IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 16, 2018, accepted September 25, 2018, date of publication October 5, 2018, date of current version October 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2874005

Hybrid Precoding With Time-Modulated Arrays for Mmwave MIMO Systems

JOSÉ P. GONZÁLEZ-COMA[1], (Member, IEEE), ROBERTO MANEIRO-CATOIRA[2], (Member, IEEE), AND LUIS CASTEDO[1], (Senior Member, IEEE)

[1]CITIC, GTEC, Department of Computer Engineering, Faculty of Informática, University of A Coruña, 15071 A Coruña, Spain
[2]GTEC, Department of Computer Engineering, Faculty of Informática, University of A Coruña, 15071 A Coruña, Spain

Corresponding author: José P. González-Coma (jose.gcoma@udc.es)

This work was supported in part by the Xunta de Galicia under Grant ED431C 2016-045, Grant ED341D R2016/012, and Grant ED431G/01, in part by the AEI of Spain under Grant TEC2015-69648-REDC and Grant TEC2016-75067-C4-1-R, and in part by ERDF funds under Grant AEI/FEDER, EU.

ABSTRACT We consider the utilization of time-modulated arrays (TMAs) as a simple and cost-effective approach to hybrid digital–analog precoding in millimeter wave (mmWave) multiple-input multiple-output (MIMO) systems. Instead of conventional variable phase shifters (VPSs), our proposed TMA hybrid precoders use radio frequency (RF) switches followed by 1-bit VPSs in the analog part. We study the insertion losses at mmWave of a TMA analog precoding network in terms of sideband radiation and hardware efficiency. In addition, we present different algorithms for the design of both the baseband and RF parts of a TMA hybrid precoder. The proposed methods exhibit different tradeoffs between performance, complexity, and power efficiency. Finally, TMA precoders are compared to those implemented with conventional VPSs in terms of insertion losses, chip area, and cost, concluding that precoding with TMAs is a competitive solution for mmWave MIMO systems.

INDEX TERMS Hybrid precoding, time-modulated arrays, mmWave, MIMO, RF precoding network.

I. INTRODUCTION

Millimeter Wave Multiple-Input Multiple-Output (MIMO) systems typically use large-scale antenna arrays to achieve high antenna gains to compensate the significant propagation losses at these bands. In such systems, precoding is a fundamental operation responsible for obtaining the radiated signals from the multiple input data streams. Precoding in millimeter wave (mmWave) MIMO is a difficult task, unavoidably constrained by the trade-off between performance, mobility and sustainability [1].

Fully digital precoding offers an unbeatable performance but at the expense of an extremely high complexity, cost, and power consumption, as a result of requiring a complete Radio Frequency (RF) chain per antenna. Digital precoding is usually implemented at baseband. Alternatively, it is possible to perform precoding at RF with a circuit network usually implemented with Variable Phase Shifters (VPSs) and adders. RF precoders are inherently fully analog and hence are also referred to as Analog Precoding Networks (APNs). APNs have the clear advantage of being feasible with a single RF chain per input data stream [2]. On the contrary, the performance of APNs is usually improvable because they limit to control the phase of the connections between the antenna array elements and the RF chains.

Hybrid digital-analog precoding is an alternative solution that combines in a single scheme the flexibility of digital baseband precoding and the hardware simplicity of RF analog precoding. Hybrid precoding schemes use a limited number of RF chains and distribute the MIMO processing between the analog and digital domains. Due to its numerous advantages in mmWave MIMO systems, hybrid precoding has been the subject of intense research activities over the past years (see [1]–[6] and references therein).

Most hybrid precoding schemes so far proposed use VPSs in the analog domain [1]. VPSs, however, are subject to severe flexibility constraints like phase quantization or insertion losses [3]. Moreover, insertion losses increase linearly with the carrier frequency and hence can be very large in mmWave [7]. Furthermore, higher losses in these devices will require higher gains at low-noise and power amplifiers resulting, at the end, not only in higher power consumption but also in larger chip areas [3], [8]. Notice also that

the number of devices employed in the implementation of APNs in mmWave systems is large, and the cost of such devices increases with the carrier frequency. To overcome these limitations, two-state switches can be used to select a subset of active antennas and hence alleviate insertion losses and quantization errors. Nevertheless, the performance of this approach is poor compared to that obtained with conventional VPSs [3].

In this work, we focus on the utilization of Time-Modulated Arrays (TMAs) for mmWave hybrid digital-analog MIMO precoding. Rather than conventional VPSs, TMAs use periodically controlled RF switches in the analog domain. The ability of TMAs to perform beamforming has already been analyzed in [9]. A remarkable feature of TMA switches, which highlights its suitability for mmWave systems, is that their performance and insertion losses are independent of the carrier frequency. To provide further flexibility, our proposed TMA hybrid precoder includes two-state or 1-bit VPSs also in the analog domain. The resulting scheme costs are much lower than available VPSs for mmWave.

TMA precoding is modeled as a time varying matrix whose entries change according to the periodic pulses of the RF switches. This causes Sideband Radiation (SR), i.e., the radiation of replicas of the transmitted signal at the TMA harmonic frequencies. Along this work, SR will be considered a harmful effect to be minimized [9]–[11]. SR losses will be characterized by the TMA power efficiency defined as the ratio between the power radiated at the carrier frequency and the total radiated power [9], [10], [12].

In addition, this work also presents a detailed analysis of the hardware efficiency, i.e., the insertion losses corresponding to the particular hardware elements involved in the TMA APN implementation [13], [14]. Furthermore, a comparative study —in terms of power efficiency, chip area and cost— between the hardware technologies of APNs with conventional VPSs and those based on TMAs is also considered.

Similarly to [2] and [15]–[17], the achievable rate will be the reference metric to design TMA hybrid precoders. The maximization of this metric, subject to the total power constraint including the SR, leads to an involved optimization problem because the precoder response at the TMA harmonic replicas is determined by the analog precoding matrix at the fundamental frequency. To overcome this limitation, we first formulate the optimization problem subject to the regular power and hardware constraints, i.e., we only consider the transmission over the fundamental frequency and ignore the SR. In a second step, we investigate techniques to increase the power efficiency.

Finally, we provide numerical results to illustrate the performance of the TMA hybrid precoder designs and evaluate the power efficiency of the APN. We compare these results with others obtained with previous solutions based on VPSs to show the feasibility of TMA techniques to perform hybrid precoding in mmWave MIMO systems. We observe larger power efficiencies than those obtained with conventional VPS hybrid precoders, while similar performance with significantly higher convergence speed is achieved with the proposed algorithms for TMA hybrid precoding.

The main contributions of this work are:

- The modeling of TMA hybrid digital-analog precoders constructed with periodically governed switching devices and 1-bit VPSs.
- The characterization of the SR and power absorption losses due to TMA precoding.
- The comparison of power losses, together with other key performance indicators such as size and cost, with those of conventional RF precoders based on VPSs.
- The study of several strategies to design TMA hybrid precoders with different trade-offs among complexity, performance, and power efficiency.
- The performance analysis of TMA hybrid precoders in terms of the usual metrics, as well as the comparison with existing schemes using VPSs.

The rest of this article is organized as follows. Sec. II presents TMA hybrid precoding and its corresponding system model. Sec. III characterizes the power efficiency of TMA hybrid precoders, distinguishing between hardware and TMA efficiencies. Sec. IV makes a hardware comparison between using TMAs and conventional VPSs in hybrid precoding. Sec. V focuses on the design of TMA hybrid precoders. Numerical results are shown in Sec. VI and conclusions are presented in Sec. VII.

II. TMA HYBRID PRECODING

We consider the utilization of an array of N_T antennas to transmit N_s streams of data, typically digital modulated signals, by means of precoding. Let $\boldsymbol{s} \in \mathbb{C}^{N_s}$ and $\boldsymbol{x} \in \mathbb{C}^{N_T}$ be the baseband representation of the input data streams and output radiated signals, respectively. MIMO transmitters equipped with a dedicated RF chain per antenna, allow for a fully digital baseband implementation of precoding. Linear precoding in such a case, is adequately modeled as $\boldsymbol{x} = \boldsymbol{P}_{\mathrm{D}}\boldsymbol{s}$, where $\boldsymbol{P}_{\mathrm{D}} \in \mathbb{C}^{N_T \times N_s}$ is the baseband representation of the fully digital linear precoding matrix. As mentioned before, this MIMO transmit architecture is not practical for mmWave and for this reason we consider the alternative hybrid digital-analog architecture in Fig. 1. Notice the APN of such architecture is constructed with TMA switches and 1-bit VPSs. We assume that the N_s data streams are transmitted synchronously using the same digital linear modulation format. The complex baseband representation of the precoder input is

$$\boldsymbol{s}(t) = \sum_{k=-\infty}^{\infty} \boldsymbol{s}_k q_T(t - kT_{\mathrm{s}}), \tag{1}$$

where $q_T(t)$ is the transmit pulse waveform, T_{s} is the transmit symbol period, and $\boldsymbol{s}_k \in \mathbb{C}^{N_s}$ is the k-th symbol vector transmitted at time kT_{s}. This signal goes across the baseband linear precoder $\boldsymbol{P}_{\mathrm{BB}} \in \mathbb{C}^{L_{\mathrm{BS}} \times N_s}$ to produce the output signal $\boldsymbol{u}(t) = \boldsymbol{P}_{\mathrm{BB}}\boldsymbol{s}(t) \in \mathbb{C}^{L_{\mathrm{BS}}}$ where $N_{\mathrm{s}} \leq L_{\mathrm{BS}} \leq N_{\mathrm{T}}$ is the number of available RF chains. Then, $\boldsymbol{u}(t)$ enters the RF chains. We will assume ideal RF chains that do not distort the

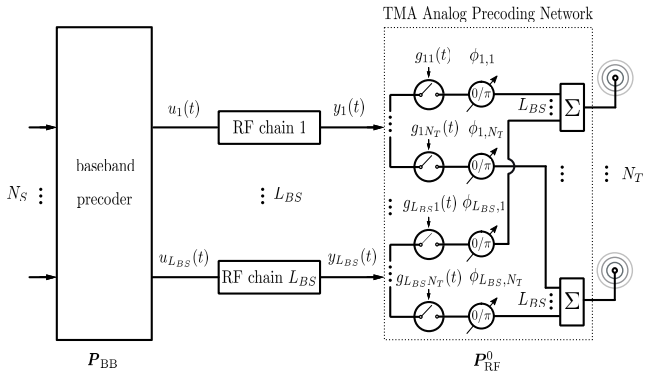

FIGURE 1. Block diagram of a hybrid digital-analog precoder with TMAs and 1-bit VPSs.

input signal and produce the passband signal $\boldsymbol{y}(t) = \boldsymbol{u}(t)\mathrm{e}^{\mathrm{j}\omega_c t}$ with ω_c the RF carrier frequency.

The RF signal $\boldsymbol{y}(t)$ next goes into the analog part of our proposed precoding scheme, as shown in Fig. 1. Due to the utilization of TMAs, the analog precoding stage is modeled by the time-varying RF analog precoding matrix $\boldsymbol{P}_{\mathrm{RF}}(t) \in \mathbb{R}^{N_{\mathrm{T}} \times L_{\mathrm{BS}}}$ and produces the radiated signal

$$\boldsymbol{x}(t) = \boldsymbol{P}_{\mathrm{RF}}(t)\boldsymbol{y}(t). \tag{2}$$

The RF analog precoding matrix has the following structure

$$\boldsymbol{P}_{\mathrm{RF}}(t) = \begin{bmatrix} g_{1,1}(t)\mathrm{e}^{\mathrm{j}\phi_{1,1}} & \cdots & g_{1,L_{\mathrm{BS}}}(t)\mathrm{e}^{\mathrm{j}\phi_{1,L_{\mathrm{BS}}}} \\ \vdots & \ddots & \vdots \\ g_{N_{\mathrm{T}},1}(t)\mathrm{e}^{\mathrm{j}\phi_{N_{\mathrm{T}},1}} & \cdots & g_{N_{\mathrm{T}},L_{\mathrm{BS}}}(t)\mathrm{e}^{\mathrm{j}\phi_{N_{\mathrm{T}},L_{\mathrm{BS}}}} \end{bmatrix}. \tag{3}$$

Each matrix entry in $[\boldsymbol{P}_{\mathrm{RF}}(t)]_{m,n} = g_{n,m}(t)e^{j\phi_{n,m}}$, $n \in \{1, \cdots, N_T\}$ and $m \in \{1, \cdots, L_{\mathrm{BS}}\}$, is the product of two factors. The first one is the real-valued periodic rectangular signal $g_{n,m}(t)$ that models the TMA switch connecting the m-th RF chain with the n-th antenna. $g_{n,m}(t)$ can be interpreted as the periodic continuation, with fundamental period T_0, of the aperiodic rectangular pulse

$$\mathrm{rect}(t/\tau_{n,m}) = \begin{cases} 1 & t \in (-\tau_{n,m}/2, \tau_{n,m}/2) \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The exponential Fourier series representation of $g_{n,m}(t)$ is

$$g_{n,m}(t) = \sum_{q=-\infty}^{\infty} G^q_{n,m}\mathrm{e}^{\mathrm{j}q\omega_0 t}, \tag{5}$$

with $\omega_0 = 2\pi/T_0$ the TMA fundamental frequency and $G^q_{n,m}$ the following exponential Fourier series coefficients

$$G^q_{n,m} = \frac{1}{T_0}G_{n,m}(q/T_0) = \xi_{n,m}\,\mathrm{sinc}(\pi q\xi_{n,m}), \tag{6}$$

where $G_{n,m}(\omega) = \tau_{n,m}\mathrm{sinc}\left(\frac{\omega\tau_{n,m}}{2}\right)$ is the Fourier Transform of $\mathrm{rect}(t/\tau_{n,m})$ and $\xi_{n,m} = \tau_{n,m}/T_0$ are the normalized TMA pulse durations. Notice that $\xi_{n,m} \in [0, 1]$ are positive real numbers less than one. As explained later, this places a severe limitation on the proposed analog precoder that can be relaxed by introducing the 1-bit VPSs $\mathrm{e}^{\mathrm{j}\phi_{n,m}}$ with $\phi_{n,m} \in \{0, \pi\}$ (see Fig. 1). Thanks to these 1-bit VPSs, the terms $\xi_{n,m}\mathrm{e}^{\mathrm{j}\phi_{n,m}} \in [-1, 1]$, i.e., they can be either positive or negative.

