H\boldsymbol{\Delta}\boldsymbol{f}x + \boldsymbol{w}^H\boldsymbol{n}, \end{aligned} \tag{5}$$

where $\boldsymbol{n} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \sigma_n^2\boldsymbol{I}_{L_r})$ is the additive noise. The ratio $\gamma \doteq \frac{\sigma_x^2}{\sigma_n^2}$ is referred to as *pre-processing* SNR. Following the guidelines[1] from [49], [53], and given that the estimation error $\boldsymbol{\Delta}$ is unknown to both transmitter and receiver, we consider a worst-case scenario in which the component due to channel estimation errors, $\boldsymbol{w}^H\boldsymbol{\Delta}\boldsymbol{f}x$, is regarded as an additional noise term uncorrelated with the useful signal component $\boldsymbol{w}^H\widehat{\boldsymbol{H}}\boldsymbol{f}x$. In this way, the *post-processing* SNR, *i.e.,* the SNR measured after receive beamforming, is defined as

$$\Gamma \doteq \frac{\left|\boldsymbol{w}^H\widehat{\boldsymbol{H}}\boldsymbol{f}\right|^2\gamma}{\boldsymbol{w}^H\boldsymbol{w} + |\boldsymbol{w}^H\boldsymbol{\Delta}\boldsymbol{f}|^2\gamma} = \frac{\left|\boldsymbol{w}^H\widehat{\boldsymbol{H}}\boldsymbol{f}\right|^2\gamma}{L_r + |\boldsymbol{w}^H\boldsymbol{\Delta}\boldsymbol{f}|^2\gamma}. \tag{6}$$

Under short packets, Shannon capacity does not adequately reflect the relationship between coding rates, PER, and the blocklength n. Following [29], [30], for a given PER ϵ and under Gaussian signaling, each channel encoder $\mathfrak{R}_q \in \mathcal{R}$, with blocklength n and coding rate R_q, satisfies:

$$R_q = C(\Gamma) - \sqrt{\frac{V(\Gamma)}{n}}Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log_2 n}{n}\right), \tag{7}$$

where $C(\Gamma) \doteq \log_2(1+\Gamma)$ is the Shannon capacity, $V(\Gamma) \doteq (1-(1+\Gamma)^{-2})(\log_2 e)^2$ is the channel dispersion, and $Q(\cdot)$ is the tail probability of a standard Gaussian distribution. From (7), the PER vs. SNR characteristic of $\mathfrak{R}_q \in \mathcal{R}$ reads

$$\mathrm{PER}_q[\Gamma] \approx Q\left(\sqrt{n}\frac{C(\Gamma) - R_q}{\sqrt{V(\Gamma)}}\right), \tag{8}$$

where Γ is given by (6).

B. Problem Formulation

EE is defined as the ratio of the information throughput τ_q achieved by the q-th encoder to the total power P_{tot} consumed by the communication process:

$$\mathrm{EE}[\mathfrak{R}_q, \boldsymbol{f}, \boldsymbol{w}] = \frac{\tau_q}{P_{\mathrm{tot}}} \quad \text{(bits/J)}. \tag{9}$$

Some details on τ_q and P_{tot} are as follows:

- *Information throughput*: We assume a packet transmission time of T seconds and linearly modulated signals of bandwidth B Hz, with each symbol transmitted over

[1] As studied in the seminal works [49], [53], a lower bound on the mutual information is given precisely when channel estimation error is regarded as an extra noise term uncorrelated with the error-free signal component. Thus, the adopted approach permits addressing the design of beamformers robust to the worst-case decoding performance.

one channel use. Packets are transmitted over $n = BT$ channel uses, for which the information throughput reads

$$\tau_q = BR_q(1 - \mathrm{PER}_q[\Gamma]) \quad \text{(bits/s)}. \tag{10}$$

- *Total power consumption*: In the architecture considered, switches are configured to deactivate transmission together with the power supply of elements in the same branch; as in [41], only their power consumption during switching transition is considered. Then

$$P_{\mathrm{tot}} = \|\boldsymbol{f}\|_2^2 + b\|\boldsymbol{f}\|_0 + cL_r + d, \tag{11}$$

where: $\|\boldsymbol{f}\|_2^2$ is the transmit power; $b \doteq P_{\mathrm{sw}} + P_{\mathrm{ps}} + P_{\mathrm{vga}}$ is the power used by the transmitter analog processing (switching transition, PS and VGA), and whose specific value depends on circuitry implementation; $c \doteq P_{\mathrm{lna}} + P_{\mathrm{ps}}$ is the power consumed by PSs and LNAs at the receiver; and $d \doteq P_{\mathrm{dac}} + P_{\mathrm{adc}} + 2P_{\mathrm{rfc}}$ is the power consumption of Digital-to-Analog Converter (DAC), Analog-to-Digital Converter (ADC), and RF chains, respectively. Note that $b, c, d > 0$.

To meet the stringent PER requirement in URLLC applications, we introduce a maximum PER constraint $\mathrm{PER}_q[\Gamma] \leq \epsilon$, so that the analog beamformer design problem can be cast as

$$\max_{\substack{\boldsymbol{w}\in\mathcal{W}^{L_r}, \boldsymbol{f}\in\mathcal{F}^{L_t} \\ \mathfrak{R}_q\in\mathcal{R}}} \quad \mathrm{EE}[\mathfrak{R}_q, \boldsymbol{f}, \boldsymbol{w}] \tag{12a}$$

$$\text{s.to} \quad \max_{\mathfrak{R}_q\in\mathcal{R}, \boldsymbol{\Delta}\in\mathcal{E}} \ \mathrm{PER}_q[\Gamma] \leq \epsilon. \tag{12b}$$

Since the coding rate of each $\mathfrak{R}_q \in \mathcal{R}$ and the blocklength are fixed, the following equivalence applies to constraint (12b):

$$\max_{\substack{\mathfrak{R}_q\in\mathcal{R} \\ \boldsymbol{\Delta}\in\mathcal{E}}} \mathrm{PER}_q[\Gamma] \leq \epsilon \iff \min_{\substack{\mathfrak{R}_q\in\mathcal{R} \\ \boldsymbol{\Delta}\in\mathcal{E}}} \Gamma \geq \mathrm{PER}_q^{-1}[\epsilon], \tag{13}$$

where the minimum SNR requirement is computed from the known PER vs SNR curve (8).

Note that, for a given encoder $\mathfrak{R}_q$, and in view of (10), the energy efficiency is bounded for all feasible PER values as $\frac{BR_q(1-\epsilon)}{P_{\mathrm{tot}}} \leq \mathrm{EE}[\mathfrak{R}_q, \boldsymbol{f}, \boldsymbol{w}] \leq \frac{BR_q}{P_{\mathrm{tot}}}$. Since ϵ will be small in practice, it follows that the optimal beamformers for a given channel encoder approximately minimize power consumption. In view of this, we proceed in two steps as follows. Since the set of available channel encoders is discrete, the optimum beamformers for each $\mathfrak{R}_q \in \mathcal{R}$ are determined first, and then the best channel encoder is selected:

- For each encoder $\mathfrak{R}_q \in \mathcal{R}$, solve

$$(\boldsymbol{f}_q, \boldsymbol{w}_q) = \arg\min_{\boldsymbol{f}\in\mathcal{F}^{L_t}, \boldsymbol{w}\in\mathcal{W}^{L_r}} P_{\mathrm{tot}}[\boldsymbol{f}] \tag{14a}$$

$$\text{s.to} \quad \min_{\boldsymbol{\Delta}\in\mathcal{E}} \ \Gamma \geq \mathrm{PER}_q^{-1}[\epsilon] \tag{14b}$$

and compute the corresponding value of $\mathrm{EE}[\mathfrak{R}_q, \boldsymbol{f}_q, \boldsymbol{w}_q]$.
- Then, choose $\{\mathfrak{R}_{q^\star}, \boldsymbol{f}_{q^\star}, \boldsymbol{w}_{q^\star}\}$ satisfying

$$q^\star = \arg\max_{1\leq q\leq p} \mathrm{EE}[\mathfrak{R}_q, \boldsymbol{f}_q, \boldsymbol{w}_q]. \tag{15}$$

III. Energy-Efficient Analog Beamforming Design

In this section, we present a computationally-efficient scheme to determine the optimal configuration according to the design problem stated in Sec. II-B. We first determine the worst-case post-processing SNR featuring in the constraint (14b), by solving

$$\Gamma_0 = \arg\min_{\boldsymbol{\Delta}} \ \Gamma \ \text{ s.to } \ \|\boldsymbol{\Delta}\|_{\mathrm{F}}^2 \leq \delta^2, \tag{16}$$

where the constraint follows from the definition of the uncertainty set (4). By virtue of the Cauchy-Schwarz inequality and the constraint in (16), one has

$$|\boldsymbol{w}^H\boldsymbol{\Delta}\boldsymbol{f}|^2 \ \leq \ \|\boldsymbol{\Delta}\|_{\mathrm{F}}^2\|\boldsymbol{w}\|_2^2\|\boldsymbol{f}\|_2^2 \ \leq \ \delta^2 L_r\|\boldsymbol{f}\|_2^2, \tag{17}$$

since $\|\boldsymbol{w}\|_2^2 = L_r$ for all $\boldsymbol{w} \in \mathcal{W}^{L_r}$. This upper bound is achieved when $\boldsymbol{\Delta} = \sigma\boldsymbol{u}\boldsymbol{v}^H$, *i.e.*, a rank-one matrix with $\boldsymbol{u} = \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$, $\boldsymbol{v} = \frac{\boldsymbol{f}}{\|\boldsymbol{f}\|}$, and $\sigma = \delta$. Thereby, recalling (6),

$$\Gamma_0(\boldsymbol{f}, \boldsymbol{w}) = \frac{\gamma}{L_r} \frac{\left|\boldsymbol{w}^H\widehat{\boldsymbol{H}}\boldsymbol{f}\right|^2}{1 + \gamma\delta^2\|\boldsymbol{f}\|_2^2}, \tag{18}$$

where we have emphasized the dependence of the worst-case SNR with the beamforming vectors. Then, problem (14) can be rewritten as follows: for each encoder $\mathfrak{R}_q \in \mathcal{R}$, solve

$$\min_{\boldsymbol{f}\in\mathcal{F}^{L_t}, \boldsymbol{w}\in\mathcal{W}^{L_r}} \quad P_{\mathrm{tot}}[\boldsymbol{f}] \tag{19a}$$

$$\text{s.to} \quad \Gamma_0(\boldsymbol{f}, \boldsymbol{w}) \geq \mathrm{PER}_q^{-1}[\epsilon]. \tag{19b}$$

Note that problem (19) is not convex, because: (i) the cost function $P_{\mathrm{tot}}[\boldsymbol{f}]$, given by (11), contains an L_0-norm, (ii) the feasible sets $\mathcal{F}^{L_t}$ and $\mathcal{W}^{L_r}$ are discrete since PSs have finite resolution, and (iii) constraint (19b) is not convex in $\boldsymbol{f}$ or $\boldsymbol{w}$. To sidestep these drawbacks, we first discuss a tractable approach for the case of unquantized PSs. Subsequently, we appropriately modify the solutions so obtained to account for finite-resolution PSs.

