Energy Efficiency of mmWave Massive MIMO Precoding with Low-Resolution DACs

Lucas N. Ribeiro, *Student Member, IEEE*, Stefan Schwarz, *Member, IEEE*, Markus Rupp, *Fellow, IEEE*,
André L. F. de Almeida, *Senior Member, IEEE*

Abstract—With the congestion of the sub-6 GHz spectrum, the interest in massive multiple-input multiple-output (MIMO) systems operating on millimeter wave spectrum grows. In order to reduce the power consumption of such massive MIMO systems, hybrid analog/digital transceivers and application of low-resolution digital-to-analog/analog-to-digital converters have been recently proposed. In this work, we investigate the energy efficiency of quantized hybrid transmitters equipped with a fully/partially-connected phase-shifting network composed of active/passive phase-shifters and compare it to that of quantized digital precoders. We introduce a quantized single-user MIMO system model based on an additive quantization noise approximation considering realistic power consumption and loss models to evaluate the spectral and energy efficiencies of the transmit precoding methods. Simulation results show that partially-connected hybrid precoders can be more energy-efficient compared to digital precoders, while fully-connected hybrid precoders exhibit poor energy efficiency in general. Also, the topology of phase-shifting components offers an energy-spectral efficiency trade-off: active phase-shifters provide higher data rates, while passive phase-shifters maintain better energy efficiency.

Index Terms—Hybrid Precoding, Millimeter Wave, Massive MIMO, Energy Efficiency, Low-Resolution DAC

I. INTRODUCTION

THE rapid increase of connected mobile terminals in the past few years has been pushing data rate requirements of 5G systems to new levels [1]. As the sub-6 GHz spectrum is congested, moving towards other ranges, such as millimeter wave (mmWave) regime, has been one of the main ideas for achieving these requirements [2]. The mmWave band is little regulated and is mostly available, allowing for mobile communication systems to operate on large bandwidths. However, the propagation characteristics on this elevated frequency range poses many engineering challenges. For instance, according to

Friis' Law, the isotropic path loss in free-space propagation is inversely proportional to the wavelength squared [3]. It implies that path loss at the mmWave range is more severe than in lower frequency ranges in general. Nevertheless, for a given physical aperture, the antenna directional gain is also inversely proportional to the wavelength squared, and, thus, employing an array of highly directional antennas more than compensates for the free-space path loss [4]. Such technology has been referred to in literature as mmWave massive multiple-input multiple-output (MIMO) [5], [6], [7].

Hybrid analog/digital (A/D) transceiver architectures have been proposed to enable mmWave massive MIMO systems [7]. They employ digital filtering (precoding/decoding) at baseband, and perform beamforming in the radio-frequency (RF) domain by analog components. The most popular implementation of this RF beamformer consists of an active phase-shifting network (PSN) connecting the outputs of the baseband filter to the antennas, which is known as fully-connected PSN. This implementation, however, is associated with a large power consumption as a considerable number of active phase-shift elements is required. As an alternative, one can employ sub-array beamforming, reducing the number of phase-shifters, and, consequently, power consumption. In this case, the PSN is said to be partially-connected. It has been claimed that hybrid precoding provides a data throughput close to that of fully-digital systems [8]. However, insertion losses of RF hardware are usually disregarded in the analysis of such hybrid systems. If these losses are not properly compensated for, then their spectral efficiency might be much smaller in practice than what is expected.

Another energy-efficient approach to mmWave massive MIMO consists of using low-resolution digital-to-analog/analog-to-digital converters (DACs/ADCs) [4]. At the receive side, the high-resolution ADC chains are the most power hungry part, motivating the application of low-resolution devices to reduce their power consumption. At the transmit side, however, power expenditure is dominated by power amplifiers (PAs), which are usually required to operate within the high linearity regime to avoid distortion of the signal constellation. Employing low-resolution DACs relaxes the linearity requirement, allowing the amplifiers to operate closer to saturation, thus increasing their efficiency.

A. Related Work

Most of recent works on massive MIMO focus on analyzing either the performance of energy-efficient full-resolution hybrid or low-resolution fully-digital transceivers. In the following, we discuss their contributions.

1) Energy Efficiency of Hybrid Systems: The work of [9] investigates the spectral and energy efficiencies of hybrid systems with switches and phase-shifters to perform analog beamforming. In order to evaluate energy efficiency, they define a power consumption model for both types of RF beamforming considering different interconnection of components. They also propose a channel estimation technique based on compressive sensing. According to simulation results, all hybrid architectures yield similar spectral efficiency for a given power consumption. In [10], the energetic performance of single/multi-carrier full-resolution hybrid transceivers is investigated. A transceiver optimization problem based on energy efficiency maximization is presented and solved by the alternating direction of multipliers method. The power consumption model proposed therein considers the computational power expenditure and RF hardware losses, which are usually ignored in most energy efficiency models. This model assumes the application of PAs and low-noise amplifiers that compensate for the analog beamforming losses. Based on this assumption, the paper claims that hybrid precoders with fully-connected PSN can be energy-efficient when the transmit power is high, even more than those with partially-connected PSN. This conclusion, however, contradicts the results of [11], where the spectral efficiency of a hybrid precoder is examined under a realistic RF model. Therein, the fully-connected PSN is modeled as a bank of RF components described by their S-parameters. The obtained results show that signal-to-noise ratio (SNR) losses are significant, going up to $25\,\mathrm{dB}$ for the given scenario. Unfortunately, with the current mmWave amplifier technology, one cannot assume that these losses are simply compensated for as in [10]. Therefore, in order to make a more realistic comparison between hybrid and digital transmitters, the effect of RF losses on the spectral efficiency has to be considered.

Analog and digital receivers employing low-resolution ADCs are studied in [12] for single-user MIMO (SU-MIMO). The authors resort to a stochastic linear quantization model referred to as additive quantization noise (AQN) model in order to simplify the analysis of systems with low-resolution ADCs. For digital combining, singular value decomposition (SVD) processing with water-filling power allocation is applied, whereas much simpler matched filtering is employed in analog beamforming. Lower bounds on data rate are derived based on the linear quantization approximation, and the AQN model is shown to be accurate in the low SNR regime. Simulation results indicate that digital combining exhibits better performance than the analog scheme. The energy and spectral efficiency trade-off for digital, analog, and hybrid receivers is extensively studied in [13]. The ADC is approximated by the AQN model and achievable rate expressions are obtained. A power consumption model similar to that of [9] is considered, allowing to calculate energy efficiency. At the transmit side,

the authors assume fully-digital precoding, whereas, at the receive side, fully-connected RF beamforming is employed in addition to digital baseband combining. The analog beamformers are computed by the alternating minimization method of [14], and the baseband filter is obtained through SVD processing with water-filling power allocation. Results indicate that analog combining is the most energy-efficient solution only at low SNR or low-rank channels, while the efficiency of hybrid and digital combining strongly depends on the assumed hardware power consumption characteristics for any other than low-rank channels. It is important to stress that [13] does not consider insertion losses at the RF domain and only investigates systems with fully-connected PSN. In [15], by contrast, hybrid receivers with partially-connected PSN are compared to digital combiners under a low-resolution ADC assumption. This contribution shows that digital receivers using low-resolution ADCs are robust to small automatic gain control (AGC) imperfections. In the considered power consumption model, a simple implementation for a 1-bit ADC is presented, leading to negligible power consumption. The considered analog combining strategy relies on beam scanning, and baseband combining consists of SVD processing with water-filling power allocation, as in the previously mentioned works. Results suggest that digital combining is more efficient than hybrid combining especially in the low SNR regime. However, the contributions mentioned above impose low-resolution only at the receiver side, assuming fully-digital or hybrid transmitters with high-resolution DACs.

2) Transmitters with Low-Resolution DACs: Recently proposed massive MIMO signal processing methods such as [16], [17] assume very large antenna arrays with dedicated RF chains either at the transmit or at the receive side. However, such implementation is not energy-efficient and thus needs to be modified in order to reduce its power requirements. An alternative to hybrid systems consists of employing low-resolution DACs/ADCs at transmitter/receiver to relax the power demands on the fully-digital transceiver.

A SU-MIMO model with low-resolution quantization at the transmit side is introduced in [18]. A linear approximation for DAC quantization based on the Bussgang Theorem [19] is presented, allowing the derivation of a minimum mean square error (MMSE) precoder optimized for tackling the quantization effects. Bit error ratio results reveal that this optimized MMSE filter outperforms the plain MMSE solution. In [20], a narrow-band multi-user MIMO (MU-MIMO) system employing low-resolution DACs at the base station is considered. The authors investigate the performance of linear precoders with coarse quantization and propose a variety of non-linear precoders based on relaxations of the MMSE-optimal downlink precoding problem. Achievable rate expressions are obtained and simulation results suggest that performance achieved with infinite-resolution DAC can be attained by using 3 or 4 bits of resolution for the given scenario. Furthermore, it was shown that the presented non-linear precoding algorithms significantly outperform the linear filtering solutions for 1-bit quantization.

B. Contribution

To the best of our knowledge, extensive performance evaluation of quantized hybrid transmitters with fully- and partially-connected PSN under realistic RF modeling has not been considered yet. We thus aim at filling this gap by this paper. Our main contributions in this direction can be summarized as follows:

- A SU-MIMO system model for quantized hybrid precoding based on the AQN model is presented. The proposed system model differs from those previously introduced in other works in the definition of the total additive noise vector, which accounts for the additive white Gaussian noise (AWGN) and an RF-filtered quantization noise.
- We define the quantized hybrid precoding problem, which can be regarded as a generalization of the quantized digital precoding and classical hybrid precoding problems. Achievable rate expressions for both schemes are derived. For hybrid precoding, we consider both fully- and partially-connected PSNs. The effects of DAC quantization on spectral efficiency are assessed through asymptotic analysis.
- A power consumption model considering device power characteristics reported in recent mmWave literature is presented. Active and passive phase-shifting topologies are considered in our model. The former has important power consumption and amplifies the shifted signals, whereas the latter has insignificant power consumption and substantial insertion losses. We properly account for power insertion losses in our signal model using a realistic RF modeling. The effects of signal processing computations are also considered in the power budget by the model presented in [21]. Simulations are conducted to evaluate the importance of this computational term and the obtained results suggest that it may be crucial for large-array systems.
- We introduce an analog beamforming method for partially-connected structures based on maximum eigenmode transmission (MET), and implement this solution with the low-complexity power method of [22, Section 7.3.1]. The computational complexity of the precoding methods are determined in order to calculate their power consumption.
- The energy-spectral efficiency trade-off is investigated to determine the most efficient precoding strategy, and to study the influence of phase-shifter implementation on performance. Our results show that hybrid precoding with partially-connected PSN and digital precoding are the most energy- and spectral-efficient solutions, respectively. Fully-connected PSN, however, exhibits poor energetic performance due to substantial power consumption and large insertion losses, which cause severe spectral efficiency degradation. Moreover, we observed that active phase-shifters favor spectral efficiency, while their passive implementations focus on energy efficiency.

This work is organized as follows: Section II describes the system model considered in this work. Signal and channel models for a narrow-band mmWave SU-MIMO system are introduced therein. The quantization operation employed in DACs is defined and approximated by the AQN model, and the power consumption and loss models are presented. The analog and digital precoding strategies adopted in this work are defined in Section III. Also, computational complexity analysis is conducted in Section III-C and achievable rate bounds are derived in Section III-D. The proposed models and precoding methods are assessed through simulations in Section IV, and the work is concluded in Section V.

C. Notation

In this paper, x denotes a scalar, $\mathbf{x}$ a vector, and $\mathbf{X}$ a matrix. The (i, j)-th entry of $\mathbf{X}$ is given by $[\mathbf{X}]_{i,j}$. The transposed and conjugated transposed (Hermitian) of $\mathbf{X}$ are denoted by $\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{H}$, respectively. The $(M \times M)$-dimensional identity matrix is represented by $\mathbf{I}_M$. The $(M \times N)$-dimensional null matrix is given by $\mathbf{0}_{M \times N}$. $\mathrm{tr}(\cdot)$ is the matrix trace, $\mathrm{diag}(\cdot)$ forms a diagonal matrix out of the main diagonal of the matrix argument, $\mathrm{Diag}(\cdot)$ transforms the input vector into a diagonal matrix, $\mathrm{Diagblk}(\cdot)$ forms a block-diagonal matrix from the matrix inputs, and $\det(\cdot)$ is the determinant. The operator $\angle(\cdot)$ extracts the angle from the argument's complex elements. The absolute value, the Frobenius and ℓ_2 norms, and the expected value operator are respectively represented by $|\cdot|$, $\|\cdot\|_\mathrm{F}$, $\|\cdot\|_2$, and $\mathbb{E}[\cdot]$. The operators $Q_b(\cdot)$ and $\mathcal{Q}_b(\cdot)$ denote scalar and vector quantization, respectively. The scalar ceiling function is given by $\lceil\cdot\rceil$. The set of circularly-symmetric jointly-Gaussian complex random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ is denoted by $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$.