Recall that all entries in $\boldsymbol{P}_{\mathrm{RF}}(t)$ are periodic signals with fundamental frequency ω_0 and thus can be represented by its exponential Fourier series expansion

$$\boldsymbol{P}_{\mathrm{RF}}(t) = \sum_{q=-\infty}^{\infty} \boldsymbol{P}^q_{\mathrm{RF}}\mathrm{e}^{\mathrm{j}q\omega_0 t}, \tag{7}$$

where $\boldsymbol{P}^q_{\mathrm{RF}}$ are the exponential Fourier series coefficients matrices with entries $\left[\boldsymbol{P}^q_{\mathrm{RF}}\right]_{n,m} = G^q_{n,m}\mathrm{e}^{\mathrm{j}\phi_{n,m}}$.

Elaborating the signal model from (2) and (7), the hybrid precoding output signal can be represented as

$$\begin{aligned} \boldsymbol{x}(t) &= \boldsymbol{P}_{\mathrm{RF}}(t)\boldsymbol{P}_{\mathrm{BB}}\boldsymbol{s}(t)\mathrm{e}^{\mathrm{j}\omega_c t} \\ &= \sum_{q=-\infty}^{\infty} \boldsymbol{P}^q_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}}\boldsymbol{s}(t)e^{j(\omega_c+q\omega_0)t}. \end{aligned} \tag{8}$$

This means that the proposed hybrid TMA precoder inherently generates replicas of the input signals $\boldsymbol{s}(t)$ at the harmonic frequencies $\omega_c + q\omega_0$. The precoding scheme of Fig. 1 has a multiple frequency response and provides a precoding matrix $\boldsymbol{P}^q_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}}$ per frequency $\omega_c + q\omega_0$. This is the key differentiating feature of TMA precoding.

The signal $\boldsymbol{x}(t)$ will be next radiated through the antenna array. Let $\boldsymbol{a}_{\mathrm{T}}(\theta^{\mathrm{T}}) \in \mathbb{C}^{N_{\mathrm{T}}}$ be the transmit steering vector at the Angle of Departure (AoD) θ^{T}. The total radiated signal over θ^{T} is the sum of the radiated signals over all harmonic frequencies

$$\begin{aligned} s^{\mathrm{rad}}_{\mathrm{tot}}(\theta^{\mathrm{T}}, t) &= \boldsymbol{a}^{\mathrm{H}}_{\mathrm{T}}(\theta^{\mathrm{T}})\boldsymbol{x}(t) \\ &= \boldsymbol{a}^{\mathrm{H}}_{\mathrm{T}}(\theta^{\mathrm{T}})\sum_{q=-\infty}^{\infty} \boldsymbol{P}^q_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}}\boldsymbol{s}(t)\mathrm{e}^{\mathrm{j}(\omega_c+q\omega_0)t}. \end{aligned} \tag{9}$$

Further elaborating the signal model, the m-th component in the RF chain output signal $\boldsymbol{y}(t)$ can be represented as $y_m(t) = \boldsymbol{p}^{\mathrm{T}}_{\mathrm{BB},m}\boldsymbol{s}(t)\mathrm{e}^{\mathrm{j}\omega_c t}$, $m \in \{1, \cdots, L_{\mathrm{BS}}\}$, with $\boldsymbol{p}^{\mathrm{T}}_{\mathrm{BB},m} = [\boldsymbol{P}_{\mathrm{BB}}]_{m,:}$ denoting the m-th row of $\boldsymbol{P}_{\mathrm{BB}}$. Thereafter, the RF $y_m(t)$ is precoded in the TMA APN with the m-th column of $\boldsymbol{P}_{\mathrm{RF}}(t)$, i.e. $\boldsymbol{p}_{\mathrm{RF},m}(t) = [\boldsymbol{P}_{\mathrm{RF}}(t)]$:, m, to obtain $\boldsymbol{x}_m(t) = \boldsymbol{p}_{\mathrm{RF},m}(t)y_m(t)$. Finally, the APN output signal is $\boldsymbol{x}(t) = \sum_{m=1}^{L_{\mathrm{BS}}} \boldsymbol{x}_m(t)$, i.e., the sum of the L_{BS} analog precoding output signals per-RF chain. This signal model is helpful to interpret TMA analog precoding as a generalization of [10], where authors considered TMA analog precoding at transmission with a single data stream. Indeed, the total radiated signal (9) can be rewritten as

$$\begin{aligned} s^{\mathrm{rad}}_{\mathrm{tot}}(\theta^{\mathrm{T}}, t) &= \boldsymbol{a}^{\mathrm{H}}_{\mathrm{T}}(\theta^{\mathrm{T}})\sum_{m=1}^{L_{\mathrm{BS}}} \boldsymbol{x}_m(t) \\ &= \boldsymbol{a}^{\mathrm{H}}_{\mathrm{T}}(\theta^{\mathrm{T}})\sum_{m=1}^{L_{\mathrm{BS}}} \boldsymbol{p}_{\mathrm{RF},m}(t)y_m(t) \\ &= \boldsymbol{a}^{\mathrm{H}}_{\mathrm{T}}(\theta^{\mathrm{T}}_p)\sum_{m=1}^{L_{\mathrm{BS}}}\sum_{q=-\infty}^{\infty} \boldsymbol{p}^q_{\mathrm{RF},m}\boldsymbol{p}^{\mathrm{T}}_{\mathrm{BB},m}\boldsymbol{s}(t)\mathrm{e}^{\mathrm{j}(\omega_c+q\omega_0)t}, \end{aligned} \tag{10}$$

where $\boldsymbol{p}^q_{\mathrm{RF},m}$ are the exponential Fourier series coefficients of $\boldsymbol{p}_{\mathrm{RF},m}(t)$.

A. MIMO CHANNEL MODEL

We will assume that another antenna array of N_{R} elements is used at reception. The radiated signal $\boldsymbol{x}(t)$ is thus the input to a MIMO channel with matrix response $\boldsymbol{H} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$. We assume block fading, i.e., the MIMO channel response remains constant during the transmission of a block of symbols. We will also assume that $\omega_0 \ll \omega_c$ which allows us to consider that the MIMO channel response is frequency flat at each of the significant harmonics of the radiated signal. Finally, we will resort to the following geometric MIMO channel model

$$\boldsymbol{H} = \sqrt{\frac{N_{\mathrm{T}} N_{\mathrm{R}}}{N_{\mathrm{p}}}} \sum_{l=1}^{N_{\mathrm{cl}}} \sum_{i=1}^{N_{\mathrm{p},l}} \alpha_{l,i} \boldsymbol{a}_{\mathrm{R}}(\theta_{l,i}^{\mathrm{R}}) \boldsymbol{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_{l,i}^{\mathrm{T}}), \tag{11}$$

where $\boldsymbol{a}_{\mathrm{R}}(\theta^{\mathrm{R}}) \in \mathbb{C}^{N_{\mathrm{R}}}$ is the receive steering vector at the Angle of Arrival (AoA) θ^{R}. This channel model results from assuming a multipath propagation with N_{cl} spatial clusters and $N_{\mathrm{p},l}$ propagation paths for the l-th cluster, $l \in \{1, \ldots, N_{\mathrm{cl}}\}$. The total number of channel paths is $N_{\mathrm{p}} = \sum_{l=1}^{N_{\mathrm{cl}}} N_{\mathrm{p},l}$. The parameters $\theta_{l,i}^{\mathrm{T}}$, $\theta_{l,i}^{\mathrm{R}}$, and $\alpha_{l,i}$ denote the AoD, the AoA, and the channel gain for the i-th path in the l-th cluster, respectively.

Finally, we assume a Uniform Linear Array (ULA), in which case the transmit steering vector is of the form

$$\boldsymbol{a}_{\mathrm{T}}(\theta_{p,l}^{\mathrm{T}}) = [\mathrm{e}^{-\mathrm{j}kz_1 \cos\theta_{p,l}^{\mathrm{T}}}, \ldots, \mathrm{e}^{-\mathrm{j}kz_{N_{\mathrm{T}}} \cos\theta_{p,l}^{\mathrm{T}}}]^{\mathrm{T}}, \tag{12}$$

where z_n, $n \in \{1, \ldots, N_{\mathrm{T}}\}$, is the position on the z-axis for the n-th array element, and $k = 2\pi/\lambda$ represents the wavenumber for the carrier wavelength $\lambda = 2\pi c/\omega_c$, where c is the speed of light. The steering vectors at reception are defined in a similar manner, i.e., $\boldsymbol{a}_{\mathrm{R}}(\theta_{p,l}^{\mathrm{R}}) = [\mathrm{e}^{-\mathrm{j}kz_1 \cos\theta_{p,l}^{\mathrm{R}}}, \ldots, \mathrm{e}^{-\mathrm{j}kz_{N_{\mathrm{R}}} \cos\theta_{p,l}^{\mathrm{R}}}]^{\mathrm{T}}$.

III. POWER EFFICIENCY OF TMA ANALOG PRECODING

In this section, we discuss the power efficiency of the APN due to the utilization of TMAs. There are two sources of power inefficiencies when handling TMAs [18]. The first one is the insertion losses associated to the particular hardware switches selected for the practical implementation of the TMAs. The power efficiency corresponding to this circumstance will be termed hardware efficiency, η_{HW}. The second one is the inefficiency due to the utilization of only the fundamental TMA mode $q = 0$ for transmission and the consideration of the power radiated over the harmonics $q \neq 0$ as wasted. This power efficiency will be termed TMA efficiency, η_{TMA}, and is inherent to the TMA operation, regardless of the hardware employed. As a result, by considering both terms of the efficiency in dBs, the overall power efficiency of the TMA APN is

$$\eta_{\mathrm{APN}}(dB) = \eta_{\mathrm{HW}}(dB) + \eta_{\mathrm{TMA}}(dB). \tag{13}$$

In the ensuing subsections, we proceed to analyze both the hardware and TMA power efficiencies.

A. HARDWARE EFFICIENCY

The hardware efficiency, η_{HW}, depends on the particular switching devices employed to implement the TMA APN. A key component of η_{HW} is that related to the amount of time that the switching devices are off, fruitlessly absorbing a fraction of the available energy. We will denote this component of η_{HW} as η_{OFF}. In order to firstly analyze η_{OFF}, we will take as a reference the following three solutions so far considered in the literature [18], [19]:

- A pessimistic scenario with single-pole-single-throw (SPST) RF switches where one switch is connected to a unique antenna. This is the TMA APN illustrated in Fig. 1. In this case, each switch will unavoidably remain in the off-state during a certain amount of time. In order to characterize η_{OFF}, let us consider the output at the m-th RF chain in Fig. 1, $y_m(t) = \boldsymbol{p}_{\mathrm{BB},m}^{\mathrm{T}} \boldsymbol{s}(t) \mathrm{e}^{\mathrm{j}\omega_c t}$, whose power is $\frac{\|q_T\|^2}{T_s} \|\boldsymbol{p}_{\mathrm{BB},m}\|_2^2$ with $\|q_T\|^2$ the energy of the transmit pulse waveform in (1). Notice that $\|\boldsymbol{p}_{\mathrm{BB},m}\|_2$ can be interpreted as an excitation amplitude common to the N_{T} switches connected to the m-th RF chain. Accordingly, the η_{OFF} of the m-th RF chain is defined as (see [13, eq. 10])

$$\eta_{\mathrm{OFF},m}(dB) = 10 \log_{10} \frac{\sum_{n=1}^{N_{\mathrm{T}}} \|\boldsymbol{p}_{\mathrm{BB},m}\|_2^2 |[\boldsymbol{P}_{\mathrm{RF}}^0]_{n,m}|}{\|\boldsymbol{p}_{\mathrm{BB},m}\|_2^2 N_{\mathrm{T}}}. \tag{14}$$

  Likewise, averaging over the L_{BS} RF chains, we obtain the overall η_{OFF}

$$\begin{aligned} \eta_{\mathrm{OFF}}(dB) &= 10 \log_{10} \frac{1}{L_{\mathrm{BS}}} \sum_{m=1}^{L_{\mathrm{BS}}} \eta_{H,m} \\ &= 10 \log_{10} \frac{\sum_{m=1}^{L_{\mathrm{BS}}} \sum_{n=1}^{N_{\mathrm{T}}} |[\boldsymbol{P}_{\mathrm{RF}}^0]_{n,m}|}{N_{\mathrm{T}} L_{\mathrm{BS}}}. \end{aligned} \tag{15}$$

- An intermediate scenario with single-pole-multiple-throw (SPMT) switches where a given switch is connected to M antennas [13], [18], thus minimizing the duration of its off state. Nevertheless, the corresponding antenna excitations are restricted to have complimentary switch-on durations and, as a consequence, the precoding design has to take into account these stringent restrictions. This scenario is out of the scope of this paper.
- An optimistic scenario recently proposed in [14] which employs PIN-diodes as switching devices. The TMA implemented with PIN-diodes is referred to as a Reconfigurable Power Divider/Combiner (RPDC) which allows for maximizing[1] η_{OFF} without losing neither flexibility nor aperture efficiency but at the expense of using some common additional elements like shunt stubs, discrete resistors, or RF chokes [20]. When utilizing a RPDC, the power absorbed during the off state is

[1] In fact, by assuming ideal conditions i.e., free of matching and metal/dielectric losses in the feeding network, a unitary value of η_{OFF} is accomplished [14].

transferred to other array elements during the on state. It is noteworthy that this technique allows to synthesize the same radiation pattern as that of a discrete SPST switched TMA. Moreover, it is demonstrated that better performance in terms of directivity and Side-Lobe Level (SLL) are achieved when compared to the SPST switched counterpart [14].