A. The Case of Infinite-Resolution PSs

With unquantized PSs, the feasible sets for the entries of transmit and receive beamforming vectors become $\mathcal{F} \doteq \{z \in \mathbb{C}, |z| \leq \sqrt{P}\}$ and $\mathcal{W} \doteq \{z \in \mathbb{C}, |z| = 1\}$. Even in this case, the coupling between $\boldsymbol{f}$ and $\boldsymbol{w}$ introduced by Γ_0 makes it difficult to find a closed-form solution to problem (19). However, if either of the transmit or receive beamforming vectors is kept fixed, the optimal value of the other vector can be found, suggesting the following cyclic design:

1) *Receive beamformer design*. Since $\boldsymbol{w}$ is composed solely of PSs, the objective in (19a) does not depend on $\boldsymbol{w}$. Thus, for a given $\boldsymbol{f}$, the optimum $\boldsymbol{w}$ can be found as

$$\max_{\boldsymbol{w}\in\mathcal{W}^{L_r}} \ \frac{1}{L_r} \frac{\left|\boldsymbol{w}^H\widehat{\boldsymbol{H}}\boldsymbol{f}\right|^2 \gamma}{1 + \gamma\delta^2\|\boldsymbol{f}\|_2^2}. \tag{20}$$

$$\lambda = \frac{\pm\sqrt{\rho_q B\left[B(1+\gamma\delta^2\ell P)+\gamma\delta^2 PA^2-(1+\gamma\delta^2\ell P)\gamma\delta^2\rho_q\right]}-\sqrt{P}AB}{B\left(B-\rho_q\gamma\delta^2\right)}, \text{ with } A=\sum_{i=1}^{\ell}|v_i| \text{ and } B=\sum_{i=\ell+1}^{k}|v_i|^2 \quad (28)$$

2) *Transmit beamformer design*. Let us define the effective channel after combining as $\boldsymbol{v} \doteq \widehat{\boldsymbol{H}}^H\boldsymbol{w}$, and let $\rho_q \doteq \mathrm{PER}_q^{-1}[\epsilon]\frac{L_r}{\gamma}$. Then, for fixed $\boldsymbol{w}$, problem (19) reads as

$$\min_{\boldsymbol{f}} \quad \|\boldsymbol{f}\|_2^2 + b\|\boldsymbol{f}\|_0 \quad (21a)$$

$$\text{s.to} \quad 0 \le |f_i|^2 \le P, \ \ 1 \le i \le L_t, \quad (21b)$$

$$\frac{|\boldsymbol{v}^H\boldsymbol{f}|^2}{1+\delta^2\gamma\|\boldsymbol{f}\|_2^2} \ge \rho_q. \quad (21c)$$

These steps are then iterated until convergence, and for each channel encoder $\mathfrak{R}_q \in \mathcal{R}$, after which the optimal configuration is obtained via (15). The overall procedure is sketched in Algorithm 1. Next, we present the solutions to (20) and (21).

Regarding subproblem (20), and recalling that the receiver can only tune the phases of each antenna, it is readily seen that the solution is an equal-gain combiner, given element-wise by

$$w_i = \exp(j\measuredangle(\widehat{\boldsymbol{H}}\boldsymbol{f})_i) \ \text{ for } \ i = 1,\ldots,L_r. \quad (22)$$

Whereas subproblem (20) has the simple closed-form solution in (22), further elaboration is needed to solve subproblem (21), as it is not convex in its current form.

Note that (21a), (21b), and the denominator of the left-hand side of (21c) depend on $\boldsymbol{f}$ through the magnitudes of its entries only. Hence, the optimal phases are those maximizing the numerator of the left-hand side of (21c), *i.e.*,

$$\measuredangle(f_i) = \measuredangle(v_i), \text{ for } 1 \le i \le L_t, \quad (23)$$

such that subproblem (21) becomes real-valued. Using (23), one has $|\boldsymbol{v}^H\boldsymbol{f}|^2 = \left(\sum_{i=1}^{L_t}|v_i||f_i|\right)^2$, so that (21c) reads as

$$\frac{\sum_{i=1}^{L_t}|v_i||f_i|}{\sqrt{1+\delta^2\gamma\sum_{i=1}^{L_t}|f_i|^2}} \ge \sqrt{\rho_q}, \quad (24)$$

which is a second-order cone constraint, and thus convex. Therefore, the real-valued version of subproblem (21) is only not convex due to the L_0-norm in the cost function. To solve it efficiently, we next present a sequential approach which explores all possible L_t cases corresponding to the number of active transmit antennas[2].

We begin by assuming the magnitudes of the elements in $\boldsymbol{v}$ arranged in non-increasing order, *i.e.*,

$$|v_1| \ge |v_2| \ge \cdots \ge |v_{L_t}|. \quad (25)$$

This is without loss of generality, since the entries of $\boldsymbol{v}$ and $\boldsymbol{f}$ can always be appropriately re-labeled.

Let $\boldsymbol{f}^\star$ be the solution to (21), and let $k^\star = \|\boldsymbol{f}^\star\|_0$. If a permutation is applied to the entries of $\boldsymbol{f}^\star$ (of which only $k^\star$ are nonzero), the objective in (21a) remains invariant, and the constraint in (21b) is also satisfied. Therefore, the optimal subset of non-zero entries of $\boldsymbol{f}^\star$ maximizes the left-hand side of inequality (24), as its right-hand side is also permutation-invariant. By the rearrangement inequality [54] and in view of (25), it follows that the subset of non-zero entries of $\boldsymbol{f}^\star$ is precisely $\{f_1^\star, f_2^\star, \ldots, f_{k^\star}^\star\}$, and that these satisfy $|f_1^\star| \ge |f_2^\star| \ge \cdots \ge |f_{k^\star}^\star|$. These observations indicate that the solution to (21a)–(21c) can be obtained by solving the following L_t convex problems, for $k = 1,\ldots,L_t$:

$$\mathbb{P}_k: \quad \min_{\{|f_i|\}_{1\le i\le k}} \quad \sum_{i=1}^{k}|f_i|^2 + bk \quad (26a)$$

$$\text{s.to} \quad 0 \le |f_i| \le \sqrt{P}, \ \ 1 \le i \le k \quad (26b)$$

$$\frac{\sum_{i=1}^{k}|v_i||f_i|}{\sqrt{1+\delta^2\gamma\sum_{i=1}^{k}|f_i|^2}} \ge \sqrt{\rho_q}. \quad (26c)$$

Then, denoting the minimum value of the objective for problem $\mathbb{P}_k$ as J_k, the optimal precoder satisfies $\|\boldsymbol{f}^\star\|_0 = k^\star = \arg\min_{1\le k\le L_t} J_k$. If problem $\mathbb{P}_k$ is unfeasible, it means that activating k transmit antennas is not enough to guarantee the desired operating PER.

As shown in Appendix A, problem $\mathbb{P}_k$ in (26) admits the following closed-form solution:

$$|f_i| = \begin{cases} \sqrt{P} & \text{if } 1 \le i \le \ell, \\ \lambda|v_i| & \text{if } \ell < i \le k, \\ 0 & \text{otherwise,} \end{cases} \quad (27)$$

where $\ell \le k$ denotes the number of active inequalities in the second constraint of $\mathbb{P}_k$, *i.e.*, the number of antennas transmitting at full power, and $\lambda \in \mathbb{R}_+$ is a scaling factor common for all $i = \ell+1,\ldots,k$, which is given in (28) on the top of this page. Observing (28), note that for each pair $\{k,\ell\}$ we have two possible values of λ, say λ_+ and λ_-, depending on the chosen sign. If both λ_+ and λ_- are negative, the pair $\{k,\ell\}$ cannot be optimal and need not be considered. Otherwise, the value of λ for a given pair $\{k,\ell\}$, namely $\lambda_{k,\ell}$, is selected according to the following rule:

$$\lambda_{k,\ell} = \begin{cases} \lambda_+ & \text{if } \lambda_+ > 0 \text{ and } \lambda_- < 0 \\ \lambda_- & \text{if } \lambda_+ < 0 \text{ and } \lambda_- > 0 \\ \min\{\lambda_-,\lambda_+\} & \text{if } \lambda_+ > 0 \text{ and } \lambda_- > 0 \end{cases} \quad (29)$$

In particular, when both λ_+ and λ_- are positive, the selected value is the one minimizing the power consumption. In any case, the scaling factor $\lambda_{k,\ell}$ must also satisfy the per-antenna transmit power constraint, *i.e.*,

$$\lambda_{k,\ell}^2 \le \frac{P}{|v_i|^2}, \ \ 1 \le i \le k. \quad (30)$$

Otherwise, the pair $\{k,\ell\}$ cannot be optimal and must not be considered.

The structure of (27) is illustrated in Fig. 3, in which ℓ antennas transmit at full power P and the remaining $k-\ell$

[2]Note that this is not the same as exploring all possible 2^{L_t} configurations of active/inactive transmit antennas by exhaustive search.

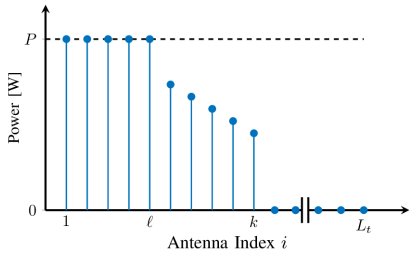


Fig. 3: General form of the power allocation coefficients $|f_i|^2$.

active antennas transmit with a power level proportional to their effective channel power gain.

To determine the number of antennas transmitting at full power (*i.e.,* ℓ), we sequentially try $\ell = 0, 1, \ldots, k$: for each pair $\{k, \ell\}$, the minimum value of the objective in $\mathbb{P}_k$ is

$$J_{k,\ell} \doteq \ell P + \lambda_{k,\ell}^2 \sum_{i=\ell+1}^{k} |v_i|^2 + bk, \tag{31}$$

so that $J_k = \min_{0 \le \ell \le k} J_{k,\ell}$.

The whole cyclic power minimization is summarized through the function CyclicMin in Algorithm 1. Note that the sequence of objective values along the iterations is non-increasing; thus, since the objective is lower bounded, this sequence must be convergent [55]. To start the procedure, the combiner is initialized based on the phases of the dominant left singular vector of the channel matrix, as this constitutes the optimum beamformer when the constraints on the feasible sets are relaxed.

Algorithm 1: Proposed Low-Complexity Energy-Efficient Analog Beamforming System Design

Input: $\widehat{\boldsymbol{H}}$, δ^2, $\mathcal{R}$, γ, ϵ, b, c, d, $\mathcal{F}$, $\mathcal{W}$
Output: $\{\mathfrak{R}^\star, \boldsymbol{f}^\star, \boldsymbol{w}^\star\}$

for $q = 1$ **to** p **do**
  $\rho_q \leftarrow (L_r/\gamma) \cdot \mathrm{PER}_q^{-1}[\epsilon]$;
  $\{\boldsymbol{f}_q, \boldsymbol{w}_q\} \leftarrow \mathrm{CyclicMin}(\widehat{\boldsymbol{H}}, \delta^2, \gamma, b, \rho_q, P)$;
  Calculate $\mathrm{EE}[\mathfrak{R}_q, \boldsymbol{f}_q, \boldsymbol{w}_q]$;
end
$q^\star \leftarrow \arg\max_q \{\mathrm{EE}[\mathfrak{R}_q, \boldsymbol{f}_q, \boldsymbol{w}_q]\}_{1 \le q \le p}$;
$\{\mathfrak{R}^\star, \boldsymbol{f}^\star, \boldsymbol{w}^\star\} \leftarrow \{\mathfrak{R}_{q^\star}, \boldsymbol{f}_{q^\star}, \boldsymbol{w}_{q^\star}\}$;

Function $\mathrm{CyclicMin}(\widehat{\boldsymbol{H}}, \delta^2, \gamma, b, \rho, P)$
  Compute economy-size SVD of $\widehat{\boldsymbol{H}}$;
  $\widehat{\boldsymbol{z}} \leftarrow$Dominant left singular vector of $\widehat{\boldsymbol{H}}$;
  Initialize $\boldsymbol{w} \leftarrow \exp(j\angle(\widehat{\boldsymbol{z}}))$;
  repeat
    $\boldsymbol{v} \leftarrow$ Sort $\{|(\widehat{\boldsymbol{H}}^H \boldsymbol{w})_i|\}_{1 \le i \le L_t}$ as in (25);
    for $k = 1$ **to** L_t **do**
      for $\ell = 0$ **to** k **do**
        Calculate λ via (28);
        Find $\lambda_{k,\ell}$ via (29);
        if $\lambda_{k,\ell} \in \mathbb{R}_+$ *and* (30) *holds* **then**
          Calculate $J_{k,\ell}$ via (31);
        else
          $J_{k,\ell} \leftarrow +\infty$;
        end
      end
    end
    $\{k^\star, \ell^\star\} \leftarrow \arg\min_k \min_\ell \; J_{k,\ell}$;
    Calculate $\boldsymbol{f}$ via (23) and (27);
    $\boldsymbol{w} \leftarrow \exp(j\angle(\widehat{\boldsymbol{H}}\boldsymbol{f}))$;
  until *convergence*;
  return $\{\boldsymbol{f}, \boldsymbol{w}\}$
end

B. The case of Finite-Resolution PSs

In Sec. III-A no quantization was assumed for the transmitter and receiver PSs. Practical PSs, however, can only synthesize a limited number of phase shifts. Therefore, we now introduce appropriate modifications to the cyclic optimization procedure developed in Sec. III-A to take into account the effect of finite-resolution PSs with quantized phases, so that the feasible sets are given by (2)–(3) with $\mathcal{S}_f$, $\mathcal{S}_w$ finite sets.