II. SYSTEM MODEL

A. Signal Model

Consider a single-user mmWave MIMO system in which the transmitter sends N_s streams of data symbols through an array of N_t antennas to the receiver equipped with N_r antennas. The discrete-time data streams, represented by the vector $\mathbf{s} = [s_1, \ldots, s_{N_s}]^\mathsf{T}$, are assumed to be independent and Gaussian distributed with zero mean and unit variance, hence their covariance matrix is $\mathbf{R}_{ss} = \mathbb{E}[\mathbf{s}\mathbf{s}^\mathsf{H}] = \mathbf{I}_{N_s}$. Before transmission, the data streams are precoded as $\mathbf{x} = \mathcal{P}(\mathbf{s}) \in \mathbb{C}^{N_t}$ satisfying the average power constraint $P_x = \mathbb{E}\left[\|\mathbf{x}\|_2^2\right] \leq P_\mathrm{max}$, where $\mathcal{P}(\cdot) : \mathbb{C}^{N_s} \to \mathbb{C}^{N_t}$ represents the precoding operator. The propagation channel is assumed to be narrow-band block-fading, so there is no inter-symbol interference, and the discrete-time representation of the received signal can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \in \mathbb{C}^{N_r}, \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ denotes the MIMO channel matrix, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_{N_r \times 1}, \sigma_n^2 \mathbf{I}_{N_r})$ is the AWGN vector. The SNR normalized with respect to the maximum average transmit power is defined as $\gamma = P_\mathrm{max}/\sigma_n^2$.

We consider digital and hybrid A/D precoding schemes. The former consists of precoding the data streams using only a digital baseband filter $\mathbf{F}_\mathrm{BB} \in \mathbb{C}^{N_t \times N_s}$, which feeds signals to N_t DAC/RF chain pairs connected PAs and then to the antenna array, as illustrated in Figure 1a. The latter

scheme aims at decreasing the precoding power consumption by employing only $N_s \leq L_t < N_t$ DAC/RF chain pairs. To achieve this, the N_s data streams are first precoded by a digital baseband precoder $\mathbf{F}_{BB} \in \mathbb{C}^{L_t \times N_s}$, which feeds L_t DAC/RF chain pairs. Next, the up-converted signals are filtered by an analog precoder, amplified and transmitted through the antenna array. We consider that the analog beamforming is implemented by low-resolution phase-shifters as in [23]. Fully- and partially-connected PSN topologies are considered for analog beamforming. In fully-connected PSNs, the output of each RF chain is split by a N_t-way power divider and connected to N_t phase-shifters. Afterwards, the shifted signals are merged by L_t-way power combiners at each PA/antenna pair. In this case, the analog precoder is represented by a full matrix $\mathbf{F}_{RF} \in \mathbb{C}^{N_t \times L_t}$ whose entries are constrained to have constant modulus and discrete phase resolution, i.e., $[\mathbf{F}_{RF}]_{m,n} \in \mathcal{F}_{RF}, m \in \{1, \ldots, N_t\}, n \in \{1, \ldots, L_t\}$. The set $\mathcal{F}_{RF} = \{1, \phi, \phi^2, \ldots, \phi^{n_{PS}-1}\}$ denotes the phase-shift set, $\phi = \exp(j2\pi/n_{PS})$ the angular resolution, $n_{PS} = 2^{b_{PS}}$ the number of phase-shifts supported by the hardware, and b_{PS} the bit resolution. This architecture employs in total $L_t N_t$ phase-shift elements and is illustrated in Figure 1b. In partially-connected PSNs, by contrast, each RF chain is linked to only a sub-array of $N_a = \lceil N_t/L_t \rceil$ phase-shifters directly connected to PA/antenna pairs, as illustrated in Figure 1c. Therefore, the PSN for this architecture employs only N_a-way power dividers and no power combiners at all. The analog beamforming matrix in this case presents the following structure: $\mathbf{F}_{RF} = \mathrm{Diagblk}(\mathbf{f}_1, \ldots, \mathbf{f}_{L_t}) \in \mathbb{C}^{N_t \times L_t}$, where $\mathbf{f}_\ell \in \mathbb{C}^{N_a}$ denotes the analog beamforming filter of the ℓ-th sub-array. The phase-shifters constrain the sub-arrays analog beamforming filters as $[\mathbf{f}_\ell]_m \in \mathcal{F}_{RF}, \ell = 1, \ldots, L_t, m \in \{1, \ldots, N_a\}$.

To make our model consistent with power constraints present in real-world implementations, we incorporate non-linearities due to the DAC stages, and RF losses caused by analog beamforming. The non-linearity introduced by DACs is modeled as a quantization stage. This is because the continuous-valued DAC output signal has its amplitude well-represented by a finite set of values generated by the DAC hold circuits. The RF losses are accounted by considering a power loss factor $\frac{1}{\sqrt{L_{RF}}}$ multiplying the analog beamforming matrix $\mathbf{F}_{RF}$, as in [11]. The insertion loss of the analog precoder is represented by L_{RF} and depends on the power characteristics of the dividers, combiners and phase-shifters. The power loss model is further detailed in Section II-D. To include DAC and RF losses in our model, we define the precoding operation as

$$\mathbf{x} = \frac{1}{\sqrt{L_{RF}}}\mathbf{F}_{RF}\mathcal{Q}_b(\mathbf{F}_{BB}\mathbf{s}) = \frac{1}{\sqrt{L_{RF}}}\tilde{\mathbf{x}}, \qquad (2)$$

where $\mathcal{Q}_b(\cdot)$ stands for the vector quantization operator with b bits of resolution per dimension, and $\tilde{\mathbf{x}} = \mathbf{F}_{RF}\mathcal{Q}_b(\mathbf{F}_{BB}\mathbf{s})$ the lossless transmitted signal. The vector quantization operator will be further detailed and linearly approximated in Section II-B. Notice that (2) refers to digital precoding by setting $\mathbf{F}_{RF} = \mathbf{I}_{N_t}, L_{RF} = 1$, and $\mathbf{F}_{BB} \in \mathbb{C}^{N_t \times N_s}$. Therefore, for notation convenience, we consider that the quantization input and output vectors are M-dimensional, where $M = N_t$ and $M = L_t$ for digital and hybrid precoding, respectively.
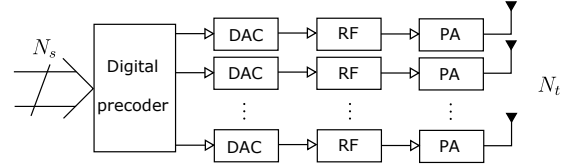
(a) Low-resolution digital precoder architecture.
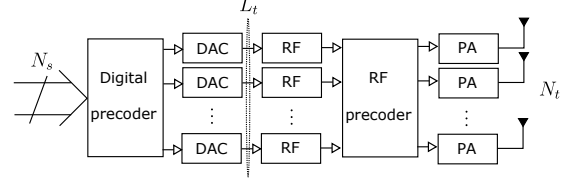
(b) Low-resolution hybrid precoder with L_t RF chains, and fully-connected PSN.
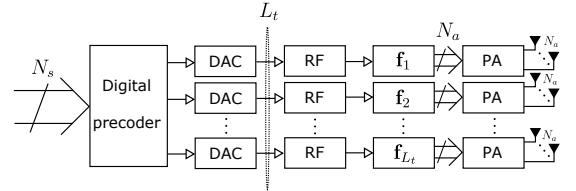
(c) Low-resolution hybrid precoder with L_t RF chains, partially-connected PSN, and L_t sub-arrays of N_a antennas.

Fig. 1: Transmitter architectures.

Thus, $\mathbf{F}_{RF}$ and $\mathbf{F}_{BB}$ are $(N_t \times M)$ and $(M \times N_s)$-dimensional matrices, respectively.

B. Quantized Signal Model

Let us define the quantization operation used in this work to model DAC and introduce its linear approximation. We define $Q_b(\cdot)$ as the uniform scalar quantizer that operates independently on both real and imaginary components of the input signal. A quantizer with b resolution bits has $N_b = 2^b$ codes for representing the quantized values. The quantizer output is $p = Q_b(u) = c_i^R + jc_\ell^I$ when $\Re[u] \in]a_{i-1}, a_i]$ and $\Im[u] \in]b_{\ell-1}, b_\ell]$, for $i, \ell \in \{1, \ldots, N_b - 1\}$. The value c_i^R (c_ℓ^I) represents the code associated with the real (imaginary) part of the i-th (ℓ-th) quantization interval, and a_i (b_ℓ) the quantization segment for the real (imaginary) component. We set $a_0 = b_0 = -\infty$ and $a_{N_b-1} = b_{N_b-1} = \infty$ to support inputs with arbitrary power. The quantization codes are chosen to minimize the MSE of Gaussian distributed signals, and they are listed in [24, Table II]. Such coding suits our modeling since the baseband precoded signals are still Gaussian. For vector inputs, we define $\mathcal{Q}_b(\cdot) : \mathbb{C}^M \to \mathbb{C}^M$ such that $\mathbf{p} = \mathcal{Q}_b(\mathbf{u}) \Rightarrow [\mathbf{p}]_m = Q_b([\mathbf{u}]_m), \quad m \in \{1, \ldots, M\}$. The vector quantizer simply consists of applying scalar quantization to each input vector entry. Unlike the Lloyd-Max quantizer [25], the considered vector quantization approach is not optimal in the MSE sense.

Conducting performance analysis of a MIMO system in terms of the exact non-linear quantization model presented above can be challenging. Among many reasons, we point to the difficulty of calculating the statistics of the transmitted signals. As an alternative, we resort to a linear approximation

of the scalar quantizer output applying the AQN model [12], [13], [26] $p = Q_b(u) \approx \sqrt{1 - \rho_b}u + e$. The input u is assumed to be Gaussian distributed, e is the quantization noise uncorrelated with u, and ρ_b is a quantization distortion factor, which is defined as the ratio of the quantization noise variance $\sigma_e^2 = \mathbb{E}\left[|e|^2\right]$ to the input signal variance $\sigma_u^2 = \mathbb{E}\left[|u|^2\right]$, i.e., $\rho_b = \sigma_e^2/\sigma_u^2$ [26]. Values for this factor are listed in [24, Table II, page 12] for resolution up to 5 bits. For larger resolution, it can be approximated as $\rho_b \approx \frac{\pi\sqrt{3}}{2}2^{-2b}$ [26]. Note that the quantization distortion factor also represents the inverse of the signal-to-quantization-noise ratio (SQNR). Thus, when the SQNR goes to infinity, the bit resolution b increases, and ρ_b converges to zero. In this quantization model, the additive noise power is proportional to the input variance and decreases as the quantization resolution grows. Such a behavior is coherent with practice and motivates the adoption of this model. Also note that the approximation definition ensures that $\sigma_p^2 \approx \sigma_u^2$. Let us extend the AQN model to MIMO systems:

$$\mathbf{p} = \mathcal{Q}_b(\mathbf{u}) \approx \boldsymbol{\Upsilon}_b\mathbf{u} + \mathbf{e} \in \mathbb{C}^M. \qquad (3)$$

Vectors $\mathbf{u} = [u_1, \ldots, u_M]^\mathsf{T}$ and $\mathbf{p} = [p_1, \ldots, p_M]^\mathsf{T}$ denote respectively the input and output of the quantizer, and $\mathbf{e} = [e_1, \ldots, e_M]^\mathsf{T}$ the noise quantization vector which satisfies $\mathbb{E}[\mathbf{u}\mathbf{e}^\mathsf{H}] = \mathbb{E}[\mathbf{e}\mathbf{u}^\mathsf{H}] = \mathbf{0}_{M \times M}$. The diagonal matrix $\boldsymbol{\Upsilon}_b = \mathrm{Diag}(\sqrt{1 - \rho_{b,1}}, \ldots, \sqrt{1 - \rho_{b,M}})$ is formed by the quantization distortion factors $\rho_{b,m}$ associated with the m-th quantizer input-output. From the definition of ρ_b, it follows that the covariance matrix of the quantization error vector is given by:

$$\mathbf{R}_{ee} = \mathrm{Diag}(\rho_{b,1}, \ldots, \rho_{b,M})\,\mathrm{diag}(\mathbf{R}_{uu}) \in \mathbb{C}^{M \times M}, \qquad (4)$$

where $\mathbf{R}_{uu} = \mathbb{E}\left[\mathbf{u}\mathbf{u}^\mathsf{H}\right]$ denotes the quantizer input covariance matrix. Knowledge of these covariance matrices will be useful in Section III, where we will derive the achievable rate for each precoder.