For an adequate analysis of the hardware efficiency, in addition to the efficiency accounting for the fraction of power absorbed in the off state, η_{OFF}, it is necessary to consider the specific power inefficiencies caused by the insertion losses introduced by each individual hardware element. Hence, both the insertion loss of the specific switching elements, η_{switch}, and that of the 1-bit VPSs, $\eta_{\text{1-bit VPS}}$, have to be considered when determining the analog precoder hardware efficiency:

$$\eta_{\text{HW}}(dB) = \eta_{\text{OFF}}(dB) + \eta_{\text{switch}}(dB) + \eta_{\text{1-bit VPS}}(dB). \quad (16)$$

B. TMA EFFICIENCY

Switching is the fundamental operation of TMAs. When applied to the RF chain outputs, it produces the radiated signal given by (9). The observation of such signal yields to the following conclusions:

- The term $q = 0$ (fundamental mode) accounts for the data signal $\boldsymbol{s}(t)$ centered at the carrier frequency ω_c. The spatial signature of this component is reconfigurable through the Fourier series coefficients $G^0_{n,m} = \xi_{n,m}$ (i.e., the normalized pulse time-widths), and the 1-bit VPS, $\phi_{n,m}$.
- The terms $q \neq 0$ (harmonic modes) represent the SR and are signal replicas located at the frequencies $\omega_c + q\omega_0$. The spatial signature of these replicas is determined by $G^q_{n,m} = \xi_{n,m}\,\text{sinc}(\pi q \xi_{n,m})$ and $\phi_{n,m}$. Notice that, since the bandwidth of the data signal, $\boldsymbol{s}(t)$, is B, the constraint $\omega_0 > B$ is necessary to avoid overlapping between the radiated signal harmonic replicas and the fundamental mode.

Although in some contexts it is possible to profitably exploit the SR, in this work we will focus on TMA analog precoding methods that effectively use the fundamental mode whereas signals radiated over the harmonic modes are useless. It is therefore apparent that, in order to increase the power efficiency of the TMA APN, we need to concentrate as much as possible the radiated power in the term $q = 0$ of (9) while minimizing the SR power in the terms $q \neq 0$.

We next determine the total power radiated in such a way that we separate the useful power at $q = 0$ from the harmonic SR at $q \neq 0$. We start considering the average power spatial density at the AoD θ^{T} given by

$$\begin{aligned}\mathcal{D}(\theta^{\text{T}}) &= \lim_{\tau\to\infty}\frac{1}{\tau}\int_{\tau/2}^{-\tau/2} |s^{\text{rad}}_{\text{tot}}(\theta^{\text{T}}, t)|^2 dt \\ &= \lim_{\tau\to\infty}\frac{1}{\tau}\int_{\tau/2}^{-\tau/2} \boldsymbol{a}^{\text{H}}_{\text{T}}(\theta^{\text{T}}) \sum_{q=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \boldsymbol{P}^q_{\text{RF}}\boldsymbol{P}_{\text{BB}} \\ &\quad \times \boldsymbol{s}(t)\boldsymbol{s}^{\text{H}}(t)\text{e}^{\text{j}(q-l)\omega_0 t}\boldsymbol{P}^{\text{H}}_{\text{BB}}(\boldsymbol{P}^l_{\text{RF}})^{\text{T}}\boldsymbol{a}_{\text{T}}(\theta^{\text{T}})dt \\ &= \frac{\|q_T\|^2}{T_s}\boldsymbol{a}^{\text{H}}_{\text{T}}(\theta^{\text{T}}) \sum_{q=-\infty}^{\infty} \boldsymbol{P}^q_{\text{RF}}\boldsymbol{P}_{\text{BB}}\boldsymbol{P}^{\text{H}}_{\text{BB}}(\boldsymbol{P}^q_{\text{RF}})^{\text{T}}\boldsymbol{a}_{\text{T}}(\theta^{\text{T}}), \end{aligned} \quad (17)$$

where we assume transmit pulsed waveforms $q_T(t)$ with power $\frac{\|q_T\|^2}{T_s}$, $\boldsymbol{s}_k$ has i.i.d. unit variance entries, and $\omega_0 > B$ makes the previous integral to be zero for $q \neq l$ [10]. Defining $\Theta^q = \boldsymbol{P}^q_{\text{RF}}\boldsymbol{P}_{\text{BB}}\boldsymbol{P}^{\text{H}}_{\text{BB}}(\boldsymbol{P}^q_{\text{RF}})^{\text{T}}$ and $\Theta = \sum_{q=-\infty}^{\infty}\Theta^q$ allows us to rewrite (17) in a more compact way as follows

$$\mathcal{D}(\theta^{\text{T}}) = \frac{\|q_T\|^2}{T_s}\boldsymbol{a}^{\text{H}}_{\text{T}}(\theta^{\text{T}})\Theta\boldsymbol{a}_{\text{T}}(\theta^{\text{T}}),$$

Finally, assuming the utilization of an ULA, the total mean radiated power, P^{rad}, is

$$\begin{aligned} P^{\text{rad}} &= \frac{2\pi\|q_T\|^2}{T_s}\sum_{i=1}^{N_{\text{T}}}\sum_{j=1}^{N_{\text{T}}}[\Theta]_{i,j}\int_0^{2\pi}\text{e}^{\text{j}k(z_j-z_i)\cos(\phi)}\sin(\phi)d\phi \\ &= \frac{4\pi\|q_T\|^2}{T_s}\sum_{i=1}^{N_{\text{T}}}\sum_{j=1}^{N_{\text{T}}}[\Theta]_{i,j}\text{sinc}\big(k(z_j-z_i)\big). \end{aligned} \quad (18)$$

Notice that the useful radiated power is the term in (18) corresponding to $q = 0$ and is given by

$$P^0 = \frac{4\pi\|q_T\|^2}{T_s}\sum_{i=1}^{N_{\text{T}}}\sum_{j=1}^{N_{\text{T}}}[\Theta^0]_{i,j}\text{sinc}\big(k(z_j-z_i)\big). \quad (19)$$

Therefore, the TMA efficiency is

$$\begin{aligned}\eta_{\text{TMA}}(dB) &= 10\log_{10}\frac{P^0}{P^{\text{rad}}} \\ &= 10\log_{10}\frac{\sum_{i=1}^{N_{\text{T}}}\sum_{j=1}^{N_{\text{T}}}[\Theta^0]_{i,j}\text{sinc}\big(k(z_j-z_i)\big)}{\sum_{i=1}^{N_{\text{T}}}\sum_{j=1}^{N_{\text{T}}}[\Theta]_{i,j}\text{sinc}\big(k(z_j-z_i)\big)}. \end{aligned} \quad (20)$$

Observe that the TMA efficiency depends on the baseband precoders. Furthermore, the columns of $\boldsymbol{P}^0_{\text{RF}}$ cannot be individually designed due the coupling through $\boldsymbol{P}_{\text{BB}}$, thus preventing the application of currently existing precoder design methods to the proposed hybrid TMA precoder.

IV. TMA VS. VPSs

In this section, we compare mmWave analog precoders with conventional VPSs to the one proposed in Fig. 1, based on TMA switches and 1-bit VPSs. Such a comparison is addressed in terms of four key performance indicators in 5G: insertion losses, size, electrical performance, and cost.

A. ANALOG PRECODING WITH VPSs

In order to perform a fair and realist comparison, we will focus on analog precoders equipped with digitally tuned passive VPSs.[2] There are five manufacturing technologies for the design of such devices: ferroelectrics,[3] Liquid

[2]Active and/or analog VPSs are power hungry devices which, in general, are not well suited to the mobility and energy efficiency requirements in 5G.

[3]The most popular ferroelectric VPSs are the Barium–Strontium–Titanate (BST)

TABLE 1. State-of-the-art of mmWave passive digital VPSs classified by manufacturing technologies [7], [21]–[30]. The maximum phase shift ($\Delta\Phi$) is considered by setting two usual performance metrics: the Figure of Merit (FoM), defined as $\Delta\Phi$/ (average insertion losses), measured in (°/dB), and the FoM/area, defined as $\Delta\Phi$/(average area), measured in (°/mm^2).

technology	Phase shifter insertion losses, $\eta_{\text{PS}}(dB)$			area(mm^2)			FoM	FoM/area
	interval	mean	standard deviation	interval	mean	standard deviation	(°/dB)	(°/mm^2)
BST	[-10.00, -3.60]	-6.63	2.30	[1.20, 33.00]	10.60	13.15	46.00	117.83
LC	[-14.00, -2.20]	-7.64	4.58	[1.50, 130.00]	36.70	47.95	23.40	20.75
CMOS	[-19.20, -4.00]	-10.09	4.53	[0.03, 3.15]	0.40	0.11	26.64	1579.26
MEMS	[-5.70, -2.70]	-3.96	1.24	[3.15, 11.90]	7.04	3.29	70.60	39.10
InGaAs pHEMT	[-8.80, -5.00]	-6.70	1.40	[1.41, 3.32]	2.14	0.79	56.00	191.80

Crystal (LC), Complementary Metal-Oxide-Semiconductor (CMOS), Micro-Electro-Mechanical Systems (MEMS) and InGaAs pseudomorphic High Electron Mobility Transistor (pHEMT). This latter is, perhaps, the more mature technology.

We start the state-of-the-art review of mmWave VPSs focusing on the manufacturing technology and exclusively comparing technical aspects. More specifically, we will pay special attention to average insertion losses (η_{PS}, that can be interpreted, somehow, as the power efficiency of such devices), chip areas and their corresponding relationship with the VPS electrical performance, quantifying the so-called Figure of Merit (FoM) and the FoM per area (see Table 1).

Although there are particular cases in each technology, we observe from the data in Table 1 that η_{PS} is poor in all technologies. The best value is achieved by MEMS (−3.96 dB), followed by Barium–Strontium–Titanate (BST)[4] (−6.63 dB) and InGAAs pHEMT (−6.70 dB) with a relatively low standard deviation in the cases of MEMS and pHEMT. Although data are quite disperse for LC and CMOS, we highlight the high average insertion losses of CMOS (−10.09 dB).

When the electrical performance is examined, we must consider the maximum phase shift ($\Delta\Phi$) accomplished by the VPSs. In this sense, the FoM, defined as $\Delta\Phi$/ (average efficiency) and measured in (°/dB), shows the best values for MEMS (70.60°/dB) followed by InGaAs pHEMT (56.00°/dB) and BST (46.00°/dB). We remark the low average FoMs of CMOS (26.64°/dB) and LC (23.40°/dB).

Regarding the sizes, LC requires the largest average area (36.70 mm^2) followed by BST (10.60 mm^2), both with very disperse values. However, such a high dispersion is not observed in CMOS and InGAAs pHEMT and, to a lesser extent, in MEMS. It is noteworthy to mention the very small average areas accomplished by CMOS technology (0.4 mm^2). Once again, we must relate the size to the electrical performance, in this case by means of the FoM/area defined as $\Delta\Phi$/(average area) and measured in (°/mm^2). We highlight the extraordinary average FoM/area achieved by CMOS (1579.26°/mm^2), while such a parameter is the pending issue of MEMS (39.10°/mm^2) and more markedly of LC (20.75°/mm^2).

In summary, LC technology shows significantly low FoM and FoM/area while BST is strongly affected by the dielectric losses at mmWave, which reduce the FoM when frequency increases. CMOS has excellent chip areas but, in general, high insertion losses. Contrarily, MEMS is the technology with the lowest insertion losses but occupies the largest chip areas. Likewise, pHEMT offers the best tradeoff between area and insertion losses.

Some recent works consider hybrid solutions to benefit from the small chip areas of CMOS, e.g. [21] which combines MEMS and CMOS achieving good insertion losses (−2.80 dB, although with a low FoM of 21.8 °/dB) and small chip areas (0.08 mm^2), or [22] which combines LC and CMOS to accomplish insertion losses of −5.35 dB (FoM of 52.00°/dB) with a 0.38 mm^2 area.

Recent advanced CMOS technologies also provide better insertion losses, e.g. [24] (−6.36 dB at 28 GHz) or [25] (−8.70 dB at 60 GHz), while employing very small chip areas (0.23 mm^2 and 0.09 mm^2, respectively).

Likewise, some InGAAs pHEMT commercial VPSs [26] offer reasonable areas (1.41 mm^2) and insertion losses (−6.00 dB at 30 GHz and −7 dB at 35 GHz) when compared to the state-of-the-art solutions.[5]

Finally, notice that the average insertion losses of VPSs increase with the number of bits in all technologies. Moreover, such increment depends on frequency: insertion losses in MEMS follows $-0.57\text{-}0.008f$ (GHz) while $-1.44\text{-}0.02f$ (GHz) in pHEMT [7].

B. ANALOG PRECODING WITH TMAs AND 1-BIT VPSs

Regarding the power efficiency η_{HW} of TMA analog precoding given by (16), we remark the following issues about the terms η_{switch} and $\eta_{\text{1-bit VPS}}$:

- The RF switching devices of a TMA analog precoder must be capable of faithfully working at frequencies higher than the signal bandwidth B [10]. We have con-

[4]Notice that most works on BST VPSs are centered in 30 GHz. Although this technology seems to be well positioned in terms of insertion losses, it suffers from dielectric losses rapidly increasing with frequency, hence damaging the insertion losses at higher frequencies.

[5]At the time of writing this article, very few commercial Monolithic Microwave Integrated Circuit (MMIC) VPSs for mmWave were in the market, being InGAAs pHEMT the most consolidated technology.

sidered 2 GHz as a realistic upper bound for the bandwidth of a 5G signal. Due to the maturity of the technology of RF switching devices, we have directly focused our discussion on two commercial examples manufactured by leading companies (see Table 2 with the corresponding values of η_{switch}): (1) a discrete switch [26] with an area of 1.21 mm^2 and 0.50 dB of insertion losses, and (2) a PIN-diode [31] with 1.28 mm^2 and 0.15 dB of η_{switch} (see Table 2);
- The insertion losses of the 1-bit VPSs, $\eta_{\text{1-bit VPS}}$, depend on the particular manufacturing technology and the working frequency (see Table 3)

TABLE 2. Features of two commercial wideband RF switching devices up to 3 GHz supplied by leading companies.

commercial wideband RF switching devices				
device	max. bandwidth (GHz)	switch insertion losses (dB)	chip area (mm^2)	ref.
discrete switch	2.7	-0.50	1.21	[26]
PIN-diode	3.0	-0.15	1.28	[31]

TABLE 3. Average insertion losses v.s frequency for 1-bit mmWave VPSs constructed with two different technologies.