Consider first the design of the receive beamformer $\boldsymbol{w}$ for a fixed transmit beamformer $\boldsymbol{f}$. The optimum combiner for problem (14) should still maximize the post-processing SNR Γ as in (20), so that the problem becomes

$$\max_{\substack{\{\psi_j\}_{1 \le j \le L_r} \\ \psi_j \in \mathcal{S}_w}} |\boldsymbol{w}^H \widehat{\boldsymbol{H}} \boldsymbol{f}| \;\; \text{s.to} \;\; w_j = e^{\jmath\psi_j}, \; 1 \le j \le L_r. \tag{32}$$

For discrete $\mathcal{S}_w$, this has to be solved by exhaustive search, which is too demanding for antenna arrays of practical sizes in mmWave. But noting that $|z| \ge \mathrm{Re}\{z\}$ for all $z \in \mathbb{C}$, we propose to maximize instead the lower bound

$$\mathrm{Re}\{\boldsymbol{w}^H \widehat{\boldsymbol{H}} \boldsymbol{f}\} = \sum_{j=1}^{L_r} |(\widehat{\boldsymbol{H}}\boldsymbol{f})_j| \cos(\angle(\widehat{\boldsymbol{H}}\boldsymbol{f})_j - \angle(w_j)), \tag{33}$$

subject to $w_j = e^{\jmath\psi_j}$, with $\psi_j \in \mathcal{S}_w$, for $j = 1, \ldots, L_r$. The lower bound in (33) is maximized by choosing

$$\angle(w_j) = \mathcal{Q}\left[\angle(\widehat{\boldsymbol{H}}\boldsymbol{f})_j \,;\, \mathcal{S}_w\right], \qquad 1 \le j \le L_r. \tag{34}$$

Consider now the design of the transmit beamformer $\boldsymbol{f}$ for a fixed receive beamformer $\boldsymbol{w}$. With finite-resolution PSs, problem (21) has to be modified into

$$\min_{\boldsymbol{f}} \quad \|\boldsymbol{f}\|_2^2 + b\|\boldsymbol{f}\|_0 \tag{35a}$$

$$\text{s.to} \quad 0 \le |f_i|^2 \le P, \;\; 1 \le i \le L_t \tag{35b}$$

$$\angle(f_i) \in \mathcal{S}_f, \; 1 \le i \le L_t \tag{35c}$$

$$\frac{|\boldsymbol{v}^H \boldsymbol{f}|^2}{1 + \delta^2 \gamma \|\boldsymbol{f}\|_2^2} \ge \rho_q \tag{35d}$$

with $\boldsymbol{v} = \widehat{\boldsymbol{H}}^H \boldsymbol{w}$. Again, the phases of $\boldsymbol{f}$ only affect the numerator of the left-hand side of constraint (35d); thus, the optimal phases should maximize $|\boldsymbol{v}^H \boldsymbol{f}|$ subject to $\angle(f_i) \in \mathcal{S}_f$, $1 \le i \le L_t$. As above, we choose to maximize instead the lower bound $\mathrm{Re}\{\boldsymbol{v}^H \boldsymbol{f}\}$; this yields

$$\angle(f_i) = \mathcal{Q}\left[\angle(v_i) \,;\, \mathcal{S}_f\right], \qquad 1 \le i \le L_t. \tag{36}$$

However, with this choice, the product $\boldsymbol{v}^H \boldsymbol{f}$ is not real-valued in general, which makes the constraint (35d) difficult to handle.

TABLE I: Average execution time (in seconds) of Algorithm 1 when each $\mathbb{P}_k$ in (26) is solved as proposed in this work (Prop) and through an interior point method (IPM). Results displayed for different number of transmit antennas L_t, $L_r = 8$ receive antennas, different number of encoders p, different uncertainty levels $\delta^2 = \alpha\|\widehat{\boldsymbol{H}}\|_F^2$, and pre-processing SNR $\gamma = -5$ dB.

	Full-Resolution PSs					
	$\alpha = 0$			$\alpha = 0.5$		
$L_t \backslash p$	3	6	9	3	6	9
16	Prop: 0.0030 IPM: 321.1383	Prop: 0.0041 IPM: 610.4178	Prop: 0.0058 IPM: 827.7133	Prop: 0.0007 IPM: 61.4557	Prop: 0.0024 IPM: 363.5050	Prop: 0.0029 IPM: 430.2995
32	Prop: 0.0031 IPM: 656.3687	Prop: 0.0056 IPM: $1.1448 \cdot 10^3$	Prop: 0.0147 IPM: $1.8197 \cdot 10^3$	Prop: 0.0014 IPM: 201.0266	Prop: 0.0045 IPM: 779.2636	Prop: 0.0060 IPM: $1.1264 \cdot 10^3$
64	Prop: 0.0056 IPM: $1.1467 \cdot 10^3$	Prop: 0.0099 IPM: $2.5337 \cdot 10^3$	Prop: 0.0186 IPM: $3.8792 \cdot 10^3$	Prop: 0.0048 IPM: 279.5878	Prop: 0.0114 IPM: $1.6307 \cdot 10^3$	Prop: 0.0149 IPM: $1.9979 \cdot 10^3$
128	Prop: 0.0127 IPM: $2.4541 \cdot 10^3$	Prop: 0.0239 IPM: $5.0474 \cdot 10^3$	Prop: 0.0396 IPM: $8.1332 \cdot 10^3$	Prop: 0.0162 IPM: 756.4053	Prop: 0.0364 IPM: $3.9966 \cdot 10^3$	Prop: 0.0492 IPM: $4.4594 \cdot 10^3$
256	Prop: 0.0379 IPM: $5.1309 \cdot 10^3$	Prop: 0.0680 IPM: $1.1351 \cdot 10^4$	Prop: 0.1078 IPM: $1.5690 \cdot 10^4$	Prop: 0.0710 IPM: $1.5723 \cdot 10^3$	Prop: 0.1615 IPM: $8.2371 \cdot 10^3$	Prop: 0.1998 IPM: $8.3257 \cdot 10^3$

	Finite-Resolution PSs with $b_t = b_r = 4$ bits					
	$\alpha = 0$			$\alpha = 0.5$		
$L_t \backslash p$	3	6	9	3	6	9
16	Prop: 0.0056 IPM: 374.5637	Prop: 0.0083 IPM: 813.0884	Prop: 0.0102 IPM: 844.9409	Prop: 0.0011 IPM: 62.5332	Prop: 0.0046 IPM: 403.3777	Prop: 0.0044 IPM: 446.4164
32	Prop: 0.0063 IPM: 833.4372	Prop: 0.0120 IPM: $1.5734 \cdot 10^3$	Prop: 0.0212 IPM: $2.4916 \cdot 10^3$	Prop: 0.0021 IPM: 246.3836	Prop: 0.0065 IPM: 982.1577	Prop: 0.0078 IPM: $1.2143 \cdot 10^3$
64	Prop: 0.0104 IPM: $1.7041 \cdot 10^3$	Prop: 0.0208 IPM: $3.3425 \cdot 10^3$	Prop: 0.0330 IPM: $5.1382 \cdot 10^3$	Prop: 0.0058 IPM: 297.7993	Prop: 0.0132 IPM: $1.8024 \cdot 10^3$	Prop: 0.0190 IPM: $2.1749 \cdot 10^3$
128	Prop: 0.0207 IPM: $3.6335 \cdot 10^3$	Prop: 0.0453 IPM: $5.9395 \cdot 10^3$	Prop: 0.0673 IPM: $1.0233 \cdot 10^4$	Prop: 0.0304 IPM: 804.4059	Prop: 0.0381 IPM: $4.1741 \cdot 10^3$	Prop: 0.0542 IPM: $4.5114 \cdot 10^3$
256	Prop: 0.1283 IPM: $7.1574 \cdot 10^3$	Prop: 0.1524 IPM: $1.4427 \cdot 10^4$	Prop: 0.1611 IPM: $2.2171 \cdot 10^4$	Prop: 0.1587 IPM: $1.0744 \cdot 10^3$	Prop: 0.2110 IPM: $8.4329 \cdot 10^3$	Prop: 0.2332 IPM: $8.5213 \cdot 10^3$

To sidestep this problem, we replace it by the *tighter* constraint $\mathrm{Re}\{\boldsymbol{v}^H\boldsymbol{f}\} \geq \sqrt{\rho_q(1+\delta^2\gamma\|\boldsymbol{f}\|_2^2)}$, which reads as

$$\frac{1}{\sqrt{\rho}_q}\sum_{i=1}^{L_t}|v_i||f_i|\cos(\measuredangle(f_i)-\measuredangle(v_i)) \geq \sqrt{1+\delta^2\gamma\sum_{i=1}^{L_t}|f_i|^2}. \tag{37}$$

Note that the cosine terms in (37) will be non-negative provided that the phase quantization errors satisfy $|\measuredangle(f_i)-\measuredangle(v_i)| \leq \frac{\pi}{2}$ (modulo 2π); for example, with uniform phase quantization this holds for all resolutions $b_t \geq 1$. Therefore, the same approach as in Section III-A can be applied now, by replacing $|v_i|$ by $|v_i|\cos(\measuredangle(f_i)-\measuredangle(v_i))$ for $1 \leq i \leq L_t$, which should be rearranged in non-increasing order, and also replacing line 27 in Algorithm 1 by eq. (34).

