ETH zürich

ETH Library

Distortion-Aware Linear Precoding for Massive MIMO Downlink Systems with Nonlinear Power Amplifiers

Working Paper

Author(s):
Aghdam, Sina Rezaei; Jacobsson, Sven; Gustavsson, Ulf; Durisi, Giuseppe; Studer, Christoph; Eriksson, Thomas

Publication date:
2021-11-14

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-b-000521458

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Distortion-Aware Linear Precoding for Massive MIMO Downlink Systems with Nonlinear Power Amplifiers

Sina Rezaei Aghdam, *Member, IEEE,* Sven Jacobsson, Ulf Gustavsson, Giuseppe Durisi, *Senior Member, IEEE,* Christoph Studer, *Senior Member, IEEE,* Thomas Eriksson

Abstract

We introduce a framework for linear precoder design over a massive multiple-input multiple-output downlink system in the presence of nonlinear power amplifiers (PAs). By studying the spatial characteristics of the distortion, we demonstrate that conventional linear precoding techniques steer nonlinear distortions towards the users. We show that, by taking into account PA nonlinearity, one can design linear precoders that reduce, and in single-user scenarios, even completely remove the distortion transmitted in the direction of the users. This, however, is achieved at the price of a reduced array gain. To address this issue, we present precoder optimization algorithms that simultaneously take into account the effects of array gain, distortion, multiuser interference, and receiver noise. Specifically, we derive an expression for the achievable sum rate and propose an iterative algorithm that attempts to find the precoding matrix which maximizes this expression. Moreover, using a model for PA power consumption, we propose an algorithm that attempts to find the precoding matrix that minimizes the consumed power for a given minimum achievable sum rate. Our numerical results demonstrate that the proposed distortion-aware precoding techniques provide significant improvements in spectral and energy efficiency compared to conventional linear precoders.

Sina Rezaei Aghdam, Giuseppe Durisi and Thomas Eriksson are with the Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden. (e-mails: {sinar, durisi, thomase}@chalmers.se).

Sven Jacobsson and Ulf Gustavsson are with Ericsson Research, 41756 Gothenburg, Sweden. (e-mails: {sven.jacobsson, ulf.gustavsson}@ericsson.com).

Christoph Studer is with the Department of Information Technology and Electrical Engineering, ETH Zurich, Zürich, Switzerland. (e-mail: studer@ethz.ch).

The work of Sina Rezaei Aghdam and Thomas Eriksson was performed within the strategic innovation program "Smarter Electronics Systems", a common venture by VINNOVA, Formas, and Energimyndigheten. The work of Sven Jacobsson was supported in part by the Swedish Foundation for Strategic Research under grant ID14-0022, and by VINNOVA within the competence center ChaseOn.

This paper was presented in part at the IEEE ICC 2019 Workshop on millimeter-wave communications for 5G and B5G [1].

Index Terms

Massive multiple-input multiple-output, nonlinear power amplifier, in-band distortion, linear precoding, energy efficiency, and out-of-band emission.

I. Introduction

Massive multiple-input multiple-output (MIMO) technology has already proven its potential for enhancing spectral efficiency in fifth generation (5G) cellular networks [2]. The core idea is to equip base stations with many individually controllable antennas to serve multiple users over the same time-frequency resource via spatial multiplexing. To make massive MIMO commercially viable, the base station should be deployed using low-complexity and inexpensive hardware. Dealing with hardware impairments is therefore one of the main challenges in making massive MIMO practically feasible; see, e.g., [3]–[6].

Low-cost and low-precision hardware can lead to different impairments, such as nonlinear amplification, quantization noise, phase noise, and in-phase/quadrature (I/Q) imbalance. How to model these impairments and how to evaluate the resulting performance loss have been the subject of extensive investigations. As far as modeling is concerned, two approaches are typically followed in the literature.

The first approach deals with the characterization of a single (or predominant) hardware impairment, which is modeled using accurate parametric models. For example, the impact of power amplifier (PA) nonlinearities on the performance of multi-antenna systems has been investigated in [7]–[9]. The work in [10] characterizes the performance of millimeter-wave multi-antenna systems in terms of spectral and energy efficiency, in the presence of PA nonlinearities and crosstalk. The effects of other hardware impairments such as phase noise, I/Q imbalance, and quantization have been investigated in, e.g., [11]–[13].

The second approach is to evaluate the joint effect of multiple hardware impairments; see, e.g., [4], [14]–[16]. In these works, the aggregate distortion is simply modeled as additive Gaussian noise, which is often assumed to be uncorrelated across the antenna array. While spatially uncorrelated models of distortion can be accurate in some scenarios (see, e.g., [16]), they are typically unsuitable to characterize the massive MIMO downlink with multiuser precoding [17], [18]. Aggregate hardware impairment models that capture the inherent spatial correlation of the distortion are therefore of great practical importance. An example of such an aggregate hardware

impairment model has been presented in [19]. This model captures the individual distortions caused by nonlinear amplifiers, as well as phase noise and quantization noise.

In scenarios with spatially correlated distortion, knowledge on the covariance matrix of the distortion can be leveraged to enhance the communication performance. For instance, distortion-aware linear receivers [20], [21] have shown improved performance with respect to conventional linear receivers by (partially) suppressing the distortion at the price of a reduced array gain. Distortion-aware channel estimation techniques have also been proposed recently [22]. These techniques are able to reduce the estimation error by exploiting the statistical properties of the hardware impairments.

In the downlink of multiuser MIMO systems, different transmit preprocessing techniques can be used to mitigate the effects of nonlinear distortion. Per-antenna digital predistortion (DPD) is widely known as an effective technique to mitigate nonlinear distortion at the output of the PAs [23]. This approach, however, requires a dedicated DPD block per antenna. Alternatively, beam-oriented DPD techniques (see, e.g., [24], [25]), perform predistortion prior to the baseband precoding block to linearize the signal received at the users. Despite their promising performance, these solutions often require oversampling and lead to high computational complexity. In this work, we show that by exploiting the excess spatial degrees of freedom in massive MIMO systems and incorporating the knowledge of distortion correlation and PA power consumption into the design of the linear precoder, one can suppress the nonlinear distortion in the direction of the users and achieve significant gains over conventional linear precoding schemes. More specifically, the main contributions of this paper are as follows:

- We study the directivity of the nonlinear PAs' distortion in a single-user setting. We model the nonlinear PAs using memoryless polynomials and analytically derive linear precoders that can null the distortion in the direction of the user, but at the cost of reducing the array gain to the user and increasing the radiation in unwanted directions.
- In view of the fact that the precoders nulling the distortion in the direction of the users are not necessarily favorable solutions in many scenarios, e.g., at low-to-moderate signal-to-noise ratio (SNR), we propose a linear precoder optimization framework that simultaneously takes into account the effect of distortion, the array gain, and the multiuser interference. We utilize Bussgang's theorem [26] to decompose the distorted transmit signal into a linear and an uncorrelated distortion term. We then derive a lower bound on the sum rate and propose an iterative procedure, which we refer to as distortion-aware beamforming (DAB).

This procedure computes a linear precoder that maximizes the sum rate under an average total power constraint. We show that the DAB precoder is able to realize considerable gains in terms of spectral efficiency compared to conventional distortion-agnostic precoding techniques, such as maximal ratio transmission (MRT) and zero-forcing (ZF).

- We generalize our framework by further considering the consumed power in the PAs. We adopt a simple PA power consumption model and formulate a precoder optimization framework that minimizes the consumed power under a constraint on the minimum required sum rate. We then propose an iterative algorithm to approximately solve this optimization problem. We refer to the output of this iterative algorithm as the energy-efficient DAB (EE-DAB) precoding matrix.
- We conduct extensive numerical experiments to evaluate the efficacy of the proposed framework in a variety of scenarios. We quantify the improvements that can be achieved by DAB and EE-DAB in terms of spectral and energy efficiency, respectively, compared to conventional distortion-unaware precoding techniques. Moreover, we study the out-of-band radiation and demonstrate that the proposed precoder optimization framework can be utilized to enforce that the spectral regrowth due to nonlinear PAs stays below given limits.

Paper Outline: The rest of the paper is organized as follows. The system model is described in Section II. In Section III, we utilize Bussgang's theorem to decompose the PA output into a linear term and an uncorrelated distortion term and use this decomposition to derive an expression for the achievable sum rate. We analyze the directivity of nonlinear distortion under different choices of linear precoders in Section IV. Our proposed distortion-aware precoder optimization techniques, i.e., DAB and EE-DAB, are detailed in Sections V and VI, respectively. Numerical examples that demonstrate the efficacy of the proposed precoders are given in Section VII and the paper is concluded in Section VIII.

Notation: Lowercase and uppercase boldface letters denote vectors and matrices, respectively. The superscripts $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote complex conjugate, transpose, and Hermitian transpose, respectively. We use $\mathbb{E}[\cdot]$ to denote expectation. We use $\|\mathbf{a}\|$ to denote the ℓ_2-norm of the vector $\mathbf{a}$. The $M \times M$ identity matrix is denoted by $\mathbf{I}_M$ and the $M \times M$ all-zeros matrix is denoted by $\mathbf{0}_{M\times M}$. We use $\mathbf{A} \odot \mathbf{B}$ to denote the Hadamard (entry-wise) product of two equally-sized matrices $\mathbf{A}$ and $\mathbf{B}$; $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Furthermore, $\mathrm{diag}(\mathbf{a})$ represents a diagonal matrix that contains the elements of the vector $\mathbf{a}$ on its diagonal, and $\mathrm{diag}(\mathbf{A})$ is the main diagonal of a square matrix $\mathbf{A}$. The element-wise magnitude of a matrix $\mathbf{A}$ is represented

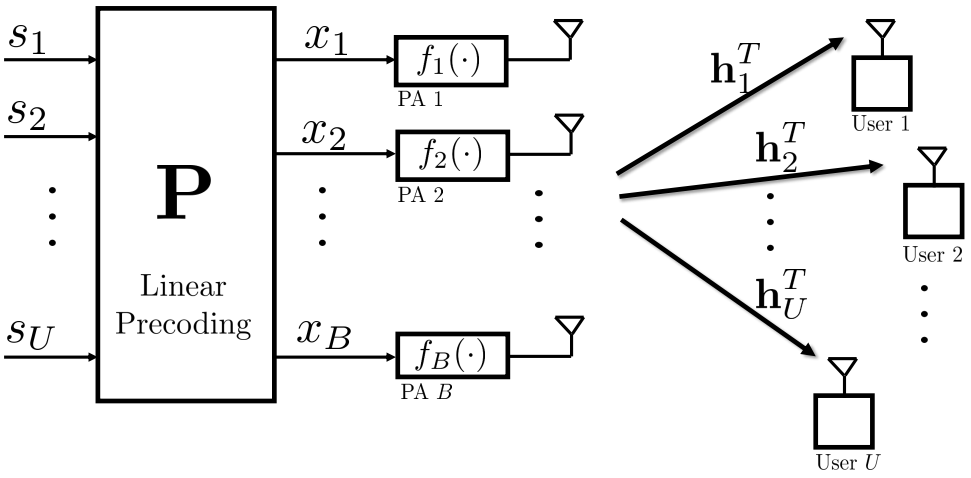


Fig. 1. Massive MIMO downlink with linear precoding and nonlinear PAs at the transmitter.

by $|\mathbf{A}|$. We use $\angle a$ to denote the angle of the complex number a. The distribution of a circularly symmetric complex Gaussian random vector with covariance matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$ is denoted by $\mathcal{CN}(\mathbf{0}, \mathbf{C})$. Finally, we use $\mathbb{1}_{\mathcal{A}}(a)$ to denote the indicator function, which is defined as $\mathbb{1}_{\mathcal{A}}(a) = 1$ for $a \in \mathcal{A}$ and $\mathbb{1}_{\mathcal{A}}(a) = 0$ for $a \notin \mathcal{A}$.

Reproducible Research: The MATLAB codes used to generate the results in this paper will be made available upon completion of the review process.

II. System Model

We consider a massive MIMO downlink scenario as depicted in Fig. 1. The base station is equipped with B antennas and serves U single-antenna users over the same time-frequency resource. Each antenna is connected to a nonlinear PA, which introduces distortion to the transmit signal.