The AQN model approximates a non-linear deterministic operation as a linear stochastic process by assuming zero correlation between the quantizer input and the quantization noise. Similar results are obtained by using the Bussgang Theorem [19], which states that the cross-correlation function between two Gaussian signals taken after one of them has been non-linearly distorted is proportional to the cross-correlation function before distortion. If a linear model, e.g. (3), is used to approximate the non-linear distortion, this Theorem implies that the approximation error is uncorrelated with the input, i.e., $\mathbb{E}[\mathbf{u}\mathbf{e}^\mathsf{H}] = \mathbb{E}[\mathbf{e}\mathbf{u}^\mathsf{H}] = \mathbf{0}_{M \times M}$. Therefore, the Bussgang Theorem can be seen as the theoretical foundation underlying the AQN model, since this non-correlation assumption is crucial. It was shown in [15] that AGC error could invalidate this assumption, but errors between -20% and 20% are still acceptable. Model accuracy has been investigated in [12] and it was found to be accurate especially in the low SNR regime, provided that the input signals are Gaussian distributed.

Now let us apply the AQN model to the precoded signal model (2). Define the DAC input signal $\mathbf{u} = \mathbf{F}_{\mathrm{BB}}\mathbf{s} \in \mathbb{C}^M$.

The lossless transmitted signal defined in (2) can now be approximated as

$$\tilde{\mathbf{x}} \approx \mathbf{F}_{\mathrm{RF}}\boldsymbol{\Upsilon}_b\mathbf{u} + \mathbf{F}_{\mathrm{RF}}\mathbf{e} \in \mathbb{C}^{N_t}. \qquad (5)$$

Since the data streams $\{s_1, \ldots, s_{N_s}\}$ are Gaussian distributed, the baseband precoded signals $\{u_1, \ldots, u_M\}$ will also be Gaussian distributed with covariance matrix $\mathbf{R}_{uu} = \mathbf{F}_{\mathrm{BB}}\mathbf{F}_{\mathrm{BB}}^\mathsf{H}$. As each quantizer output is Gaussian distributed and has b bits of resolution, it is reasonable to claim that all DACs undergo the same distortion ρ_b, i.e., $\rho_{b,1} = \rho_{b,2} = \ldots = \rho_{b,M} = \rho_b$, and, consequently, it follows that $\boldsymbol{\Upsilon}_b = \sqrt{1 - \rho_b}\mathbf{I}_M$. Now the error covariance matrix (4) simplifies to

$$\mathbf{R}_{ee} = \rho_b \, \mathrm{diag}(\mathbf{R}_{uu}). \qquad (6)$$

Using (2) and (5), the received signal model (1) can be approximated as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \approx \tfrac{1}{\sqrt{L_{\mathrm{RF}}}}\mathbf{H}'\mathbf{u} + \mathbf{n}', \qquad (7)$$

where $\mathbf{H}' = \mathbf{H}_{\mathrm{eq}}\boldsymbol{\Upsilon}_b$ denotes the $(N_r \times M)$-dimensional total channel matrix, $\mathbf{H}_{\mathrm{eq}} = \mathbf{H}\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_r \times M}$ the equivalent channel formed by the cascade of the analog precoding and the transmission channel, and $\mathbf{n}' = \frac{1}{\sqrt{L_{\mathrm{RF}}}}\mathbf{H}_{\mathrm{eq}}\mathbf{e} + \mathbf{n} \in \mathbb{C}^{N_r}$ the total additive noise component. The total channel matrix $\mathbf{H}'$ is formed by the cascade of $\mathbf{H}_{\mathrm{eq}}$ and the quantization scaling $\boldsymbol{\Upsilon}_b$. The former factor models the effect of the physical channel $\mathbf{H}$ and the lossless analog precoder $\mathbf{F}_{\mathrm{RF}}$ (for hybrid structures). Of course, one must have already designed $\mathbf{F}_{\mathrm{RF}}$ to form $\mathbf{H}_{\mathrm{eq}}$, and, indeed, computing $\mathbf{F}_{\mathrm{RF}}$ is the first step of the hybrid precoding strategies we present in Section III. The factor $\boldsymbol{\Upsilon}_b$, on the other hand, reflects the reduction of the useful transmit power due to quantization. For high-resolution quantization, $\rho_b \to 0$ and $\mathbf{H}' \to \mathbf{H}_{\mathrm{eq}}$, thus, in this case, the quantization noise has no impact on the total channel power. By contrast, for low- and mid-resolution quantization, $\boldsymbol{\Upsilon}_b$ becomes important and the total channel is then expressed as $\mathbf{H}' = \sqrt{1 - \rho_b}\mathbf{H}_{\mathrm{eq}}$. Since $\sqrt{1 - \rho_b}$ is always smaller than one, low-resolution quantization decreases the total channel power. Another interesting feature of model (7) is that $\mathbf{n}'$ comprises the AWGN plus an attenuated channel-filtered quantization noise term. Therefore, $\mathbf{n}'$ is not guaranteed to be Gaussian, as nothing is assumed about the distribution of $\mathbf{e}$. The covariance matrix of the total additive noise is

$$\mathbf{R}_{n'n'} = \mathbb{E}[\mathbf{n}'\mathbf{n}'^\mathsf{H}] = \tfrac{1}{L_{\mathrm{RF}}}\mathbf{H}_{\mathrm{eq}}\mathbf{R}_{ee}\mathbf{H}_{\mathrm{eq}}^\mathsf{H} + \mathbf{R}_{nn} \in \mathbb{C}^{N_r \times N_r}, \qquad (8)$$

where $\mathbf{R}_{nn} = \sigma_n^2\mathbf{I}_{N_r}$, and the quantization noise covariance matrix $\mathbf{R}_{ee}$ is given by Equation (6). One can easily see that $\mathbf{R}_{n'n'}$ is in general not a diagonal matrix, therefore the total additive noise vector entries are correlated. Note that we assume $\mathbb{E}\left[\mathbf{n}\mathbf{e}^\mathsf{H}\right] = \mathbf{0}_{N_r \times M}$, which is reasonable in practice, since the quantization process at the transmitter has no influence whatsoever on the receiver sensor noise.

Unfortunately, Equation (7) is still problematic when it comes to obtaining achievable rate expressions. It is difficult to separate the total MIMO channel $\mathbf{H}'$ into orthogonal subchannels due to the structure of the total noise vector $\mathbf{n}'$. There are no guarantees that $\mathbf{n}'$ is Gaussian distributed, complicating the achievable rate maximization. Moreover, its covariance

matrix (8) depends on $\mathbf{R}_{ee}$, which varies according to the diagonal elements of $\mathbf{R}_{uu}$, bringing a causality problem for the precoder design. In view of these difficulties, we make some assumptions to simplify our analysis and to obtain a lower bound for the achievable rate. In [27], it was shown that the Gaussian noise distribution minimizes the mutual information for a given noise covariance matrix. This result motivates us to approximate the additive quantization noise $\mathbf{n}'$ as a jointly Gaussian distributed noise vector $\mathbf{n}_G = \frac{1}{\sqrt{L_{\text{RF}}}}\mathbf{H}_{\text{eq}}\mathbf{e} + \mathbf{n}$ with covariance matrix $\mathbf{R}_{n_G n_G} = \mathbf{R}_{n'n'}$ in order to obtain an achievable rate lower bound as in [26]. The received signal model (7) is now be approximated as

$$\mathbf{y} \approx \frac{1}{\sqrt{L_{\text{RF}}}}\mathbf{H}'\mathbf{u} + \mathbf{n}_G = \frac{1}{\sqrt{L_{\text{RF}}}}\mathbf{H}_{\text{eq}}\mathbf{\Upsilon}_b\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{n}_G. \quad (9)$$

To simplify the analysis of this signal model, the colored noise vector $\mathbf{n}_G$ is decorrelated. Since $\mathbf{R}_{n_G n_G}$ is a Hermitian matrix, it admits an eigenvalue decomposition $\mathbf{R}_{n_G n_G} = \mathbf{J}\mathbf{L}\mathbf{J}^{\mathsf{H}}$, where $\mathbf{J}^{\mathsf{H}}\mathbf{J} = \mathbf{J}\mathbf{J}^{\mathsf{H}} = \mathbf{I}_{N_r}$. The whitening filter is thus given by $\mathbf{R}_{n_G n_G}^{-1/2} = \mathbf{L}^{-1/2}\mathbf{J}^{\mathsf{H}}$ [28], and the received signal vector $\mathbf{y}$ is pre-multiplied by this filter, yielding:

$$\mathbf{R}_{n_G n_G}^{-1/2}\mathbf{y} = \frac{1}{\sqrt{L_{\text{RF}}}}\mathbf{R}_{n_G n_G}^{-1/2}\mathbf{H}'\mathbf{u} + \mathbf{R}_{n_G n_G}^{-1/2}\mathbf{n}_G. \quad (10)$$

Notice that after noise whitening, the covariance matrix of the noise vector in (10) is an $(N_r \times N_r)$-dimensional identity matrix. We further apply model (10) in Section III to define the quantized hybrid precoding problem.

C. Channel Model

Experiments conducted in [29], [30] indicate that mmWave massive MIMO channels present a high degree of spatial and angular sparsity due to the large path loss, and thus there are only a few dominant multipaths. Such channels can be modeled by a narrow-band clustered channel model contributing with L propagation paths [4], [8] resulting in a channel matrix

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}}\sum_{\ell=1}^{L}\alpha_\ell \mathbf{a}_r(\theta_\ell^r, \phi_\ell^r)\mathbf{a}_t(\theta_\ell^t, \phi_\ell^t)^{\mathsf{H}} \in \mathbb{C}^{N_r \times N_t}, \quad (11)$$

where $\alpha_\ell \in \mathbb{C}$ denotes the small-scale fading, $\mathbf{a}_r(\theta_\ell^r, \phi_\ell^r) \in \mathbb{C}^{N_r}$ and $\mathbf{a}_t(\theta_\ell^t, \phi_\ell^t) \in \mathbb{C}^{N_t}$ the array response and steering vectors evaluated at elevation θ_ℓ^r (θ_ℓ^t) and azimuth ϕ_ℓ^r (ϕ_ℓ^t) arrival (departure) angles at the receiver (transmitter), respectively. Each channel matrix realization is normalized such that $\mathbb{E}\left[\|\mathbf{H}\|_{\text{F}}^2\right] = N_r N_t$. We assume that each path has the same average power, i.e., α_ℓ is modeled as circular symmetric Gaussian random variables with zero mean and unit variance. The departure/arrival elevation and azimuth angles are uniformly distributed in the intervals $]-\pi/2, \pi/2[\,\text{rad}$ and $[0, 2\pi]\,\text{rad}$, respectively.

The transmit and receive arrays are assumed to be linear and uniformly spaced with inter-element spacing of $\lambda/2$, where λ denotes the carrier wavelength. The methods discussed in this paper can be easily applied to arrays with arbitrary geometry and element beam-pattern. Assuming that the uniform linear array is distributed along the x-axis, the steering and array response vectors follow the Vandermonde structure

$$\mathbf{a}_x(\theta, \phi) = \frac{1}{\sqrt{N_x}}\left[1, \dots, e^{j\pi(N_x-1)\sin\theta\,\cos\phi}\right]^{\mathsf{T}} \in \mathbb{C}^{N_x},$$

for $x \in \{t, r\}$.

D. Power Consumption and Loss Models

Energy efficiency is an important concern in the design of mmWave massive MIMO systems. Large antenna arrays are employed to compensate for the increased path loss of mmWave channels and to achieve high spectral efficiency. However, several electronic components such as DACs, phase-shifters, and power amplifiers become inefficient when operating in the millimeter wave range over large bandwidth, making radio systems equipped with such arrays expensive due to large power consumption and losses of their RF front-end. Therefore, there has been important research efforts on designing systems that not only maximize their spectral efficiency, but also their energy efficiency, which can be defined as the ratio of spectral efficiency to power consumption [13].