1-bit mmWave VPSs		
technology	$\eta_{\text{1-bit PS}}$ (dB) (f is frequency in GHz)	ref.
MEMS	-0.57-0.008f	[7]
pHEMT	-1.44-0.02f	[7]

In this work, we have not considered the losses introduced by the transmission lines connecting the different hardware elements in a TMA APN. In the literature, such losses are usually specified by designers and manufacturers in dB/mm, e.g., 0.35 dB/mm at 60 GHz or 0.55 dB/mm at 110 GHz [32], [33, pp. 17–18]. Therefore, a meaningful characterization of such losses demands the knowledge of the physical layout or geometry of the specific circuit that implements the TMA APN, and this is out of the scope of the present work.

Additionally, notice that components are not ideal due to parameter tolerances, temperature drifts, and transient errors, among others. All these component errors unavoidably affect the reliability of the whole TMA APN and, ultimately, the precoder performance. The idea of performing an analysis of component errors in order to subsequently calibrate the TMA APN is relevant and necessary, because it will be possible to correct such errors, making their effects negligible. A detailed analysis of component errors is left as a future work.

V. TMA HYBRID PRECODER DESIGN

In this section, we consider the problem of determining the digital baseband precoder, $\boldsymbol{P}_{\text{BB}}$, and the TMA RF precoder, $\{\boldsymbol{P}_{\text{RF}}^q\}_{q=-\infty}^{\infty}$. Let us start by charactering the constraints imposed by the proposed hybrid TMA-based architecture. First of all, note that the precoder response at the harmonics modes $\boldsymbol{P}_{\text{RF}}^q$, $q \neq 0$, are totally determined by the response at the fundamental mode $\boldsymbol{P}_{\text{RF}}^0$. Indeed, $\boldsymbol{P}_{\text{RF}}^q$, $q \neq 0$, can be written in terms of $\boldsymbol{P}_{\text{RF}}^0$ as follows

$$\boldsymbol{P}_{\text{RF}}^q = \boldsymbol{P}_{\text{RF}}^0 \odot \boldsymbol{S}^q, \tag{21}$$

where $[\boldsymbol{S}^q]_{n,m} = \text{sinc}(\pi q|[\boldsymbol{P}_{\text{RF}}^0]_{n,m}|)$ according to (6) and (7), and $\odot$ represents the Hadamard product. Furthermore, the hardware limitations of the proposed analog precoding network in Sec. II establish the constraints

$$[\boldsymbol{P}_{\text{RF}}^0]_{n,m} \in [-1, 1]. \tag{22}$$

Finally, the available power budget is limited to a given value, that is, the total radiated power should be less than P_{TX}, i.e.

$$\sum\nolimits_{q=-\infty}^{\infty} \|(\boldsymbol{P}_{\text{RF}}^0 \odot \boldsymbol{S}^q)\boldsymbol{P}_{\text{BB}}\|_{\text{F}}^2 \leq P_{\text{TX}}. \tag{23}$$

As done in Sec. III-B, we split the total power between the fundamental and the harmonic modes as follows

$$\sum\nolimits_{q\neq 0} \|(\boldsymbol{P}_{\text{RF}}^0 \odot \boldsymbol{S}^q)\boldsymbol{P}_{\text{BB}}\|_{\text{F}}^2 + \|\boldsymbol{P}_{\text{RF}}^0\boldsymbol{P}_{\text{BB}}\|_{\text{F}}^2 \leq P_{\text{TX}}. \tag{24}$$

In summary, the constrained optimization problem to be considered when designing the baseband and analog parts of a TMA hybrid precoder is

$$\begin{aligned} &\max_{\boldsymbol{P}_{\text{RF}}^0, \boldsymbol{P}_{\text{BB}}} \; f(\boldsymbol{P}_{\text{RF}}^0, \boldsymbol{P}_{\text{BB}}) \\ &\text{s.t. } [\boldsymbol{P}_{\text{RF}}^0]_{n,m} \in [-1, 1] \\ &\quad \sum\nolimits_{q=-\infty}^{\infty} \|(\boldsymbol{P}_{\text{RF}}^0 \odot \boldsymbol{S}^q)\boldsymbol{P}_{\text{BB}}\|_{\text{F}}^2 \leq P_{\text{TX}}, \end{aligned} \tag{25}$$

where $f(\boldsymbol{P}_{\text{RF}}^0, \boldsymbol{P}_{\text{BB}})$ is the performance metric considered in the precoder design.

Notice that (25) is a rather involved problem because the transmit power constraint incorporates the harmonic SR losses. To circumvent this issue, the power radiated over the harmonic modes will be initially ignored in order to obtain tractable problem formulations. That is, $\boldsymbol{P}_{\text{RF}}^q = 0$, for $q \neq 0$, and the power constraint is assumed to be $\|\boldsymbol{P}_{\text{RF}}^0\boldsymbol{P}_{\text{BB}}\|_{\text{F}}^2 \leq P_{\text{TX}}$. The total radiated power (23) will be considered in Sec. V-C and the SR losses will be evaluated in Sec. VI.

Observe that the hardware efficiency was not considered in the problem formulation since it depends on the switching devices and can be thereby treated as insertion losses.

A. PERFECT FACTORIZATION

Consider a fully digital linear precoder $\boldsymbol{P}_D \in \mathbb{C}^{N_{\text{T}} \times N_{\text{s}}}$ designed to maximize (minimize) a given performance metric $f(\boldsymbol{P}_D)$. A way to determine the baseband and analog precoders is to obtain an accurate approximation, in terms of a distance metric meaningful for $f(\cdot)$, of the reference digital precoder by means of the factorization $\boldsymbol{P}_{\text{RF}}^0\boldsymbol{P}_{\text{BB}} \approx \boldsymbol{P}_{\text{D}}$. In other words, we try to achieve a performance similar to that of the digital solution, i.e. $f(\boldsymbol{P}_{\text{RF}}^0\boldsymbol{P}_{\text{BB}}) \approx f(\boldsymbol{P}_D)$.

$$\boldsymbol{P}_{\mathrm{d}} = \underbrace{\begin{bmatrix} \frac{\Re[\boldsymbol{P}_{\mathrm{d}}]_{1,1}}{\beta_1} & \frac{\Im[\boldsymbol{P}_{\mathrm{d}}]_{1,1}}{\beta_1} & \cdots & \frac{\Re[\boldsymbol{P}_{\mathrm{d}}]_{1,N_{\mathrm{s}}}}{\beta_{N_{\mathrm{s}}}} & \frac{\Im[\boldsymbol{P}_{\mathrm{d}}]_{1,N_{\mathrm{s}}}}{\beta_{N_{\mathrm{s}}}} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \frac{\Re[\boldsymbol{P}_{\mathrm{d}}]_{N_{\mathrm{T}},1}}{\beta_1} & \frac{\Im[\boldsymbol{P}_{\mathrm{d}}]_{N_{\mathrm{T}},1}}{\beta_1} & \cdots & \frac{\Re[\boldsymbol{P}_{\mathrm{d}}]_{N_{\mathrm{T}},N_{\mathrm{s}}}}{\beta_{N_{\mathrm{s}}}} & \frac{\Im[\boldsymbol{P}_{\mathrm{d}}]_{N_{\mathrm{T}},N_{\mathrm{s}}}}{\beta_{N_{\mathrm{s}}}} \end{bmatrix}}_{\boldsymbol{P}^0_{\mathrm{RF}}} \underbrace{\left(\mathrm{diag}(\beta_1, \ldots, \beta_{N_{\mathrm{s}}}) \otimes \begin{bmatrix} 1 \\ j \end{bmatrix}\right)}_{\boldsymbol{P}_{\mathrm{BB}}} \tag{26}$$

Recall that in TMA precoding $\boldsymbol{P}^0_{\mathrm{RF}} \in \mathbb{R}^{N_{\mathrm{T}} \times L_{\mathrm{BS}}}$ with $[\boldsymbol{P}^0_{\mathrm{RF}}]_{n,m} \in [-1, 1]$, $\forall m, n$, whereas $\boldsymbol{P}_{\mathrm{BB}} \in \mathbb{C}^{L_{\mathrm{BS}} \times N_{\mathrm{s}}}$. Following a procedure similar to that in [34], we show in the sequel that at most $L_{\mathrm{BS}} = 2N_{\mathrm{s}}$ RF chains are necessary to achieve $\boldsymbol{P}^0_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}} = \boldsymbol{P}_{\mathrm{D}}$.

First, note that it is not possible to construct a complex column vector $[\boldsymbol{P}_{\mathrm{D}}]_{:,i} = \boldsymbol{p}_i \in \mathbb{C}^{N_{\mathrm{T}}}$ as $c\boldsymbol{v}$, with c a complex scalar and $\boldsymbol{v} \in \mathbb{R}^{N_{\mathrm{T}}}$, unless all components in $\boldsymbol{p}_i$ have the same phase. On the contrary, with two RF chains for each linearly independent column vector, we decompose the reference digital precoder $\boldsymbol{P}_{\mathrm{D}}$ as in (26), as shown at the top of this page, where $\otimes$ represents the Kronecker product and $\beta_i = \max(\max_j \Re[\boldsymbol{P}_{\mathrm{d}}]_{j,i}, \max_j \Im[\boldsymbol{P}_{\mathrm{d}}]_{j,i})$. Therefore, the number of RF chains necessary to achieve a perfect factorization is $L_{\mathrm{BS}} = 2N_{\mathrm{s}}$, i.e., the same as that for the architecture employing infinite resolution VPSs [34]. Despite the above, in the ensuing subsections the condition $L_{\mathrm{BS}} = 2N_{\mathrm{s}}$ is relaxed to $L_{\mathrm{BS}} = N_{\mathrm{s}}$, a more realistic assumption in practical systems.

B. ALGORITHMIC SOLUTIONS FOR TMA HYBRID PRECODING

In this subsection, and similarly to [15], [16], and [35], we will consider the mutual information as the performance metric for precoder design. Recall that, in order to have tractable optimization problems, SR will be ignored and precoder design limits to $\boldsymbol{P}^0_{\mathrm{RF}}$. Hence, the optimization problem (25) reduces to

$$\begin{aligned} &\max_{\boldsymbol{P}^0_{\mathrm{RF}}, \boldsymbol{P}_{\mathrm{BB}}} \ \mathcal{I}(\boldsymbol{P}^0_{\mathrm{RF}}, \boldsymbol{P}_{\mathrm{BB}}) \\ &\text{s.t. } \|\boldsymbol{P}^0_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}}\|_{\mathrm{F}}^2 \le P_{\mathrm{TX}} \quad [\boldsymbol{P}^0_{\mathrm{RF}}]_{n,m} \in [-1, 1], \end{aligned} \tag{27}$$

where $\mathcal{I}(\boldsymbol{P}^0_{\mathrm{RF}}, \boldsymbol{P}_{\mathrm{BB}})$ is the mutual information. For the sake of notation simplicity we will use $\mathcal{I}$ hereafter. The mutual information is given by

$$\mathcal{I} = \log_2 \det(\mathbf{I}_{N_{\mathrm{R}}} + \frac{1}{N_{\mathrm{s}}N_0} \boldsymbol{H}\boldsymbol{P}^0_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}}\boldsymbol{P}^{\mathrm{H}}_{\mathrm{BB}}(\boldsymbol{P}^0_{\mathrm{RF}})^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}), \tag{28}$$

where N_0 is the received noise power per antenna. In spite of the relaxation on the power constraint, (27) is still hard to solve due to the coupling between the analog and the baseband precoders. Moreover, the number of RF chains imposes a restriction on the precoder sizes.

In the sequel, three different TMA precoder design approaches will be considered. Similar techniques have already been successfully employed to design analog precoders based on VPSs. It is intriguing to evaluate the performance and the computational complexity, while analyzing the power losses incurred by the TMA, and compare them to that of designs based on VPSs. Notice that, although precoders proposed in the following might not fulfill $[\boldsymbol{P}^0_{\mathrm{RF}}]_{n,m} \in [-1, 1]$, the baseband and analog precoders can be scaled such that the constraint holds. Furthermore, this scaling operation is transparent in terms of achievable rate and transmit power. In the interest of notation simplicity, we will denote $\boldsymbol{P}^0_{\mathrm{RF}}$ as $\boldsymbol{P}_{\mathrm{RF}}$ in the remaining of the subsection.

1) Alternating Optimization (AO)

The first hybrid approach we will consider is similar to the one described in [2], where the Frobenius norm is used to compute the distance between the digital and the hybrid precoders. In particular, we seek for $\boldsymbol{P}_{\mathrm{BB}}$ and $\boldsymbol{P}_{\mathrm{RF}}$ that solve the following optimization problem

$$\begin{aligned} &\min_{\boldsymbol{P}_{\mathrm{BB}}, \boldsymbol{P}_{\mathrm{RF}}} \ \|\boldsymbol{P}_{\mathrm{D}} - \boldsymbol{P}_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}}\|_{\mathrm{F}}^2 \\ &\text{s.t. } \|\boldsymbol{P}_{\mathrm{RF}}\|_{\mathrm{F}}^2 \le P_{\mathrm{TX}}, \quad \boldsymbol{P}_{\mathrm{BB}} \in \mathcal{U}, \ \boldsymbol{P}_{\mathrm{RF}} \in \mathbb{R}^{N_{\mathrm{T}} \times L_{\mathrm{BS}}}, \end{aligned} \tag{29}$$

where $\boldsymbol{P}_{\mathrm{D}}$ is the fully digital precoder maximizing the mutual information, and $\mathcal{U}$ is the set of unitary matrices. For analog precoders based on VPSs, this heuristic approach has shown excellent performance results [17], [35]. Therein, a solution for $\boldsymbol{P}_{\mathrm{BB}}$ is provided for given $\boldsymbol{P}_{\mathrm{RF}}$, whereas $\boldsymbol{P}_{\mathrm{RF}}$ is computed according to a heuristic that, in turn, depends on $\boldsymbol{P}_{\mathrm{BB}}$. The updates of $\boldsymbol{P}_{\mathrm{RF}}$ and $\boldsymbol{P}_{\mathrm{BB}}$ define an iterative procedure which alternates between the computation of the baseband and the analog precoding matrices. Similarly, we define an AO algorithm for TMA hybrid precoding design. The baseband part is the solution to the orthogonal Procrustes problem [36] as in [17] and [35], that is, $\boldsymbol{P}_{\mathrm{BB}} = \boldsymbol{V}\boldsymbol{U}^{\mathrm{H}}$ with $\boldsymbol{P}^{\mathrm{H}}_{\mathrm{D}}\boldsymbol{P}_{\mathrm{RF}} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{H}}$. However, contrarily to analog precoding using VPSs, the feasible set of analog precoders in the proposed TMA architecture is tractable. Hence, we obtain a closed-form solution for $\boldsymbol{P}_{\mathrm{RF}}$, given $\boldsymbol{P}_{\mathrm{BB}}$. Indeed, considering a fixed $\boldsymbol{P}_{\mathrm{BB}}$, we rewrite (29) as

$$\begin{aligned} &\max_{\boldsymbol{P}_{\mathrm{RF}} \in \mathbb{R}^{N_{\mathrm{T}} \times L_{\mathrm{BS}}}} \ \mathrm{tr}\left(\boldsymbol{P}_{\mathrm{RF}}(\boldsymbol{P}_{\mathrm{BB}}\boldsymbol{P}^{\mathrm{H}}_{\mathrm{D}} + \boldsymbol{P}^{*}_{\mathrm{BB}}\boldsymbol{P}^{\mathrm{T}}_{\mathrm{D}})\right) \\ &\text{s.t. } \|\boldsymbol{P}_{\mathrm{RF}}\|_{\mathrm{F}}^2 \le P_{\mathrm{TX}}, \end{aligned} \tag{30}$$

whose solution is readily obtained as

$$\boldsymbol{P}_{\mathrm{RF}} = \frac{\sqrt{P_{\mathrm{TX}}}}{\|\boldsymbol{P}^{\mathrm{H}}_{\mathrm{BB}}\boldsymbol{P}_{\mathrm{D}} + \boldsymbol{P}^{*}_{\mathrm{BB}}\boldsymbol{P}^{\mathrm{T}}_{\mathrm{D}}\|_{\mathrm{F}}} \left(\boldsymbol{P}^{\mathrm{H}}_{\mathrm{BB}}\boldsymbol{P}_{\mathrm{D}} + \boldsymbol{P}^{*}_{\mathrm{BB}}\boldsymbol{P}^{\mathrm{T}}_{\mathrm{D}}\right). \tag{31}$$