A. Channel Input-Output Relationship

The received signal at the uth user is given by

$$y_u = \mathbf{h}_u^T f(\mathbf{x}) + w_u, \tag{1}$$

for $u = 1, \ldots, U$, where $\mathbf{h}_u \in \mathbb{C}^B$ represents the channel between the base station and the uth user. The function $f(\mathbf{x}) = [f_1(x_1), \ldots, f_B(x_B)]^{\mathrm{T}}$ describes the output of the PAs. Here,

$f_b(\cdot) : \mathbb{C} \to \mathbb{C}$ characterizes the memoryless nonlinear behavior of the PA at the bth antenna, $b = 1, \ldots, B$. We consider linear precoding where the precoded signal is given by $\mathbf{x} = \mathbf{P}\mathbf{s}$. Here, $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_U] \in \mathbb{C}^{B \times U}$ denotes the precoding matrix and $\mathbf{s} = [s_1, \ldots, s_U]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_U)$ models the transmitted symbols. Finally, $w_u \sim \mathcal{CN}(0, N_0)$ in (1) is additive white Gaussian noise (AWGN). Throughout the paper, unless stated otherwise, we assume that perfect channel state information (CSI) is known at both the transmitter and the users.[1]

B. PA Nonlinearity Model

We model the PAs using a polynomial. Specifically, the nonlinear behavior of the bth PA is modeled by the following $(2K + 1)$th order polynomial:[2]

$$f_b(x_b) = \sum_{k=0}^{K} \beta_{2k+1}^{(b)} x_b |x_b|^{2k} = \beta_1^{(b)} x_b + \beta_3^{(b)} x_b |x_b|^2 + \cdots + \beta_{2K+1}^{(b)} x_b |x_b|^{2K}, \tag{2}$$

for $b = 1, \ldots, B$, where $\beta_1^{(b)}, \beta_3^{(b)}, \ldots, \beta_{2K+1}^{(b)}$ are complex-valued model parameters that capture both amplitude-to-amplitude modulation (AM/AM) and amplitude-to-phase modulation (AM/PM) distortions [27]. Nonlinear memory effects are typically much weaker than the direct nonlinearities. Therefore, we have neglected them throughout this work and the analysis of their impact on precoder design is left for future studies.

C. PA Power Consumption Model

To quantify the power consumption of PAs, we use the model proposed in [28, Eq. (4)]. Let us denote the output power of the bth PA by $\rho_{\text{tx},b}$ and represent the consumed power at that PA by $\rho_{\text{cons},b}$. Then the power efficiency can be expressed as

$$\eta_b = \frac{\rho_{\text{tx},b}}{\rho_{\text{cons},b}}, \tag{3}$$

for $b = 1, \ldots, B$. Through measurements, the authors in [28] have verified that the following equation accurately describes the power efficiency for a variety of PAs:

$$\eta_b = \eta_{\text{max},b} \sqrt{\frac{\rho_{\text{tx},b}}{\rho_{\text{max},b}}}. \tag{4}$$

[1]In Fig. 6, we study the impact of imperfect channel state information at the transmitter.

[2]Since we are mainly interested in the effect of in-band distortion in this paper, even-order terms (which only introduce power components outside the operating frequency band) are neglected in this model. These terms will, however, be reintroduced in a numerical example in Fig. 10, where we study the out-of-band radiation.

Here, $\eta_{\text{max},b} \in [0,1]$ is the maximum power efficiency obtained when $\rho_{\text{tx},b} = \rho_{\text{max},b}$. From (3) and (4), the consumed power of the bth PA can be expressed as

$$\rho_{\text{cons},b} = \frac{1}{\eta_{\text{max},b}} \sqrt{\rho_{\text{tx},b} \rho_{\text{max},b}}, \tag{5}$$

for $b = 1, \ldots, B$ where $\rho_{\text{tx},b} \leq \rho_{\text{max},b}$. In words, the consumed power is proportional to the square root of the output power, which is compatible with results of several other studies; see, e.g., [29, Eq. (6.93)].

III. Bussgang Decomposition and Spectral Efficiency

To analyze the impact of the PA distortion on the performance of the system, we shall use Bussgang's theorem [26] to decompose the nonlinear signal into a scaled linear term and an uncorrelated distortion term. This enables us to derive a lower bound on the sum rate, which we will use as a metric for evaluating the spectral efficiency.

A. Bussgang's Theorem

By applying Bussgang's theorem [26], we can, similarly to, e.g., [10], [19], [20], rewrite the distorted signal $f(\mathbf{x})$ as

$$f(\mathbf{x}) = \mathbf{G}\mathbf{x} + \mathbf{e}, \tag{6}$$

where $\mathbf{G} \in \mathbb{C}^{B \times B}$ is a diagonal matrix containing the Bussgang gains and $\mathbf{e} \in \mathbb{C}^B$ is the distortion term which is, by definition, uncorrelated with $\mathbf{x}$, i.e., $\mathbb{E}[\mathbf{x}\mathbf{e}^H] = \mathbf{0}_{B \times B}$. The entries of $\mathbf{G}$ are given by

$$[\mathbf{G}]_{b,b} = \frac{\mathbb{E}[f(x_b) x_b^*]}{\mathbb{E}[|x_b|^2]} = \frac{\sum_{k=0}^{K} \beta_{2k+1}^{(b)} \mathbb{E}\left[|x_b|^{2k+2}\right]}{\mathbb{E}[|x_b|^2]}. \tag{7}$$

For the $(2K+1)$th order polynomial model in (2) and for $\mathbf{x} = \mathbf{P}\mathbf{s}$, we obtain $\mathbf{G}$ in (7) using the moments of the complex Gaussian random variables [30] as

$$[\mathbf{G}]_{b,b} = \frac{\sum_{k=0}^{K} \beta_{2k+1}^{(b)} (k+1)! \ (\sigma_{x_b}^2)^{k+1}}{\sigma_{x_b}^2} = \sum_{k=0}^{K} (k+1)! \ \beta_{2k+1}^{(b)} (\sigma_{x_b}^2)^k, \tag{8}$$

where $\sigma_{x_b}^2$ is the variance of the precoded signal x_b at the bth antenna. Accordingly, the matrix $\mathbf{G}$ can be expressed as a function of the precoding matrix $\mathbf{P}$ as

$$\mathbf{G}(\mathbf{P}) = \sum_{k=0}^{K} (k+1)!\ \mathbf{A}_{2k+1}\text{diag}(\mathbf{C}_\mathbf{x})^k \tag{9}$$

$$= \mathbf{A}_1\mathbf{I}_B + 2\mathbf{A}_3\text{diag}(\mathbf{C}_\mathbf{x}) + \cdots + (K+1)!\ \mathbf{A}_{2K+1}\text{diag}(\mathbf{C}_\mathbf{x})^K, \tag{10}$$

where

$$\mathbf{A}_{2k+1} = \text{diag}\big(\beta_{2k+1}^{(1)}, \ldots, \beta_{2k+1}^{(B)}\big), \tag{11}$$

for $k = 1, \ldots, K$ and $\mathbf{C}_\mathbf{x} = \mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{P}\mathbf{P}^H$ is the covariance matrix of the input. Using (9) and following an approach similar to that used in [20, Eq. 24] and [10, Eq. 11], the covariance matrix of the distortion $\mathbf{e}$ can be derived as

$$\mathbf{C}_\mathbf{e}(\mathbf{P}) = \sum_{k=1}^{K} \mathbf{L}_k\mathbf{C}_\mathbf{x} \odot |\mathbf{C}_\mathbf{x}|^{2k}\mathbf{L}_k^H, \tag{12}$$

where

$$\mathbf{L}_k = \frac{1}{\sqrt{k+1}} \sum_{l=k}^{K} \binom{l}{k} (l+1)!\ \mathbf{A}_{2l+1}\text{diag}(\mathbf{C}_\mathbf{x})^{l-k}. \tag{13}$$

It can be seen from (12) that if the channel input $\mathbf{x}$ has correlated entries, then the distortion will be spatially correlated. In other words, non-zero off-diagonal elements in $\mathbf{C}_\mathbf{x}$ (as is generally the case for precoded signal) lead to non-zero off-diagonal elements in the distortion covariance matrix $\mathbf{C}_\mathbf{e}$.

B. Achievable Sum Rate

By inserting (6) into (1), we can write the received signal at the uth user as

$$y_u = \mathbf{h}_u^T\mathbf{G}(\mathbf{P})\mathbf{p}_u s_u + \sum_{r\neq u} \mathbf{h}_u^T\mathbf{G}(\mathbf{P})\mathbf{p}_r s_r + \mathbf{h}_u^T\mathbf{e} + w_u. \tag{14}$$

Here, s_u is the desired symbol at the uth user. Moreover, $\mathbf{p}_u$ and $\mathbf{p}_r$ denote the uth and the rth columns of the precoding matrix $\mathbf{P}$, respectively. The second, third, and fourth terms on the right-hand side of (14) are the inter-user interference, the received nonlinear distortion, and the AWGN, respectively. The sum of these three terms can be considered as the effective noise $w_{\text{eff},u} = \sum_{r\neq u} \mathbf{h}_u^T\mathbf{G}(\mathbf{P})\mathbf{p}_r s_r + \mathbf{h}_u^T\mathbf{e} + w_u$. It should be noted that $w_{\text{eff},u}$ is, in general, not Gaussian due to the nonlinearity of the PA, and thus determining the exact sum rate for

the input-output model in (14) is not straightforward [31]. A common approach to evaluate the spectral efficiency is to use the so-called "auxiliary-channel lower bound" (see, e.g., [32, Sec. VI]) to derive a lower bound on the achievable rate. More specifically, in the auxiliary channel, the effective noise $w_{\text{eff},u}$ is replaced by the auxiliary noise $\tilde{w}_{\text{eff},u}$, which is a circularly symmetric complex Gaussian random variable whose variance is the same as that of $w_{\text{eff},u}$. The capacity of the auxiliary channel can be given in closed form as

$$R_{\text{sum}}(\mathbf{P}) = \sum_{u=1}^{U} \log_2 \left(1 + \text{SINDR}_u(\mathbf{P})\right), \tag{15}$$

where

$$\text{SINDR}_u(\mathbf{P}) = \frac{|\mathbf{h}_u^T \mathbf{G}(\mathbf{P}) \mathbf{p}_u|^2}{\sum\limits_{r \neq u} |\mathbf{h}_u^T \mathbf{G}(\mathbf{P}) \mathbf{p}_r|^2 + \mathbf{h}_u^T \mathbf{C}_{\mathbf{e}}(\mathbf{P}) \mathbf{h}_u^* + N_0} \tag{16}$$

is the signal-to-interference-noise and distortion ratio (SINDR) at the uth user. This lower bound corresponds to the sum rate achieved with a Gaussian codebook, assuming that the channel is perfectly known to the users and using a decoder that operates according to the nearest neighbor principle, i.e., that returns the codeword closest to the received signal in the Euclidean norm [33].

C. Precoding Matrix Normalization

We utilize the decomposition in (6) to introduce a procedure for precoding matrix normalization, which is needed to control the average total transmit power under different choices of the precoding matrix. In particular, our objective is to find a real-valued normalization factor α such that

$$\mathbb{E}\left[\|f(\alpha \mathbf{P}\mathbf{s})\|^2\right] = \rho_{\text{tot}}, \tag{17}$$

where ρ_{tot} is a positive real value denoting the total power budget. Let us define $\mathbf{z} = f(\alpha \mathbf{P}\mathbf{s})$ and $\mathbf{C}_{\mathbf{z}} = \mathbb{E}[\mathbf{z}\mathbf{z}^H]$. Then, the normalization problem can be rewritten as finding α such that $\text{tr}(\mathbf{C}_{\mathbf{z}}) - \rho_{\text{tot}} = 0$, i.e., solving

$$\text{tr}\left(\mathbf{C}_{\mathbf{e}}(\alpha \mathbf{P}) + \alpha^2 \mathbf{G}(\alpha \mathbf{P}) \mathbf{P}\mathbf{P}^H \mathbf{G}(\alpha \mathbf{P})^H\right) - \rho_{\text{tot}} = 0, \tag{18}$$

where $\mathbf{G}(\alpha \mathbf{P})$ and $\mathbf{C}_{\mathbf{e}}(\alpha \mathbf{P})$ are the Bussgang gain and covariance matrix of distortion corresponding to the decomposition of $f(\alpha \mathbf{P}\mathbf{s})$ as in (6). Using (9)–(11) and (12)–(13), we can

express (18) as

$$\sum_{k=1}^{K} \alpha^{2k+4} \mathrm{tr}\left(\mathbf{L}_k \left(\mathbf{P}\mathbf{P}^H \odot |\mathbf{P}\mathbf{P}^H|^{2k} \right) \mathbf{L}_k^H \right)$$
$$+ \alpha^2 \mathrm{tr}\left(\left(\sum_{k=0}^{K} \alpha^k (k+1)!\ \mathbf{A}_{2k+1} \mathrm{diag}(\mathbf{P}\mathbf{P}^H)^k \right) \mathbf{P}\mathbf{P}^H \left(\sum_{k=0}^{K} \alpha^k (k+1)!\ \mathbf{A}_{2k+1} \mathrm{diag}(\mathbf{P}\mathbf{P}^H)^k \right)^H \right)$$
$$- \rho_{\text{tot}} = 0, \tag{19}$$

where $\mathbf{L}_k$ is given in (13). Note that the left hand side of (19) is a $(K+2)$th order polynomial in $\xi = \alpha^2$. Therefore, to solve (19), all we need to do is to find the roots of this polynomial and then pick the positive real solution for ξ, which yields the desired scalar $\alpha = \sqrt{\xi}$.