Hybrid systems seek to decrease power consumption by employing only a few RF chains and an analog beamforming stage. Although hybrid architectures are expected to exhibit reduced power consumption compared to fully-digital systems, they shall also present important power losses due to phase-shifters, power dividers, and power combiners. To compensate for these losses, one could simply adopt high-gain PAs that would cover the insertion losses. However, high-gain and high-efficiency PAs are not available for mmWave yet, and, thus, these losses cannot be easily compensated for, incurring in energy efficiency degradation. In order to consider these power losses in our modeling, we adopt a realistic RF formulation based on the insertion loss of power dividers, power combiners, and phase-shifters.

Recently, active and passive phase-shifting elements for mmWave have been reported in [31]. The former present non-negligible power consumption and moderate gain, whereas the latter has insignificant power consumption and considerable insertion loss. Therefore, the system designer can either maximize the transmit power with increased power consumption by active phase-shifters, or reduce power consumption with decreased transmit power by passive phase-shifters. Which circuit implementation leads to the most energy-efficient hybrid system is not clear. Thus, in order to compare the energy performance of the different transmitter architectures, we define power consumption models, and we introduce the power loss modeling in the following.

The influence of signal processing on system power consumption may be important in massive MIMO systems. It is well-known that the computational demand of standard signal processing methods significantly grows with the number of antenna elements. Since these methods are typically implemented in power hungry devices, it is thoughtful to avoid extensive computations in order to save energy. As pointed out in [32], this aspect is usually overlooked and should be considered during system optimization. Therefore, in our modeling, we

consider the *computational* power consumption corresponding to precoder optimization, and the *static* power consumption of the RF front-end.

Let us first discuss the static power consumption of hybrid and digital transmitters. The transmitter RF front-end is formed by a DAC for each I/Q channel, RF chains, PSN (hybrid systems), and PAs. The considered direct conversion RF chain consists of two low-pass filters, two mixers, a local oscillator shared among all chains, and a $90°$ hybrid with buffers. Denoting the power consumption of low-pass filter, mixer, local oscillator, hybrid with buffer as P_{LP}, P_{M}, P_{LO}, P_{H}, respectively, the power consumption P_{RF} of a single RF chain is then

$$P_{\text{RF}} = 2P_{\text{LP}} + 2P_{\text{M}} + P_{\text{H}}. \tag{12}$$

A transmitter with digital precoding employs N_t RF chains, PAs, and pairs of DACs, thus its power consumption is given by

$$P_{\text{D}}(N_t, b_{\text{DAC}}, F_s) = P_{\text{LO}} + P_{\text{PA}} + N_t[2P_{\text{DAC}}(b_{\text{DAC}}, F_s) + P_{\text{RF}}],$$

where $P_{\text{DAC}}(b_{\text{DAC}}, F_s)$ denotes the power consumption of a DAC sampling at F_s Hertz with b_{DAC} bits of resolution. P_{PA} stands for the power consumed by all PAs. A hybrid transmitter with fully-connected PSN, by contrast, employs L_t DAC/RF chain pairs. A total of $N_t L_t$ phase-shifters are used in this architecture, thus its static power consumption is

$$P_{\text{FPSN}}(N_t, L_t, b_{\text{DAC}}, b_{\text{PS}}, F_s) =$$
$$P_{\text{LO}} + P_{\text{PA}} + L_t[2P_{\text{DAC}}(b_{\text{DAC}}, F_s) + P_{\text{RF}}] + N_t L_t P_{\text{PS}}(b_{\text{PS}}),$$

where $P_{\text{PS}}(b_{\text{PS}})$ denotes power consumption of a single phase-shift element with b_{PS} bits of resolution. In general, the power consumption of dividers and combiners is negligible. Transmitters with partially-connected PSN have N_a phase-shifters per sub-array, i.e., there are $N_a L_t$ phase-shifters in total, hence their static power consumption is given by

$$P_{\text{PPSN}}(N_t, L_t, b_{\text{DAC}}, b_{\text{PS}}, F_s) =$$
$$P_{\text{LO}} + P_{\text{PA}} + L_t[2P_{\text{DAC}}(b_{\text{DAC}}, F_s) + P_{\text{RF}}] + N_a L_t P_{\text{PS}}(b_{\text{PS}}).$$

The need of efficient mmWave systems fueled the development of new electronic components. One can find in the literature different parameters for the electronic devices considered in our models. We adopt an optimistic parameter selection approach, in which we choose the most efficient implementation reported. Results obtained with this approach ought to provide an idea of what to expect for future mmWave massive MIMO systems. Regarding the power consumption of the RF chain components, we consider values reported in [33, Chapter 5] for the $90°$ hybrid with buffers, [34] for the mixers, [35] for the low-pass filters, and [36] for the local oscillator. We employ DACs with a binary-weighted current-steering topology. Its power consumption is a function of the effective number of bits b_{DAC} and the sampling frequency F_s [37]:

$$P_{\text{DAC}}(b_{\text{DAC}}, F_s) = 1.5 \times 10^{-5} \cdot 2^{b_{\text{DAC}}} + 9 \times 10^{-12} \cdot b_{\text{DAC}} \cdot F_s. \tag{13}$$

Equation (13) was obtained by the same parameter setup as in [37]. According to [38, Section 1], this type of DAC is well-adapted for high-speed conversion since no buffer is required

and thus switching can be done fast. PAs are the most power hungry devices on the transmit side, because the high linearity requirements render them inefficient. The power consumed by the set of PAs with power-added efficiency (PAE) η is given by $P_{\text{PA}} = P_x/\eta$ [39], where P_x is the actual transmit power considering RF losses defined in Section III. Peak PAE values for state-of-art PAs are listed in [40], and vary between $6.5-27\%$. According to our optimistic approach, we set $\eta = 27\%$. State-of-art mmWave active phase-shifters have been listed in [9] and their power consumption lies in the range $15-108\,\text{mW}$ with moderate peak gains. As an alternative to active phase-shifters, passive implementations with negligible power consumption and significant insertion loss have been reported in [31]. We spend $P_{\text{PS}} = 21.6\,\text{mW}$ for active phase-shifters [41] and consider zero power consumption for passive phase-shifters [31]. The assumed power consumption values of the RF front-end components are summarized in Table I.

Let us now present the power loss model considered in this work. As in [11], RF power losses in the analog beamformer lead to the following representation of the analog precoding matrix: $\mathbf{F}_{\text{RF}}/\sqrt{L_{\text{RF}}^{\{\text{FPSN,PPSN}\}}}$, where $L_{\text{RF}}^{\{\text{FPSN,PPSN}\}}$ denotes the loss factor for fully- and partially-connected PSNs. We will derive expressions for $L_{\text{RF}}^{\{\text{FPSN,PPSN}\}}$ taking into account the insertion loss of each RF component in the PSN. We consider that K-way devices are formed by concatenating $\lceil \log_2(K) \rceil$ stages of two-way devices [11], thus the power loss for the total dividing and combining stages is given by $L_{\{\text{D,C}\}}(K)\,[\text{dB}] = \bar{L}_{\{\text{D,C}\}}\lceil \log_2(K) \rceil$, where $\bar{L}_{\{\text{D,C}\}}$ denotes the static power loss in decibels of a single two-way device [11]. We consider that two-way power dividers lose $0.6\,\text{dB}$ [11], and two-way power combiners $0.6\,\text{dB}$ plus extra $3\,\text{dB}$ due to amplitude and phase mismatches. Regarding phase-shifting, we assume that active phase-shifters have a gain of $2.3\,\text{dB}$, i.e., loss of $L_{\text{PS}} = -2.3\,\text{dB}$ [41], while passive phase-shifters have a loss of $L_{\text{PS}} = 8.8\,\text{dB}$ [31]. Values for $\bar{L}_{\text{D}}$, $\bar{L}_{\text{C}}$, and L_{PS} are listed in Table I.

Let us consider a fully-connected PSN fed with P_{in} Watts. The signal power at the input of each N_t-way power divider is P_{in}/L_t. After power splitting and phase-shifting, the signal power at each network branch is $(P_{\text{in}}/L_t)/(N_t L_{\text{D}}(N_t)L_{\text{PS}})$. The power level at the output of each L_t-way power combiner is $[(P_{\text{in}}/L_t)L_t]/(N_t L_{\text{D}}(N_t)L_{\text{PS}}L_{\text{C}}(L_t))$. Since there are N_t power combiners, the PSN output power is $P_{\text{out}} = P_{\text{in}}/(L_{\text{D}}(N_t)L_{\text{PS}}L_{\text{C}}(L_t))$, and, hence, the loss factor for fully-connected PSNs is

$$L_{\text{RF}}^{\text{FPSN}} = \frac{P_{\text{in}}}{P_{\text{out}}} = L_{\text{D}}(N_t)L_{\text{PS}}L_{\text{C}}(L_t).$$

The loss factor for partially-connected PSNs is obtained analogously, considering N_a-way power dividers and no power combining:

$$L_{\text{RF}}^{\text{PPSN}} = L_{\text{D}}(N_a)L_{\text{PS}}.$$

The computational analysis of signal processing methods involves counting of arithmetical operations, memory overhead analysis, among other factors [22]. Here, we consider the simple computational power consumption model of [21], in which the power spent with precoder optimization is proportional to the number N_{flops} of floating-point operations (flops)

Table I: Power consumption and loss of the RF front-end components.

Component	Notation	Value
Power amplifier [40]	P_{PA}	P_x/η, $\eta = 27\%$
Phase-shifter (active [41] ; passive [31])	P_{PS}	21.6 ; $0\,\mathrm{mW}$
DAC [37]	P_{DAC}	Equation (13)
Local oscillator [36]	P_{LO}	$22.5\,\mathrm{mW}$
$90°$ hybrid with buffers [33]	P_{H}	$3\,\mathrm{mW}$
Mixer [34]	P_{M}	$0.3\,\mathrm{mW}$
Low-pass filter [35]	P_{LP}	$14\,\mathrm{mW}$
RF chain (Equation (12))	P_{RF}	$31.6\,\mathrm{mW}$
Two-way power divider [11]	$\bar{L}_{\mathrm{D}}$	$0.6\,\mathrm{dB}$
Two-way power combiner [11]	$\bar{L}_{\mathrm{C}}$	$0.6\,\mathrm{dB} + 3\,\mathrm{dB}$
Phase-shifter (active [41] ; passive [31])	L_{PS}	-2.3 ; $8.8\,\mathrm{dB}$

demanded by the optimization algorithm, the number C of coherence blocks per second, and inversely proportional to the transmitter computational efficiency E_c [flops/s/W], i.e., it is given by

$$P_{\mathrm{comp}} = C N_{\mathrm{flops}}/E_c.$$

The number C of coherence blocks determines how many times the system updates its precoding filters per second. It is defined as the ratio of the transmit bandwidth B to the number of coherence blocks, which is given by the product of the channel coherence bandwidth B_C and the channel coherence time T_C, i.e., $C = B/(B_C T_C)$ [21]. Naturally, the computational power consumption in practice strongly depends on the hardware implementation. However, distinct implementations are characterized by different computational efficiency E_c. Therefore, the considered model is general enough to provide insights on the computational power consumption at the transmitter.

III. Precoding Strategies

In this section, we introduce the quantized hybrid precoding problem, and present precoding strategies to solve it. Analog precoding methods for fully- and partially-connected PSN are presented in Sec. III-A, baseband precoding is defined in Sec. III-B, computational complexity analysis is conducted in Sec. III-C, and achievable rate lower bounds are derived in Sec. III-D.

The *quantized hybrid precoding problem* is based on system model (10) assuming perfect and instantaneous channel state information at both transmitter and receiver. It consists of finding the precoding matrices $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ that maximize the achievable instantaneous rate R [42], i.e.,

$$\underset{\mathbf{F}_{\mathrm{RF}},\, \mathbf{F}_{\mathrm{BB}}}{\mathrm{maximize}} \quad \log_2 \det\left(\mathbf{I}_{N_r} + \frac{1}{L_{\mathrm{RF}}}\mathbf{R}_{n_G n_G}^{-1/2}\mathbf{H}'\mathbf{R}_{uu}\mathbf{H}'^{\mathsf{H}}\mathbf{R}_{n_G n_G}^{-1/2,\mathsf{H}}\right)$$
$$\text{subject to} \quad [\mathbf{F}_{\mathrm{RF}}]_{u,v} \in \mathcal{F}_{\mathrm{RF}}, \forall u \forall v, \; \mathbb{E}\left[\|\tilde{\mathbf{x}}\|_2^2\right] \le P_{\max}. \tag{14}$$

The achievable rate expression in (14) provides some insights on the performance of the quantized hybrid precoder. In order to discuss them, consider the actual transmitted power $P_x = \mathbb{E}\left[\|\mathbf{x}\|_2^2\right] = \frac{1}{L_{\mathrm{RF}}}P_{\tilde{x}}$, where $P_{\tilde{x}} = \mathbb{E}\left[\|\tilde{\mathbf{x}}\|_2^2\right] =$

$(1 - \rho_b)\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_{\mathrm{F}}^2 + \mathrm{tr}(\mathbf{F}_{\mathrm{RF}}\mathbf{R}_{ee}\mathbf{F}_{\mathrm{RF}}^{\mathsf{H}})$ stands for the lossless transmit signal power. In the appendix, we show that $P_{\tilde{x}} = P_{\max}$ when the baseband precoder presented in Section III-B is employed. Therefore, the actual transmit power is $P_x = P_{\max}/L_{\mathrm{RF}}$, satisfying the average power constraint. As discussed in Section II-D, if the RF losses are not properly compensated for, then the actual transmit power is reduced, decreasing the achievable rate.

