

Article

Low Complexity Hybrid Precoding Designs for Multiuser mmWave/THz Ultra Massive MIMO Systems

João Pedro Pavia ^{1,2,*}, Vasco Velez ^{1,2}, Renato Ferreira ^{1,2}, Nuno Souto ^{1,2}, Marco Ribeiro ^{1,2}, João Silva ^{1,2} and Rui Dinis ^{2,3}

¹ Department of Information Science and Technology, ISCTE-Instituto Universitário de Lisboa, 1649-026 Lisboa, Portugal; Vasco_Velez@iscte-iul.pt (V.V.); Renato_Ferreira@iscte-iul.pt (R.F.); Nuno.Souto@iscte-iul.pt (N.S.); Marco.Ribeiro@iscte-iul.pt (M.R.); Joao.Silva@iscte-iul.pt (J.S.)

² Instituto de Telecomunicações, 1049-001 Lisboa, Portugal

³ Department of Electrical and Computer Engineering, Faculty of Science and Technology, Universidade Nova de Lisboa, 1099-085 Lisboa, Portugal; rdinis@fct.unl.pt

* Correspondence: jpavia@lx.it.pt

Citation: Pavia, J.P.; Velez, V.; Ferreira, R.; Souto, N.; Ribeiro, M.; Silva, J.; Dinis, R. Low Complexity Hybrid Precoding Designs for Multiuser mmWave/THz Ultra Massive MIMO Systems. *Sensors* **2021**, *21*, 6054. <https://doi.org/10.3390/s21186054>

Academic Editor: Luís Castedo Ribas

Received: 14 July 2021

Accepted: 8 September 2021

Published: 9 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: Millimeter-wave and terahertz technologies have been attracting attention from the wireless research community since they can offer large underutilized bandwidths which can enable the support of ultra-high-speed connections in future wireless communication systems. While the high signal attenuation occurring at these frequencies requires the adoption of very large (or the so-called ultra-massive) antenna arrays, in order to accomplish low complexity and low power consumption, hybrid analog/digital designs must be adopted. In this paper we present a hybrid design algorithm suitable for both mmWave and THz multiuser multiple-input multiple-output (MIMO) systems, which comprises separate computation steps for the digital precoder, analog precoder and multiuser interference mitigation. The design can also incorporate different analog architectures such as phase shifters, switches and inverters, antenna selection and so on. Furthermore, it is also applicable for different structures, namely fully-connected structures, arrays of subarrays (AoSA) and dynamic arrays of subarrays (DAoSA), making it suitable for the support of ultra-massive MIMO (UM-MIMO) in severely hardware constrained THz systems. We will show that, by using the proposed approach, it is possible to achieve good trade-offs between spectral efficiency and simplified implementation, even as the number of users and data streams increases.

Keywords: millimeter wave (mmWave); Terahertz (THz); multiuser ultra-massive-MIMO; hybrid precoding and combining; antenna arrays

1. Introduction

Over the last few years, significant advances have been made to provide higher-speed connections to users in wireless networks, with several novel technologies being proposed to achieve this objective. However, future generations of communication systems will have to fulfil more demanding requirements that cannot be met by the methods adopted in today's communications systems. This motivates the exploration of other candidate technologies, like the millimeter wave (mmWave) and Terahertz (THz) bands, where many applications that require ultra-high data rates can be designed. These bands offer great underutilized bandwidths and also allow for a simplified implementation of large antenna arrays, which are crucial to combat the severe signal attenuation and path losses that occurs at these frequencies [1–4]. The first applications of the THz band were limited to imaging and sensing due to the unavailability of efficient devices that can work on these frequencies. However, recent advances in the field of THz devices give us strong indications that THz communications will be feasible in the near future [3]. These technologies (THz systems in particular), are expected to ease the

spectrum limitations of today's systems. They face several issues, such as the reflection and scattering losses through the transmission path, the high dependency between distance and frequency of channels at the THz band and the need for controllable time-delay phase shifters, since the phase shift will vary with frequencies based on the signal traveling time, which will also affect the system performance. These limitations require not only the proper system design, but also the definition of a set of strategies to enable communications [5,6].