IV. Spatial Direction of Transmitted Distortion

In this section, we analyze the directivity of the nonlinear distortion term $\mathbf{e}$ for different choices of the precoding matrix $\mathbf{P}$. To do so, we focus our attention on line-of-sight (LoS) channels. Let us assume that the transmit antennas in the system model depicted in Fig. 1 are arranged as a uniform linear array (ULA) with $\lambda_c/2$ spacing (where λ_c is the carrier wavelength). Consider a single-path LoS channel between the transmitter and the uth user with an angle of departure (AoD) of ψ_u. Then, the channel coefficients for the plane-wave model can be expressed as[3]

$$[\mathbf{h}(\psi_u)]_b = e^{-j\pi(b-1)\cos(\psi_u)}, \tag{20}$$

for $b = 1, \ldots, B$.

A. Directivity of Distortion with MRT Precoding

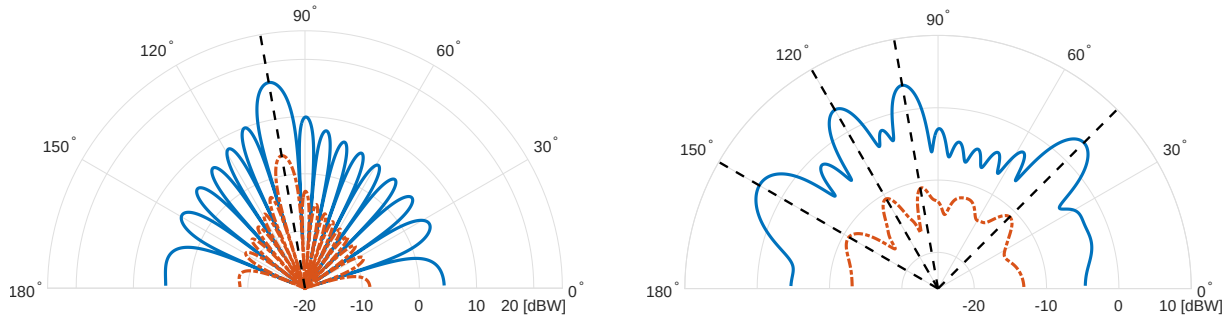
To study the directivity of the useful signal and distortion, we calculate the power of the linear and distortion terms in different directions $\psi \in [0, 2\pi)$. In particular, for a given precoding matrix $\mathbf{P}$, we evaluate

$$\rho_{\text{lin}}(\psi) = \mathbf{h}^T(\psi)\mathbf{G}\mathbf{P}\mathbf{P}^H\mathbf{G}^H\mathbf{h}^*(\psi), \tag{21}$$

and

$$\rho_{\text{dist}}(\psi) = \mathbf{h}^T(\psi)\mathbf{C}_{\mathbf{e}}\mathbf{h}^*(\psi), \tag{22}$$

[3]In this section, for simplicity, we ignore path-loss effects.



(a) MRT precoding; $U = 1$ user and $\psi = 100°$.

(b) MRT precoding; $U = 4$ users, $\psi_1 = 45°$, $\psi_2 = 100°$, $\psi_3 = 120°$, and $\psi_4 = 150°$.

Fig. 2. Far-field radiation pattern for MRT precoding and equal PAs; $B = 16$ antennas, $U \in \{1, 4\}$ users. PA coefficients $\beta_1^{(b)} = 0.98$ and $\beta_3^{(b)} = -0.02 - 0.01j$ for $b = 1, \ldots, B$. The precoding matrix is normalized such that the total transmit power is set to $\rho_{\text{tot}} = 43\,\text{dBm}$. The blue curves correspond to the radiated linear power $\rho_{\text{lin}}(\psi)$ and the red curves correspond to the distortion power $\rho_{\text{dist}}(\psi)$. MRT precoding steers the distortion in the direction of user(s).

which correspond to the radiated linear power and the distortion power, respectively. For the sake of simplicity, throughout this section, we consider only a third-order nonlinearity, namely, $K = 1$ in (2), for which the Bussgang gain in (9)–(11) and the distortion covariance matrix in (12)–(13) reduce to [20]

$$\mathbf{G} = \mathbf{A}_1 \mathbf{I}_B + 2\mathbf{A}_3 \text{diag}(\mathbf{P}\mathbf{P}^H), \tag{23}$$

and

$$\mathbf{C}_{\mathbf{e}}(\mathbf{P}) = 2\mathbf{A}_3 \left(\mathbf{P}\mathbf{P}^H \odot |\mathbf{P}\mathbf{P}^H|^2\right) \mathbf{A}_3^*, \tag{24}$$

respectively. Let us first consider the special case with $U = 1$ user and equal PAs at different antennas. In this case, the MRT precoder is given by $\mathbf{p} = \alpha \mathbf{h}^*/\|\mathbf{h}\|$ where α is the normalization factor that has been described in Section III-C. Note that, under these assumptions, we get $\mathbf{p}\mathbf{p}^H \odot |\mathbf{p}\mathbf{p}^H|^2 = c\mathbf{p}\mathbf{p}^H$ where c is a real scalar. As a result, by replacing (23) and (24) in (21) and (22), respectively, we observe that the directivity of the distortion matches exactly that of the linear term. This is illustrated in Fig. 2a, where we have plotted the far-field radiation pattern for MRT precoding in a setup with $U = 1$ user and $B = 16$ antennas with equal PAs, $\beta_1^{(b)} = 0.98$ and $\beta_3^{(b)} = -0.02 - 0.01j$ for $b = 1, \ldots, B$.

Similarly, by evaluating (21) and (22) for MRT precoding in multiuser scenarios, i.e., $U > 1$, where $\mathbf{p}_u = \alpha \mathbf{h}_u^*/\|\mathbf{h}_u\|$ for $u = 1, \ldots, U$, we can see that this precoding technique steers the nonlinear distortion in the direction of the users. This phenomenon is demonstrated for the case $U = 4$ in Fig. 2b.

The above analysis reveals that in the presence of nonlinear PAs, MRT precoding can lead to a considerable performance degradation due to the beamforming of distortion.[4] Motivated by this, in what follows, we seek precoder design approaches that take explicitly into account the directivity of distortion. In Section IV-B, we focus on the case of possibly different PAs at different antennas and demonstrate that for single-user scenarios, properly designed precoders can fully eliminate the distortion in the direction of the user at the price of a reduced array gain. We refer to these precoders as zero-distortion precoders. Later on, in Section V, we study the precoder design in a general scenario with multiple users and arbitrary channel models.

B. Zero-Distortion Precoding in Single-User Scenarios

We start by considering a simple scenario with a single user, i.e., $U = 1$ and $B = 2$ transmit antennas. To reduce the complexity further, let us consider a third-order nonlinearity model with $\beta_1^{(1)} = \beta_1^{(2)} = 1$. In the absence of AWGN, we can write the received signal as

$$\begin{aligned} y &= \begin{bmatrix} h_1 & h_2 \end{bmatrix} \begin{bmatrix} p_1 s + \beta_3^{(1)} p_1 s |p_1 s|^2 \\ p_2 s + \beta_3^{(2)} p_2 s |p_2 s|^2 \end{bmatrix} \\ &= (h_1 p_1 + h_2 p_2) s + (\beta_3^{(1)} h_1 p_1 |p_1|^2 + \beta_3^{(2)} h_2 p_2 |p_2|^2) s |s|^2, \end{aligned} \tag{25}$$

where p_1 and p_2 are the precoding coefficients associated with the first and the second antenna at the base station, respectively. It follows from (25) that, to have zero distortion at the UE, the precoding coefficients must satisfy

$$\beta_3^{(1)} h_1 p_1 |p_1|^2 + \beta_3^{(2)} h_2 p_2 |p_2|^2 = 0. \tag{26}$$

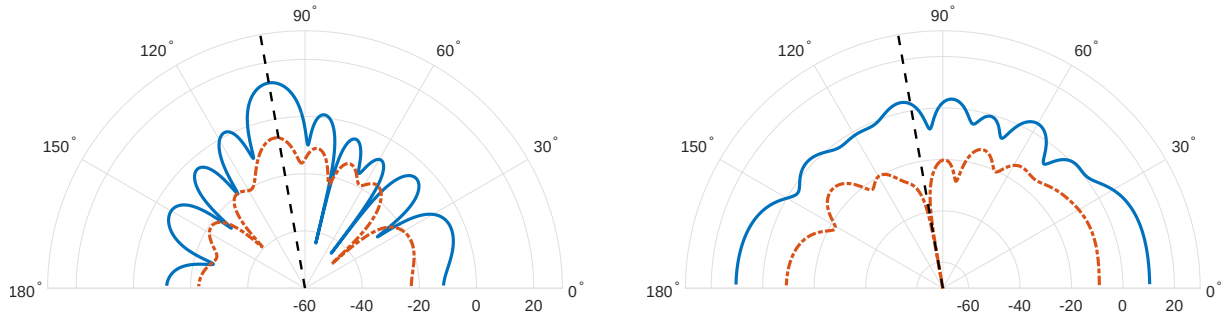
For the LoS channel given in (20) with a single user at angle ψ, the following precoder is a solution of (26):[5]

$$\mathbf{p} = \alpha \begin{bmatrix} 1 \\ \left| \frac{\beta_3^{(1)}}{\beta_3^{(2)}} \right|^{\frac{1}{3}} e^{j\left(\pi \cos(\psi) + \pi + \angle \frac{\beta_3^{(1)}}{\beta_3^{(2)}}\right)} \end{bmatrix}, \tag{27}$$

where α is the normalization factor (as described in Section III-C). Note that if $\beta_3^{(1)} = \beta_3^{(2)}$, namely, when the PAs at the two antenna ports are identical, the received useful (linear) signal

[4]The same statement is true for ZF precoding.

[5]This solution can also be found by setting $\rho_{\text{dist}}(\psi)$ to zero.



(a) MRT; $U = 1$ user and $\psi = 100°$.

(b) Zero-distortion precoding in (28); $U = 1$ user and $\psi = 100°$.

Fig. 3. Far-field radiation pattern for MRT and zero-distortion precoding with unequal PAs; $B = 10$ antennas and $U = 1$ user at $\psi = 100$. Total transmit power is set to $\rho_{\text{tot}} = 43\,\text{dBm}$. The nonlinearity coefficients are as in Table I. The blue curve corresponds to the linear term of the transmitted signal and the red curve corresponds to the distortion term. Zero-distortion precoder nulls the distortion in the direction of the user while reducing array gain by about 25 dB compared to MRT.

is also zero. Thus, no useful signal is received by the user. Therefore, in order for this precoder to yield positive achievable rates over a LoS channel, PAs with different distortion profiles are needed at different antenna ports.