Problem (14) is sufficiently general to model other precoding schemes. For example, when $\rho_b = 0$, we have $\mathbf{R}_{ee} = \mathbf{0}_{M \times M} \Rightarrow \mathbf{R}_{n_G n_G}^{-1/2} = \sigma_n^{-1}\mathbf{I}_{N_r}$, and the objective function is rewritten as $\log_2 \det\left(\mathbf{I}_{N_r} + \frac{1}{L_{\mathrm{RF}}}\mathbf{R}_{n_G n_G}^{-1}\mathbf{H}_{\mathrm{eq}}\mathbf{R}_{uu}\mathbf{H}_{\mathrm{eq}}^{\mathsf{H}}\right)$, which refers to the hybrid precoding problem [8] with lossy RF hardware. Moreover, when $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{N_t \times N_s}$, $\mathbf{F}_{\mathrm{RF}} = \mathbf{I}_{N_t}$, $L_{\mathrm{RF}} = 1$, and $\rho_b = 0$, (14) falls back to the classical SU-MIMO digital precoding problem. The corresponding quantized problems are obtained for $\rho_b \ne 0$.

To solve (14), we adopt the sub-optimal strategy of decoupling the precoder design problem into two subproblems for optimizing $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$, separately. We first tackle the analog precoding subproblem for obtaining $\mathbf{F}_{\mathrm{RF}}$, allowing us to form the equivalent channel matrix $\mathbf{H}_{\mathrm{eq}}$. Subsequently, we design our baseband precoder based on the SVD of $\mathbf{H}_{\mathrm{eq}}$.

A. Analog Precoding Strategies

In the following, we present analog precoding methods for the fully- and partially-connected PSN topologies. Since $\mathbf{F}_{\mathrm{RF}}$ may have very large dimensions, the computational efforts of precoding design can be significant, and, thus, we focus on low computational complexity instead of high spectral efficiency.

1) Fully-connected PSN: As in [13], we design the RF precoder with the alternating projection method of [14, Section 3]. It consists of initially selecting the first L_t right singular vectors of the MIMO channel matrix $\mathbf{H}$ and forming a semi-unitary precoder $\mathbf{F}_{\mathrm{SU}}$. Next, this matrix is projected onto the constant modulus space, and the result is projected back to the semi-unitary matrix $\mathbf{F}_{\mathrm{SU}}$. This alternating projection procedure is repeated until convergence, which is achieved when the Frobenius norm residual between two consecutive iterations is smaller than a tolerance value. After convergence, the phase of each element in $\mathbf{F}_{\mathrm{RF}}$ is quantized to the closest value in the phase resolution set $\mathcal{F}_{\mathrm{RF}}$. The fully-connected PSN hybrid precoder is summarized in Algorithm 1.

Algorithm 1 Hybrid precoding (fully-connected PSN)

Require: $\mathbf{H}$, L_t, b_{PS}
1: Compute SVD of $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}$
2: Initialize $\mathbf{F}_{\mathrm{SU}} \leftarrow [\mathbf{V}]_{:,1:L_t}$
3: **repeat**
4:     $\mathbf{F}_{\mathrm{RF}} \leftarrow \exp(j\angle\mathbf{F}_{\mathrm{SU}})$
5:     Compute SVD of $\mathbf{F}_{\mathrm{RF}} = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^{\mathsf{H}}$
6:     $\mathbf{F}_{\mathrm{SU}} \leftarrow [\tilde{\mathbf{U}}]_{:,1:L_t}\tilde{\mathbf{V}}^{\mathsf{H}}$
7: **until** convergence criterion triggers
8: $\mathbf{F}_{\mathrm{RF}} \leftarrow \exp(j\mathcal{Q}_{b_{\mathrm{PS}}}(\angle\mathbf{F}_{\mathrm{RF}}))$

---

2) Partially-connected PSN: For the partially-connected RF precoding, we define the sub-channel matrix $\mathbf{H}_\ell = [\mathbf{H}]_{:,n_a+(\ell-1)N_a} \in \mathbb{C}^{N_r \times N_a}$, $n_a = 1, \ldots, N_a$, that contains the N_a columns of $\mathbf{H}$ belonging to the ℓ-th antenna sub-array.

The proposed design employs quantized MET for each subarray, i.e., $\mathbf{f}_\ell = \exp(j\mathcal{Q}_{b_{PS}}(\angle \mathbf{v}_\ell^{\max}))$, $\ell = 1, \ldots, L_t$, where $\mathbf{v}_\ell^{\max}$ denotes the right singular vector corresponding to the largest (dominant) singular value of $\mathbf{H}_\ell$. The projection step enforces the constant modulus constraint imposed by the phase-shifters. Since we are interested only in $\mathbf{v}_\ell^{\max}$, we can use the simple power method [22, Section 8.2.1] to calculate it. This algorithm is significantly less expensive than, for example, the R-SVD algorithm [22] which is useful for computing the full SVD. The power method achieves convergence when the residual error between two consecutive iterations is smaller than a tolerance value. It is guaranteed to converge if the dominant singular value is larger than all the other singular values in modulus and if the initial guess for $\mathbf{v}_\ell^{\max}$ has a non-zero component in the direction of the corresponding right singular vector [22], which occurs for the model presented in Section II-C. The partially-connected hybrid precoding method is outlined in Algorithm 2.

Algorithm 2 Analog beamforming for partially-connected PSN via the power method

---

Require: $\mathbf{H}$, L_t, N_a, b_{PS}
1: **for** $\ell = 1, \ldots, L_t$ **do**
2:     Form $\mathbf{H}_\ell \leftarrow [\mathbf{H}]_{:,n_a+(\ell-1)N_a}$, $n_a = 1, \ldots, N_a$
3:     Randomly initialize $\mathbf{v}_\ell \in \mathbb{C}^{N_a}$
4:     **repeat**
5:         $\mathbf{v}_\ell \leftarrow \mathbf{H}_\ell^{\mathsf{H}}\mathbf{H}_\ell \mathbf{v}$
6:         $\mathbf{v}_\ell \leftarrow \mathbf{v}_\ell/\|\mathbf{v}_\ell\|_2$
7:     **until** convergence criterion triggers
8: **end for**
9: $\mathbf{F}_{RF} \leftarrow \mathrm{Diagblk}\left[\exp(j\mathcal{Q}_{b_{PS}}(\angle\mathbf{v}_1)), \ldots, \exp(j\mathcal{Q}_{b_{PS}}(\angle\mathbf{v}_{L_t}))\right]$

B. Digital Precoding Strategy

Unfortunately, finding the digital precoder $\mathbf{F}_{BB}$ that maximizes (14) for fixed $\mathbf{F}_{RF}$ is not straightforward. The SVD precoding with water-filling power allocation is not guaranteed anymore to decompose the MIMO system into orthogonal sub-channels due to the structure of the total noise vector $\mathbf{n}'$. Furthermore, in order to form $\mathbf{H}'$, one needs knowledge of the diagonal elements of $\mathbf{R}_{uu}$, which depends on $\mathbf{F}_{BB}$.

To solve these problems, notice that as the quantization resolution increases, $\rho_b \to 0$, $\mathbf{R}_{n_G n_G} \to \mathbf{R}_{nn}$, $\mathbf{H}' \to \mathbf{H}_{eq}$, and the instantaneous rate in (14) goes to $\log_2 \det\left(\mathbf{I}_{N_r} + \frac{1}{L_{RF}}\mathbf{R}_{nn}^{-1}\mathbf{H}_{eq}\mathbf{R}_{uu}\mathbf{H}_{eq}^{\mathsf{H}}\right)$. In this case, SVD precoding and water-filling power allocation with respect to $\mathbf{H}_{eq}$ and $\mathbf{R}_{nn}$ become optimal. Note that this strategy still maximizes the achievable rate regardless of $\frac{1}{L_{RF}}$, thus we ignore this power loss factor in the transmitted signal model when designing $\mathbf{F}_{BB}$ by setting $L_{RF} = 1$. In general, this strategy becomes sub-optimal under low-resolution quantization. Nevertheless, it is still expected to be close to optimal, particularly at low SNR. In such regime, AWGN overruns the quantization noise, motivating the use of the SVD precoding. At high SNR, however, the quantization noise dominates and this solution is only sub-optimal. Since massive MIMO systems will likely operate in low SNR regimes, this choice is reasonable in practice.

Following our discussion, we design $\mathbf{F}_{BB}$ as the optimal precoder in infinite-resolution DAC scenarios. In this sense, we assume $\rho_b = 0$ and knowledge of $\mathbf{F}_{RF}$ (see Section III-A), allowing $\mathbf{H}_{eq}$ to be formed[1]. The baseband precoder is then obtained by solving

$$\begin{aligned}\underset{\mathbf{F}_{BB}}{\text{maximize}} \quad & \log_2 \det\left(\mathbf{I}_{N_r} + \mathbf{R}_{nn}^{-1}\mathbf{H}_{eq}\mathbf{F}_{BB}\mathbf{F}_{BB}^{\mathsf{H}}\mathbf{H}_{eq}^{\mathsf{H}}\right) \\ \text{subject to} \quad & \mathbb{E}\left[\|\tilde{\mathbf{x}}\|_2^2\right] \leq P_{\max}.\end{aligned} \quad (15)$$

It is straightforward to show that the average power constraint in (15) can be rewritten as $\|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 \leq P_{\max}$ when $\rho_b = 0$. Let the SVD of the equivalent channel be denoted by $\mathbf{H}_{eq} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}$, in which $\mathbf{U} \in \mathbb{C}^{N_r \times N_s}$, $\boldsymbol{\Sigma} \in \mathbb{C}^{N_s \times N_s}$, and $\mathbf{V} \in \mathbb{C}^{M \times N_s}$. We employ SVD baseband precoding with water-filling power allocation, and normalize it so that the average power constraint in (15) is obeyed:

$$\mathbf{F}_{BB} = \frac{\sqrt{P_{\max}}}{\|\mathbf{F}_{RF}\mathbf{Q}\|_F}\mathbf{Q}, \quad (16)$$

where $\mathbf{Q} = \mathbf{V}\boldsymbol{\Lambda}^{1/2} \in \mathbb{C}^{N_t \times N_s}$ represents the SVD precoder with the diagonal power allocation matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{N_s \times N_s}$. In the appendix, we show that the average power constraint is always satisfied regardless of DAC quantization resolution, and that $\|\mathbf{F}_{RF}\mathbf{Q}\|_F^2 = N_t P_{\max}$, which gives $\mathbf{F}_{BB} = \frac{1}{\sqrt{N_t}}\mathbf{Q}$.

For fully-digital transmitters, Problem (15) becomes

$$\begin{aligned}\underset{\mathbf{F}_{BB}}{\text{maximize}} \quad & \log_2 \det\left(\mathbf{I}_{N_r} + \mathbf{R}_{nn}^{-1}\mathbf{H}\mathbf{F}_{BB}\mathbf{F}_{BB}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\right) \\ \text{subject to} \quad & \|\mathbf{F}_{BB}\|_F^2 \leq P_{\max}.\end{aligned}$$

The solution in this case is $\mathbf{F}_{BB} = \bar{\mathbf{V}}\bar{\boldsymbol{\Lambda}}^{-1/2}$, where $\bar{\mathbf{V}} \in \mathbb{C}^{N_t \times N_s}$ represents the right singular vector matrix of $\mathbf{H}$, and $\bar{\boldsymbol{\Lambda}} \in \mathbb{C}^{N_s \times N_s}$ the corresponding water-filling power allocation matrix.

C. Computational Complexity Analysis

The computational analysis of the proposed precoding methods is important to evaluate the impact of signal processing on the transmitter power budget with the power consumption modeling presented in Section II-D. In the following analysis, we consider that the matrix product of $\mathbf{A} \in \mathbb{C}^{M \times R}$ and $\mathbf{B} \in \mathbb{C}^{R \times N}$ requires $2MNR$ flops as in [22].