Due to the optimality of the updates of $\boldsymbol{P}_{\mathrm{RF}}$ and $\boldsymbol{P}_{\mathrm{BB}}$, the below-bounded metric reduces at each iteration and convergence to a local optimum is therefore ensured. Note that, in contrast to [17] and [35], the update of the analog precoder satisfies the power constraint with equality.

2) GRADIENT

An alternative method to calculate the precoders based on the Frobenius norm distance is a variation of the gradient-projection algorithm in [37] and [38]. The problem formulation is slightly different from (29), and reads as

$$\begin{aligned}&\min_{\boldsymbol{P}_{\mathrm{BB}},\boldsymbol{P}_{\mathrm{RF}}} \|\boldsymbol{P}_{\mathrm{D}}-\boldsymbol{P}_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}}\|_{\mathrm{F}}^2\\ &\text{s.t. } \|\boldsymbol{P}_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}}\|_{\mathrm{F}}^2\le P_{\mathrm{TX}},\quad \boldsymbol{P}_{\mathrm{RF}}\in\mathbb{R}^{N_{\mathrm{T}}\times L_{\mathrm{BS}}}.\end{aligned}\tag{32}$$

The baseband precoder is first chosen to minimize $\|\boldsymbol{P}_{\mathrm{D}}-\boldsymbol{P}_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{BB}}\|_{\mathrm{F}}^2$, i.e., $\boldsymbol{P}_{\mathrm{BB}}=(\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{RF}})^{-1}\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{D}}$. Next, in order to fulfill the power constraint with equality, a scaling factor multiplies the baseband precoder to eventually obtain

$$\boldsymbol{P}_{\mathrm{BB}}=\frac{\sqrt{P_{\mathrm{TX}}}}{\|\boldsymbol{P}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{P}_{\mathrm{RF}}(\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{RF}})^{-1/2}\|_{\mathrm{F}}}(\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{RF}})^{-1}\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{D}}.\tag{33}$$

Since $\boldsymbol{P}_{\mathrm{BB}}$ in (33) is defined as a function of the analog precoder $\boldsymbol{P}_{\mathrm{RF}}$, it is possible to reduce the problem formulation in (32) to the minimization of the following distance metric

$$d=\|\boldsymbol{P}_{\mathrm{D}}\|_{\mathrm{F}}^2-\sqrt{P_{\mathrm{TX}}}\sqrt{\mathrm{tr}\left(\boldsymbol{P}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{P}_{\mathrm{RF}}(\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{RF}})^{-1}\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{D}}\right)}.\tag{34}$$

We propose to perform the minimization of this distance metric with a gradient algorithm of the form $\boldsymbol{P}_{\mathrm{RF}}=\boldsymbol{P}_{\mathrm{RF}}-\mu\frac{\partial d}{\partial \boldsymbol{P}_{\mathrm{RF}}}$, where

$$\begin{aligned}\frac{\partial d}{\partial \boldsymbol{P}_{\mathrm{RF}}}=&\frac{\sqrt{P_{\mathrm{TX}}}\left(\mathbf{I}_{N_{\mathrm{s}}}-(\boldsymbol{P}_{\mathrm{RF}}(\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{RF}})^{-1}\boldsymbol{P}_{\mathrm{RF}})\right)}{\|\boldsymbol{P}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{P}_{\mathrm{RF}}(\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{RF}})^{-1/2}\|_{\mathrm{F}}}\\&\times\left(\boldsymbol{P}_{\mathrm{D}}\boldsymbol{P}_{\mathrm{D}}^{\mathrm{H}}+\boldsymbol{P}_{\mathrm{D}}^{*}\boldsymbol{P}_{\mathrm{D}}^{\mathrm{T}}\right)\boldsymbol{P}_{\mathrm{RF}}(\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{RF}})^{-1},\end{aligned}\tag{35}$$

and μ is the step size Notice that this approach jointly computes the TMA analog and baseband precoders. Moreover, for typical architectures employing VPSs, a subsequent projection to the feasible set is needed after the gradient step [37], [38]. In contrast, this operation is not performed for TMA analog precoding, thus leading to a faster convergence to a local minimum.

3) DECOUPLED ANALOG AND BASEBAND DESIGNS

A third method is based on decoupling the design of the analog and baseband TMA precoders. Such method relies on $\boldsymbol{P}_{\mathrm{BB}}\boldsymbol{P}_{\mathrm{BB}}^{\mathrm{H}}=\gamma\mathbf{I}$, as in [15] and [35]. This way the power constraint is fulfilled as long as $\|\boldsymbol{P}_{\mathrm{RF}}\|_{\mathrm{F}}^2\le\gamma^{-1}P_{\mathrm{TX}}$. Thus, the optimization problem (27) reduces to

$$\begin{aligned}&\max_{\boldsymbol{P}_{\mathrm{RF}}}\ \mathcal{I}=\log_2\det(\mathbf{I}_{N_{\mathrm{R}}}+\frac{\gamma}{N_0N_{\mathrm{s}}}\boldsymbol{H}\boldsymbol{P}_{\mathrm{RF}}\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{H}})\\&\text{s.t. } \|\boldsymbol{P}_{\mathrm{RF}}\|_{\mathrm{F}}^2\le\gamma^{-1}P_{\mathrm{TX}}.\end{aligned}\tag{36}$$

In order to solve (36), let us define the channel Gramian $\boldsymbol{G}=\frac{\gamma}{N_0N_{\mathrm{s}}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$, which is split up as $\boldsymbol{G}=\boldsymbol{G}^{\Re}+\mathrm{j}\boldsymbol{G}^{\Im}$ with $\boldsymbol{G}^{\Re}$ and $\boldsymbol{G}^{\Im}$ being symmetric and antisymmetric real matrices, respectively. We now define the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ entry-wise as $[\boldsymbol{A}]_{i,j}=1+\boldsymbol{p}_{\mathrm{RF},i}^{\mathrm{T}}\boldsymbol{G}^{\Re}\boldsymbol{p}_{\mathrm{RF},j}$ for $i=j$ and $[\boldsymbol{A}]_{i,j}=\boldsymbol{p}_{\mathrm{RF},i}^{\mathrm{T}}\boldsymbol{G}^{\Re}\boldsymbol{p}_{\mathrm{RF},j}$ for $i\neq j$, and $[\boldsymbol{B}]_{i,j}=\mathrm{j}\boldsymbol{p}_{\mathrm{RF},i}^{\mathrm{T}}\boldsymbol{G}^{\Im}\boldsymbol{p}_{\mathrm{RF},j}$ for all i,j. Therefore, the mutual information metric in (36) can be rewritten as

$$\mathcal{I}=\log_2\det(\boldsymbol{A}+\boldsymbol{B}).\tag{37}$$

The optimal precoder $\boldsymbol{P}_{\mathrm{RF}}$ is the one that makes $\boldsymbol{A}+\boldsymbol{B}$ to be a diagonal matrix and suppresses the inter-stream interference [39], i.e., $\mathcal{I}=\sum_{m=1}^{L_{\mathrm{BS}}}\log_2(1+[\boldsymbol{A}]_{m,m}+[\boldsymbol{B}]_{m,m})$. This is achieved with fully digital precoding, and can be asymptotically reached for large antenna arrays with a VPS analog precoder since the steering vectors for different AoDs become mutually orthogonal [40]. However, in our case, the entries of $\boldsymbol{P}_{\mathrm{RF}}$ are real-valued and it is not possible to diagonalize the matrix summation since the equality $\boldsymbol{p}_{\mathrm{RF},i}^{\mathrm{T}}\boldsymbol{G}^{\Re}\boldsymbol{p}_{\mathrm{RF},j}=\mathrm{j}\boldsymbol{p}_{\mathrm{RF},i}^{\mathrm{T}}\boldsymbol{G}^{\Im}\boldsymbol{p}_{\mathrm{RF},j}$, for $i\neq j$, cannot be achieved.

To gain further insight and better motivate our initial precoder design of (38), we characterize in Appendix the set of all possible values for the mutual information $\mathcal{I}$ under the assumption considered in this section. Based on the discussion in App., a simple yet reasonable design for the analog precoder is given by

$$\boldsymbol{P}_{\mathrm{RF}}=[\boldsymbol{U}]_{:,1:L_{\mathrm{BS}}},\tag{38}$$

where $\boldsymbol{U}$ are the eigenvectors of $\boldsymbol{G}^{\Re}$ corresponding to the L_{BS} largest eigenvalues. An immediate consequence of this choice is that the inter-stream interference corresponding to $\boldsymbol{A}$ is removed whereas the inter-stream interference in $\boldsymbol{B}$ will be mitigated with the baseband precoder. To that end, we define an effective channel $\boldsymbol{H}_{\mathrm{eff}}=\boldsymbol{H}\boldsymbol{P}_{\mathrm{RF}}(\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{RF}})^{-1/2}$ and we decompose the baseband precoder as $\boldsymbol{P}_{\mathrm{BB}}=(\boldsymbol{P}_{\mathrm{RF}}^{\mathrm{T}}\boldsymbol{P}_{\mathrm{RF}})^{-1/2}\bar{\boldsymbol{P}}_{\mathrm{BB}}$. Hence, it is possible to reformulate the optimization problem (36) as (see [15])

$$\begin{aligned}&\max_{\bar{\boldsymbol{P}}_{\mathrm{BB}}}\ \mathcal{I}=\log_2\det(\mathbf{I}_{N_{\mathrm{R}}}+\frac{1}{N_0N_{\mathrm{s}}}\boldsymbol{H}_{\mathrm{eff}}\bar{\boldsymbol{P}}_{\mathrm{BB}}\bar{\boldsymbol{P}}_{\mathrm{BB}}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{eff}}^{\mathrm{H}})\\&\text{s.t. } \|\bar{\boldsymbol{P}}_{\mathrm{BB}}\|_{\mathrm{F}}\le P_{\mathrm{TX}}.\end{aligned}\tag{39}$$

The former problem formulation can be solved as in [39]. Notice that in the resulting baseband precoder, the unitary constraint relaxes to $\boldsymbol{P}_{\mathrm{BB}}\boldsymbol{P}_{\mathrm{BB}}^{\mathrm{H}}=\Gamma$, with diagonal Γ, and it has to be properly scaled to fulfill the power limitation.

The bottleneck of the previous approach comes from the remaining interference due to the reduced number of RF chains. Notice, however, that $\boldsymbol{A}\to\mathbf{I}$ and $\boldsymbol{B}\to\mathbf{0}$ when $N_0\to\infty$, hence showing the suitability of this strategy in the low SNR regime. This method will be denoted as 2-step SVD (2S-SVD).

We next present a different design of the analog precoder which incorporates interference mitigation. Recall that the precoders for different streams are coupled and they have to be jointly designed. Aligned with this objective, we propose

to follow an iterative approach similar to that in [41]. Therein, the contribution of each precoder column is isolated, thus allowing to independently update one column at each iteration. Likewise, we define the matrix

$$\boldsymbol{G}_m = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}(\mathbf{I} - \frac{\gamma}{N_0 N_{\mathrm{s}}}\boldsymbol{P}_{\mathrm{RF},\bar{m}}\boldsymbol{C}_m^{-1}\boldsymbol{P}_{\mathrm{RF},\bar{m}}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}), \quad (40)$$

where $\boldsymbol{C}_m = \mathbf{I} + \frac{\gamma}{N_0}\boldsymbol{P}_{\mathrm{RF},\bar{m}}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{P}_{\mathrm{RF},\bar{m}}$ and $\boldsymbol{P}_{\mathrm{RF},\bar{m}}$ is the result of removing the m-th column from $\boldsymbol{P}_{\mathrm{RF}}$. Then, the m-th column of $\boldsymbol{P}_{\mathrm{RF}}$, $[\boldsymbol{P}_{\mathrm{RF}}]_{:,m} = \boldsymbol{p}_{\mathrm{RF},m} \in \mathbb{C}^{N_{\mathrm{T}}}$, is chosen to maximize

$$\log_2(1 + \frac{\gamma}{N_0 N_{\mathrm{s}}}\boldsymbol{p}_{\mathrm{RF},m}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{p}_{\mathrm{RF},m}) \quad \text{s.t. } \|\boldsymbol{p}_{\mathrm{RF},m}\|_2^2 = 1. \quad (41)$$

In order to solve (41), we exploit that $\boldsymbol{G}_m \succeq 0$ [41]. As a consequence, $\boldsymbol{G}_m^{\Im} = \Im\{\boldsymbol{G}_m\}$ is antisymmetric, i.e., $\boldsymbol{G}_m^{\Im,\mathrm{T}} = -\boldsymbol{G}_m^{\Im}$. Thus, for any real vector $\boldsymbol{x}$ we have

$$\begin{aligned}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}^{\Im}\boldsymbol{x} &= \mathrm{tr}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}^{\Im}\boldsymbol{x}) = \mathrm{tr}(\boldsymbol{G}^{\Im}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}) \\ &= \mathrm{tr}(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}^{\Im,\mathrm{T}}) = -\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}^{\Im}\boldsymbol{x}, \end{aligned} \quad (42)$$

meaning that $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}_m^{\Im}\boldsymbol{x} = 0$. Therefore, (41) is optimally solved by taking $\boldsymbol{p}_{\mathrm{RF},m}$ as the dominant eigenvector of $\boldsymbol{G}_m^{\Re} = \Re\{\boldsymbol{G}_m\}$. It is important to highlight that this method converges to the precoding strategy in (38) for $N_0 \to \infty$. Finally, the baseband precoder $\boldsymbol{P}_{\mathrm{BB}}$ is the solution to (39). In general, digital performance will be reached only when the baseband precoder compensates the loss caused by the analog counterpart.