C. Computational Complexity

Consider the CyclicMin function in Algorithm 1. The complexity of the initialization step (lines 9-11) is dominated by the singular value decomposition (SVD), which is $\mathcal{O}(\max(L_t, L_r)(\min(L_t, L_r))^2)$. Since only the left dominant singular vector is needed, this complexity can be further reduced by computing a reduced-rank SVD. Then, some steps (lines 13-27) are iterated until convergence. Finding the power allocation coefficients is the most onerous operation, involving the computation of λ (which is $\mathcal{O}(2L_t)$ since common parameters only need to be computed once) and the determination of the number of antennas that transmit at full power. Hence, the complexity of calculating the power allocation coefficients[3], *i.e.,* of solving problem $\mathbb{P}_k$ in (26), is $\mathcal{O}(2L_t^2)$. This has to be repeated for each possible number of active antennas and until convergence, *i.e.*, ML_t times, where M is the number of iterations to achieve convergence. Thus, the complexity of the CyclicMin function is $\mathcal{O}(2ML_t^3)$. With finite-resolution PSs, since (34) and (36) have complexity $\mathcal{O}(2L_rL_t+L_r+2^{b_r+1}L_r)$

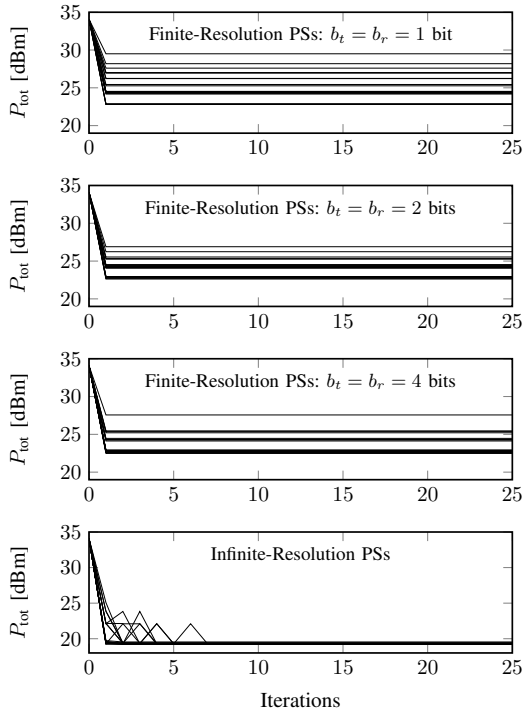
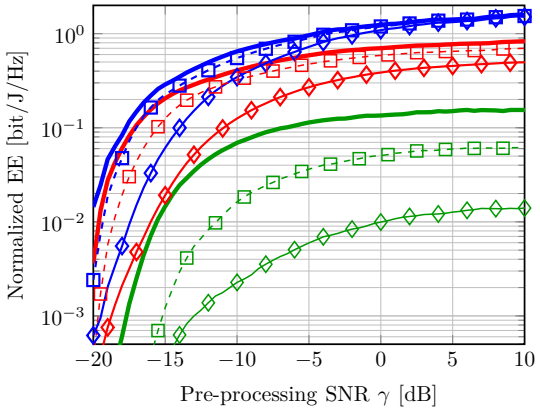
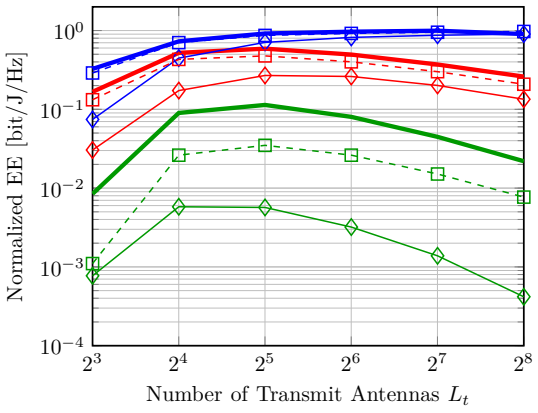


Fig. 4: Convergence of the proposed algorithm for different PS resolutions, with $L_t = 128$, $L_r = 16$, $\gamma = 5$ dB, and $\delta^2 = 0.5\|\widehat{\boldsymbol{H}}\|_F^2$.

[3]Note that each $\mathbb{P}_k$ in (26) is a second-order cone program (SOCP), which could be alternatively solved through standard interior point methods with worst-case complexity of $\mathcal{O}(L_t^3K)$, where K, which is about $\mathcal{O}(\sqrt{L_t})$, denotes the required number of iterations to solve each $\mathbb{P}_k$ [56].



(a) EE vs. γ for $L_t = 32$.



(b) EE vs. L_t for $\gamma = -5$ dB.

Fig. 5: Performance evaluation of the worst-case design proposed in Sec. III: (a) EE vs pre-processing SNR γ; (b) EE vs number of transmit antennas L_t. In both cases, the EE is plotted for different relative uncertainties: $\alpha = 0$ (blue), $\alpha = 0.1$ (red), and $\alpha = 0.5$ (green). Different PSs resolutions are assessed: infinite (thick solid), $b_t = b_r = 2$ bits (dashed with square markers), and $b_t = b_r = 1$ bit (thin solid with diamond markers).

and $\mathcal{O}(2L_tL_r + L_t + 2^{b_t+1}L_t)$, these operations do not incur additional complexity for practical values of b_t and b_r.

Regarding the main block of Algorithm 1, some operations are repeated for each channel encoder in $\mathcal{R}$. Since the PER vs. SNR characteristics of channel encoders are known, the values of $\{\rho_q\}_{1\le q\le p}$ can be computed offline and then selected from a lookup table. The calculation of EE requires $\mathcal{O}(3L_t)$, and the selection of the optimum configuration only requires p comparisons, where p is the cardinality of $\mathcal{R}$. Overall, the computational complexity of Algorithm 1 is $\mathcal{O}(2MpL_t^3)$.

To illustrate the convergence speed of the proposed iterative scheme, we depict in Fig. 4 the evolution of the cost function for 25 different random channels (see Sec. IV-A). It is seen that, in all cases, convergence is fast, taking only a few iterations.

Finally, we compare in Table I the average execution time of Algorithm 1 when each problem $\mathbb{P}_k$ is solved as proposed in this work and through an interior point method (implemented via CVX). These experiments have been run on the same machine for different transmit array sizes L_t, different number of channel encoders p, and different uncertainty levels. Modeling channels as in Sec. IV-A, we show results for both full-resolution and finite-resolution (with $b_t = b_r = 4$ bits) PSs. It is clear that the the proposed scheme is much faster (by several orders of magnitude) than the interior point-based approach, regardless of the parameter values. Remarkably, even with $L_t = 256$ transmit antennas, $p = 9$ channel encoders, and 4-bit PSs, the proposed scheme obtains the solution in much less than one second.

IV. Simulation Results

We evaluate the proposed scheme for a setting in the 60 GHz band with bandwidth $B = 100$ MHz. Both nodes are equipped with uniformly-spaced linear arrays (ULAs) with half-wavelength element separation. The array size at the receiver is $L_r = 8$; the transmitter will be specified afterward. We consider uniformly quantized PSs at both nodes. The per-antenna power constraint is $P = 20$ mW. The power consumption model is the same as in [41], with $P_{\text{ref}} = 20$ mW: $P_{\text{sw}} = 0.25P_{\text{ref}}$, $P_{\text{ps}} = 1.50P_{\text{ref}}$, $P_{\text{vga}} = P_{\text{lna}} = P_{\text{ref}}$, $P_{\text{adc}} = P_{\text{dac}} = 10P_{\text{ref}}$ and $P_{\text{rfc}} = 2P_{\text{ref}}$.

TABLE II: Coding rates in bits per channel use (bit/c.u.) and target SNR for each channel encoder with $n = 512$.

Encoder $\mathfrak{R}$	Coding Rate R [bit/c.u.]	Target SNR $\text{PER}^{-1}[\epsilon]$
$\mathfrak{R}_1$	$2/3$	-0.6505 dB
$\mathfrak{R}_2$	$4/3$	3.0322 dB
$\mathfrak{R}_3$	2	5.8090 dB

Unless otherwise stated, we consider data packets consisting of $n = 512$ complex symbols and a target PER of $\epsilon = 10^{-5}$, so that the target SNR can be obtained from (8) as $\text{PER}_q^{-1}[\epsilon]$.

A. Worst-Case Analysis

We first assess the performance of the proposed scheme in the worst case described in Sec. III, *i.e.,* the error matrix $\boldsymbol{\Delta}$ is always assumed to be the most unfavorable one, *i.e.,* (17), and the post-combining SNR is evaluated as $\Gamma_0(\boldsymbol{f}, \boldsymbol{w})$ in (18).

We consider a transmitter equipped with $p = 3$ channel encoders $\mathcal{R} = \{\mathfrak{R}_1, \mathfrak{R}_2, \mathfrak{R}_3\}$. The coding rates and the target SNR for $n = 512$ symbols and $\epsilon = 10^{-5}$ are given in Table II. To model the imperfect CSI available at both nodes, we adopt a narrowband clustered channel representation based on the extended Saleh-Valenzuela model [57], *i.e.*,

$$\widehat{\boldsymbol{H}} = \sum_{m=1}^{N_{\text{cl}}} \sum_{n=1}^{N_{\text{ray}}} \widehat{\beta}_{mn} \boldsymbol{a}_{\text{R}}\left(\widehat{\theta}_{mn}\right) \boldsymbol{a}_{\text{T}}^H\left(\widehat{\phi}_{mn}\right), \tag{38}$$

where $\boldsymbol{a}_{\text{T}}(\phi)$ and $\boldsymbol{a}_{\text{R}}(\theta)$ are the transmit and receive array steering vectors at directions ϕ and θ, respectively, and N_{cl}, N_{ray}, $\widehat{\beta}_{mn}$, $\widehat{\phi}_{mn}$, and $\widehat{\theta}_{mn}$ stand for the number of clusters, the number of rays per cluster, the estimated complex path gains, the estimated angles of departure (AoDs), and the estimated AoAs. In the simulations, we consider $N_{\text{cl}} = 3$ and $N_{\text{ray}} = 7$. AoDs/AoAs are Gaussian distributed with a standard deviation of 0.4 rad and mean cluster angles uniformly distributed within $[0, 2\pi]$. Path gains are i.i.d. complex Gaussian distributed with mean 1 dB and standard deviation 0.5 dB. The *nominal* channel matrix in (38) is normalized so that $\|\widehat{\boldsymbol{H}}\|_{\text{F}}^2 = L_tL_r$. Results are averaged over 10^4 independent realizations.

TABLE III: EE relative to the unquantized upper-bound for different quantization bits (b_t and b_r), uncertainty levels $\delta^2 = \alpha\|\widehat{\boldsymbol{H}}\|_F^2$, and pre-processing SNRs γ. $L_t = 32$, $L_r = 8$.

$\alpha = 0$, $\gamma = -5$ dB			
$b_r \backslash b_t$	1	2	4
1	77.40%	87.58%	92.21%
2	87.68%	94.78%	98.41%
4	92.85%	98.24%	100%

$\alpha = 0.1$, $\gamma = -5$ dB			
$b_r \backslash b_t$	1	2	4
1	45.95%	67.91%	77.43%
2	65.01%	80.92%	91.27%
4	72.47%	89.32%	98.79%

$\alpha = 0.5$, $\gamma = -5$ dB			
$b_r \backslash b_t$	1	2	4
1	5.51%	27.83%	51.76%
2	7.79%	31.89%	65.43%
4	14.21%	53.73%	98.39%

$\alpha = 0$, $\gamma = 0$ dB			
$b_r \backslash b_t$	1	2	4
1	87.83%	91.23%	93.79%
2	93.82%	95.95%	97.74%
4	97.12%	98.85%	100%

$\alpha = 0.1$, $\gamma = 0$ dB			
$b_r \backslash b_t$	1	2	4
1	54.22%	69.31%	78.41%
2	69.45%	82.86%	92.06%
4	77.36%	91.24%	98.82%

$\alpha = 0.5$, $\gamma = 0$ dB			
$b_r \backslash b_t$	1	2	4
1	7.05%	30.66%	51.55%
2	10.51%	37.19%	69.39%
4	17.77%	59.79%	100%

$\alpha = 0$, $\gamma = 5$ dB			
$b_r \backslash b_t$	1	2	4
1	93.89%	94.90%	96.16%
2	97.23%	97.77%	98.48%
4	99.80%	100%	100%

$\alpha = 0.1$, $\gamma = 5$ dB			
$b_r \backslash b_t$	1	2	4
1	58.80%	70.55%	79.50%
2	72.49%	84.62%	92.62%
4	80.49%	92.38%	99.28%

$\alpha = 0.5$, $\gamma = 5$ dB			
$b_r \backslash b_t$	1	2	4
1	8.06%	30.71%	49.60%
2	11.47%	37.39%	67.42%
4	19.40%	59.01%	95.63%

Regarding the uncertainty level δ^2, similarly to [50], we set

$$\delta^2 = \alpha\|\widehat{\boldsymbol{H}}\|_{\mathrm{F}}^2 = \alpha L_t L_r, \quad \alpha > 0, \tag{39}$$

such that the Frobenius norm of the error matrix $\boldsymbol{\Delta}$ is proportional to the Frobenius norm of the nominal channel; α is referred to as *relative uncertainty*.