For $B > 2$, the problem becomes underdetermined and various precoders may be found that result in zero distortion. For example, when B is even, the following precoder yields zero distortion:

$$\mathbf{p} = \alpha \begin{bmatrix} 1 \\ \left|\frac{\beta_3^{(1)}}{\beta_3^{(2)}}\right|^{\frac{1}{3}} e^{j\left(\pi\cos(\psi)+\pi+\angle\frac{\beta_3^{(1)}}{\beta_3^{(2)}}\right)} \\ 1 \\ \left|\frac{\beta_3^{(3)}}{\beta_3^{(4)}}\right|^{\frac{1}{3}} e^{j\left(\pi\cos(\psi)+\pi+\angle\frac{\beta_3^{(3)}}{\beta_3^{(4)}}\right)} \\ \vdots \\ 1 \\ \left|\frac{\beta_3^{(B-1)}}{\beta_3^{(B)}}\right|^{\frac{1}{3}} e^{j\left(\pi\cos(\psi)+\pi+\angle\frac{\beta_3^{(B-1)}}{\beta_3^{(B)}}\right)} \end{bmatrix}. \tag{28}$$

Fig. 3 depicts the far-field radiation pattern for MRT as well as the zero-distortion precoding given in (28) for a setup with a single user and $B = 10$ antennas with different PAs at different antenna ports. The nonlinearity parameters for the PAs are given in Table I. It can be seen from Fig. 3b that the zero-distortion precoder nulls the distortion in the direction of the user at the price of a reduced array gain— about 25 dB reduction compared to that of MRT precoding as

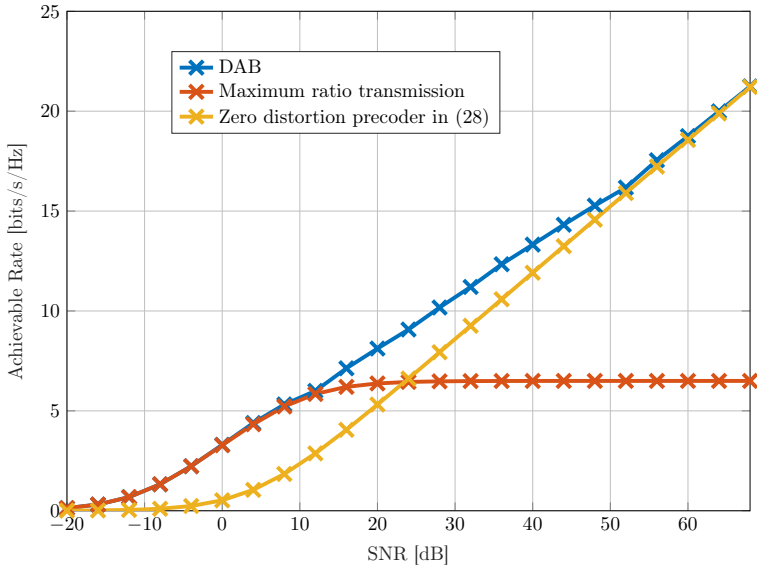


Fig. 4. Achievable rates averaged over 500 realizations of LoS channels in (20) with uniformly randomly distributed AoDs $\psi \in [0, 180)$, $B = 10$ antennas at base station and $U = 1$ user. The nonlinearity coefficients are as in Table I. With distortion-aware precoding, the achievable rate becomes a strictly increasing function of the SNR.

depicted in Fig. 3a.

In Fig. 4, for different SNR values, where we define SNR $= \rho_{\text{tot}}/N_0$, we evaluate the sum rate in (15) for the precoder in (28) as well as the MRT precoder, over LoS channels with $B = 10$ and $U = 1$. It can be seen that the MRT precoder can achieve a higher rate than the zero-distortion precoder at low SNR values where AWGN dominates the nonlinear distortion. However, at higher SNRs, where the nonlinear distortion is dominant, the achievable rate for MRT saturates, whereas the achievable rate for the precoder in (28) is a strictly increasing function of SNR.

As a "sneak peek" of what we are going to describe in the next section, we have also included the results achieved by DAB precoding scheme, which achieves an enhanced performance over

TABLE I
PA NONLINEARITY PARAMETERS IN THE EXAMPLES DEMONSTRATED IN FIGS 3 AND 4.

$\beta_1^{(1)}$	$\beta_1^{(2)}$	$\beta_1^{(3)}$	$\beta_1^{(4)}$	$\beta_1^{(5)}$	$\beta_1^{(6)}$	$\beta_1^{(7)}$	$\beta_1^{(8)}$	$\beta_1^{(9)}$	$\beta_1^{(10)}$
1.047	1.026	0.984	1.027	1.003	0.994	1.000	1.011	0.991	1.029
$\beta_3^{(1)}$	$\beta_3^{(2)}$	$\beta_3^{(3)}$	$\beta_3^{(4)}$	$\beta_3^{(5)}$	$\beta_3^{(6)}$	$\beta_3^{(7)}$	$\beta_3^{(8)}$	$\beta_3^{(9)}$	$\beta_3^{(10)}$
-0.024 $-0.015j$	-0.012 $-0.035j$	-0.043 $-0.023j$	-0.013 $-0.014j$	-0.022 $-0.028j$	-0.038 $-0.026j$	-0.051 $-0.012j$	-0.042 $-0.021j$	-0.030 $-0.019j$	-0.014 $-0.038j$

the whole range of SNR and converges to the zero-distortion precoder at high SNR values. The algorithm explained in the next section is based on a general precoder design framework that is able to simultaneously take into account the effect of the linear precoder on the array gain, distortion, and multiuser interference.

V. Distortion-Aware Precoding

The choice of the precoding matrix $\mathbf{P}$ affects the sum rate as it controls the array gain, the multiuser interference, and the distortion radiated in the direction of the users (as we saw in Section IV). Motivated by this observation, we now present a framework for optimizing the precoding matrix $\mathbf{P}$ that uses knowledge of the nonlinearity parameters of PA in addition to CSI and aims to maximize the lower bound on the sum rate in (15). In particular, we formulate this problem as

$$\begin{aligned} \underset{\mathbf{P}\in\mathbb{C}^{B\times U}}{\text{maximize}} \quad & R_{\text{sum}}(\mathbf{P}) \\ \text{subject to} \quad & \mathbb{E}\left[\|f(\mathbf{P}\mathbf{s})\|^2\right] = \rho_{\text{tot}}. \end{aligned} \tag{29}$$

Here, ρ_{tot} denotes the average total transmit power. The power constraint in $\mathbb{E}\left[\|f(\mathbf{P}\mathbf{s})\|^2\right] = \rho_{\text{tot}}$ corresponds to a scenario with fixed total transmit power.[6] Note that (29) is a non-convex optimization problem since $R_{\text{sum}}(\mathbf{P})$ is a non-convex function of $\mathbf{P}$. Inspired by the solution in [34, Sec. V], in what follows, we solve this problem approximately using an iterative projected gradient ascent algorithm. We refer to this algorithm as DAB. In particular, DAB updates the precoding matrix by taking steps along the steepest ascent direction of the objective function $R_{\text{sum}}(\mathbf{P})$. The resulting precoding matrix is then normalized to ensure the feasibility of the solution. This procedure can be formulated as [34, Eq. (72)]

$$\widetilde{\mathbf{P}} = \left[\mathbf{P}^{(i-1)} + \mu^{(i-1)}\nabla_{\mathbf{P}} R_{\text{sum}}\left(\mathbf{P}^{(i-1)}\right)\right]^{+}_{\mathbb{E}\left\|f\left(\mathbf{P}^{(i-1)}\mathbf{s}\right)\right\|^2=\rho_{\text{tot}}}. \tag{30}$$

Here, $i = 1, \ldots, I$ is the iteration index, I is the maximum number of iterations, $\mu^{(i)}$ is the step size of the ith iteration, and $[\cdot]^{+}_{\mathbb{E}\left\|f\left(\mathbf{P}^{(i-1)}\mathbf{s}\right)\right\|^2=\rho_{\text{tot}}}$ denotes the normalization of the updated precoding matrix, such that the power constraint $\mathbb{E}\left[\left\|f(\mathbf{P}^{(i-1)}\mathbf{s})\right\|^2\right] = \rho_{\text{tot}}$ is satisfied. The details of this normalization have been explained in Section III-C. If $R_{\text{sum}}(\widetilde{\mathbf{P}}) > R_{\text{sum}}(\mathbf{P}^{(i-1)})$, we update the precoding matrix to $\mathbf{P}^{(i)} = \widetilde{\mathbf{P}}$ and reset the step size $\mu^{(i)} = \mu^{(0)}$. Otherwise, we

[6]Later on in Fig. 7, we demonstrate how this algorithm can be employed along with a power control strategy.

Algorithm 1 Algorithm for computing the distortion-aware beamforming (DAB) precoding matrix.

Inputs: $\mathbf{h}_1, \ldots, \mathbf{h}_U$, $\beta_1^{(b)}, \ldots, \beta_K^{(b)}$ for $b = 1, \ldots, B$, ρ_{tot}, and N_0

Output: $\mathbf{P}_{\text{DAB}}$

Initialization: $\mu^{(0)}$ and $\mathbf{P}^{(0)}$

1: $R_{\text{sum}}^{(0)} \leftarrow R_{\text{sum}}(\mathbf{P}^{(0)})$
2: **for** $i = 1, \ldots, I$ **do**
3: $\quad \widetilde{\mathbf{P}} \leftarrow \left[\mathbf{P}^{(i-1)} + \mu^{(i-1)} \nabla_{\mathbf{P}} R_{\text{sum}}\left(\mathbf{P}^{(i-1)}\right)\right]^{+}_{\mathbb{E}\left\|f\left(\mathbf{P}^{(i-1)}\mathbf{s}\right)\right\|^2 = \rho_{\text{tot}}}$
4: $\quad \widetilde{R}_{\text{sum}} \leftarrow R_{\text{sum}}(\widetilde{\mathbf{P}})$
5: $\quad$ **if** $\widetilde{R}_{\text{sum}} > R_{\text{sum}}^{(i-1)}$ **then**
6: $\quad\quad \mathbf{P}^{(i)} \leftarrow \widetilde{\mathbf{P}}$, $R_{\text{sum}}^{(i)} \leftarrow \widetilde{R}_{\text{sum}}$, and $\mu^{(i)} \leftarrow \mu^{(0)}$
7: $\quad$ **else**
8: $\quad\quad \mathbf{P}^{(i)} \leftarrow \mathbf{P}^{(i-1)}$, $R_{\text{sum}}^{(i)} \leftarrow R_{\text{sum}}^{(i-1)}$, and $\mu^{(i)} \leftarrow \frac{1}{2}\mu^{(i-1)}$
9: $\quad$ **end if**
10: **end for**
11: $\mathbf{P}_{\text{DAB}} \leftarrow \mathbf{P}^{(I)}$

do not update the precoding matrix, i.e., $\mathbf{P}^{(i)} = \mathbf{P}^{(i-1)}$, and decrease the step size $\mu^{(i)} = \frac{1}{2}\mu^{(i-1)}$. Finally, we choose $\mathbf{P}_{\text{DAB}} = \mathbf{P}^{(I)}$ as the output of the algorithm. We refer to this solution as the DAB precoding matrix. In Algorithm 1, we summarize the steps required for computing the DAB precoding matrix using a projected gradient ascent approach.

The computational complexity of Algorithm 1 is dominated by Step 3 and in particular the gradient calculation. We now explain the procedure for gradient calculation in (30). The gradient calculation can be carried out either numerically or through evaluation of the closed-form expression. In order to compute $\nabla_{\mathbf{P}} R_{\text{sum}}(\mathbf{P})$ numerically, we shall evaluate

$$\left[\nabla_{\mathbf{P}} R_{\text{sum}}(\mathbf{P})\right]_{b,u} \approx \frac{\left(R_{\text{sum}}\left(\mathbf{P} + \boldsymbol{\Delta}_{\text{r}}^{(b,u)}\right) - R_{\text{sum}}\left(\mathbf{P}\right)\right) + j\left(R_{\text{sum}}\left(\mathbf{P} + \boldsymbol{\Delta}_{\text{i}}^{(b,u)}\right) - R_{\text{sum}}\left(\mathbf{P}\right)\right)}{\delta}, \tag{31}$$

for $b = 1, \ldots, B$ and $u = 1, \ldots, U$, where δ is a small positive constant and $\boldsymbol{\Delta}_{\text{r}}^{(b,u)}$ is a $B \times U$ matrix with δ on the bth row and the uth column and zero entries on all other positions. Similarly, $\boldsymbol{\Delta}_{\text{i}}^{(b,u)}$ is a $B \times U$ matrix whose only non-zero entry is $j\delta$ on the bth row and the uth column. A more accurate way of calculating this gradient is via deriving and evaluating closed-form expressions. This case is discussed for the special case of third-order nonlinearity ($K = 1$) in Appendix A.