Because of the interference mitigation feature, we can alternatively rely only on the analog precoder of (41). One drawback of this assumption is that the same transmit power is allocated to all the data streams. Although performance improves by employing the baseband precoder resulting from (39), which also includes a smarter power allocation, this decision saves complexity in the implementation and was adopted in [42]. Moreover, this assumption makes it possible to handle the total power constraint in (25), as discussed in the subsequent section. Also, as observed in our numerical experiments, it reduces the switches' OFF-state and TMA SR losses compared to the hybrid design that incorporates the baseband precoder resulting from (39).

C. SWITCHING EFFICIENCY AND SR POWER CONSTRAINT

TMA hybrid precoding solutions in the previous subsection ignore SR, i.e., switching efficiency is not taken into account while deriving the solutions. The obtained TMA hybrid precoders thus exhibit good behavior in terms of achievable throughput and computational cost but their performance in terms of switching efficiency remains unknown. This issue will be considered in this subsection.

Observe that the consideration of SR into the power restriction in (25) leads to a rather involved problem formulation. A more tractable constraint can be obtained if we restrict the baseband precoders as $\boldsymbol{P}_{\mathrm{BB}}\boldsymbol{P}_{\mathrm{BB}}^{\mathrm{H}} = \Gamma$ and we recall that $[\boldsymbol{P}_{\mathrm{RF}}^0]_{n,m} \in [-1, 1]$. Accordingly, the power constraint can be rewritten as

$$\begin{aligned}&\sum_{q=-\infty}^{\infty} \|\boldsymbol{P}_{\mathrm{RF}}^q\boldsymbol{P}_{\mathrm{BB}}\|_{\mathrm{F}}^2 \\ &= \sum_{q=-\infty}^{\infty}\sum_{m=1}^{L_{\mathrm{BS}}}\gamma_m\sum_{n=1}^{N_{\mathrm{T}}}[\boldsymbol{P}_{\mathrm{RF}}^q]_{n,m}^2 \\ &= \sum_{m=1}^{L_{\mathrm{BS}}}\sum_{n=1}^{N_{\mathrm{T}}}\gamma_m\xi_{n,m}^2\sum_{q=-\infty}^{\infty}\mathrm{sinc}^2(\pi q\xi_{n,m}) \quad (43) \\ &= \sum_{m=1}^{L_{\mathrm{BS}}}\sum_{n=1}^{N_{\mathrm{T}}}\gamma_m\xi_{n,m} = \mathbf{1}^{\mathrm{T}}|\boldsymbol{P}_{\mathrm{RF}}^0|\Gamma\mathbf{1} \le P_{\mathrm{TX}}, \quad (44)\end{aligned}$$

where Parseval's theorem was employed to get (44) from (43). We now introduce $\xi'_{n,m} = \psi_m\xi_{n,m}$ and $\gamma'_m = \gamma_m/\psi_m^2$. Observe that the value of $\Psi = \mathrm{diag}(\psi_1, \ldots, \psi_{L_{\mathrm{BS}}})$ is transparent for the achievable rate and the regular power constraint. Hence, we can reduce the SR by choosing $\psi_m = \min_n \xi_{n,m}^{-1} \ge 1$. By updating the analog precoders corresponding to the fundamental mode as $\boldsymbol{P}_{\mathrm{RF}}^{0,\prime} = \boldsymbol{P}_{\mathrm{RF}}\Psi$ and following the steps in (44) we obtain

$$\mathbf{1}^{\mathrm{T}}|\boldsymbol{P}_{\mathrm{RF}}^{0,\prime}|\Gamma\Psi^{-1}\mathbf{1} \le \mathbf{1}^{\mathrm{T}}|\boldsymbol{P}_{\mathrm{RF}}^0|\Gamma\mathbf{1}. \quad (45)$$

Since the term $q = 0$ is not affected, we conclude that this scaling reduces the SR and, consequently, increases the switching efficiency. A similar reasoning applies to the power absorption inherent to the hardware given by η_{OFF} as in (15). Moreover, the equivalent power constraint in (44) makes it possible to bound the transmit power including the SR for the proposed strategies that initially restrict the baseband precoding to $\boldsymbol{P}_{\mathrm{BB}}\boldsymbol{P}_{\mathrm{BB}}^{\mathrm{H}} = \gamma\mathbf{I}$, that is, AO and the analog scheme of (41).

In addition to the scaling matrix, it is possible to further improve the efficiency of the iterative algorithm presented in Sec. V-B.3 by not employing the columns of the analog precoder exhibiting large inner products. Thus, we define the baseband precoder $\boldsymbol{P}'_{\mathrm{BB}} = \boldsymbol{Z}\boldsymbol{P}_{\mathrm{BB}}$, where $\boldsymbol{Z} = \sum_{i\notin\mathcal{R}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}$ and $\mathcal{R}$ is a set containing the indices corresponding to the columns for which $(\boldsymbol{p}_{\mathrm{RF},i}^0)^{\mathrm{T}}\boldsymbol{p}_{\mathrm{RF},j}^0 \ge \beta$ for any j such that $j > i$. In terms of efficiency, this is equivalent to set the column of the analog precoder to the all ones vector and, accordingly, to reduce the SR and the insertion losses. This strategy is particularly useful in the low SNR regime where the power is concentrated on the largest modes of the channel. Also, it can take advantage of scenarios where the number of channel paths is small, or when there is a dominant Line-of-sight (LOS) direction [43]. On the contrary, in the high SNR regime and large number of channel paths, the correlation of the columns of $\boldsymbol{P}_{\mathrm{RF}}$ is small to take advantage of the channel spatial gains. As a consequence, it is likely that $\mathcal{R} = \emptyset$.

VI. PERFORMANCE ANALYSIS

In this section we present the results of several numerical experiments carried out to evaluate the performance of the proposed TMA hybrid precoder. The considered setup consists of a $N_{\mathrm{T}} = 128$ antennas transmitter and a $N_{\mathrm{R}} = 16$

antennas receiver. The number of streams is equal to that of the RF chains $N_s = L_{\text{BS}} = 8$. The transmit power is set to 55 dBm $\approx$ 24.8 dB [44] with an average SNR ranging from -20 dB to 10 dB, where SNR $= \frac{P_{\text{TX}}}{N_0 N_s}$ for $\mathbb{E}[|\alpha_{i,i}|^2] = 1$. We assume a ULA with the antenna elements at positions $z_i = (i-1)\lambda/2$. The channel is generated according to (11) with $N_{\text{p},l} = 2$ propagation paths for each of the $N_{\text{cl}} = 6$ spatial clusters. The results are averaged over 1000 channel realizations of (11).

A. NUMERICAL EXPERIMENTS

Fig. 2 shows the achievable rates for the precoding strategies presented in this work. Moreover, we include for comparison the fully digital precoder with $N_s = 8$, and the hybrid design using the algorithm HD-LSR (hybrid design by least squares relaxation) in [17]. Notice the satisfactory performance of the iterative methods. The curve labeled Iterative NLD refers to the method where the columns presenting high correlation are removed (see Sec. V-C). The worst performance corresponds to the curve satisfying the total power constraint in (44), marked as Analog TotPwr. Analog represents the curve with only analog precoding, as presented in Sec. (V-B.3), where the baseband precoder fulfills $\boldsymbol{P}_{\text{BB}}\boldsymbol{P}_{\text{BB}}^{\text{H}} = \gamma \mathbf{I}$. Notice that even when the computational complexity is very small, the performance is equal to that obtained with the AO strategy. Finally, the steepest descent procedure and 2S-SVD method perfectly match the fully digital precoder in the mid-to-low SNR regime.

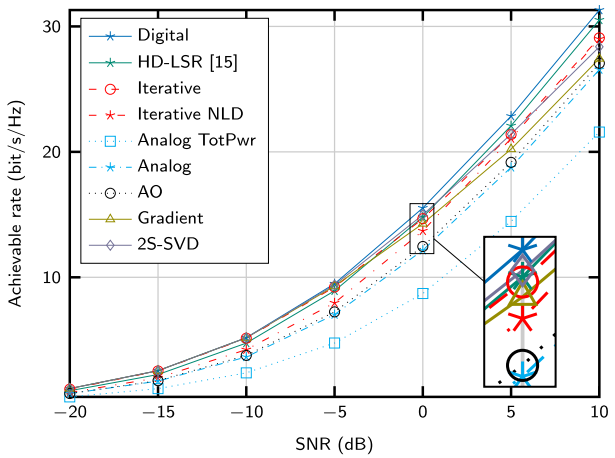

FIGURE 2. Achievable rate for different hybrid precoding strategies, $N_{\text{T}} = 128$, $N_{\text{R}} = 16$, and $L_{\text{BS}} = 8$.

Figure 3 shows the Normalized Mean-Square-Error (NMSE) for the strategies aiming at minimizing the distance metric $\|\boldsymbol{P}_{\text{D}} - \boldsymbol{P}_{\text{RF}}^0\boldsymbol{P}_{\text{BB}}\|_{\text{F}}^2$. The NMSE is defined as $\|\boldsymbol{P}_{\text{D}} - \boldsymbol{P}_{\text{RF}}^0\boldsymbol{P}_{\text{BB}}\|_{\text{F}}^2/\|\boldsymbol{P}_{\text{D}}\|_{\text{F}}^2$. As observed in Figure 3, the performance of the Gradient approach is better than that of the AO method. However, the accuracy of the approximation using the Gradient strategy degrades as the SNR increases. The main reason behind this observation is the number of data streams allocated by $\boldsymbol{P}_{\text{D}}$ for different SNRs, which

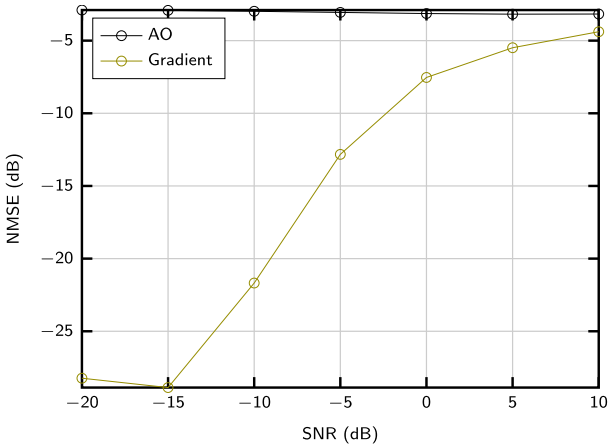

FIGURE 3. NMSE for strategies using the Frobenius norm as figure of merit, $N_{\text{T}} = 128$, $L_{\text{BS}} = 8$.

translates to a smaller number of complex-valued entries to be approximated in the low SNR regime. On the contrary, the AO procedure achieves slightly better approximations as the SNR increases.

In Fig. 4, the number of iterations for the different precoding designs is plotted. The 2S-SVD method and the iterative algorithm are especially attractive for its fast convergence speed and low computational complexity. The gradient approach works very well for low SNRs but the number of iterations quickly increases with the SNR. Finally, the AO procedure also employs more steps as the SNR increases, but the number of iterations is smaller compared to that of the gradient method for high SNRs. Regarding the comparison with hybrid precoding approaches using VPSs, the computational complexity is similar to that of [17, HD-AM] and lower to that of the evaluated method [17, HD-LSR]. However, the gradient performs the search in a convex set and a projection to the feasible set is not necessary afterwards. This way, the number of gradient steps and the computational complexity of each of them reduces in comparison to that proposed in [37], [38].

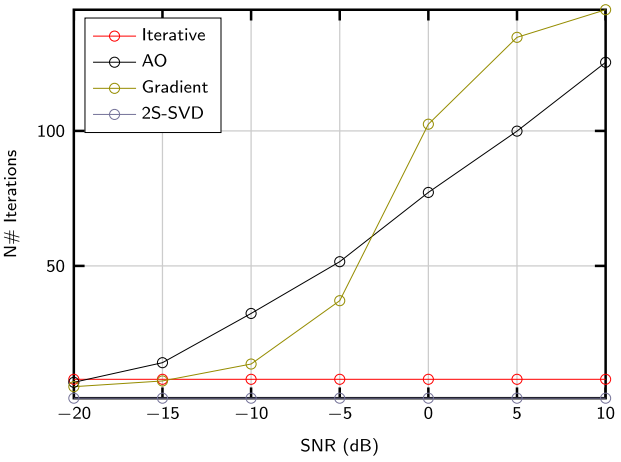

FIGURE 4. Number of iterations of the proposed algorithms for different SNRs.

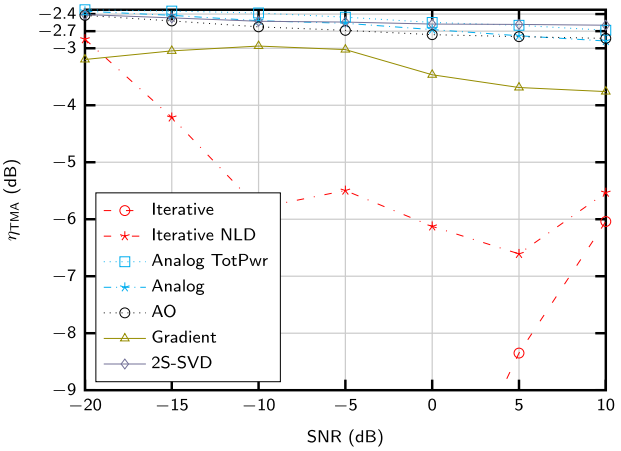

FIGURE 5. TMA efficiency considering the harmonic modes as SR for different hybrid precoding strategies.