Fig. 5 illustrates the performance of the proposed scheme in terms of EE normalized by system bandwidth vs. pre-processing SNR γ and the number of transmit antennas L_t. Focusing on the case $\alpha = 0$ (corresponding to perfect CSI), the EE performance with finite-resolution PSs is seen to be close to that obtained in the unquantized case as long as PS resolution is at least 2 bits; even with 1-bit PSs, the loss becomes negligible for sufficiently high SNR and/or a sufficiently large transmit array. Note that, on one hand, as the SNR is reduced, more transmit power is needed to satisfy the reliability requirement, and consequently the EE degrades. On the other hand, for high SNR the attained EE saturates due to the finite number of channel encoders available (see Table II): in this regime, the channel encoder with the highest coding rate sets the limit to the achievable performance.

For $\alpha > 0$, CSI errors induce a loss on the worst-case post-processing SNR Γ_0, incurring an EE penalty. It is seen from (18) that this loss is given by

$$\frac{\Gamma_0|_{\delta^2=0}}{\Gamma_0} = 1 + \gamma\delta^2\|\boldsymbol{f}\|^2 = 1 + \gamma\alpha L_r L_t\|\boldsymbol{f}\|^2. \tag{40}$$

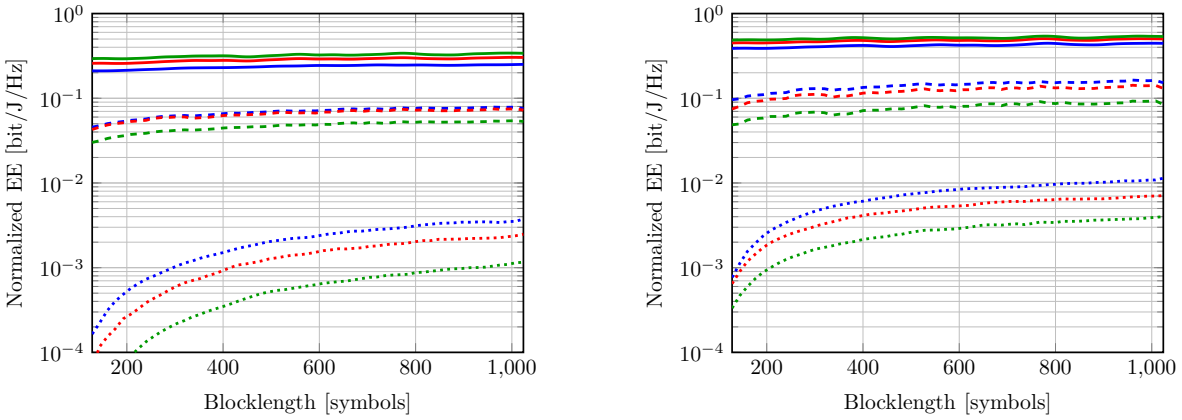
As this loss increases, a higher power consumption will become necessary to achieve the target SNR required by the channel encoders, even with unquantized PSs. The presence of the pre-processing SNR γ in (40) explains the behavior of the curves in Fig. 5a for unquantized PSs: for high SNR, the EE saturates at a lower value than that with no CSI errors. Similarly, the presence of the term $L_t\|\boldsymbol{f}\|^2$ in (40) is the cause of the degradation of the EE as the array size increases in the presence of inaccurate CSI, seen in Fig. 5b.

It is observed that the EE penalty due to CSI errors becomes more severe the coarser PS resolution is. For instance, consider in Fig. 5a the case $\alpha = 0.5$ with high SNR: with 2-bit resolution PSs, the achieved EE is half that of the unquantized case, whereas with 1-bit PSs the EE is reduced by a factor of 10. The reason is that the use of low-resolution PSs decreases the beamforming gain, so that it may become necessary to activate more antennas and/or use more transmit power to satisfy the minimum PER requirement. This increased power consumption results in the EE degradation with respect to the unquantized case observed in Fig. 5.

More insights on the impact of the phase quantization bits can be extracted from Table III. As CSI becomes less accurate, more bits are required to attain full-resolution performance even at high SNR. For low values of α, *i.e.,* small CSI uncertainty, increasing only b_t or b_r yields similar benefits. However, as α grows, increasing b_t appears to be more beneficial than increasing b_r. This is likely due to the fact that with improved resolution in the transmit beamformer, the PER requirement can be met with less transmit power, thus ameliorating the impact of channel estimation uncertainties. The impact of the blocklength n on the EE is depicted in Fig. 6, for different relative uncertainty levels, different number of transmit antennas L_t, and different PSs resolutions. It is seen that, under perfect CSI ($\alpha = 0$), EE remains almost constant. Since decreasing the blocklength increases the target SNR of each encoder, the proposed scheme balances the coding rate-power consumption tradeoff to maximize the EE. When the blocklength increases, although the target SNR for each channel encoder is relaxed, the system is limited by the encoder with the highest rate. For $\alpha > 0$, since CSI errors induce a loss on the worst-case post-processing SNR Γ_0, EE improves as blocklength increases; nevertheless, performance saturates once the encoder with highest rate is adopted.

The proposed scheme tunes the beamforming direction, the per-antenna power control, and the rate selection to maximize the EE given the PER requirement, all in the presence of channel estimation errors. To illustrate the importance of these three optimization dimensions, we compare next the EE performance vs. the pre-processing SNR γ of the proposed scheme and that of the following benchmarks:

(i) *Maximum Information Rate* (MaxRate): The per-antenna power control and the beamforming direction are optimized; only the encoder with the highest rate (R_3 in this setting) is used. This approach can be regarded as a short-packet adaptation of [58].



(a) $b_t = b_r = 1$ bit. (b) $b_t = b_r = 2$ bits.

Fig. 6: EE versus the blocklength with quantized PSs at pre-processing SNR $\gamma = -5$ dB. Results plotted for different relative uncertainties: $\alpha = 0$ (solid), $\alpha = 0.1$ (dashed), and $\alpha = 0.5$ (dotted); and for different number of transmit antennas: $L_t = 32$ (blue), $L_t = 64$ (red), and $L_t = 128$ (green).

(ii) *Most Conservative Precoder* (MCP): Similar to MaxRate, but now only the encoder with the lowest rate (R_1 in this setting) is used.

(iii) *Adaptive Uniform Power Allocation* (AUPA): All transmitting antennas are active and the power necessary to attain the minimum reliability constraint for encoder q

$$p_q \doteq \frac{\rho_q}{\left(\sum_{i=1}^{L_t} |v_i| \cos |\measuredangle(f_i) - \measuredangle(v_i)|\right)^2 - \rho_q \delta^2 \gamma L_t} \quad (41)$$

is uniformly allocated; rate allocation is optimized. This approach is a short-packet adaptation of the CMAB approach typically used in the literature (see, *e.g.*, [12]).

(iv) *Fully-Digital Beamforming*: Both transmitter and receiver are implemented with fully-digital arrays consisting in a baseband processor connected to each antenna through a dedicated RF chain and a DAC/ADC (see, *e.g.*, [14]). As in [41], the power consumption model for this case reads

$$P_{\text{tot}} = \|\boldsymbol{f}\|^2 + 2P_{\text{bb}} + L_t a_t + L_r a_r, \quad (42)$$

where P_{bb} is the power consumed by the baseband processor (assumed $P_{\text{bb}} = 10P_{\text{ref}}$ as in [41]), $a_t = P_{\text{rfc}} + P_{\text{dac}}$ and $a_r = P_{\text{rfc}} + P_{\text{adc}}$. Note that $a_t, a_r > 0$. The fully-digital transmit and receive beamformers (whose design is sketched in Appendix B) aim at minimizing power consumption for each channel encoder $\mathfrak{R}_q \in \mathcal{R}$; then, EE is maximized via rate allocation.

Fig. 7 shows the results for the different schemes in terms of EE, as a function of the pre-processing SNR γ, for a transmit array with $L_t = 32$ elements and perfect CSI ($\alpha = 0$). The proposed design with unquantized PSs provides an upper bound to the performance of the remaining methods with quantized phases, for which 2-bit phase resolution was considered at both nodes. The performance of MCP is close to that of the proposed design in the low SNR regime, but it saturates at high SNR since it is limited by the attainable spectral efficiency R_1. AUPA performs similarly to MCP, although saturation at high SNR is due to larger power consumption rather than limited spectral efficiency. MaxRate practically achieves the performance attained by the full-resolution configuration at high SNR, but can only transmit under favorable channel conditions, as expected. This illustrates the additional benefits of optimally exploiting the availability of multiple encoders and per-antenna power control over optimizing phase shifters only, as in conventional CMAB approaches. Even with low-resolution PSs, the proposed scheme outperforms the fully-digital beamforming design, except at very low SNR, where the possibility of implementing an SNR-optimal Maximal Ratio Combiner (MRC) at the receiver side almost compensates for the increased power consumption. In this limiting regime, the proposed scheme with unquantized PSs and the fully-digital design exhibit the same performance.

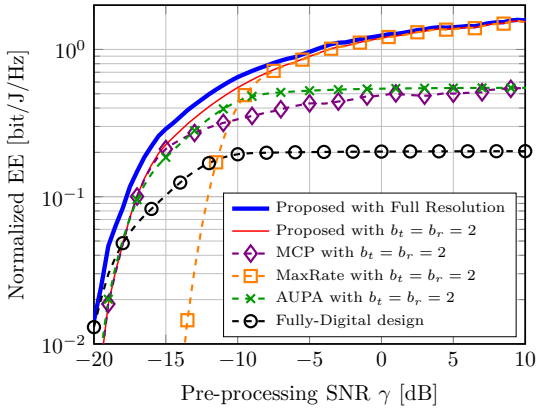


Fig. 7: Comparison of the EE performance vs. pre-processing SNR γ achieved with the proposed design with full and finite resolution PSs, and that exhibited by MCP, MaxRate, AUPA, and Fully-Digital benchmarks for $\alpha = 0$.

Figs. 8 and 9 show the results for $\alpha = 0.1$ and $\alpha = 0.5$, respectively. In both cases, the full-resolution performance for $\alpha = 0$ (upper-bound) is depicted as reference. The most noticeable difference concerning the ideal case ($\alpha = 0$) is that MaxRate significantly underperforms as α increases, until becoming useless. Since a relative uncertainty $\alpha > 0$ results in an SNR loss, the fully-digital design is able to yield better performance than the proposed scheme with 2-bit resolution at low-to-moderate SNR γ, since the additional SNR gain offered by the MRC receiver justifies the additional power consumption. AUPA performance worsens with α due

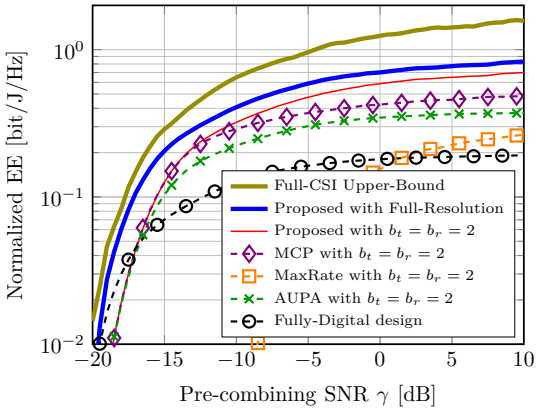


Fig. 8: Comparison of the EE performance vs. pre-processing SNR γ achieved with the proposed design with full and finite resolution PSs, and that exhibited by MCP, MaxRate, AUPA, and Fully-Digital benchmarks for $\alpha = 0.1$.