Due to the non-convexity of the problem, Algorithm 1 does not provide global optimality guarantees, i.e., it can return a local maximum. To increase the likelihood of Algorithm 1 to converge to the global maximum, we repeat it with multiple initializations and pick the solution that achieves the highest sum rate. Including the conventional precoding matrices, i.e., MRT and ZF, among the set of initializations can help accelerate the convergence, and it guarantees that the DAB precoder does not perform worse than these precoders. It is worth noting that Algorithm 1 can be optimized in several ways to improve convergence[7] or reduce computational complexity. However, since the main goal of this paper is to highlight the performance gain that can be achieved by distortion-aware linear precoding, rather than proposing an optimal transmission scheme, we shall not discuss such optimizations.

VI. Energy-Efficient Distortion-Aware Precoding

In Section V, we studied the precoder design with the objective of maximizing the sum rate under a given power constraint. The average total power constraint in (29) does not take into account the maximum output power constraint of the individual PAs. In this section, we consider a precoding matrix optimization problem, in which the precoding matrix is subjected to a per-antenna maximum radiated power constraint. Furthermore, we adopt the model in (5) to quantify the total base station side consumed power. This model takes into account the dissipated power in addition to the sum radiated power, which allows for the evaluation and optimization of energy efficiency.

A common way for quantifying energy efficiency is via the evaluation of the ratio between throughput and consumed power [36]

$$\eta_{\text{EE}} = \frac{W R_{\text{sum}}(\mathbf{P})}{\rho_{\text{cons,tot}}(\mathbf{P})}, \tag{32}$$

measured in bits per joule. Here, W denotes the bandwidth and $\rho_{\text{cons,tot}}(\mathbf{P}) = \sum_{b=1}^{B} \rho_{\text{cons},b}(\mathbf{P})$ is the total consumed power which can be evaluated using (5). In an attempt to maximize energy efficiency, we seek the precoding matrix $\mathbf{P}$ that minimizes the consumed power while guaranteeing a given minimum required sum rate, i.e.,

$$R_{\text{sum}}(\mathbf{P}) \geq R_0, \tag{33}$$

[7]For example, by choosing the step sizes $\mu^{(i)}$ according to Armijo's step size rule, improved convergence properties can be achieved [35].

and under a per-antenna power constraint given by

$$\rho_{\text{tx},b} = \mathbb{E}\left[|f_b(x_b)|^2\right] \leq \rho_{\text{max},b}, \tag{34}$$

for $b = 1, \ldots, B$. In (34), x_b is the precoded signal input to the bth PA and can be obtained by multiplying the bth row of the precoding matrix $\mathbf{P}$ by the transmitted symbols $\mathbf{s}$. The energy efficiency maximization problem can be formulated as

$$\begin{aligned} \underset{\mathbf{P}\in\mathbb{C}^{B\times U}}{\text{minimize}} \quad & \rho_{\text{cons,tot}}(\mathbf{P}) \\ \text{subject to} \quad & \text{(33) and (34)}. \end{aligned} \tag{35}$$

We solve the nonconvex problem in (35) approximately in two steps. First, we find the precoding matrix that maximizes the sum rate under a per-antenna power constraint, namely, we maximize $R_{\text{sum}}(\mathbf{P})$ subject to (34). We refer to this problem as ***S1***. Then, starting from the solution of ***S1***, we minimize the consumed power by updating the precoding matrix along the descent direction of the consumed power $\rho_{\text{cons,tot}}(\mathbf{P})$ until (33) is violated or a maximum number of iterations is reached. This problem is referred to as ***S2*** and its output, which is an approximate solution of (35), is called energy-efficient DAB (EE-DAB) precoding matrix.

With a slight modification, Algorithm 1 in Section V can be used to solve ***S1***. In particular, we replace (30) (Step 3 in Algorithm 1) with

$$\widetilde{\mathbf{P}} = \left[\mathbf{P}^{(i-1)} + \mu^{(i-1)}\nabla_{\mathbf{P}} R_{\text{sum}}\left(\mathbf{P}^{(i-1)}\right)\right]^{+}_{\mathbb{E}[|f_b(x_b)|^2]\leq\rho_{\text{max},b}}, \tag{36}$$

where $[\cdot]^{+}_{\mathbb{E}[|f_b(x_b)|^2]\leq\rho_{\text{max},b}}$ denotes *occasional* normalization of the bth row of the precoding matrix $\mathbf{P}$ such that (34) is satisfied. Specifically, after updating the precoding matrix (line 3 of Algorithm 1), normalization is carried out only if the per-antenna power constraint is violated at one or more antennas. For these antennas, the normalization problem is equivalent to finding the values of the scalars $\alpha_{\hat{b}}$ such that $\mathbb{E}[|f(\alpha_{\hat{b}}x_{\hat{b}})|^2] = \rho_{\text{max},\hat{b}}$ where $\hat{b}$ denotes the index of an antenna for which (34) is violated. The normalization problem can be solved using the same procedure described in (18)–(19). Other than the normalization, all other steps are implemented in the same way as explained in Section V.

Once ***S1*** is solved, its solution $\mathbf{P}^{(\text{S1})}$ is used to initialize the algorithm for solving ***S2***. This algorithm seeks the precoding matrix that minimizes the consumed power for a given sum rate no smaller than R_0. Specifically, it updates the precoding matrix using the projected gradient

Algorithm 2 Algorithm for computing the energy-efficient DAB (EE-DAB) precoding matrix.

Inputs: R_0, $\mathbf{h}_1, \ldots, \mathbf{h}_U$, $\beta_1^{(b)}, \ldots, \beta_K^{(b)}$, $\rho_{\max,b}$, $\eta_{\max,b}$ for $b = 1, \ldots, B$, and N_0

Output: $\mathbf{P}_{\text{EE-DAB}}$

Initialization: $\mu^{(0)}$ and $\mathbf{P}^{(\text{S1})}$ from Algorithm 1 (with per-antenna normalization)

$\rho_{\text{cons,tot}}^{(0)} \leftarrow \rho_{\text{cons,tot}}(\mathbf{P}^{(\text{S1})})$ and $i = 1$

while $R_{\text{sum}}(\mathbf{P}^{(i)}) \geq R_0 \quad$ and $\quad i \leq I$ **do**

  $\widetilde{\mathbf{P}} \leftarrow \left[\mathbf{P}^{(i-1)} - \mu^{(i-1)} \nabla_{\mathbf{P}} \rho_{\text{cons,tot}}\left(\mathbf{P}^{(i-1)}\right)\right]^{+}_{\mathbb{E}\|f(x_b)\|^2 \leq \rho_{\max,b}}$

  $\widetilde{\rho}_{\text{cons,tot}} \leftarrow \rho_{\text{cons,tot}}(\widetilde{\mathbf{P}})$

  if $\widetilde{\rho}_{\text{cons,tot}} < \rho_{\text{cons,tot}}^{(i-1)}$ **then**

    $\mathbf{P}^{(i)} \leftarrow \widetilde{\mathbf{P}}$, $\rho_{\text{cons,tot}}^{(i)} \leftarrow \widetilde{\rho}_{\text{cons,tot}}$, and $\mu^{(i)} \leftarrow \mu^{(0)}$

  else

    $\mathbf{P}^{(i)} \leftarrow \mathbf{P}^{(i-1)}$, $\rho_{\text{cons,tot}}^{(i)} \leftarrow \rho_{\text{cons,tot}}^{(i-1)}$, and $\mu^{(i)} \leftarrow \frac{1}{2}\mu^{(i-1)}$

  end if

  $i = i + 1$

end while

$\mathbf{P}_{\text{DAB}} \leftarrow \mathbf{P}^{(\hat{i})}$ where $\hat{i} = i - 1$ is the final iteration where the stopping criterion is met.

descent, namely, in each iteration,

$$\widetilde{\mathbf{P}} = \left[\mathbf{P}^{(i-1)} - \mu^{(i-1)} \nabla_{\mathbf{P}} \rho_{\text{cons,tot}}\left(\mathbf{P}^{(i-1)}\right)\right]^{+}_{\mathbb{E}[|f_b(x_b)|^2] \leq \rho_{\max,b}}, \tag{37}$$

is calculated for $i = 1, \ldots, I$, where $\mathbf{P}^{(0)} = \mathbf{P}^{(\text{S1})}$. The gradient $\nabla_{\mathbf{P}} \rho_{\text{cons,tot}}\left(\mathbf{P}^{(i-1)}\right)$ can be evaluated similarly to (31) as

$$\begin{aligned}\left[\nabla_{\mathbf{P}} \rho_{\text{cons,tot}}(\mathbf{P})\right]_{b,u} \\ \approx \frac{\left(\rho_{\text{cons,tot}}\left(\mathbf{P} + \boldsymbol{\Delta}_{\text{r}}^{(b,u)}\right) - \rho_{\text{cons,tot}}\left(\mathbf{P}\right)\right) + j\left(\rho_{\text{cons,tot}}\left(\mathbf{P} + \boldsymbol{\Delta}_{\text{i}}^{(b,u)}\right) - \rho_{\text{cons,tot}}\left(\mathbf{P}\right)\right)}{\delta}\end{aligned} \tag{38}$$

for $b = 1, \ldots, B$ and $u = 1, \ldots, U$. The normalization in (37), similarly to that of (36), is carried out occasionally and on the rows of the precoding matrix that violate (34). At each iteration, if $\rho_{\text{cons,tot}}(\widetilde{\mathbf{P}}) < \rho_{\text{cons,tot}}(\mathbf{P}^{(i-1)})$, the precoding matrix is updated to $\mathbf{P}^{(i)} = \widetilde{\mathbf{P}}$ and the step size is reset to $\mu^{(i)} = \mu^{(0)}$. Otherwise, the step size is decreased according to $\mu^{(i)} = \frac{1}{2}\mu^{(i-1)}$. This procedure is repeated until either the sum rate goes below R_0 or the maximum number of iteration I, is reached. This procedure is summarized in Algorithm 2.

VII. Numerical Results

We verify the efficacy of the proposed framework by means of numerical simulations. In our simulations, we focus on millimeter-wave communication scenarios without loss of generality. We first describe the channel model we use for our simulations (except for the example in Section VII-F). Then, we compare the spectral and energy efficiencies achieved by DAB and EE-DAB with those achieved by MRT and ZF precoding. Finally, we evaluate the out-of-band radiation associated with the proposed precoding techniques.

A. Geometric Channel Model

In what follows, unless stated otherwise, we adopt a channel model that captures the sparse scattering characteristics of millimeter-wave channels in non-LoS (nLoS) environments, namely, when there is no dominant path. In this model, which is typically referred to as the geometric channel model, each scatterer contributes to a single path and the channel coefficients can be expressed as [37]

$$\mathbf{h}_u = \sqrt{\frac{B}{N_{\text{path}}}} \sum_{\ell=1}^{N_{\text{path}}} \zeta_{u,\ell} \mathbf{a}(\psi_{u,\ell}), \tag{39}$$

for $u = 1, \ldots, U$, where N_{path} represents the number of paths/scatterers. Here, $\psi_{u,\ell}$ is the AoD for the ℓth path, and $\mathbf{a}(\psi_{u,\ell})$ is the corresponding array response vector. For a ULA with $\lambda_c/2$ antenna spacing, the bth entry of $\mathbf{a}(\psi_{u,\ell})$ is given by

$$[\mathbf{a}(\psi_{u,\ell})]_b = \frac{1}{\sqrt{B}} e^{-j\pi(b-1)\cos(\psi_{u,\ell})} \tag{40}$$

for $b = 1, \ldots, B$. Furthermore, $\zeta_{u,\ell} \sim \mathcal{CN}(0, \gamma_u^2)$ is the independent and identically distributed (i.i.d.) channel gain (including path loss) corresponding to the ℓth path.