From our discussion in Section III-B, fully-digital precoding consists of SVD and water-filling power allocation. Its computational complexity is dominated by the algorithm that decomposes $\mathbf{H}$. According to [22, Figure 8.6.1], one needs $N_{\text{flops}}^{\text{SVD}} = 4P^2 Q + 22Q^3$ flops to compute the full SVD of a $(P \times Q)$-dimensional matrix with the R-SVD algorithm. Therefore, the number of flops necessary to obtain the digital precoder is $N_{\text{flops}}^{\text{D}} = 4N_r^2 N_t + 22N_t^3$.

Signal processing of hybrid precoding is carried out in two stages: baseband and analog precoding. The computational complexity of the former stage is dominated by the SVD of $\mathbf{H}_{eq}$ which demands $N_{\text{flops}}^{\text{SVD}} = 4N_r^2 L_t + 22L_t^3 + 2N_r N_t L_t$ flops. The term $2N_r N_t L_t$ in $N_{\text{flops}}^{\text{SVD}}$ refers to the matrix product that forms $\mathbf{H}_{eq}$. The computational effort of the latter stage corresponds to the number of flops demanded by the analog

---

[1]In a practical setup, $\mathbf{H}_{eq}$ could be estimated at the receiver and sent back to the transmitter via a feedback channel.

Table II: Flop counting for Algorithm 1.

Line	Flops count
1	$4N_r^2 N_t + 22N_t^3$
4	$4N_t L_t$
5	$4N_t^2 L_t + 22L_t^3$
6	$2N_t L_t^2$

Table III: Flop counting for Algorithm 2.

Line	Flops count
5	$4N_t N_a$
6	$2N_a + 1$

precoding algorithm. The flop counting for Algorithms 1 and 2 is detailed in Tables II and III, respectively. According to Table II, $N_{\text{flops}}^{\text{FPSN}} = 4N_r^2 N_t + 22N_t^3 + I(4N_t L_t + 4N_t^2 L_t + 22L_t^3 + 2N_t L_t^2)$ flops are necessary to compute the fully-connected analog precoder, where I denotes the number of iterations of the alternating minimization method. The total cost is then given by the baseband and analog precoders computation cost $N_{\text{flops}}^{\text{F}} = N_{\text{flops}}^{\text{SVD}} + N_{\text{flops}}^{\text{FPSN}}$. According to Table III, the total number of flops necessary for computing the partially-connected analog precoder is $N_{\text{flops}}^{\text{PPSN}} = J(4N_t N_a + 2N_a + 1)$, where J represents the number of iterations of the power method. The total computation cost of the partially-connected hybrid precoder accounts for the calculation of $\mathbf{H}_{\text{eq}}$ and its SVD. Therefore, it is given by $N_{\text{flops}}^{\text{P}} = N_{\text{flops}}^{\text{SVD}} + N_{\text{flops}}^{\text{PPSN}}$. Comparing $N_{\text{flops}}^{\text{F}}$ and $N_{\text{flops}}^{\text{P}}$, we observe that the latter is less complex than the former because the partially-connected structure is exploited to reduce the analog precoding computational complexity.

D. Achievable Rate Bounds

Now we conduct an asymptotic performance assessment of the quantized hybrid precoder. Let us first consider the received signal model (9) with Gaussian noise approximation and hybrid precoding, then we employ SVD combining (with respect to $\mathbf{H}_{\text{eq}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}$) at the receiver, and finally decorrelate the Gaussian noise vector $\mathbf{n}_G$ to obtain an achievable rate lower bound. From Equation (16), it follows that the rotated received signal can be written as

$$\tilde{\mathbf{y}} = \mathbf{U}^{\mathsf{H}}\mathbf{y} = \frac{1}{\sqrt{L_{\text{RF}}N_t}}\mathbf{U}^{\mathsf{H}}\mathbf{H}_{\text{eq}}\boldsymbol{\Upsilon}_b\mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{s} + \frac{1}{\sqrt{L_{\text{RF}}}}\mathbf{U}^{\mathsf{H}}\mathbf{H}_{\text{eq}}\mathbf{e} + \mathbf{U}^{\mathsf{H}}\mathbf{n}$$

$$= \sqrt{\frac{1-\rho_b}{L_{\text{RF}}N_t}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{1/2}\mathbf{s} + \tilde{\mathbf{n}} \in \mathbb{C}^{N_s},$$

where $\tilde{\mathbf{n}} = \frac{1}{\sqrt{L_{\text{RF}}}}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}\mathbf{e} + \mathbf{U}^{\mathsf{H}}\mathbf{n} \in \mathbb{C}^{N_s}$ represents the rotated total additive Gaussian noise component with covariance matrix $\mathbf{R}_{\tilde{n}\tilde{n}} = \frac{1}{L_{\text{RF}}}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}\mathbf{R}_{ee}\mathbf{V}\boldsymbol{\Sigma} + \sigma_n^2\mathbf{I}_{N_s}$. After the whitening filter, we have

$$\mathbf{R}_{\tilde{n}\tilde{n}}^{-1/2}\tilde{\mathbf{y}} = \sqrt{\frac{1-\rho_b}{L_{\text{RF}}N_t}}\mathbf{R}_{\tilde{n}\tilde{n}}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{1/2}\mathbf{s} + \mathbf{R}_{\tilde{n}\tilde{n}}^{-1/2}\tilde{\mathbf{n}}.$$

This equation provides the following lower bound for the achievable rate:

$$R \geq \log_2 \det\left(\mathbf{I}_{N_s} + \frac{1-\rho_b}{L_{\text{RF}}N_t}\mathbf{R}_{\tilde{n}\tilde{n}}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}\boldsymbol{\Sigma}\mathbf{R}_{\tilde{n}\tilde{n}}^{-1/2,\mathsf{H}}\right). \quad (17)$$

For low-resolution quantization and high SNR, water-filling power allocation yields $\boldsymbol{\Lambda} = (P_{\max}/N_s)\mathbf{I}_{N_s}$ and (17) goes to

$$R \geq \log_2 \det\left(\mathbf{I}_{N_s} + \frac{P_{\max}(1-\rho_b)}{L_{\text{RF}}N_t N_s}\mathbf{R}_{\text{high}}^{-1/2}\boldsymbol{\Sigma}^2\mathbf{R}_{\text{high}}^{-1/2,\mathsf{H}}\right), \quad (18)$$

where $\mathbf{R}_{\text{high}} = \boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}\mathbf{R}_{ee}\mathbf{V}\boldsymbol{\Sigma}$. At low SNR, $\mathbf{R}_{\tilde{n}\tilde{n}} \to \sigma_n^2\mathbf{I}_{N_s}$, and the achievable rate is lower bounded by $R \geq \log_2\left[1 + (1-\rho_b)\frac{\gamma}{L_{\text{RF}}N_t}\sigma_{\max}^2\right]$, where $\sigma_{\max}$ denotes the largest singular value of $\mathbf{H}_{\text{eq}}$ and γ the SNR. Inequality (18) shows that quantization noise dominates over AWGN in high SNR regime, leading to saturation of the system capacity. The obtained achievable rate expression at low SNR shows that quantized system capacity is always smaller than that of unquantized systems due to the factor $(1 - \rho_b)$ inside the logarithm.

IV. SIMULATION RESULTS

In this section, we present the results of numerical simulations conducted to evaluate the spectral and energy efficiencies of the proposed precoding strategies. Also, the power consumption model introduced in Section II-D is assessed. Hereafter, hybrid precoding with fully- and partially-connected PSN are referred to as HPF and HPP, respectively. The obtained results are averaged over 1000 independent experiments. For each experiment realization, a channel matrix with $L = 5$ paths is generated according to Equation (11). We assume that both transmitter and receiver have perfect and instantaneous channel state information. We set the average transmit power constraint to $P_{\max} = 1\text{W}$, which is reasonable for base stations with small coverage. Also, in order to reduce costs, the number L_t of RF chains is equal to the number N_s of streams. We assume phase-shifters with phase range of $360°$ and shift resolution of $b_{\text{PS}} = 5$ bits, as in the hardware implementation of [31]. The DAC sampling rate is set to $F_s = 1\,\text{GHz}$, which should be sufficient for mmWave systems to provide high data rates. Regarding the calculation of computational power consumption, we consider the same parameters as in [21]: computational efficiency of $E_c = 12.8$ Gflops/s/W, transmission bandwidth of $B = 20\,\text{MHz}$, channel coherence bandwidth of $B_C = 180\,\text{kHz}$, and channel coherence time of $T_C = 10\,\text{ms}$, giving $C = B/(B_C T_C) = 11111$ coherence blocks per second. The convergence tolerance value of the iterative algorithms is set to 10^{-6}. Preliminary simulations have shown that the iterative algorithms used to compute HPF and HPP converge within 37 and 11 iterations on average, respectively.

A. Spectral Efficiency

Figures 2, 3, and 4 show the average spectral efficiency of the digital and hybrid precoders as a function of the SNR for DAC resolution of 1 and 8 bits. The curves depicted in Figure 2 indicate that the quantized digital precoder does not lose much spectral efficiency at low SNR, whereas its performance saturates at high SNR. This is because quantization distortion does not drop at high SNR, as discussed in Section III-D. Since the distortion caused by 1-bit DACs is important, the spectral efficiency already saturates at 20 dB,
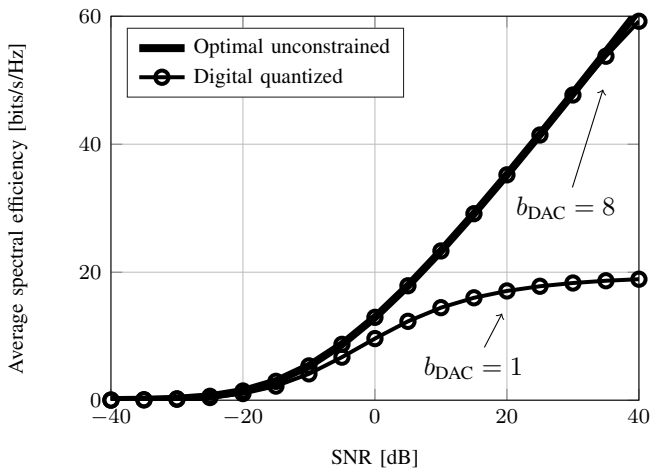
Fig. 2: Spectral efficiency of digital precoders. $N_t = 64$, $N_r = 4$, $L_t = N_s = 4$, $b_{\text{DAC}} \in \{1, 8\}$.

Fig. 3: Spectral efficiency of hybrid precoders (ignoring RF hardware losses). $N_t = 64$, $N_r = 4$, $L_t = N_s = 4$, $b_{\text{DAC}} \in \{1, 8\}$.

whereas it occurs only at $> 40\,\text{dB}$ for 8-bit DACs. The spectral efficiency of lossless hybrid precoders using active and passive phase-shifters is compared to that of the unquantized digital precoder in Figure 3. As observed in several hybrid precoding works, HPF outperforms HPP when RF hardware losses are ignored. In this case, the type of phase-shifting circuit does not have any impact on the spectral efficiency as no insertion losses are considered yet. Comparing Figures 2 and 3, one recognizes that hybrid precoding saturates at lower data rates compared to digital precoding. This is because the quantization distortion $\mathbf{e}$ is filtered by analog precoding matrix $\mathbf{F}_{\text{RF}}$, as visible in Equation (5), increasing the total noise $\mathbf{n}'$ power. In general, digital precoding provides higher data rate than hybrid precoding for fixed number of bits per DAC and SNR.

In Figure 4, the average spectral efficiency of the hybrid precoders is shown considering RF hardware losses. In this scenario, HPF performs similarly to HPP because its insertion loss is larger than that of the other architecture, reducing the effective transmit power and, consequently, the spectral efficiency. When RF losses are taken into account, the phase-shifting implementation becomes important. As shown in Figure 4, precoders with active phase-shifters are more spectral-efficient since these components do not attenuate the signals like passive phase-shifters, however they come with increased power consumption. We observe in Figure 4 SNR losses of $10\,\text{dB}$ for the active PSN and $20\,\text{dB}$ for the passive PSN when $b_{\text{DAC}} = 8$. Similar losses have been reported in other works [11], [43] for hybrid precoders with fully-connected PSN. Such important power losses occur due to the large number of lossy RF components employed in this mmWave massive MIMO setup.