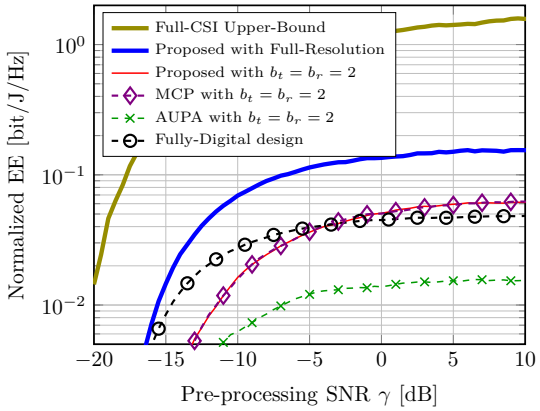


Fig. 9: Comparison of the EE performance vs pre-processing SNR γ achieved with the proposed design with full and finite resolution PSs and that exhibited by MCP and AUPA benchmarks for the case $\alpha = 0.5$.

to larger power consumption, and MCP tends to achieve the same performance as the proposed design. This is explained by noting that, under the SNR loss induced by imperfect CSI, encoders with lower rates eventually become preferable, given their reduced SNR requirement to satisfy the fixed PER constraint. This is illustrated in Fig. 10, which shows the probability of using each encoder as a function of the pre-processing SNR γ for $L_t = 32$. In the low-SNR limit, the available power is not sufficient to meet the PER constraint even when using the encoder with the lowest rate, so that all antennas are deactivated and no transmission occurs, yielding an outage event. As the SNR increases, the target PER can be attained with more ease, allowing the use of encoders with progressively higher rate. Since the post-processing SNR worsens as the relative uncertainty α increases, outage events occur even at moderate pre-processing SNRs, making encoders with higher rates useless. This observation reveals that having large codebooks is not necessarily useful unless the channel encoders are suitable for each particular scenario and the specific operating conditions. For instance, for the case $\alpha = 0.5$ shown in Fig. 10, having encoders with rates higher than that of $\mathfrak{R}_1$ would be useless, since they could not satisfy the minimum SNR requirement under these conditions leading to outage events. In contrast, system performance would improve if encoders with lower rate than $\mathfrak{R}_1$ were available.

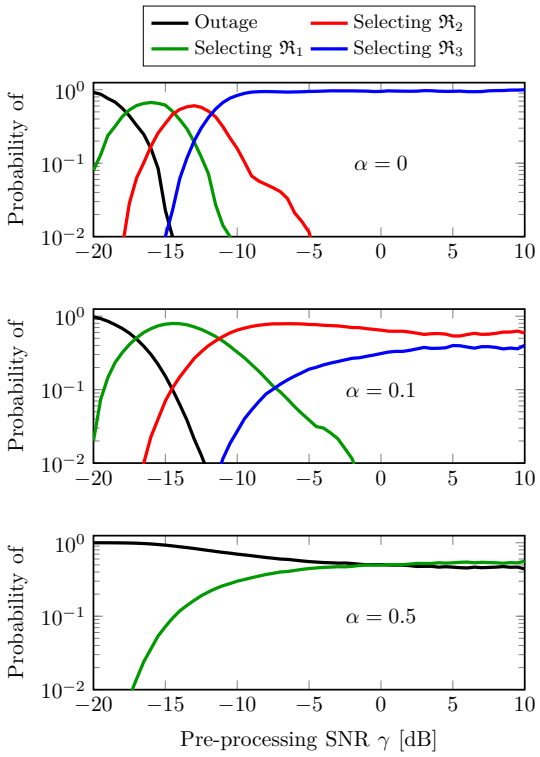


Fig. 10: Probability of using each encoder $\mathfrak{R}_{1\leq q\leq 3}$ and probability of outage vs. pre-processing SNR γ with full-resolution PSs for different values of α.

B. Robustness Analysis

Next, we evaluate the robustness of the proposed design under practical channel estimation errors. We assume that the transmitter is equipped with $L_t = 32$ antennas and with the set of $p = 6$ channel encoders described in Table IV.

TABLE IV: Coding rates [bit/c.u.] and target SNR for each channel encoder for $n = \{256, 512\}$ symbols and $\epsilon = 10^{-5}$.

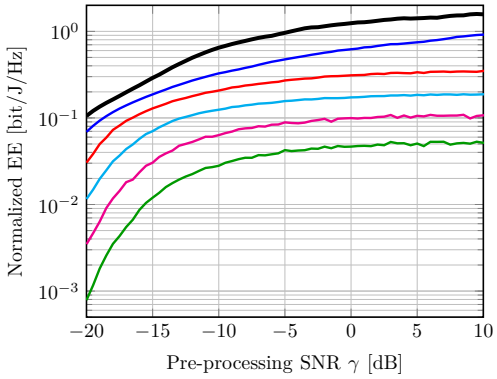
Encoder $\mathfrak{R}$	R [bit/c.u.]	Target SNR (512)	Target SNR (256)
$\mathfrak{R}_1$	1/4	−4.5632 dB	−3.5578 dB
$\mathfrak{R}_2$	1/3	−3.5326 dB	−2.6474 dB
$\mathfrak{R}_3$	1/2	−1.9227 dB	−1.1850 dB
$\mathfrak{R}_4$	2/3	−0.6505 dB	−0.0019 dB
$\mathfrak{R}_5$	4/3	3.0322 dB	3.5164 dB
$\mathfrak{R}_6$	2	5.8090 dB	6.2282 dB

Since in typical mmWave scenarios CSI acquisition involves the estimation of the parameters of the narrowband clustered channel, *i.e.,* the complex channel gains and AoDs/AoAs, we model the *actual* channel matrix as

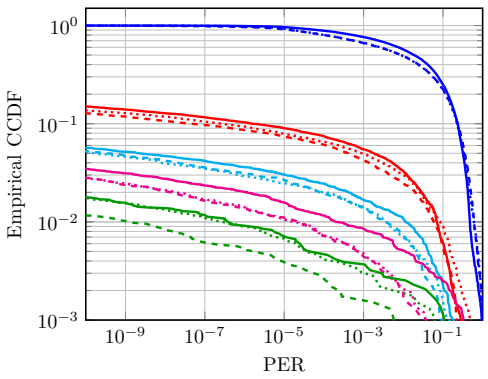
$$\boldsymbol{H} = \sum_{m=1}^{N_{\text{cl}}} \sum_{n=1}^{N_{\text{ray}}} \left[\left(\widehat{\beta}_{mn} + \Delta\beta_{mn} \right) \boldsymbol{a}_{\text{R}} \left(\widehat{\theta}_{mn} + \Delta\theta_{mn} \right) \times \boldsymbol{a}_{\text{T}}^{H} \left(\widehat{\phi}_{mn} + \Delta\phi_{mn} \right) \right],$$

where $\Delta\beta_{mn}$, $\Delta\phi_{mn}$, and $\Delta\theta_{mn}$ are the estimation errors of the complex channel gains, AoDs, and AoAs, respectively. The *actual* channel matrix $\boldsymbol{H}$ is normalized so that $\|\boldsymbol{H}\|_{\text{F}}^2 = L_t L_r$. Then, the error matrix $\boldsymbol{\Delta}$ is defined as

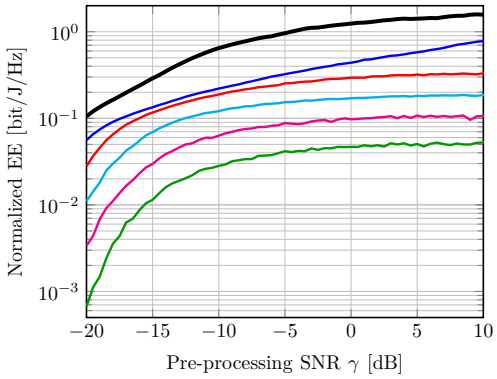
$$\boldsymbol{\Delta} \doteq \boldsymbol{H} - \widehat{\boldsymbol{H}}, \tag{43}$$