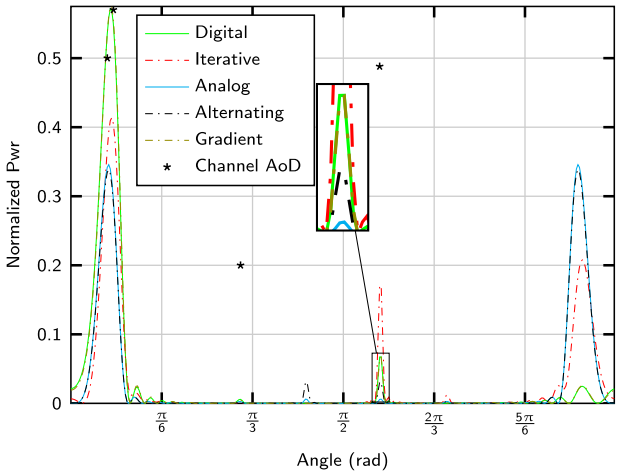

FIGURE 7. Beam pattern example of the different precoding strategies at 0dB of SNR.

The precoder efficiency is illustrated in figures 5 and 6. In particular, the TMA efficiency, η_{TMA}, is represented in Fig. 5. Analog, AO, and 2S-SVD methods present losses smaller than 3 dB for all SNR values. The losses of the Gradient method range between 3 dB and 4 dB. The hybrid iterative procedures present the poorest efficiency. We want to remark that the scaling matrix to reduce the SR (see Sec. V-C) is employed in all the approaches. Moreover, the effectiveness of removing the high correlated columns can be appreciated in this figure. As expected, it greatly increases the efficiency for low SNRs when comparing Iterative and Iterative NLD designs. On the other hand, Fig. 6 shows the efficiency related to the off-state time, η_{OFF}, for the SPST switches in Sec III-A. Again, removing the highly correlated columns is crucial to improve the efficiency in the low SNR regime. AO and Analog TotPwr approaches present losses below 3.5 dB in all cases. Finally, the gradient approach obtains the worst result under this metric. It is clear from Fig. 6 that the

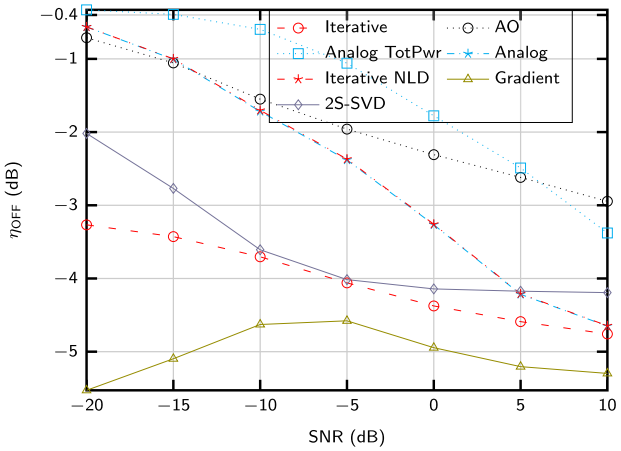

FIGURE 6. Efficiency of switching devices for the pessimistic scenario with SPST and different hybrid precoding strategies.

efficiency decreases when the number of spatial dimensions to be exploited increases, like in the high SNR regime.

Fig. 7 shows the beam pattern for a particular channel realization with $N_{\text{p},l} = 1$ and $N_{\text{cl}} = 4$. The digital curve establishes a reference to be approximated by the hybrid approaches. The only analog approach is clearly limited due to the symmetry of the beam pattern. This comes from the antisymmetric property of the cosine function in the $\{0, \pi\}$ interval, and the real entries of the analog precoder. By using hybrid precoding, it is possible to avoid this behavior up to certain extent, as observed in the Iterative and Gradient curves. In the limit case where the efficiency of the TMA is 1, the beam pattern would be that of the ULA with uniform weighting. Recall that this naive pattern focuses on the $\pi/2$ angle. This observation again exhibits the aforementioned trade-off between performance and efficiency.

In general, the performance achieved by the proposed architecture meets the one achieved with the analog precoder based on VPSs in the low and mid SNR regimes. Moreover, for these regimes, the efficiency of the TMA is larger than in the high SNR regime where all the available spatial directions are exploited to allocate the maximum number of data steams. In any case, the number of data streams to be allocated is bounded by the number of RF chains, or the number of channel paths. In typical mmWave scenarios, none of the former numbers are large. In the Non Line-of-sight (NLOS) case, low SNR values are to be expected due to blockage and large propagation losses [45], thereby leading to a small number of allocated data streams. On the other hand, for LOS scenarios like in [43], there is a dominant path with better SNR. As a consequence, one could obtain larger efficiencies by simply employing the best channel propagation cluster.

Table 4 summarizes the numerical results presented in this subsection.

B. COMPARATIVE EXAMPLE

With the aim of comparing conventional APNs based on VPSs with the herein proposed with TMAs, we have

TABLE 4. Numerical results summary.

	Achievable Rate		η_{TMA}		η_{OFF}		Convergence speed	
Hybrid Prec. Method	low-to-mid SNR	high SNR	low-to-mid SNR	high SNR	low-to-mid SNR	high SNR	low-to-mid SNR	high SNR
4-bit VPS HD-LSR [17]	high	high	n/a	n/a	n/a	n/a	mid	low
Iterative	high	high	very low	low	mid	low	high	high
Iterative NLD	mid	high	mid-to-low	low	high-to-mid	low	high	high
Analog Tot. Power	low	low	ideal$^{(*)}$	ideal$^{(*)}$	high	mid	high	high
Analog	mid	mid	high-to-mid	high-to-mid	high-to-mid	low	high	high
AO	mid	mid	high-to-mid	high-to-mid	high-to-mid	mid	mid	mid
Gradient	high	mid	mid	mid	low	low	high	mid
2S-SVD	high	mid	high-to-mid	high-to-mid	mid	mid-to-low	very high	very high

($*$) The SR has been already included in the power constraint to design the hybrid precoder.

TABLE 5. Comparative example between APNs based on VPSs and TMAs.

Comparative examples of RF analog precoding networks							
scenario	**type of APN**	**device**	**insertion losses (dB)**	**chip area (mm^2)**	**cost**	**η_{APN}(dB) (†)**	**ref.**
I (W-band)	**TMA**	1-bit VPS MEMS (60 GHz)	-1.05	<0.35	nc $^{(*)}$	-6.25	[7], [21]
		wideband RF switch (<2.7 GHz)	-0.50	1.21	very low		[26]
	VPSs	4-bit CMOS VPS (60 GHz)	-8.10	0.09	nc	-8.10	[25]
II(a)/(b) (Ka-band)	**TMA**	1-bit VPS: InGaAs pHEMT/GaAs MEMS (33 GHz)	-3.50/-1.10	1.21/0.27	medium/nc	-6.15/-3.75	[26]/ [29]
		wideband RF PIN-diode (<3 GHz)	-0.15	1.28	very low		[31]
	VPSs	5-bit VPS InGaAs pHEMT (33 GHz)	-7.00	1.21	very high$^{(**)}$	-7.00	[26]
III (W-band)	**TMA**	1-bit VPS MEMS (65 GHz)	-1.09	<0.35	nc	-3.74	[7], [21]
		wideband RF PIN-diode (<3 GHz)	-0.15	1.28	very low		[31]
	VPSs	3-bit VPS MEMS (65 GHz)	-3.20	9.60	nc	-3.20	[30]

($*$) the term very high implies that this cost is around a 500% of the cost designated as medium.
($**$) not cataloged as commercial products at these frequencies.
(†) we have set an SNR of $= -5$dB to obtain both components of η_{APN} (η_{TMA} and η_{HW}) for the different algorithms in Fig. 5 and Fig. 6, respectively. In particular, for scenarios II(a)/(b) and III we have considered the 2S-SVD algorithm and, for the scenario I, the AO algorithm.

considered four different scenarios. We have selected two different frequency bands and some of the technologies described in Section IV for each of the analog precoders (see Table 5). More specifically:

- Scenario I considers the W-band (60 GHz) and compares an APN with a recently published 4-bit CMOS VPSs to a TMA APN equipped with RF (up to 2.7 GHz) switches and 1-bit MEMS VPSs. Although the sizes of CMOS VPSs are unbeatable, the TMA solution, using a reasonable size, has better overall efficiency η_{APN} (-6.25 dB vs -8.10 dB). At this frequency band, integrated VPSs are not commercially available. Hence, in terms of cost, we only have the certainty that RF switches are rather cheap.
- Scenario II(a) focuses on the Ka-band (33 GHz). In these frequencies -and exclusively for pHEMT- it is feasible to find commercial VPSs. As a matter of fact, we compare an APN with 5-bit InGaAs pHEMT VPSs to a TMA APN equipped with RF (up to 3 GHz) PIN-diodes as switching devices and 1-bit InGaAs pHEMT VPSs. Although the insertion losses of the 1-bit VPSs are higher than the average in this technology at these frequencies [7], the equivalent η_{APN} is slightly better for the TMA case. The chip areas for the devices of the TMA analog precoder is 2.42 mm^2 while for the VPSs analog precoder is 1.21 mm^2. Nevertheless, the great difference between the previous schemes is the cost. In this case, the price (for the same manufacturer [26]) of a 5-bit VPS

is five times that of a 1-bit one. Thus, for this scenario, the TMA precoder is clearly the better solution in terms of cost.

- Scenario II(b) differs from II (a) in that the TMA APN is equipped with a 1-bit GaAs MEMS VPS (not cataloged; hence, the cost is not available) instead of the InGaAs pHEMT one. Such a 1-bit VPS -with better insertion losses and chip area- allows for accomplishing similar sizes for both schemes while providing the TMA precoder with a significantly better equivalent η_{APN} (-3.75 dB vs -7.00 dB).
- Scenario III, again in the W-band (65 GHz, with no integrated VPS cataloged), reflects an optimistic scenario for VPS APNs in terms of efficiency using a 3-bit MEMS VPS. In this case, the TMA solution is more competitive in terms of area while η_{APN} in both schemes are very close (-3.74 vs -3.20 dB) although slightly worse for the TMA case.

Therefore, in view of the above, we conclude that TMA APNs accomplish the best trade off between power efficiency, cost and size. The former discussion establishes the trade-off between performance and power efficiency. Observe that, contrarily to analog precoding with VPSs, where a Power Amplifier (PA) follows each VPS, independently of the precoding strategy, the proposed architecture can be adapted to the channel conditions in order to improve either the efficiency or the performance. This is particularly interesting in scenarios where the channel conditions change rapidly like, e.g., high mobility scenarios. Recall also that some of the strategies proposed to compute the hybrid precoders present very low computational complexity, an attractive feature for this scenario. Finally, the low cost of the proposed architecture, together with the mentioned flexibility that allows to compensate the inherent losses under some particular setup conditions, makes it possible to consider MIMO transmitters using larger numbers of RF chains.

VII. CONCLUSIONS

Hybrid digital-analog precoding for mmWave MIMO systems using TMAs has been considered. Given the importance of insertion losses in the design of mmWave systems, we have accurately characterized and quantified the total losses of the analog part of TMA hybrid precoders. In addition, three strategies for the design of TMA hybrid precoders have been studied. While the performance is similar to that obtained with VPS hybrid precoders, algorithm's convergence is faster. Moreover, power efficiency can be improved by adequately scaling the columns of the analog precoding matrix. We have shown, by comparing TMA and conventional VPS APNs, that the proposed scheme is a competitive solution in terms of cost, size, insertion losses, and performance. This work considers for the first time hybrid precoding using TMAs. Therefore, there is still room for improvement towards several directions like the use of other kind of TMA periodic pulses, the profitable exploitation of the TMA harmonic patterns, or the extension to multiuser and/or wideband scenarios.

APPENDIX
BOUNDS FOR MUTUAL INFORMATION

Based on the eigenvalues of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ in (37), we will derive a set containing the feasible values for the mutual information. We start by defining the diagonal matrix $\tilde{\boldsymbol{A}} = \text{diag}(\lambda_1, \ldots, \lambda_{L_{\text{BS}}}) \in \mathbb{R}_{+}^{L_{\text{BS}} \times L_{\text{BS}}}$, and the antisymmetric matrix $\tilde{\boldsymbol{B}} = \sum_{m=1}^{L_{\text{BS}}/2} \beta_{2m}(\boldsymbol{e}_{2m-1}\boldsymbol{e}_{2m}^{\text{T}} - \boldsymbol{e}_{2m}\boldsymbol{e}_{2m-1}^{\text{T}}) \in \mathbb{R}^{L_{\text{BS}} \times L_{\text{BS}}}$, with $\boldsymbol{e}_i$ being the canonical vector, containing the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. The decomposition of $\boldsymbol{B}$ is that in [46]. We assume that $\lambda_1 \geq \lambda_2 \geq \lambda_{L_{\text{BS}}}$ and $\beta_1 \geq \beta_2 \geq \beta_{L_{\text{BS}}/2}$. We now introduce the matrices $\boldsymbol{A}_{\text{AP}}$ and $\boldsymbol{B}_{\text{AP}}$, which are Hermitian and complex antisymmetric, respectively, as

$$\boldsymbol{A}_{\text{AP}} = \boldsymbol{U}\tilde{\boldsymbol{A}}\boldsymbol{U}^{\text{H}} \quad \text{and} \quad \boldsymbol{B}_{\text{AP}} = \boldsymbol{V}\tilde{\boldsymbol{B}}\boldsymbol{V}^{\text{H}}.$$

The former matrices are constructed with the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$ for some arbitrary orthonormal basis $\boldsymbol{U}, \boldsymbol{V} \in \mathcal{U}$. From [47, Th. 2.1] it follows that the mutual information $\mathcal{I} = \log_2 \det(\boldsymbol{A} + \boldsymbol{B})$ belongs to the set

$$\mathcal{A} = \{|\log_2 \det(\boldsymbol{A}_{\text{AP}} + \boldsymbol{B}_{\text{AP}})| : \boldsymbol{U}, \boldsymbol{V} \in \mathcal{U}\}, \tag{46}$$

which determines upper and lower bounds for the mutual information as follows

$$|\det(\tilde{\boldsymbol{A}} + \text{j}\tilde{\boldsymbol{B}})| \leq \mathcal{I} \leq \det(\tilde{\boldsymbol{A}} + \tilde{\boldsymbol{B}}). \tag{47}$$

The lower bound is valid when $\beta_1^2 < \lambda_{L_{\text{BS}}-1}\lambda_{L_{\text{BS}}}$ or $\beta_{L_{\text{BS}}/2}^2 > \lambda_1\lambda_2$ are satisfied. Otherwise, $|\det(\tilde{\boldsymbol{A}} + \text{j}\tilde{\boldsymbol{B}})|$ in (47) has to be replaced by the trivial lower bound 0.