B. Power Consumption

The static power consumption is plotted as a function of the number N_t of antennas for $b_{\text{DAC}} \in \{1, 3, 5, 7\}$ in the left part of Figure 5. Despite the reduced number of RF chains, active HPF can still be power hungry, even more
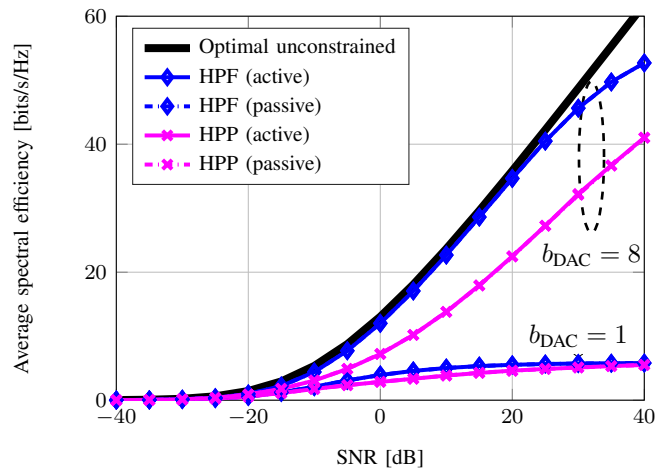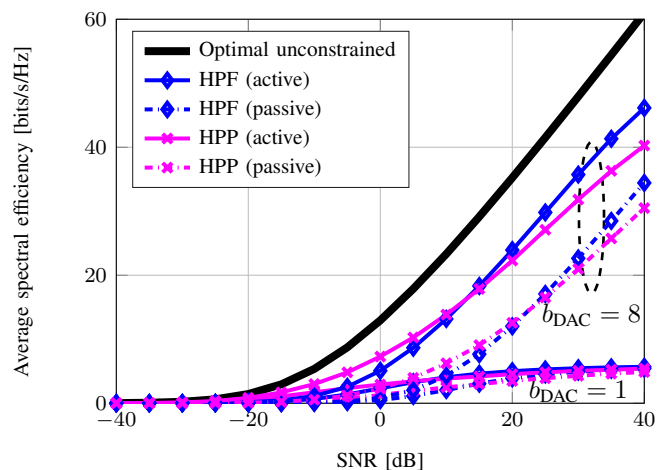

Fig. 4: Spectral efficiency of hybrid precoders (considering RF hardware losses). $N_t = 64$, $N_r = 4$, $L_t = N_s = 4$, $b_{\text{DAC}} \in \{1, 8\}$.

than the digital precoder. This is mainly due to the power consumption of the fully-connected PSN. To solve this issue, one can employ passive phase-shifters, drastically decreasing the power demand of HPF, as shown in Figure 5. However, as observed in Figure 4, the use of such passive components incurs in spectral efficiency degradation. Interestingly, passive HPP consumes more power than passive HPF. This is because the PAs in the former scheme consume more power than in the latter, as partially-connected PSNs lose less power, hence delivering more power to the amplifiers. One can also observe in Figure 5 the effect of DAC resolution on the static power consumption. Digital precoding is more sensitive to DAC resolution because it uses N_t DACs, while hybrid precoding uses only L_t.

In practice, computational power consumption plays an important role on system energy efficiency. The plot in the right part of Figure 5 suggests that it is negligible when the antenna array is comparatively small (up to 64 antennas). By
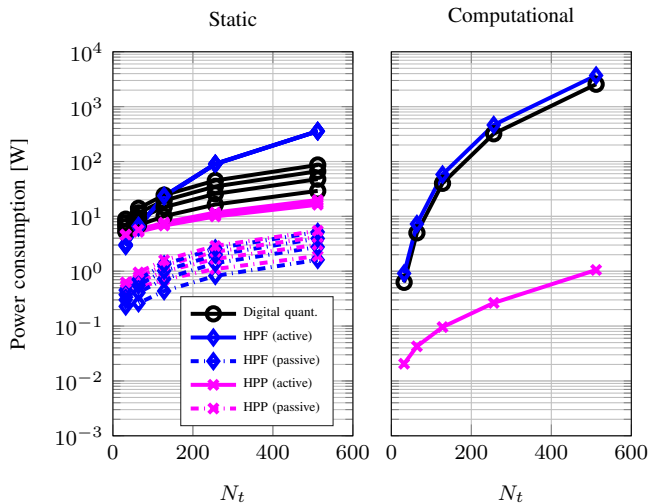
Fig. 5: $N_r = 4$, $L_t = N_s$, $b_{\text{DAC}} \in \{1, 3, 5, 7\}$, $(N_t, N_s) \in \{(32, 2), (64, 4), (128, 8), (256, 16), (512, 32)\}$.

Fig. 6: Energy-spectral efficiency curves for varying DAC resolution. SNR $= 0\,\text{dB}$, $N_t = 64$, $N_r = 4$, $L_t = N_s = 4$, $b_{\text{DAC}} \in \{1, \ldots, 8\}$.

contrast, it becomes substantial for HPF and digital precoder with growing N_t. HPP is not as power hungry since the number of flops necessary to run Algorithm 2 scales linearly with the number N_t of transmit antennas, whereas the other schemes contain quadratic and cubic terms of N_t. We stress that the proposed power consumption model mainly serves to provide energy efficiency estimates, allowing us to compare different precoding structures. In practice, optimized hardware and software designs may reduce the values provided by our model. For example, one can simplify PA, DAC, and modulation design when employing 1-bit quantization. However, considering such specialized implementations is out of the scope of this work.

C. Energy Efficiency

Numerical experiments were also carried out to study the energy efficiency of the proposed precoding strategies and its trade-off with spectral efficiency when the RF hardware losses are considered. The compromise between active and passive phase-shifters is investigated as well. Mathematically, we define the energy efficiency of a system as the ratio of its spectral efficiency to its static power consumption [13]. Figures 6 and 7 show the energy-spectral efficiency curves of different precoding schemes for varying DAC resolution at high and low SNR, respectively. A system designer should aim at maximizing both figures of merit, approaching the top right corner of the chart in Figs. 6 and 7.

In contrast to what is usually claimed by several hybrid precoding papers, Figures 6 and 7 indicate that HPF is not necessarily an efficient scheme. This is because the combination of large insertion losses due to phase-shifters and power combiners with the increased power consumption of fully-connected PSN degrades the energy efficiency. Such insertion losses are disregarded in most hybrid precoding works, leading to the erroneous conclusion that fully-connected PSN is the most spectral-efficient hybrid topology in general. As an alternative to HPF, transmitters equipped with partially-connected
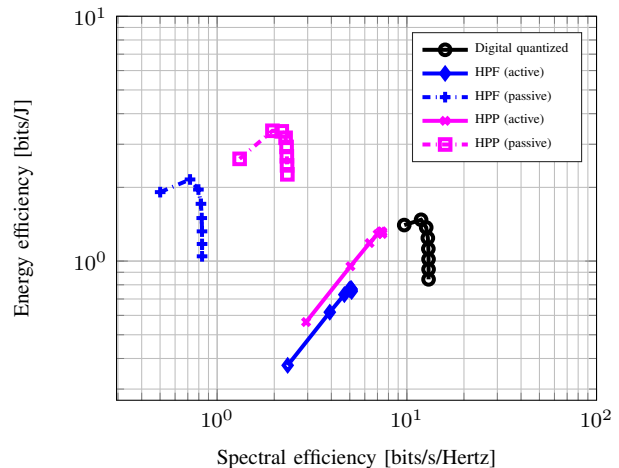
PSNs lose less power by employing a reduced number of phase-shifters and no power combiners at all. Therefore, as observed in Figures 6 and 7, HPP can be more energy- and spectral-efficient than HPF. These figures also reveal that phase-shifter topology offers an energy-spectral efficiency trade-off. As active phase-shifting amplifies the shifted signals, it favors spectral efficiency. On the other hand, passive phase-shifting offers negligible power consumption, promoting energy efficiency. Among the studied precoding methods, digital precoding is the most spectral-efficient solution for obvious reasons. Figure 7 indicates that digital precoding is as energy-efficient as HPP at low SNR. In this scenario, digital precoding compensates for the important power consumption by high data throughput, thus increasing energy efficiency. Regarding DAC resolution, Figures 6 and 7 also show that using $b_{\text{DAC}} > 3$ does not necessarily ameliorate energy efficiency. In fact, passive HPF and HPP and digital precoding become less efficient for higher bit resolution.

Do the obtained energy efficiency results hold for other scenarios? For example, do they still hold for antenna arrays with more elements? To answer this question, consider Figures 8 and 9, where the energy efficiency is plotted for different (N_t, N_s) tuples at high and low SNR, respectively. These plots were obtained by setting $b_{\text{DAC}} = 3$, motivated by the discussion above. In general, HPP is the most energy-efficient scheme, followed by digital precoding, and HPF at both high and low SNR. For $N_t = 32$ antennas and $0\,\text{dB}$ SNR, passive HPF becomes the most efficient solution as it provides enough data throughput and its PSN does not dissipate too much power due to the comparatively small number of phase-shifting elements. As the number N_t of antennas increase, however, its energy efficiency quickly drops. These results also suggest that asymptotically expanding the transmit array size does not improve energy efficiency, as static power consumption becomes asymptotically large with N_t for fixed SNR, while spectral efficiency does not grow as fast.
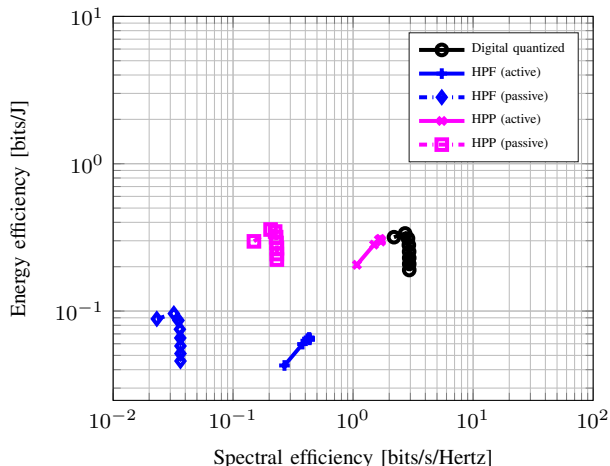
Fig. 7: Energy-spectral efficiency curves for varying DAC resolution. SNR $= -15$ dB, $N_t = 64$, $N_r = 4$, $L_t = N_s = 4$, $b_{\text{DAC}} \in \{1, \ldots, 8\}$.
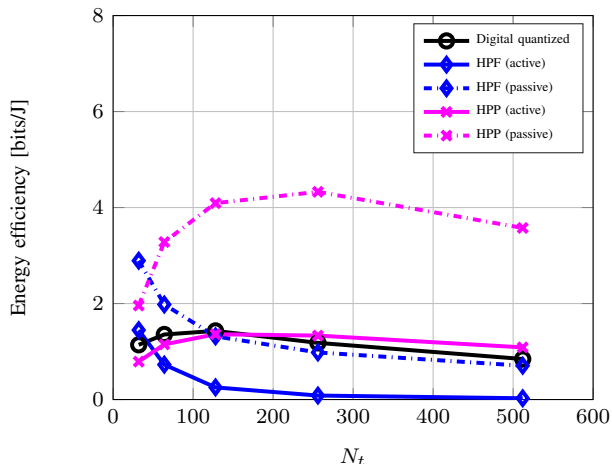

Fig. 8: SNR $= 0$ dB, $b_{\text{DAC}} = 3$, $N_r = L_t = N_s$, and $(N_t, N_s) \in \{(32, 2), (64, 4), (128, 8), (256, 16), (512, 32)\}$.
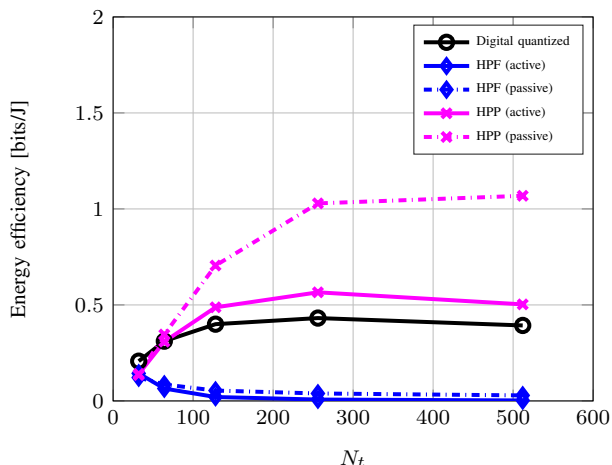

Fig. 9: SNR $= -15$ dB, $b_{\text{DAC}} = 3$, $N_r = L_t = N_s$, and $(N_t, N_s) \in \{(32, 2), (64, 4), (128, 8), (256, 16), (512, 32)\}$.