Throughout our simulations, we set $N_{\text{path}} = 4$ and assume that the AoD $\psi_{u,\ell}$ is uniformly distributed over the interval $[0^\circ, 180^\circ)$. We adopt the nLoS path-loss model presented in [38, Tbl. I] and, assuming that the system operates at a carrier frequency $f_c = 28\,\text{GHz}$ (such that $\lambda_c \approx 10.7\,\text{mm}$), the path-loss for the user u (at a distance of d_u meters) is calculated using

$$\gamma_u^2 = -72 - 29.2 \log_{10}(d_u) \qquad [\text{dB}]. \tag{41}$$

We further assume that the users are uniformly distributed in a disk-shaped area with the base station at its center. The minimum and maximum distances from the base station are set to

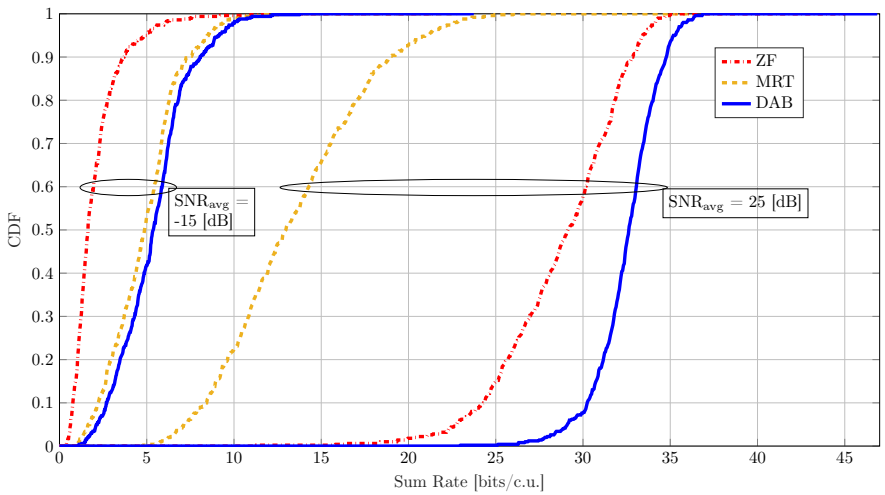


Fig. 5. CDF of sum rates with different linear precoders; geometric channel model, $B = 32$ antennas, $U = 4$ users, equal PAs with $\beta_1^{(b)} = 1$, and $\beta_3^{(b)} = -0.049 - 0.023j$ for $b = 1, \ldots, B$, $\rho_{\text{tot}} = 43\,\text{dBm}$, and $\text{SNR}_{\text{avg}} \in \{-15, 25\}\,\text{dB}$.

$d_{\text{min}} = 5$ and $d_{\text{max}} = 40$ meters, respectively. For this setting, the average path loss is about $\gamma_{\text{avg}}^2 = -110\,\text{dB}$, which corresponds to a user at the distance of 19.8 meters from the base station. Finally, we use the following definition for the average SNR:

$$\text{SNR}_{\text{avg}} = \gamma_{\text{avg}}^2 \frac{\rho_{\text{tot}}}{N_0}. \tag{42}$$

B. Spectral Efficiency: DAB versus Conventional Precoders

We compare the spectral efficiency of Algorithm 1, namely DAB, with that of conventional precoding techniques including ZF and MRT. To this end, we fix the total transmit power to $\rho_{\text{tot}} = 43\,\text{dBm}$ and characterize the statistics of the achievable sum rate in (15) for two different average SNR values, namely -15 and $25\,\text{dB}$. In Fig. 5, the cumulative distribution function (CDF) of the achievable sum rate is plotted for DAB, ZF, and MRT. In this example, $B = 32$ antennas and $U = 4$ users are considered. We assume that all the antennas at the transmitter are equipped with equal PAs with $\beta_1^{(b)} = 1$, and $\beta_3^{(b)} = -0.049 - 0.023j$ for $b = 1, \ldots, B$. We run Algorithm 1 with 20 initializations and set the number of iterations per initialization to $I = 50$. Moreover, the initial step size is set to $\mu^{(0)} = 0.1$. From Fig. 5, it can be seen that the DAB precoding technique outperforms the conventional linear precoders, which do not take into account the distortion introduced by the nonlinear PAs. The improvement in performance is more

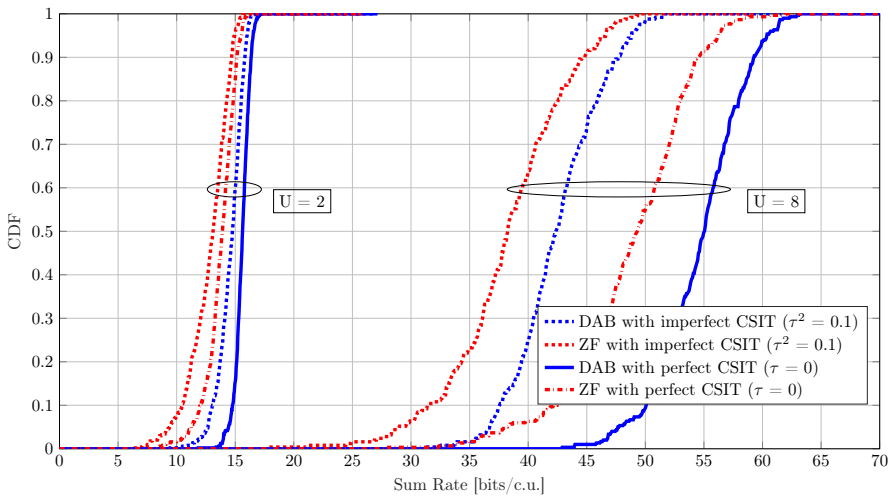


Fig. 6. CDF of the sum rate with ZF and DAB precoders with perfect and imperfect CSI at the transmitter; geometric channel model, $B = 32$ antennas, $U \in \{2, 8\}$ users, equal PAs with $\beta_1^{(b)} = 1$, and $\beta_3^{(b)} = -0.049 - 0.023j$ for $b = 1, \ldots, B$, $\rho_{\text{tot}} = 43\,\text{dBm}$, $N_0 = -85\,\text{dBm}$ (corresponding to $\text{SNR}_{\text{avg}} = 18\,\text{dB}$).

significant at high SNR values. This is because, at low SNR (i.e., $\text{SNR}_{\text{avg}} = -15\,\text{dB}$), the thermal noise dominates the nonlinear distortion. As a result, dealing with distortion is not necessary and the DAB precoder converges to the MRT precoder (as was also shown in Fig. 4). In the high-SNR regime, however, the performance of both MRT and ZF is limited by the nonlinear distortion. As a result, DAB provides considerable gains over these two conventional precoding techniques.

In Fig. 6, we consider a setup with $B = 32$ antennas and $U \in \{2, 8\}$ users and illustrate the performance gains achieved by DAB over ZF at $\text{SNR}_{\text{avg}} = 18\,\text{dB}$, with both perfect and imperfect CSI at the transmitter (CSIT). For this purpose, we model the channel estimation error as an additive independent random error term. Accordingly, the estimated CSIT is modeled by

$$\widehat{\mathbf{h}}_u = \sqrt{1 - \tau^2}\mathbf{h}_u + \tau\mathbf{v}, \tag{43}$$

where $\widehat{\mathbf{h}}_u$ and $\mathbf{h}_u$ denote the estimated and actual channel corresponding to the uth user, respectively, $\tau \in [0, 1]$ is a parameter that reflects the accuracy of channel estimation and the elements of the CSI error $\mathbf{v}$ are distributed according to $\mathcal{CN}\left(0, \sigma^2_{\mathbf{h}_u}\right)$. Fig. 6 shows the CDF of the sum rate for ZF and DAB with $\tau = 0$ (perfect CSIT) and $\tau = \sqrt{0.1}$ (imperfect CSIT). Our results show that once computed with imperfect CSIT, DAB and ZF experience roughly equal performance degradation in terms of achievable sum rate. For example, at CDF $= 0.9$,

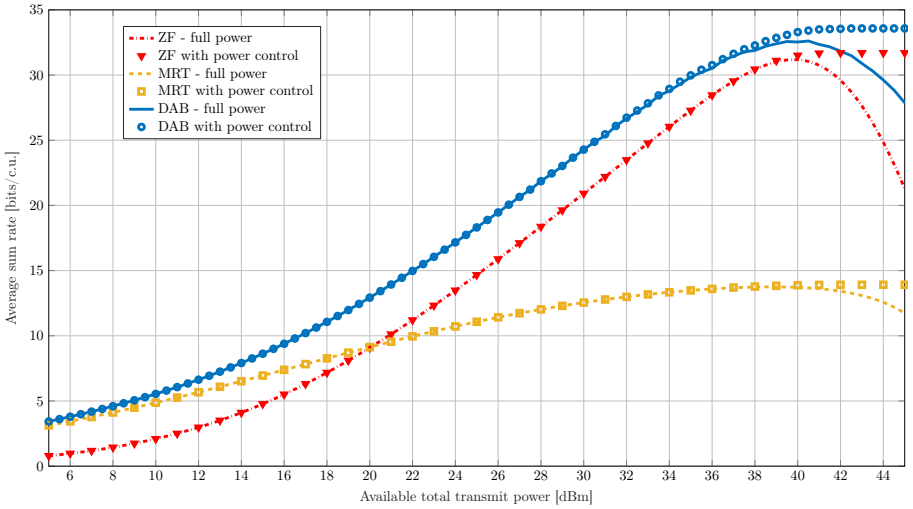


Fig. 7. Average sum rate with different linear precoders with and without power control; geometric channel model, $B = 32$ antennas, $U = 4$ users, equal PAs with $\beta_1^{(b)} = 1$, and $\beta_3^{(b)} = -0.049 - 0.023j$ for $b = 1, \ldots, B$, $N_0 = -85$ dBm.

this degradation is about 5% and 20% for both precoding schemes in the scenarios with $U = 2$ and $U = 8$ users, respectively.

C. DAB versus Conventional Precoders in Presence of Power Control

In Figs. 5 and 6, we focused on scenarios with fixed total transmit power and demonstrated the advantages of distortion-aware precoding in distortion-limited regimes. We now consider the case in which the transmitter adopts a power control strategy along with the linear precoding matrix calculation. In particular, we consider a maximum total transmit power $\rho_{\text{tot,max}}$ and similar to [39], find precoding matrices that maximize the sum rate under the following power constraint

$$\mathbb{E}\left[\|f(\mathbf{Ps})\|^2\right] \leq \rho_{\text{tot,max}}. \tag{44}$$

Fig. 7 shows the average sum rate values achieved by MRT, ZF, and DAB precoding techniques under different available total transmit power values for a setup with $B = 32$ antennas and $U = 4$ users. The channel model and the PA nonlinearity parameters are the same as in Fig. 6. The full power transmission corresponds to $\rho_{\text{tot}} = \rho_{\text{tot,max}}$ and in the precoding with power control, we calculate the precoding matrices for different values of $\rho_{\text{tot}} \in [0, \rho_{\text{tot,max}}]$ and pick the one that gives the highest sum rate value. The results in Fig. 7 show how DAB performs compared to the conventional linear precoding techniques in different power regimes. It can be seen that, in the low-power regime, DAB converges to the MRT precoding due to operating

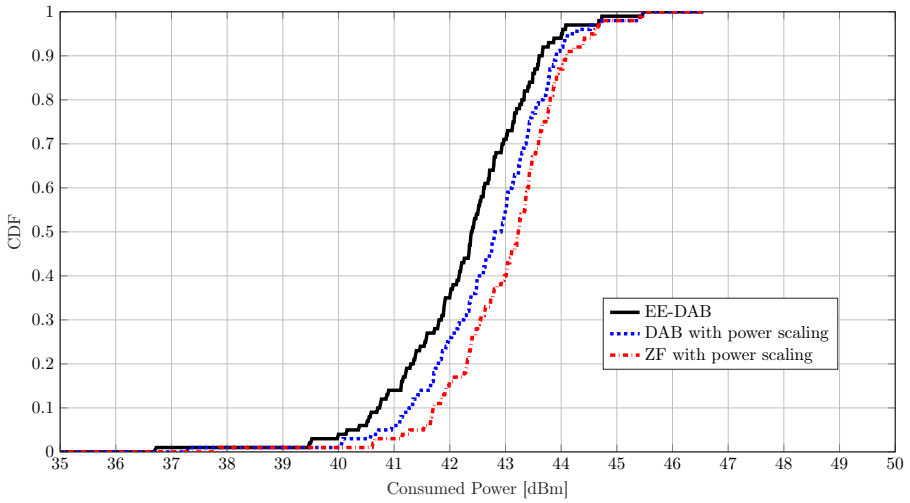
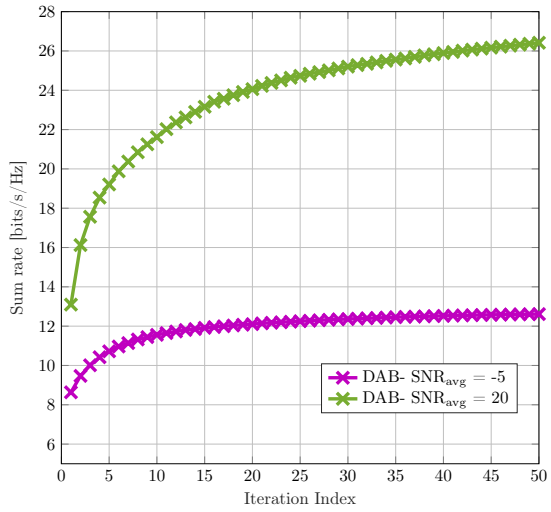


Fig. 8. CDF of total base station side consumed power with ZF, DAB and EE-DAB precoders for $R_0 = 35$ (bits/c.u.); geometric channel model, $B = 64$ antennas, $U = 4$ users, equal PAs with $\beta_1^{(b)} = 1$, and $\beta_3^{(b)} = -0.049 - 0.023j$ for $b = 1, \ldots, B$, $N_0 = -85\,\text{dBm}$, power consumption model parameters $\eta_{\text{max},b} = 0.55$ and $\rho_{\text{max},b} = 25\,\text{dBm}$.