The exploration of the potentialities of millimeter and sub-millimeter wavelengths is closely related to the paradigm of using very large arrays of antennas in beamforming architectures. This gives rise to so-called ultra-massive multiple-input multiple-output (UM-MIMO) systems. Still, to achieve the maximum potential of these systems, it is necessary to consider the requirements and the challenges related not only to the channel characteristics but also to the hardware component, especially regarding THz circuits [5,7,8]. Considering that high complexity and power usage are pointed out as major constraints of large-antenna systems, it is unfeasible to implement UM-MIMO schemes with a dedicated RF chain per antenna element in the mmWave and THz bands. Therefore, instead of fully-digital precoders and combiners, it becomes crucial to adopt hybrid digital/analog architectures as these require a reduced number of dedicated RF chains. By adopting this type of design, the signal processing is split into two separate parts: a low-dimensional digital part and a complementary analog part. This approach can enable a substantial reduction in the overall circuit complexity and power consumption [9]. By adopting a proper problem formulation, the analog design part can then be reduced to a simple projection operation in a flexible precoding or combining algorithm that can cope with different architectures, as we proposed in [10,11]. Despite the ultra-wide bandwidths available in mmWave and THz bands, and besides considering the problem of distance limitation, MIMO systems should take into account the operation in frequency selective channels [12]. To make the development of hybrid schemes for these systems a reality, it is necessary to handle the fading caused by multiple propagation paths typical in these types of channels [13]. Therefore, solutions inspired on multi-carrier schemes, such as orthogonal frequency division multiplexing (OFDM), are often adopted to address such problems [14].

Spectral Efficiency (SE) of point-to-point transmissions is a major concern in SingleUser (SU) and MultiUser (MU) systems. To achieve good performances, it is necessary to develop algorithms that are specially tailored to the architecture of these systems. Several hybrid precoding schemes have been proposed in the literature [15–19]. The authors of [15] proposed two algorithms for low-complexity hybrid precoding and beamforming for MU mmWave systems. Even though they assume only one stream per user, i.e., the number of data streams (N_s) is equal to the number of users (N_u), it is shown that the algorithms achieve interesting results when compared to the fully-digital solution. The concept of precoding based on adaptive RF-chain-to-antenna was only introduced in [16] for SU scenarios, but showed promising results. In [17], a nonlinear hybrid transceiver design relying on Tomlinson–Harashima precoding was proposed. Their approach only considers Fully-Connected (FC) architectures but can achieve a performance close to the fully-digital transceiver. A Kalman-based Hybrid Precoding method was proposed for MU scenarios in [18]. While designed for systems with only one stream per user and based on fully-connected structures, the performance of the algorithm is competitive with other existing solutions. A hybrid MMSE-based precoder and combiner design with low complexity was proposed in [19]. The algorithm is designed for MU-MIMO systems in narrowband channels, and it presents lower complexity and better results when compared to Kalman's precoding. Most of the hybrid solutions for mmWave systems aim to achieve near-optimal performance using FC structures, resorting to phase shifters or switches. However, the difficulty of handling the hardware constraint imposed by the analog phase shifters or by switches in the THz band is an issue that limits the expected performance in terms of SE.

Array-of-Subarrays (AoSAs) structures have gained particular attention over the last few years as a more practical alternative to FC structures, especially for the THz band. In contrast to FC structures, in which every RF chain is connected to all antennas via an individual group of phase shifters (which is prohibitive for higher frequencies), the AoSA approach allows us to have each RF chain connected to only a reduced subset of antennas. The adoption of a disjoint structure with fewer phase shifters reduces the system complexity, the power consumption and the signal power loss. Moreover, all of the signal processing can be easily carried out at the subarray level by using an adequate number of antennas [6].

Following the AoSA approach, it was shown in [20] that, to balance SE and power consumption in THz communications, adaption and dynamic control capabilities should be included in the hybrid precoding design. Therefore, Dynamic Arrays-of-Subarrays (DAoSAs) architectures could be adopted. The same authors proposed a DAoSA hybrid precoding architecture which can intelligently adjust the connections between RF chains and subarrays through a network of switches. Their results showed that it is possible to achieve a good trade-off for the balancing between the SE and power consumption.