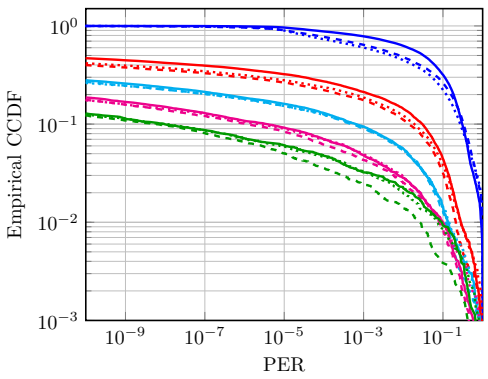
(a) $\sigma_\phi^2 = 0.0005$, $n = 512$



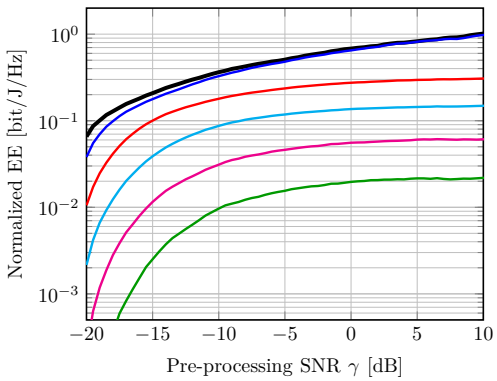
(b) $\sigma_\phi^2 = 0.0005$, $n = 512$



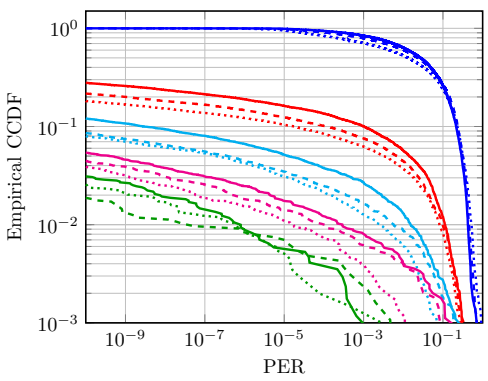
(c) $\sigma_\phi^2 = 0.005$, $n = 512$



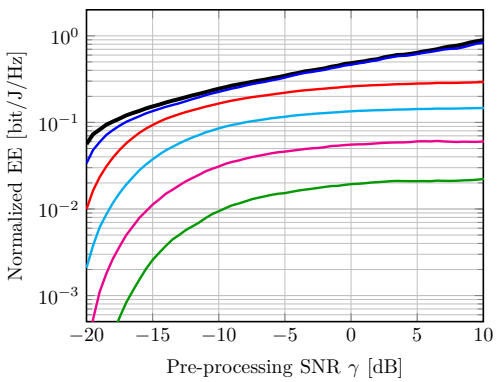
(d) $\sigma_\phi^2 = 0.005$, $n = 512$



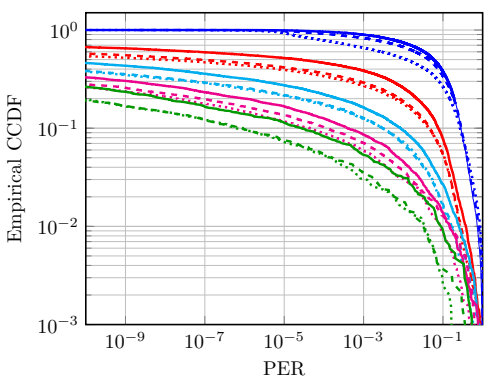
(e) $\sigma_\phi^2 = 0.0005$, $n = 256$



(f) $\sigma_\phi^2 = 0.0005$, $n = 256$

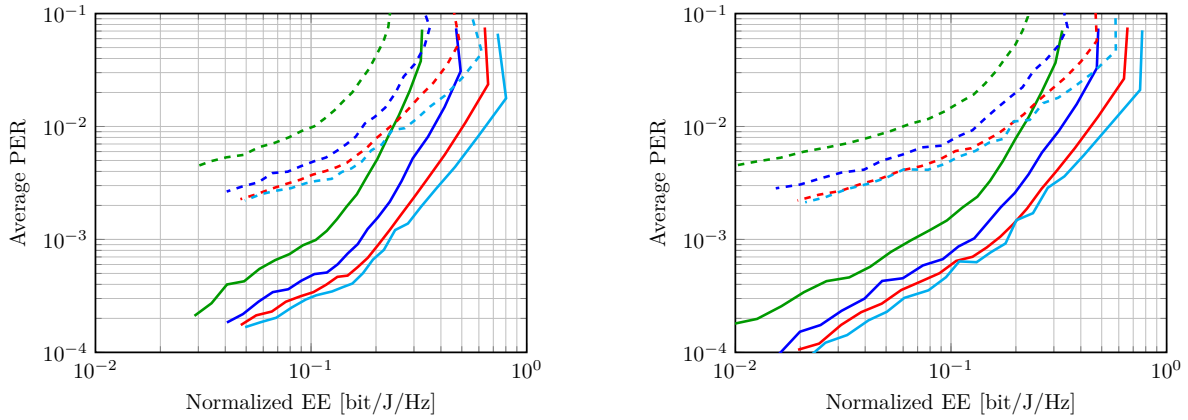


(g) $\sigma_\phi^2 = 0.005$, $n = 256$



(h) $\sigma_\phi^2 = 0.005$, $n = 256$

Fig. 11: Robustness evaluation of the proposed design with $L_t = 32$ and $b_t = b_r = 2$ bits, for packet size $n = \{256, 512\}$, $\sigma_\beta^2 = 10^{-1}$, $\sigma_\theta^2 = 0.005$, and $\sigma_\phi^2 \in \{0.005, 0.0005\}$. In all cases, the performance of the full-resolution design under perfect CSI is depicted in black, and different values of the relative uncertainty have been tested: $\alpha = 0$ (blue), $\alpha = 0.25$ (red), $\alpha = 0.5$ (cyan), $\alpha = 0.75$ (magenta), and $\alpha = 1$ (green). The empirical CCDFs are shown at different pre-processing SNRs: $\gamma = -5$ dB (solid), $\gamma = 0$ dB (dashed), and $\gamma = 5$ dB (dotted). Results averaged over 10^4 Monte Carlo runs.



(a) $n = 512$ symbols. (b) $n = 256$ symbols.

Fig. 12: EE-PER tradeoff curves for $L_t = 32$ and $b_t = b_r = 2$ bits, for $\sigma_\beta^2 = 10^{-1}$; $\sigma_\theta^2 = 0.005$; different DoD error variance: $\sigma_\phi^2 = 0.0005$ (solid) and $\sigma_\phi^2 = 0.005$ (dashed); and different pre-processing SNR: $\gamma = -10$ dB (green) $\gamma = -5$ dB (blue), $\gamma = 0$ dB (red), and $\gamma = 5$ dB (cyan).

with $\widehat{\boldsymbol{H}}$ the nominal channel given in (38). The beamformers are designed following the same worst-case approach of Sec. III, and then the post-combining SNR is evaluated as in (6), using the error matrix from (43). Note that δ^2 (or equivalently α) becomes a design parameter determining the tradeoff between performance and robustness, to be tuned depending on the expected quality of channel parameter estimates; and that the attained PER becomes a random variable, depending on the channel realization.

In the simulations, we consider again $N_{\text{cl}} = 3$ and $N_{\text{ray}} = 7$. The nominal channel parameters $\widehat{\beta}_{mn}$, $\widehat{\phi}_{mn}$, and $\widehat{\theta}_{mn}$ are generated as described below (38). The estimation errors $\Delta\beta_{mn}$, $\Delta\phi_{mn}$, and $\Delta\theta_{mn}$ are assumed independent, zero-mean Gaussian distributed with variances σ_β^2, σ_ϕ^2, and σ_θ^2, respectively. Taking as reference [11], [59], we have numerically tested different error variances: $\sigma_\beta^2 = \{0, 0.01, 0.1\}$, $\sigma_\phi^2 = \{0, 0.0005, 0.005\}$, and $\sigma_\theta^2 = \{0, 0.0005, 0.005\}$; although only illustrative cases are shown for the sake of brevity. These experiments showed that angular errors have a much larger impact than amplitude errors. For instance, whereas for $\sigma_\phi^2 = \sigma_\theta^2 = 0$, small values of α suffice to counteract the impact of amplitude errors ($\sigma_\beta^2 > 0$), larger values of α are needed to guarantee the desired PER in the presence of angular errors. Regarding these, AoD estimation errors were seen to have a larger impact than AoA errors as far as $L_t \gg L_r$. Moreover, numerical results showed that even small AoD estimation errors have a devastating impact in terms of PER.

To illustrate these observations, Fig. 11 shows the EE vs. pre-processing SNR γ, as well as the empirical complementary cumulative distribution function (CCDF) of the achieved PER for $\sigma_\beta^2 = 10^{-1}$, $\sigma_\theta^2 = 0.005$, $\sigma_\phi^2 \in \{0.005, 0.0005\}$, $b_t = b_r = 2$, $n = \{256, 512\}$ and different values of the relative uncertainty α. In the EE vs. γ plots, the proposed design with full-resolution PSs and perfect CSI (thus $\alpha = 0$) is depicted in black as reference.

With 2-bit resolution PSs at both transmitter and receiver, the non-robust design ($\alpha = 0$) exhibits an EE performance close to the upper bound for $\sigma_\phi^2 = 0.0005$, with a somewhat larger gap for $\sigma_\phi^2 = 0.005$. With the robust design ($\alpha > 0$), EE degrades as the value of α is increased. This is the price paid for robustness in terms of achieved PER: Note that, even for $\sigma_\phi^2 = 0.0005$, the non-robust design ($\alpha = 0$) delivers a PER that will be almost surely larger than the 10^{-5} requirement. In fact, it will be smaller than 10^{-3} only with probability 0.3, irrespective of the pre-processing SNR γ, and further degrading as σ_ϕ^2 increases. The importance of appropriately setting the robustness parameter α is highlighted in Figs. 11b and 11d. Let us consider, for instance, the design with $\alpha = 0.75$ (magenta line). Note that the achieved EE at $\gamma = 0$ dB is roughly the same regardless of the value of σ_ϕ^2; however, for $\sigma_\phi^2 = 0.0005$, the robust design with $\alpha = 0.75$ guarantees a PER equal to or smaller than the 10^{-5} requirement with probability ~ 0.99, and with probability ~ 0.91 for $\sigma_\phi^2 = 0.005$. Similar conclusions can be drawn when the blocklength is halved (*i.e.*, $n = 256$ complex symbols). In this case, since the target SNRs shown in Table IV are higher than for $n = 512$, higher values of α are needed to guarantee the desired operating PER. For instance, observing Figs. 11d and 11h, to guarantee a PER below the 10^{-5} requirement with probability ~ 0.91, α has to be increased from 0.75 to 1 as the blocklength is halved.

This EE-PER tradeoff is further illustrated in Fig. 12. Selecting a particular value of α fixes the operation point on the corresponding curve. In this vein, higher values of α guarantee a lower PER at the expense of worse EE. Note that, as the AoD error variance increases, certain degradation of EE may be necessary to reduce the average PER. We can conclude that the choice of the robustness parameter α should be based on the confidence in the AoD/AoA estimator, since these errors have the greatest impact on error decoding performance.

V. Conclusions

We have addressed the design of energy-efficient analog beamformers for short-packet communication at mmWave bands, and adopting per-antenna power constraints and a finite encoder set. It has been shown that directly maximizing EE may not meet the required PER target. To satisfy reliability demands, we posed the design as the minimization of power consumption under worst-case CSI uncertainty. This admits a

$$\mathcal{L}(\bar{\boldsymbol{f}},\{t_i\},\{\eta_i\},\mu)=\bar{\boldsymbol{f}}^T\bar{\boldsymbol{f}}+bk+\sum_{i=1}^{k}\eta_i(P-\boldsymbol{e}_i^T\bar{\boldsymbol{f}}\bar{\boldsymbol{f}}^T\boldsymbol{e}_i-t_i^2)+\mu\left(1+\delta^2\gamma\bar{\boldsymbol{f}}^T\bar{\boldsymbol{f}}-\frac{1}{\rho_q}(\bar{\boldsymbol{f}}^T\bar{\boldsymbol{v}})^2\right). \tag{45}$$

closed-form solution yielding a power allocation scheme in which (i) only the minimum number of transmit antennas to attain the target SNR are active; (ii) antennas having strong channel gain are allocated maximum power; (iii) antennas having weak channel gain remain silent; and (iv) power allocation in the remaining antennas is proportional to their channel gain. Then, maximum EE is achieved by appropriately selecting the best channel encoder. The impact of different parameters was numerically analyzed in terms of EE and PER, showing that high-resolution VGAs can almost compensate for the loss incurred with low-resolution PSs, as well as the relevance of designing the degree of robustness to balance the EE-PER tradeoff.

Future work will consider extensions to multi-user settings, which require incorporating at least as many RF chains as users to be served. Potential approaches include subarray-based fully-analog beamforming and hybrid analog-digital beamforming. While the former suffers from a low per-user beamforming gain, the latter offers better spectral efficiency at the expense of extra power consumption and computational cost. Understanding the single-user solution derived in this work should lay the foundation for developing scalable, sub-optimal multi-user solutions.

APPENDIX A

We define the real-valued vectors $\bar{\boldsymbol{f}} \doteq [|f_1|,\ldots,|f_k|]^T \in \mathbb{R}^{k\times 1}$ and $\bar{\boldsymbol{v}} \doteq [|v_1|,\ldots,|v_k|]^T \in \mathbb{R}^{k\times 1}$ such that $\mathbb{P}_k$ in (26) can be rewritten as

$$\min_{\bar{\boldsymbol{f}},\{t_i\}_{1\le i\le k}} \quad \bar{\boldsymbol{f}}^T\bar{\boldsymbol{f}}+bk \tag{44a}$$

$$\text{s.to} \quad P-\boldsymbol{e}_i^T\bar{\boldsymbol{f}}\bar{\boldsymbol{f}}^T\boldsymbol{e}_i-t_i^2=0,\ \ 1\le i\le k \tag{44b}$$

$$\frac{1}{\rho_q}\left(\bar{\boldsymbol{f}}^T\bar{\boldsymbol{v}}\right)^2\ge 1+\delta^2\gamma\bar{\boldsymbol{f}}^T\bar{\boldsymbol{f}} \tag{44c}$$

where $\boldsymbol{e}_i$ is the i-th column of $\boldsymbol{I}_k$ and $\{t_i\}$ are slack variables. The Lagrangian associated to (44) is given by (45) on the top of this page, with $\{\eta_i\}_{1\le i\le k}$ and μ the Lagrange multipliers associated with (44b) and (44c), respectively.

The stationary point equation $\frac{\partial\mathcal{L}}{\partial t_i}=0$ reveals that, whenever the i-th per-antenna power constraint is active, *i.e.*, $\eta_i\neq 0$, then $\bar{f}_i^\star=\sqrt{P}$. Otherwise, $\eta_i=0$ and the stationary point equation $\frac{\partial\mathcal{L}}{\partial \bar{f}_i}=0$ leads to

$$\bar{f}_i^\star=\frac{\mu\bar{\boldsymbol{f}}^T\bar{\boldsymbol{v}}}{(1+\mu\delta^2\gamma)\rho_q}\bar{v}_i=\lambda\bar{v}_i, \tag{46}$$

where λ is implicitly defined. Since $\bar{\boldsymbol{f}}^T\bar{\boldsymbol{f}}$ is permutation-invariant, and in view of this result, we note that the left-hand side of (44c) is maximized when full power is allocated to the strongest *channels*, say 1 through ℓ, whereas we must allocate power according to (46) in the remaining $k-\ell$ ones. Thus, the solution to (44a)–(44c) can be compactly written as

$$\bar{\boldsymbol{f}}^\star=\sqrt{P}\begin{bmatrix}\mathbf{1}_{\ell\times 1}\\ \mathbf{0}_{(k-\ell)\times 1}\end{bmatrix}+\lambda\bar{\boldsymbol{v}}\odot\begin{bmatrix}\mathbf{0}_{\ell\times 1}\\ \mathbf{1}_{(k-\ell)\times 1}\end{bmatrix}. \tag{47}$$

Next, we need to find the scaling parameter λ, for a given ℓ. Note that the optimal precoder $\bar{\boldsymbol{f}}^\star$ must satisfy (44c) with equality; otherwise, a scalar $0<\alpha<1$ would exist such that $\alpha\bar{\boldsymbol{f}}^\star$ is feasible and provides a lower value of the objective (44a) than $\bar{\boldsymbol{f}}^\star$. Taking this into account, the scalar parameter λ can be found by substituting (47) into (44c) and solving a second-order equation, which leads to (28).