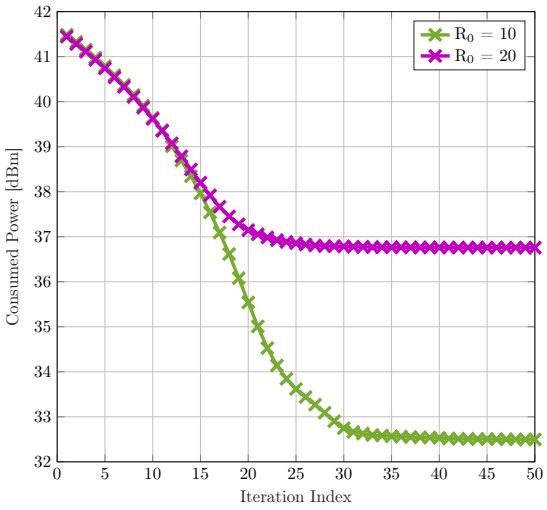
in the AWGN-dominated region (similar to the case with $\text{SNR}_{\text{avg}} = -15\,\text{dB}$ in Fig. 5). Once the transmit power is increased, DAB yields improved performance compared to ZF and MRT. This improved performance follows because the array gain, multiuser interference, and nonlinear distortion are jointly considered in the precoding matrix optimization procedure. In this example, with full power transmission, MRT, ZF, and DAB reach their maximum average sum rate roughly at $\rho_{\text{tot}} = 38.2\,\text{dBm}$, $\rho_{\text{tot}} = 39.7\,\text{dBm}$ and $\rho_{\text{tot}} = 40.5\,\text{dBm}$, respectively. With power control, the average sum rate remains constant even when $\rho_{\text{tot,max}}$ is increased further. Accordingly, DAB provides a constant improvement over ZF and MRT for $\rho_{\text{tot}} \geq 40.5\,\text{dBm}$.

D. Energy Efficiency: EE-DAB versus ZF

To evaluate the energy efficiency of EE-DAB precoding, we plot the CDF of the consumed power $\sum_{b=1}^{B} \rho_{\text{cons},b}$ for different minimum required sum rate values R_0. We consider a setup with $B = 64$ and $U = 4$ and assume that the channel model and the PA nonlinearity parameters are the same as in Section VII-B. The parameters associated with PA power consumption are set to $\eta_{\text{max},b} = 0.55$ and $\rho_{\text{max},b} = 25\,\text{dBm}$ for all $b = 1, \ldots, B$. The noise power is $N_0 = -85\,\text{dBm}$ and the consumed-power values are calculated for the precoding matrices acquired via Algorithm 2 for $R_0 = 35$ bits per channel use (bits/c.u.). As benchmarks, we also evaluate the power consumption with the ZF and DAB precoders with a total transmit power normalization such that R_0 is achieved. The resulting CDF is given in Fig. 8.



(a) DAB, $\rho_{\text{tot}} = 43\,\text{dBm}$.



(b) EE-DAB, $\eta_{\text{max},b} = 0.55$ and $\rho_{\text{max},b} = 25\,\text{dBm}$ for $b = 1, \ldots, B$.

Fig. 9. Average convergence behavior of Algorithms 1 and 2; geometric channel model, $B = 32$ antennas, $U = 4$ users, $\beta_1^{(b)} = 1$, $\beta_3^{(b)} = -0.049 - 0.023j$ for $b = 1, \ldots, B$ and $N_0 = -85\,\text{dBm}$. Both algorithms converge in a relatively low number of iterations.

It can be seen from Fig. 8 that, by taking the power consumption into account (in addition to the effect of nonlinear distortion), EE-DAB can achieve a given sum rate with a lower consumed power than ZF and DAB precoding schemes. This improvement is achieved via optimizing the transmit power in addition to the directivity of the linear signal and distortion using Algorithm 2.

E. Convergence Speed of the Proposed Algorithms

The improved performance of the DAB and EE-DAB precoders compared to conventional linear precoders is achieved at the cost of increased computational complexity. Here we *empirically* study the convergence of Algorithms 1 and 2, which are used to obtain the DAB and EE-DAB precoders, respectively. This provides insight into the number of iterations required to achieve a given performance.

We consider a setup with $B = 32$ and $U = 4$. In Fig. 9a, we show the sum rate achieved by the DAB precoder as a function of the number of iterations. It can be seen that the convergence of Algorithm 1 is faster at low SNR compared to high SNR. Indeed, at low SNR, conventional precoding schemes are as good as DAB precoding (see Fig. 5). Therefore, since the MRT and ZF precoding matrices are included in the initialization, fast convergence is expected. At high SNR, however, the structure of the optimal precoder is significantly different from that of the MRT and ZF precoders, which results in slower convergence. Fig. 9b depicts the convergence behavior of

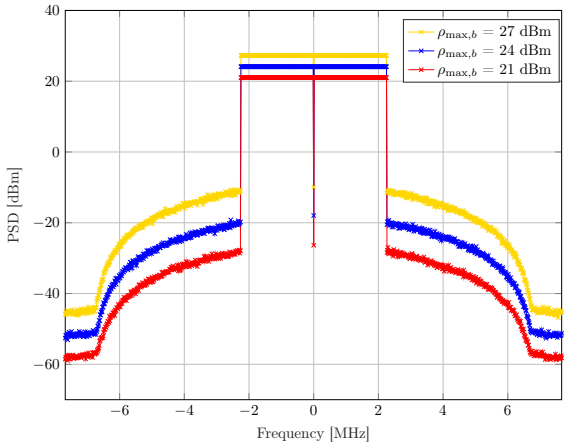


Fig. 10. PSD of transmitted signal from the antenna with the highest spectral regrowth with DAB under per-antenna power constraint in (34); LoS channel, $B = 16$ antennas, $U = 1$ users, equal PAs as in (45). By controlling $\rho_{\max,b}$, we can bound the spectral regrowth within a given mask limit.

Algorithm 2. This demonstrates that with the two proposed algorithms, it is possible to achieve a certain performance in a relatively small number of iterations, keeping the computational complexity within a manageable range.

F. Out-of-Band Emissions

In addition to analyses of the in-band distortion caused by the nonlinear PAs, which has been the main focus of this paper, analyses of out-of-band emissions are also of significant practical importance. With out-of-band emission, we refer to emissions at frequencies outside the allocated frequency band due to the PA nonlinearities. These emissions, which are undesirable because they interfere with concurrent transmission in adjacent bands, are strictly regulated in wireless standards.

We evaluate the spectral emissions with DAB precoding under a per-antenna power constraint (34) for different $\rho_{\max,b}$ values. In Fig. 10, we plot the power spectral density (PSD) of the transmitted signal from the antenna with the highest spectral regrowth (worst case) in the array. We assume a LoS channel with $B = 16$ and $U = 1$ and consider the following nonlinearity model for all PAs in the array

$$f_b(x_b) = \beta_1^{(b)} x_b + \beta_2^{(b)} x_b|x_b| + \beta_3^{(b)} x_b|x_b|^2, \tag{45}$$

with $\beta_1^{(b)} = 1$, $\beta_2^{(b)} = -0.02 - 0.01j$ and $\beta_3^{(b)} = -0.049 - 0.023j$ for $b = 1, \ldots, B$. The precoding matrix is calculated using Algorithm 1 (while replacing Step 3 with (36) to satisfy per-antenna

transmit power constraints) and applied to an OFDM signal with a one-side bandwidth of 2.25 MHz. The total number of subcarriers is 1024, among which 300 subcarriers are occupied (the occupied carriers are the first 150 to the left and to the right of the DC carrier). The results in Fig. 10 indicate that by controlling $\rho_{\text{max},b}$, one can limit the spectral regrowth outside a given frequency band. It is indeed possible to generalize the proposed precoder optimization framework to account explicitly for out-of-band emission on top of in-band distortion. A detailed study of this more general framework is left for future work.

VIII. Conclusions

We have shown that the choice of precoding matrix has a direct impact on the amount and directivity of distortion emitted by nonlinear power amplifiers in the downlink of massive MIMO systems. Accordingly, we have introduced precoding optimization algorithms that incorporate knowledge of the nonlinearity characteristics of the PAs as well as their power consumption. Using numerical examples, we have shown that our proposed solutions can achieve significant improvements in spectral and energy efficiency compared to conventional precoding techniques, i.e., zero-forcing and maximal ratio transmission. Moreover, by studying out-of-band emission, we have shown that our proposed precoder optimization framework can also be used to limit the spectral regrowth due to nonlinear power amplifiers outside a given frequency band.

Appendix A
Closed-Form Expression for $\nabla_{\mathbf{P}} R_{\text{SUM}}(\mathbf{P})$

We derive a closed-form expression for the gradient $\nabla_{\mathbf{P}} R_{\text{sum}}(\mathbf{P})$, which facilitates a more accurate evaluation of the update in (30) compared to its approximate numerical evaluation in (31). To this end, let $n_u(\mathbf{P}) = |\mathbf{h}_u^T \mathbf{G}(\mathbf{P})\mathbf{p}_u|^2$ and $d_u(\mathbf{P}) = d_u^{\text{mui}}(\mathbf{P}) + d_u^{\text{dist}}(\mathbf{P}) + N_0$ denote the numerator and denominator of the SINDR in (16), respectively, where $d_u^{\text{mui}}(\mathbf{P}) = \sum_{r \neq u} |\mathbf{h}_u^T \mathbf{G}(\mathbf{P})\mathbf{p}_r|^2$ and $d_u^{\text{dist}}(\mathbf{P}) = \mathbf{h}_u^T \mathbf{C}_{\mathbf{e}}(\mathbf{P})\mathbf{h}_u^*$ is the part of the denominator corresponding to multiuser interference and nonlinear distortion, respectively. With these definitions, the gradient $\nabla_{\mathbf{P}} R_{\text{sum}}(\mathbf{P})$ can be written as

$$\nabla_{\mathbf{P}} R_{\text{sum}}(\mathbf{P}) = \sum_{u=1}^{U} \frac{2\log_2(e)}{d_u^2(\mathbf{P})(1 + n_u(\mathbf{P})/d_u(\mathbf{P}))} \left(d_u(\mathbf{P}) \frac{\partial n_u(\mathbf{P})}{\partial \mathbf{P}^*} - n_u(\mathbf{P}) \frac{\partial d_u(\mathbf{P})}{\partial \mathbf{P}^*} \right), \tag{46}$$