We now highlight that, for fixed $\tilde{\boldsymbol{B}}$, an increase on the elements of $\tilde{\boldsymbol{A}}$ leads to a larger upper bound in (47). Furthermore, when the lower bound is not trivial, we distinguish two cases:

A. $\beta_{L_{BS}/2}^2 > \lambda_1\lambda_2$

This condition means that the analog precoders allocate the power in the subspace of $\boldsymbol{B}$, leading to high inter-stream interference since it is not possible to diagonalize $\boldsymbol{B}$ using real unitary basis. This situation reveals analog precoders which are not well designed.

B. $\beta_1^2 < \lambda_{L_{BS}-1}\lambda_{L_{BS}}$

If $\lambda_{L_{\text{BS}}} > \beta_1$ holds, then $\tilde{\boldsymbol{A}} + \text{j}\tilde{\boldsymbol{B}}$ is positive definite and increasing the entries in $\tilde{\boldsymbol{A}}$ for fixed $\tilde{\boldsymbol{B}}$ increases the lower bound. An example of this situation may arise in the low SNR regime. When $\lambda_{L_{\text{BS}}} < \beta_1$, $\tilde{\boldsymbol{A}} + \text{j}\tilde{\boldsymbol{B}}$ is in general indefinite and the latter reasoning does not apply. However, larger values of diagonal elements in $\tilde{\boldsymbol{A}}$ make it possible to achieve larger absolute values for the eigenvalues of $\tilde{\boldsymbol{A}} + \text{j}\tilde{\boldsymbol{B}}$, according to the relationship in [48]

$$|\lambda(\boldsymbol{X})| \leq \min\{\max_n \sum_{m=1}^{L_{\text{BS}}} |[\boldsymbol{X}]_{m,n}|, \max_n \sum_{m=1}^{L_{\text{BS}}} |[\boldsymbol{X}]_{n,m}|\},$$

where $\lambda(\boldsymbol{X})$ denotes the eigenvalue of $\boldsymbol{X}$.

We conclude that a simple and intuitive strategy is to design precoding matrices $\boldsymbol{P}_{\mathrm{RF}}$ such that the diagonal elements of $\tilde{\boldsymbol{A}}$ are large. The effect of this choice can be also appreciated by rewriting (46) as

$$\mathcal{A} = \{|\log_2 \det(\tilde{\boldsymbol{A}}) + \log_2 \det(\mathbf{I} + \tilde{\boldsymbol{A}}^{-1}\boldsymbol{C})| : \boldsymbol{U}, \boldsymbol{V} \in \mathcal{U}\},$$

with $\boldsymbol{C} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{V}\tilde{\boldsymbol{B}}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{U}$. Moreover, large entries in $\tilde{\boldsymbol{A}}$ correspond to analog precoders aiming at maximizing the diagonal elements of $\boldsymbol{A}$, whereas the effect over the off-diagonal elements in $\boldsymbol{B}$ is not taken into account [cf. (37)].

REFERENCES

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

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

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

[4] Y. Dong, C. Chen, and Y. Jin, "Joint beamforming with low-resolution PSs for millimetre-wave communications," *Electron. Lett.*, vol. 52, no. 18, pp. 1541–1543, 2016.

[5] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog–digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.

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

[7] M. C. Scardelletti and G. E. Ponchak, "RF MEMS phase shifters and their application in phase array antennas," in *Proc. IEEE Annu. Conf. Wireless Microw. Technol.*, Apr. 2005, p. 37.

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

[9] R. Maneiro-Catoira, J. Brégains, J. A. García-Naya, and L. Castedo, "Time modulated arrays: From their origin to their utilization in wireless communication systems," *Sensors*, vol. 17, no. 3, p. 590, 2017.

[10] R. Maneiro-Catoira, J. Brégains, J. García-Naya, and L. Castedo, "On the feasibility of time-modulated arrays for digital linear modulations: A theoretical analysis," *IEEE Trans. Antennas Propag.*, vol. 62, no. 12, pp. 6114–6122, Dec. 2014.

[11] J. Fondevila, J. C. Bregains, F. Ares, and E. Moreno, "Optimizing uniformly excited linear arrays through time modulation," *IEEE Antennas Wireless Propag. Lett.*, vol. 3, no. 1, pp. 298–301, Dec. 2004.

[12] J. C. Bregains, J. Fondevila-Gomez, G. Franceschetti, and F. Ares, "Signal radiation and power losses of time-modulated arrays," *IEEE Trans. Antennas Propag.*, vol. 56, no. 6, pp. 1799–1804, Jun. 2008.

[13] Q. Zhu, S. Yang, R. Yao, and Z. Nie, "Gain improvement in time-modulated linear arrays using SPDT switches," *IEEE Antennas Wireless Propag. Lett.*, vol. 11, pp. 994–997, 2012.

[14] J. Chen *et al.*, "Efficiency improvement of time modulated array with reconfigurable power divider/combiner," *IEEE Trans. Antennas Propag.*, vol. 65, no. 8, pp. 4027–4037, Aug. 2017.

[15] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 2929–2933.

[16] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[17] C. Rusu, R. Mendez-Rial, N. González-Prelcic, and R. W. Heath, Jr., "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8380–8393, Dec. 2016.

[18] R. Maneiro-Catoira, J. C. Brégains, J. A. García-Naya, L. Castedo, P. Rocca, and L. Poli, "Performance analysis of time-modulated arrays for the angle diversity reception of digital linear modulated signals," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 2, pp. 247–258, Mar. 2017.

[19] R. Maneiro-Catoira, J. C. Brégains, J. A. García-Naya, and L. Castedo, "Enhanced time-modulated arrays for harmonic beamforming," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 2, pp. 259–270, Mar. 2017.

[20] H. Fan, J. Geng, X. Liang, R. Jin, and X. Zhou, "A three-way reconfigurable power divider/combiner," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 3, pp. 986–998, Mar. 2015.

[21] V. Nasserddine, "Millimeter-wave phase shifters based on tunable transmission lines in MEMS technology post-CMOS process," Ph.D. dissertation, Institut de Microélectronique, Electromagnétisme, Photonique–Laboratoire Hyperfréquences et Caractérisation, Univ. Grenoble Alpes, Grenoble, France, 2016.

[22] A.-L. Franc, O. H. Karabey, G. Rehder, E. Pistono, R. Jakoby, and P. Ferrari, "Compact and broadband millimeter-wave electrically tunable phase shifter combining slow-wave effect with liquid crystal technology," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 11, pp. 3905–3915, Nov. 2013.

[23] S. Dey, S. K. Koul, A. K. Poddar, and U. L. Rohde, "Reliable and compact 3- and 4-bit phase shifters using MEMS SP4T and SP8T switches," *J. Microelectromech. Syst.*, vol. 27, no. 1, pp. 113–124, 2018.

[24] G. S. Shin *et al.*, "Low insertion loss, compact 4-bit phase shifter in 65 nm CMOS for 5G applications," *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 1, pp. 37–39, Jan. 2016.

[25] C. W. Byeon and C. S. Park, "A low-loss compact 60-GHz phase shifter in 65-nm CMOS," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 7, pp. 663–665, Jul. 2017.

[26] Qorvo. Accessed: Jan. 1, 2018. [Online]. Available: http://www.qorvo.com

[27] C. F. Campbell and S. A. Brown, "A compact 5-bit phase-shifter MMIC for K-band satellite communication systems," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 1, Jun. 2000, pp. 217–220.

[28] Q. Zheng *et al.*, "Design and performance of a wideband Ka-band 5-b MMIC phase shifter," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 5, pp. 482–484, May 2017.

[29] R. Malmqvist, C. Samuelsson, A. Gustafsson, D. Smith, T. Vähä-Heikkilä, and R. Baggen, "Monolithic integration of millimeter-wave RF-MEMS switch circuits and LNAs using a GaAs MMIC foundry process technology," in *Proc. IEEE MTT-S Int. Microw. Workshop Millim. Wave Integr. Technol.*, Sep. 2011, pp. 148–151.

[30] J.-J. Hung, L. Dussopt, and G. M. Rebeiz, "Distributed 2- and 3-bit W-band MEMS phase shifters on glass substrates," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 2, pp. 600–606, Feb. 2004.

[31] Infineon. Accessed: Jan. 1, 2018. [Online]. Available: http://www.infineon.com

[32] H. Issa, P. Ferrari, E. Hourdakis, and A. G. Nassiopoulou, "On-chip high-performance millimeter-wave transmission lines on locally grown porous silicon areas," *IEEE Trans. Electron Devices*, vol. 58, no. 11, pp. 3720–3724, Nov. 2011.

[33] T. Tanigawa, Y. Kondou, Y. Shimayama, and T. Irino, "Hitachi chemical working on wonders," Tokyo, Japan, Tech. Rep. 58, 2016. [Online]. Available: http://www.hitachi-chem.co.jp/english/report/058/58.pdf

[34] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

[35] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[36] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, Mar. 1966, doi: 10.1007/BF02289451.

[37] J. P. González-Coma, J. Rodríguez-Fernández, N. González-Prelcic, and L. Castedo, "Channel estimation and hybrid precoding/combining for frequency selective multiuser mmWave systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[38] J. P. González-Coma, J. Rodríguez-Fernández, N. González-Prelcic, L. Castedo, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 353–367, May 2018.

[39] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[40] S. Payami, M. Ghoraishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258–7271, Nov. 2016.
[41] Z. Pi, "Optimal transmitter beamforming with per-antenna power constraints," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, Canada, Jun. 2012, pp. 3779–3784.
[42] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation*, document 3GPP TS 36.211, 2013.
[43] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, Jul. 2017.
[44] V. Raghavan *et al.* (2017). "Millimeter-wave MIMO prototype: Measurements and experimental results." [Online]. Available: http://arxiv.org/abs/1710.09449
[45] S. Sun, G. R. MacCartney, and T. S. Rappaport, "A novel millimeter-wave channel simulator and applications for 5G wireless communications," in *Proc. Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7. [Online]. Available: http://arxiv.org/abs/1703.08232
[46] D. C. Youla, "A normal form for a matrix under the unitary congruence group," *J. Can. Math.*, vol. 13, no. 4, pp. 694–704, Jan. 1961.
[47] N. Bebiano, C.-K. Li, and J. da Providência, "Determinant of the sum of a symmetric and a skew-symmetric matrix," *SIAM J. Matrix Anal. Appl.*, vol. 18, no. 1, pp. 74–82, Jan. 1997.
[48] H. Wolkowicz and G. P. H. Styan, "Bounds for eigenvalues using traces," *Linear Algebra Appl.*, vol. 29, pp. 471–506, Feb. 1980.

JOSÉ P. GONZÁLEZ-COMA (M'17) was born in Marín, Spain, in 1986. He received the Engineering Computer and Ph.D. degrees from the University of A Coruña, Spain, in 2009 and 2015, respectively. Since 2017, he has been with the Research Center on Information and Communication Technologies (CITIC), University of A Coruña. He was awarded with a FPI Grant from the Ministerio de Ciencia e Innovación. He was appointed as a Visiting Researcher at the Associate Institute for Signal Processing (TUM), Germany, during 2012, and at the Signal Processing in Communications Group (UVIGO), Spain, in 2017. He is currently a Post-Doctoral Researcher with CITIC. His main research interests are in the designs of limited feedback, robust precoding in MIMO systems, and mmWave communications.

ROBERTO MANEIRO-CATOIRA (GS'15–M'16) received the Telecommunications Engineering degree from the University of Vigo, Spain, in 1995, and the Master of Information and Telecommunications Technologies for Mobile Networks degree and the Ph.D. degree from the University of A Coruña, Spain, in 2012 and 2017, respectively. From 1996 to 1997, he was at Egatel Company focused on the research and development in the field of TV and radio digital communications. From 1997 to 2000, he was with Siemens Mobile Networks as a GSM Access Network Deployment Manager. From 2000 to 2003, he was at Nortel Networks as an UMTS Network Integration Manager. Since 2003, he has been dedicated to teaching Siemens Simatic Programmable Logic Devices as well as mathematics for different levels, both for private and public organizations. Since 2017, he has been a Researcher with the Electronic Technology and Communications Group, University of A Coruña. His research is focused on smart antenna technologies.

LUIS CASTEDO (S'92–A'95–M'03–SM'14) received the Ph.D. degree in telecommunications engineering from the Technical University of Madrid, Spain, in 1993. Since 1994, he has been a Faculty Member with the Department of Computer Engineering, University of A Coruña (UDC), Spain, where he became a Professor in 2001 and acted as Chairman from 2003 to 2009. He is currently a Professor with UDC. His research interests are signal processing, coding, hardware prototyping, and experimental evaluation in wireless communications engineering. Prof. Castedo has co-authored over 250 papers in peer-reviewed international journals and conferences. He has also been principal investigator in more than 50 research projects funded by public organizations and private companies. His papers have received three Best Student Paper Awards at the IEEE/ITG Workshop on Smart Antennas in 2007, the IEEE International Workshop on Signal Processing Advances in Wireless Communications in 2013, and the IEEE International Conference on Internet of Things in 2017. He is the General Co-Chair of the 8th IEEE Sensor Array and Multichannel Signal Processing Workshop in 2014 and the 27th European Signal Processing Conference in 2019.