V. Conclusion

MmWave Massive MIMO systems using hybrid precoding with fully-connected PSN exhibit important power losses mostly due to the large number of phase-shifters, power dividers and combiners. Such losses are not easily compensated for with present mmWave PA technology, causing severe spectral and energy efficiency degradation. As an alternative, hybrid precoders equipped with partially-connected PSN are shown to be energy-efficient, as they consume less power and exhibit limited insertion loss. Our results revealed that phase-shifting topology offers an energy-spectral efficiency trade-off. Active phase-shifters favor higher data throughput, while passive elements aim at energy efficiency. Moreover, we observed that DACs with more than 3 bits of resolution do not significantly improve spectral efficiency, and may lead to energy efficiency degradation. Finally, the proposed power consumption model suggests that computational power spending should be considered in the power budget of very large array systems. The present work considered a point-to-point MIMO system for simplicity reasons. Introducing interfering users, performing transceiver optimization, and considering imperfect channel state information are envisioned as future work.

Appendix

We demonstrate that the power of the lossless transmitted signal with DAC approximation (5) follows the average power constraint for hybrid precoding. This power is defined as $P_{\tilde{x}} = \mathbb{E}\left[\|\tilde{\mathbf{x}}\|_2^2\right] = \text{tr}(\mathbf{F}_{\text{RF}} \boldsymbol{\Upsilon}_b \mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^{\mathsf{H}} \boldsymbol{\Upsilon}_b \mathbf{F}_{\text{RF}}^{\mathsf{H}}) + \text{tr}(\mathbf{F}_{\text{RF}} \mathbf{R}_{ee} \mathbf{F}_{\text{RF}}^{\mathsf{H}})$. From Equation (6), it follows that $P_{\tilde{x}}$ can be rewritten as

$$P_{\tilde{x}} \approx (1 - \rho_b) \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_{\text{F}}^2 + \rho_b \, \text{tr}(\mathbf{F}_{\text{RF}} \, \text{diag}(\mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^{\mathsf{H}}) \mathbf{F}_{\text{RF}}^{\mathsf{H}}).$$

We need to calculate the diagonal matrix $\text{diag}(\mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^{\mathsf{H}})$. To do so, consider $\mathbf{A} = \theta \mathbf{I}_M$. One can readily see that $\text{diag}(\mathbf{A}) = \text{Diag}([\text{tr}(\mathbf{A})/M, \ldots, \text{tr}(\mathbf{A})/M])$. Although the diagonal elements of $\mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^{\mathsf{H}}$ are not necessarily equal, they should be similar due to power allocation, yielding the following approximation $\text{diag}(\mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^{\mathsf{H}}) \approx \text{Diag}(\text{tr}(\mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^{\mathsf{H}})/L_t, \ldots, \text{tr}(\mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^{\mathsf{H}})/L_t) \approx \|\mathbf{F}_{\text{BB}}\|_{\text{F}}^2/L_t$. This estimate gives

$$P_{\tilde{x}} \approx (1 - \rho_b) \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_{\text{F}}^2 + \rho_b \frac{\|\mathbf{F}_{\text{BB}}\|_{\text{F}}^2 \|\mathbf{F}_{\text{RF}}\|_{\text{F}}^2}{L_t}.$$

The baseband precoder normalization (16) leads to $\|\mathbf{F}_{\text{BB}}\|_{\text{F}}^2 = \frac{P_{\max}}{\|\mathbf{F}_{\text{RF}} \mathbf{Q}\|_{\text{F}}^2} \text{tr}(\mathbf{Q}\mathbf{Q}^{\mathsf{H}})$. From water-filling power allocation, $\text{tr}(\mathbf{Q}\mathbf{Q}^{\mathsf{H}}) = P_{\max}$, and the denominator can be written as $\|\mathbf{F}_{\text{RF}} \mathbf{Q}\|_{\text{F}}^2 = \text{tr}(\mathbf{F}_{\text{RF}} \mathbf{V} \boldsymbol{\Lambda} \mathbf{V}^{\mathsf{H}} \mathbf{F}_{\text{RF}}^{\mathsf{H}})$. Notice that we can approximate $\mathbf{Q}\mathbf{Q}^{\mathsf{H}} \approx \frac{P_{\max}}{L_t} \mathbf{I}_{L_t}$, therefore

$$\|\mathbf{F}_{\text{RF}} \mathbf{Q}\|_{\text{F}}^2 \approx \frac{P_{\max}}{L_t} \text{tr}(\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{RF}}^{\mathsf{H}}) = N_t P_{\max}.$$

We finally get $\|\mathbf{F}_{\text{BB}}\|_{\text{F}}^2 = P_{\max}/N_t$ and $P_{\tilde{x}} \approx [(1 - \rho_b) + \rho_b] P_{\max} = P_{\max}$. This result can be easily extended to digital precoding by making $\mathbf{F}_{\text{BB}} \in \mathbb{C}^{N_t \times N_s}$ and $\mathbf{F}_{\text{RF}} = \mathbf{I}_{N_t}$.

ACKNOWLEDGMENTS

REFERENCES

[1] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[2] 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 Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] A. B. Constantine, *Antenna Theory: Analysis and Design*, 3rd ed. Wiley, 2005.

[4] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[6] S. Schwarz and M. Rupp, "Society in motion: challenges for LTE and beyond mobile communications," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 76–83, May 2016.

[7] R. W. Heath, "Millimeter wave: the future of commercial wireless systems," in *Proc. 2016 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, Austin, TX, USA, Oct. 2016, pp. 1–4.

[8] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[9] R. Méndez-Rial, C. Rusu, N. Gonzlez-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, 2016.

[10] C. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten, "On the energy-efficiency of hybrid analog-digital transceivers for single- and multi-carrier large antenna array systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1980–1995, Sep. 2017.

[11] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, "Hybrid analog-digital precoding revisited under realistic RF modeling," *IEEE Wireless Communications Letters*, vol. 5, no. 5, pp. 528–531, Oct. 2016.

[12] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *Proc. 2015 Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, Feb. 2015, pp. 191–198.

[13] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter wave receiver efficiency: a comprehensive comparison of beamforming schemes with low resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8131–8146, Dec. 2017.

[14] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 188–209, Jan. 2005.

[15] K. Roth and J. A. Nossek, "Achievable rate and energy efficiency of hybrid and digital beamforming receivers with low resolution ADC," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2056–2068, Sep. 2017.

[16] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[17] L. N. Ribeiro, S. Schwarz, M. Rupp, A. L. F. de Almeida, and J. C. M. Mota, "A Low-Complexity Equalizer for Massive MIMO Systems Based on Array Separability," in *Proc. 25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, Aug. 2017, pp. 2522–2526.

[18] A. Mezghani, R. Ghiat, and J. A. Nossek, "Transmit processing with low resolution D/A-converters," in *Proc. 16th IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, Yasmine Hammamet, Tunisia, Dec. 2009, pp. 683–686.

[19] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Research Laboratory of Electronics, Massachusetts Institute of Technology, Tech. Rep. 216, Mar. 1952. [Online]. Available: https://dspace.mit.edu/handle/1721.1/4847

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

[21] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.

[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. JHU Press, 2012.

[23] F. Sohrabi and W. Yu, "Hybrid beamforming with finite-resolution phase shifters for large-scale MIMO systems," in *Proc. IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Stockholm, Sweden, Jul. 2015, pp. 136–140.

[24] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory*, vol. 6, no. 1, pp. 7–12, Mar. 1960.

[25] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proceedings of the IEEE*, vol. 73, no. 11, pp. 1551–1588, Nov. 1985.

[26] A. Mezghani and J. A. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *Proc. IEEE International Symposium on Information Theory Proceedings (ISIT)*, Cambridge, MA, USA, Jul. 2012.

[27] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.

[28] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[29] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "mmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 13–21, Oct. 2015.

[30] E. Zöchmann, M. Lerch, S. Caban, C. F. Mecklenbräuker, and M. Rupp, "Directional evaluation of receive power, Rician K-factor and RMS delay spread obtained from power measurements of 60 GHz indoor channels," in *Proc. IEEE APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC)*, Cairns, Australia, 2016, pp. 246–249.

[31] D. Huang, L. Zhang, L. Zhang, and Y. Wang, "A 60 GHz 360° 5-bit digitally controlled high-pass/direct-pass T-type phase shifter using nMOS body-floating switch," *Analog Integrated Circuits and Signal Processing*, vol. 90, no. 1, pp. 93–100, Jan. 2017.

[32] K. S. V. Prasad, E. Hossain, and V. K. Bhargava, "Energy efficiency in massive MIMO-based 5G networks: Opportunities and challenges," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 86–94, Jun. 2017.

[33] C. Marcu, "LO Generation and Distribution for 60 GHz Phased Array Transceivers," Ph.D. dissertation, EECS Department, University of California, Berkeley, 2011. [Online]. Available: http://www2.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-132.html

[34] Y. Jin, J. R. Long, and M. Spirito, "A 7 dB NF 60 GHz-band millimeter-wave transconductance mixer," in *Proc. 2011 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, Baltimore, MD, USA, Jun. 2011, pp. 1–4.

[35] S. Rangan, T. Rappaport, E. Erkip, Z. Latinovic, M. R. Akdeniz, and Y. Liu, "Energy efficient methods for millimeter wave picocellular systems," in *Proc. IEEE Communications Theory Workshop*, Phuket, Thailand, Jun. 2013, pp. 1–25.

[36] K. Scheir, S. Bronckers, J. Borremans, P. Wambacq, and Y. Rolain, "A 52 GHz phased-array receiver front-end in 90 nm digital CMOS," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 12, pp. 2651–2659, Dec. 2008.

[37] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.

[38] E. Olieman, "Time-interleaved high-speed D/A converters," Ph.D. dissertation, University of Twente, 2016. [Online]. Available: http://doc.utwente.nl/99674/

[39] K. Gao, N. J. Estes, B. Hochwald, J. Chisum, and J. N. Laneman, "Power-performance analysis of a simple one-bit transceiver," in *Proc. 2017 Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, Feb. 2017, pp. 1–10.

[40] K. Greene, A. Sarkar, and B. Floyd, "A 60-GHz Dual-Vector Doherty Beamformer," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 5, pp. 1373–1387, May 2017.

[41] D. Pepe and D. Zito, "A 78.8-92.8 GHz 4-bit 0-360° active phase shifter in 28nm FDSOI CMOS with 2.3 dB average peak gain," in *Proc. 41st European Solid-State Circuits Conference (ESSCIRC)*, Graz, Austria, Sep. 2015, pp. 64–67.

[42] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. Cambridge University press, 2007.

[43] E. Zöchmann, S. Schwarz, and M. Rupp, "Comparing antenna selection and hybrid precoding for millimeter wave wireless communications," in *Proc. 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Rio de Janeiro, Brazil, Jul. 2016, pp. 1–5.

Lucas N. Ribeiro received his Bachelor degree in Teleinformatics Engineering from the Federal University of Ceará, Brazil, in 2014, and his Master degree in Informatics from the University of Nice Sophia Antipolis, France, in 2015. He is currently pursuing the Ph.D. degree in Teleinformatics Engineering in the Wireless Telecommunications Research Group, Federal University of Ceará. In 2017, he was a visiting researcher at the Institute of Telecommunications of Technische Universität Wien.

Stefan Schwarz received his Dr. techn. degree in telecommunications engineering in 2013 at Technische Universität (TU) Wien. In 2010 he received the honorary price of the Austrian Minister of Science and Research and in 2014 he received the INiTS ICT award. From 2008 to 2014 he was working at the Institute of Telecommunications (ITC) of TU Wien. Since 2015 he is heading the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion.

Markus Rupp received his Dr.-Ing. degree in 1993 at Technische Universität Darmstadt. From 1993 until 1995, he had a postdoctoral position at the University of Santa Barbara, California. From October 1995 until August 2001 he was a member of Technical Staff in the Wireless Technology Research Department of Bell-Labs at Crawford Hill, NJ. Since October 2001 he is a full professor for Digital Signal Processing in Mobile Communications at Technische Universität Wien.

André L. F. de Almeida received the B.Sc. and M.Sc. degrees in Electrical Engineering from the Federal University of Ceará, Brazil, in 2001 and 2003, respectively, and the double Ph.D. degree in Sciences and Teleinformatics Engineering from the University of Nice Sophia Antipolis, France, and the Federal University of Ceará, Brazil, in 2007. He is currently an Associate Professor with the Department of Teleinformatics Engineering of the Federal University of Ceará.