where $\partial n_u(\mathbf{P})/\partial \mathbf{P}^* = [\partial n_u(\mathbf{P})/\partial \mathbf{p}_1^*, \ldots, \partial n_u(\mathbf{P})/\partial \mathbf{p}_u^*]$ and $\partial d_u(\mathbf{P})/\partial \mathbf{P}^* = [\partial d_u(\mathbf{P})/\partial \mathbf{p}_1^*, \ldots, \partial d_u(\mathbf{P})/\partial \mathbf{p}_U^*]$. Hence, to compute the gradient $\nabla_{\mathbf{P}} R_{\text{sum}}(\mathbf{P})$, we need to compute the derivatives

of $n_u(\mathbf{P})$ and $d_u(\mathbf{P})$ for $u = 1, \ldots, U$. We consider the third-order nonlinearity, i.e., $K = 1$ in (2) and accordingly $\mathbf{G}(\mathbf{P})$ and $\mathbf{C}_{\mathbf{e}}(\mathbf{P})$ are as in (23) and (24), respectively. Starting with the numerator, it can be shown that the derivative with respect to $\mathbf{p}_{u'}^*$ can be written as

$$\frac{\partial n_u(\mathbf{P})}{\partial \mathbf{p}_{u'}^*} = \left(\mathbf{\Gamma}_u(\mathbf{P})\mathbb{1}_{\{u'=u\}}(u') + \mathbf{\Upsilon}_{u,u'}(\mathbf{P})\right)\mathbf{p}_{u'}, \tag{47}$$

for $u' = 1, \ldots, U$, where $\mathbf{\Gamma}_u(\mathbf{P}) \in \mathbb{C}^{B\times B}$ is given by

$$\begin{aligned}\mathbf{\Gamma}_u(\mathbf{P}) =& \mathbf{A}_1^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_1 + 2\left(\mathbf{A}_1^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_3\text{diag}(\mathbf{P}\mathbf{P}^H) + \text{diag}(\mathbf{p}_u\mathbf{p}_u^H\mathbf{A}_1^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_3)\right)\\ &+ 2\left(\text{diag}(\mathbf{P}\mathbf{P}^H)\mathbf{A}_3^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_1 + \text{diag}(\mathbf{A}_3^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_1\mathbf{p}_u\mathbf{p}_u^H)\right)\\ &+ 4\big(\text{diag}(\mathbf{P}\mathbf{P}^H)\mathbf{A}_3^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_3^*\text{diag}(\mathbf{P}\mathbf{P}^H) + \text{diag}(\mathbf{A}_3^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_3\text{diag}(\mathbf{P}\mathbf{P}^H)\mathbf{p}_u\mathbf{p}_u^H\\ &+ \mathbf{p}_u\mathbf{p}_u^H\text{diag}(\mathbf{P}\mathbf{P}^H)\mathbf{A}_3^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_3)\big).\end{aligned} \tag{48}$$

Furthermore, $\mathbf{\Upsilon}_u(\mathbf{P}) \in \mathbb{C}^{B\times B}$ can be expressed as

$$\begin{aligned}\mathbf{\Upsilon}_{u,u'}(\mathbf{P}) =& 2(\text{diag}(\mathbf{p}_{u'}\mathbf{p}_{u'}^H\mathbf{A}_1^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_3) + \text{diag}(\mathbf{A}_1^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_3\mathbf{p}_{u'}\mathbf{p}_{u'}^H))\\ &+4(\text{diag}(\mathbf{A}_3^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_3\text{diag}(\mathbf{P}\mathbf{P}^H)\mathbf{p}_{u'}\mathbf{p}_{u'}^H) + \text{diag}(\mathbf{p}_{u'}\mathbf{p}_{u'}^H\text{diag}(\mathbf{P}\mathbf{P}^H)\mathbf{A}_3^*\mathbf{h}_u^*\mathbf{h}_u^T\mathbf{A}_3)).\end{aligned} \tag{49}$$

The derivative with respect to $\mathbf{p}_{u'}^*$ of the denominator can be written as

$$\frac{\partial d_u(\mathbf{P})}{\partial \mathbf{p}_{u'}^*} = \frac{\partial d_u^{\text{mui}}(\mathbf{P})}{\partial \mathbf{p}_{u'}^*} + \frac{\partial d_u^{\text{dist}}(\mathbf{P})}{\partial \mathbf{p}_{u'}^*}, \tag{50}$$

where the derivative of the multiuser interference term in the denominator is given by

$$\frac{\partial d_u^{\text{mui}}(\mathbf{P})}{\partial \mathbf{p}_{u'}^*} = \left(\mathbf{\Gamma}_u(\mathbf{P})\mathbb{1}_{\{u'\neq u\}}(u') + \sum_{r\neq u}\mathbf{\Upsilon}_{u,r}(\mathbf{P})\right)\mathbf{p}_{u'} \tag{51}$$

for $u' = 1, \ldots, U$. Furthermore, the bth entry of the derivative of the nonlinear distortion term in the denominator is given by

$$\frac{\partial d_u^{\text{dist}}(\mathbf{P})}{\partial p_{u',b}^*} = 2\left|\beta_3^{(b)}\right|^2\left(2h_{u,b}^*\sum_{b'=1}^{B}h_{u,b'}p_{u',b'}\left[\left|\mathbf{P}\mathbf{P}^H\right|^2\right]_{b',b} + h_{u,b}\sum_{b'=1}^{B}h_{u,b'}^*p_{u',b'}\left[\left(\mathbf{P}\mathbf{P}^H\right)^2\right]_{b,b'}\right) \tag{52}$$

for $u' = 1, \ldots, U$ and $b = 1, \ldots, B$, where $h_{b,u} = [\mathbf{h}_u]_b$, $p_{b,u} = [\mathbf{p}_u]_b$ and $\beta_3^{(b)}$ is the coefficient corresponding to the bth PA in the array. Finally, by inserting (47), (50), (51), and (52) into (46), we obtain a closed-form expression for the gradient $\nabla_{\mathbf{P}}R_{\text{sum}}(\mathbf{P})$.

REFERENCES

[1] S. Rezaei Aghdam, S. Jacobsson, and T. Eriksson, "Distortion-aware linear precoding for millimeter-wave multiuser MISO downlink," in *IEEE Int. Conf. Comm. (ICC) Workshops*, Shanghai, China, May 2019.

[2] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, no. 3–4, pp. 154–655, Nov. 2017.

[3] U. Gustavsson, C. Sanchéz-Perez, T. Eriksson, F. Athley, G. Durisi, P. Landin, K. Hausmair, C. Fager, and L. Svensson, "On the impact of hardware impairments on massive MIMO," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Austin, TX, USA, Dec. 2014, pp. 294–300.

[4] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 11, no. 60, pp. 7112–7139, Nov. 2014.

[5] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, "Rate-splitting to mitigate residual transceiver hardware impairments in massive MIMO systems," *IEEE Trans. on Veh. Technol.*, vol. 66, no. 9, pp. 8196–8211, Sep. 2017.

[6] M. Abdelghany, A. A. Farid, U. Madhow, and M. J. W. Rodwell, "A design framework for all-digital mmwave massive MIMO with per-antenna nonlinearities," Dec. 2019. [Online]. Available: https://arxiv.org/abs/1912.11643

[7] J. Qi and S. Aissa, "On the power amplifier nonlinearity in MIMO transmit beamforming systems," *IEEE Trans. Commun.*, vol. 60, no. 3, pp. 876–887, Mar. 2012.

[8] S. Blandino, C. Desset, A. Bourdoux, L. Van der Perre, and S. Pollin, "Analysis of out-of-band interference from saturated power amplifiers in massive MIMO," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Oulu, Finland, Jul. 2017.

[9] C. Mollén, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Spatial characteristics of distortion radiated from antenna arrays with transceiver nonlinearities," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6663–6679, Oct. 2018.

[10] N. N. Moghadam, G. Fodor, M. Bengtsson, and D. J. Love, "On the energy efficiency of MIMO hybrid beamforming for millimeter-wave systems with nonlinear power amplifiers," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7208–7221, Nov. 2018.

[11] M. R. Khanzadi, G. Durisi, and T. Eriksson, "Capacity of SIMO and MISO phase-noise channels with common/separate oscillators," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3218–3231, Sep. 2015.

[12] N. Kolomvakis, M. Matthaiou, and M. Coldrey, "IQ imbalance in multiuser systems: Channel estimation and compensation," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 3039–3051, Jul. 2016.

[13] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized precoding for massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, Nov. 2017.

[14] X. Zhang, M. Matthaiou, M. Coldrey, and E. Björnson, "Impact of residual transmit RF impairments on training-based MIMO systems," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2899–2911, Aug. 2015.

[15] A. Papazafeiropoulos, S. K. Sharma, T. Ratnarajah, and S. Chatzinotas, "Impact of residual additive transceiver hardware impairments on rayleigh-product MIMO channels with linear receivers: Exact and asymptotic analyses," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 105–118, Jan. 2018.

[16] C. Studer, M. Wenk, and A. Burg, "MIMO transmission with residual transmit-RF impairments," in *Proc. Int. ITG Workshop on Smart Antennas (WSA)*, Bremen, Germany, Feb. 2010, pp. 189–196.

[17] E. G. Larsson and L. Van der Perre, "Out-of-band radiation from antenna arrays clarified," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 610–613, Feb. 2018.

[18] L. Anttila, A. Brihuega, and M. Valkama, "On antenna array out-of-band emissions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1653–1656, Dec. 2019.

[19] S. Jacobsson, U. Gustavsson, G. Durisi, and C. Studer, "Massive MU-MIMO-OFDM uplink with hardware impairments: Modeling and analysis," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Oct. 2018, pp. 1829–1835.

[20] E. Björnson, L. Sanguinetti, and J. Hoydis, "Hardware distortion correlation has negligible impact on UL massive MIMO spectral efficiency," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1085–1098, Feb. 2019.

[21] S. R. Aghdam and T. Eriksson, "On the performance of distortion-aware linear receivers in uplink massive MIMO systems," in *Proc. IEEE Int. Symp. Wirel. Comm. Syst. (ISWCS)*, Oulu, Finland, Aug. 2019, pp. 208–212.

[22] O. T. Demir and E. Björnson, "Channel estimation in massive MIMO under hardware non-linearities: Bayesian methods versus deep learning," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 109–124, Jan. 2020.

[23] A. Sheikhi, F. Rusek, and O. Edfors, "Massive MIMO with per-antenna digital predistortion size optimization: Does it help?" in *Proc. IEEE Int. Conf. Commun. (ICC)*, Montreal, QC, Canada, Jun. 2021, pp. 1–6.

[24] A. Brihuega, L. Anttila, M. Abdelaziz, and M. Valkama, "Digital predistortion in large-array digital beamforming transmitters," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Feb. 2018, pp. 611–618.

[25] M. Abdelaziz, L. Anttila, and M. Valkama, "Reduced-complexity digital predistortion for massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, New Orleans, LA, USA, Jun. 2017, pp. 6478–6482.

[26] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Res. Lab. Elec., Cambridge, MA, USA, Tech. Rep. 216, Mar. 1952.

[27] T. Schenk, *RF imperfections in high-rate wireless systems: Impact and digital compensation*. Dordrecht, The Netherlands: Springer-Verlag, 2008.

[28] D. Persson, T. Eriksson, and E. G. Larsson, "Amplifier-aware multiple-input multiple-output power allocation," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1112–1115, May 2013.

[29] A. Grebennikov, *RF and microwave power amplifier design*. McGraw-Hill New York, 2005.

[30] I. S. Reed, "On a moment theorem for complex Gaussian processes," *IEEE Trans. Inf. Theory*, vol. 8, no. 3, pp. 194–195, Apr. 1962.

[31] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.

[32] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[33] A. Lapidoth, "Mismatched decoding and the multiple access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1439–1452, Sep. 1996.

[34] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Aug. 2005.

[35] D. P. Bertsekas, "Nonlinear programming," *J. Oper. Res. Soc.*, vol. 48, no. 3, pp. 334–334, 1997.

[36] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Commun. Surv. Tut.*, vol. 15, no. 1, pp. 167–178, Feb. 2013.

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

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

[39] J. Jee, G. Kwon, and H. Park, "Precoding design and power control for SINR maximization of MISO system with nonlinear power amplifiers," *IEEE Trans. on Veh. Technol.*, vol. 69, no. 11, pp. 14 019–14 024, Sep. 2020.