Finally, note that to completely define the power allocation policy, we must determine the value of ℓ that minimizes (44a)–(44c). This value cannot be obtained in closed form, and has to be sequentially found as described in Algorithm 1.

APPENDIX B

For a fair comparison, we adopt per-antenna power constraints as in (2) so that the feasible set for $\boldsymbol{f}$ is $\mathcal{F}_{\text{fd}}^{L_t}=\{\boldsymbol{f}\in\mathbb{C}^{L_t}\mid 0\le|f_i|\le\sqrt{P},\ \forall i\}$. Concerning $\boldsymbol{w}$, its feasible set is $\mathcal{W}_{\text{fd}}^{L_r}=\{\boldsymbol{w}\in\mathbb{C}^{L_r}\mid\|\boldsymbol{w}\|^2=1\}$. With these constraints, as per (16), the worst-case post-processing SNR now reads as

$$\Gamma_0(\boldsymbol{f},\boldsymbol{w})=\gamma\frac{\left|\boldsymbol{w}^H\widehat{\boldsymbol{H}}\boldsymbol{f}\right|^2}{1+\gamma\delta^2\|\boldsymbol{f}\|_2^2}. \tag{48}$$

Since (48) couples $\boldsymbol{f}$ and $\boldsymbol{w}$, similarly to (20)–(21), for each $\mathfrak{R}_q\in\mathcal{R}$ we undertake the following cyclic design:

1) For fixed $\boldsymbol{f}$, the optimal $\boldsymbol{w}$ can be obtained as

$$\widetilde{\boldsymbol{w}}=\underset{\boldsymbol{w}\in\mathcal{W}_{\text{fd}}^{L_r}}{\arg\max}\ \Gamma_0(\boldsymbol{f},\boldsymbol{w})=\|\widehat{\boldsymbol{H}}\boldsymbol{f}\|_2^{-1}\widehat{\boldsymbol{H}}\boldsymbol{f}. \tag{49}$$

2) For fixed $\boldsymbol{w}$, the design of $\boldsymbol{f}$ can be cast as

$$\min_{\boldsymbol{f}\in\mathcal{F}_{\text{fd}}^{L_t}}\|\boldsymbol{f}\|_2^2\ \ \text{s.to}\ \ \Gamma_0(\boldsymbol{f},\boldsymbol{w})\ge\text{PER}_q^{-1}[\epsilon]. \tag{50}$$

These steps are iterated until convergence. Regarding (50), since $\|\boldsymbol{f}\|_2^2$ is phase-invariant, the phase of each f_i, for $i=1,\ldots,L_t$, is given by $\measuredangle(f_i)=\measuredangle([\widehat{\boldsymbol{H}}^H\boldsymbol{w}]_i)$, so that (50) becomes a SOCP that can be solved analogously to (26).

REFERENCES

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

[2] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1616–1653, 2018.

[3] X. Shen, Y. Liu, L. Zhao, G.-L. Huang, X. Shi, and Q. Huang, "A miniaturized microstrip antenna array at 5G millimeter-wave band," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 8, pp. 1671–1675, 2019.

[4] M. Vaezi, A. Azari, S. R. Khosravirad, M. Shirvanimoghaddam, M. M. Azari, D. Chasaki, and P. Popovski, "Cellular, Wide-Area, and Non-Terrestrial IoT: A Survey on 5G Advances and the Road Toward 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1117–1174, 2022.

[5] K. Zrar Ghafoor, L. Kong, S. Zeadally, A. S. Sadiq, G. Epiphaniou, M. Hammoudeh, A. K. Bashir, and S. Mumtaz, "Millimeter-Wave Communication for Internet of Vehicles: Status, Challenges, and Perspectives," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8525–8546, 2020.

[6] X. Yu, L. Tu, Q. Yang, M. Yu, Z. Xiao, and Y. Zhu, "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.

[7] J. Yang, B. Ai, I. You, M. Imran, L. Wang, K. Guan, D. He, Z. Zhong, and W. Keusgen, "Ultra-reliable communications for industrial internet of things: Design considerations and channel modeling," *IEEE Netw.*, vol. 33, no. 4, pp. 104–111, 2019.

[8] B. S. Khan, S. Jangsher, A. Ahmed, and A. Al-Dweik, "URLLC and eMBB in 5G Industrial IoT: A Survey," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1134–1163, 2022.

[9] H. Bo Marr, "Fundamental energy limits of digital phased arrays," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 7, pp. 2775–2783, 2019.

[10] N. T. Nguyen and K. Lee, "Unequally sub-connected architecture for hybrid beamforming in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1127–1140, 2020.

[11] L. Jiang and H. Jafarkhani, "Multi-user analog beamforming in millimeter wave MIMO systems based on path angle information," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 608–619, 2019.

[12] Z. Wang, Q. Liu, M. Li, and W. Kellerer, "Energy efficient analog beamformer design for mmWave multicast transmission," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 2, pp. 552–564, 2019.

[13] A. Arora, C. G. Tsinos, M. R. B. Shankar, S. Chatzinotas, and B. Ottersten, "Efficient algorithms for constant-modulus analog beamforming," *IEEE Trans. Signal Process.*, vol. 70, pp. 756–771, 2022.

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

[15] M.-C. Lee and W.-H. Chung, "Adaptive multimode hybrid precoding for single-RF virtual space modulation with analog phase shift network in MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2139–2152, 2017.

[16] S. S. Ioushua and Y. C. Eldar, "A family of hybrid analog–digital beamforming methods for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3243–3257, 2019.

[17] S. Buzzi and C. D'Andrea, "Energy efficiency and asymptotic performance evaluation of beamforming structures in doubly massive MIMO mmWave systems," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 2, pp. 385–396, 2018.

[18] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, 2016.

[19] S. Payami, M. Ghoraishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258–7271, 2016.

[20] X. Gao, L. Dai, S. Han, C. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[21] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 298–312, 2018.

[22] D. Zhang, Y. Wang, X. Li, and W. Xiang, "Hybridly connected structure for hybrid beamforming in mmWave massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 662–674, 2018.

[23] R. Zi, X. Ge, J. Thompson, C.-X. Wang, H. Wang, and T. Han, "Energy efficiency optimization of 5G radio frequency chain systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758–771, 2016.

[24] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5289–5304, 2017.

[25] S. Payami, N. Mysore Balasubramanya, C. Masouros, and M. Sellathurai, "Phase shifters versus switches: An energy efficiency perspective on hybrid beamforming," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 13–16, Feb. 2019.

[26] V. N. Ha, D. H. N. Nguyen, and J.-F. Frigon, "System energy-efficient hybrid beamforming for mmWave multi-user systems," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 4, pp. 1010–1023, 2020.

[27] J. Gao, W. Zhuang, M. Li, X. Shen, and X. Li, "MAC for machine-type communications in industrial IoT—part i: Protocol design and analysis," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9945–9957, 2021.

[28] J. Zeng, T. Wu, Y. Song, Y. Zhong, T. Lv, and S. Zhou, "Achieving energy-efficient massive URLLC over cell-free massive MIMO," *IEEE Internet Things J.*, vol. 11, no. 2, pp. 2198–2210, 2024.

[29] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[30] 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.

[31] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Short-packet downlink transmission with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4550–4564, 2018.

[32] J. Guo, D. Zhang, I. Lee, Y. Li, and M. Shirvanimoghaddam, "Partitioned analog fountain codes for short packet communications," *IEEE Commun. Lett.*, vol. 28, no. 6, pp. 1248–1252, 2024.

[33] T.-H. Vu, T.-V. Nguyen, Q.-V. Pham, D. B. da Costa, and S. Kim, "Hybrid long- and short-packet based NOMA systems with joint power allocation and beamforming design," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 4079–4084, 2023.

[34] S. Liu, Z. Sheng, P. Zhu, D. Wang, Y. Jiang, and Y. Wang, "Hybrid precoding with low-resolution PSs for URLLC users in cell-free mmWave MIMO systems," in *Int. Conf. Wireless Commun. Signal Process. (WCSP)*, 2023, pp. 1168–1172.

[35] X. Zhang, L. Xiang, J. Wang, P. Zhu, D. W. K. Ng, and X. Gao, "Hybrid Precoding for mmWave Massive MIMO with Finite Blocklength," *IEEE Trans. Commun.*, pp. 1–1, 2025.

[36] P. Popovski, C. Stefanovic, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A.-S. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783–5801, 2019.

[37] 3GPP, "Study on scenarios and requirements for next generation access technologies (Release 17)," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.913, Jun. 2022, version 17.0.0.

[38] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 70–78, 2017.

[39] F. Gholam, J. Via, and I. Santamaria, "Beamforming design for simplified analog antenna combining architectures," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2373–2378, 2011.

[40] I. P. Roberts, S. Vishwanath, and J. G. Andrews, "LoneSTAR: Analog beamforming codebooks for full-duplex millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 9, pp. 5754–5769, 2023.

[41] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.

[42] J. Borras, F. Molina, R. López-Valcarce, and J. Sala-Álvarez, "Energy-efficient analog beamforming with short packets in millimeter-wave MIMO systems," in *54th Asilomar Conf. Signals, Syst. Comput.*, 2020, pp. 1–5.

[43] K. Hassan, M. Masarra, M. Zwingelstein, and I. Dayoub, "Channel estimation techniques for millimeter-wave communication systems: Achievements and challenges," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1336–1363, 2020.

[44] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, 2017.

[45] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.

[46] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 949–973, 2016.

[47] S. Roger, M. Cobos, C. Botella-Mascarell, and G. Fodor, "Fast channel estimation in the transformed spatial domain for analog millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5926–5941, 2021.

[48] Z. Guo, W. Wang, X. Wang, and X. Zeng, "Hardware-efficient beamspace direction-of-arrival estimator for unequal-sized subarrays," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 3, pp. 1044–1048, 2022.

[49] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, 2000.

[50] J. Wang, M. Bengtsson, B. Ottersten, and D. P. Palomar, "Robust MIMO precoding for several classes of channel uncertainty," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3056–3070, 2013.

[51] H. Shen, J. Wang, B. C. Levy, and C. Zhao, "Robust optimization for amplify-and-forward MIMO relaying from a worst-case perspective," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5458–5471, 2013.
[52] Y. Zhang, E. DallAnese, and G. B. Giannakis, "Distributed optimal beamformers for cognitive radios robust to channel uncertainties," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6495–6508, 2012.
[53] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect side information" be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, 2002.
[54] G. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*. Cambridge University Press, 1952.
[55] P. Stoica and Y. Selen, "Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 112–114, 2004.
[56] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, no. 1-3, pp. 193–228, 1998.
[57] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.
[58] D. Persson, T. Eriksson, and E. G. Larsson, "Amplifier-aware multiple-input single-output capacity," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 913–919, 2014.
[59] K. K. Kota and P. Ubaidulla, "Sum-rate maximization in NOMA-based mmWave analog beamforming under imperfect CSI," in *2021 IEEE 93rd Veh. Technol. Conf. (VTC2021-Spring)*, 2021, pp. 1–7.