Within the context of multiuser downlink scenarios, the authors of [21] studied some precoding schemes considering THz massive MIMO systems for Beyond 5th Generation (B5G) networks. Besides showing the impact on EE and SE performance, carrier frequency, bandwidth and antenna gains, three different precoding schemes were evaluated and compared. It was observed that the hybrid precoding approach with baseband Zero Forcing for multiuser interference mitigation (HYB-ZF) achieved much better results than an Analog-only Beamsteering (AN-BST) scheme with no baseband precoder. In fact, this approach was capable of better approaching the upper bound defined by the singular value decomposition precoder (SVD-UB). The other relevant conclusion is that the design of precoding algorithms should be adapted to the communication schemes. While considering all the specific constraints may allow the maximization of the system performance of the system, formulating and solving the corresponding optimization problem may not be so simple. Motivated by the work above, in this paper we developed an algorithm for hybrid precoding design which can accommodate different low-complexity architectures suitable for both mmWave and THz MU-MIMO systems. It is based on the idea of accomplishing a near-optimal approximation of the fully-digital precoder for any configuration of antennas, RF chains and data streams through the application of the alternating direction method of multipliers (ADMM) [22]. ADMM is a well-known and effective method for solving convex optimization problems but can also be a powerful heuristic for several non-convex problems [22,23]. To use it effectively within the context of MU-MIMO, THE proper formulation of the hybrid design problem as a multiple constrained matrix factorization problem is first presented. Using the proposed formulation, an iterative algorithm comprising several reduced complexity steps is obtained.

The main contributions of this paper can be summarized as follows:

- We propose a hybrid design algorithm with near fully-digital performance, where the digital precoder, analog precoder and multiuser interference mitigation are computed separately through simple closed-form solutions. Even though the hybrid design algorithm is developed independently of a specific channel or antenna configuration, it is particularly suitable for mmWave and THz systems where, on the one hand, very large antenna arrays are required to overcome distance limitations but, on the other hand, current hardware constraints in terms of cost and power consumption make the adoption fully-digital precoders/combiners with one dedicated RF chain per antenna element unviable. Whereas our previous work [10] also proposed a hybrid design algorithm for mmWave, it did not address multiuser systems, and in particular the MIMO broadcast channel. Therefore, it does not include any step for inter-user interference mitigation within its design. As we show here, for this multiuser channel, the hybrid design method must also deal with the

residual inter-user interference as it can degrade system performance, particularly at high Signal Noise Ratios (SNRs);

- Due to the separability of the different steps (analog precoder, digital precoder and interference suppression), the proposed algorithm can incorporate different architectures, making it suitable for supporting UM-MIMO in severely hardware-constrained systems typical in the THz band. Unlike [10], where we only considered the adoption of phase shifters, in this paper we present explicit solutions for some of the most common architectures, namely FC, AoSA and DAoSA structures based in either Unquantized Phase Shifters (UPS), Quantized Phase Shifters (QPS), Switches (Swi), Switches and Inverters (SI), Antenna Selection (AS) or Double Phase Shifters (DPS);
- To cope with the large bandwidths available in mmWave/THz bands, where practical MIMO systems likely have to operate in frequency selective channels, the proposed hybrid design considers the application in a multicarrier context, where the same analog precoder is applied at different frequencies;
- We explicitly show how the proposed design can be applied to a DAoSAs approach where a reduced number of switches are inserted at each AoSA panel, which allows the connections to the RF chains to be dynamically adjusted. Through extensive simulations, it is shown that our proposed solution is capable of achieving good trade-offs between spectral efficiency, hardware complexity and power consumption, proving to be a suitable solution for the deployment of UM-MIMO, especially in hardware-constrained THz systems.

The paper is organized as follows: Section II presents the adopted system model. The adopted formulation of the hybrid design problem for the MU-MIMO scenario and the proposed algorithm are described in detail in Section III, which includes the implementation of the algorithm for different analog architectures. Performance results are then presented in Section IV. Finally, the conclusions are outlined in Section V.

Notation: Matrices and vectors are denoted by uppercase and lowercase boldface letters, respectively. The superscript $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and conjugate transpose of a matrix/vector, $\|\cdot\|_p$ is the ℓ_p -norm of a vector, $\|\cdot\|_0$ is its cardinality (i.e., the number of non-zero elements in a vector which is sometimes referred to as the ℓ_0 -norm in the literature) and \mathbf{I}_n is the $n \times n$ identity matrix.

2. System Model

In this section, we present the system and channel models adopted for the design of the hybrid precoding algorithm. Let us consider the OFDM base system illustrated in Figure 1. In this case, we have a mmWave/THz hybrid multiuser MIMO system, where a base station (BS) is equipped with N_{tx} antennas and transmits to N_u users equipped with N_{rx} antennas over F carriers, as can be seen in Figure 1. On each subcarrier, N_s data streams are transmitted to each user which are represented as $\mathbf{s}_k = [\mathbf{s}_{k,1}^T \dots \mathbf{s}_{k,N_u}^T]^T$, with $\mathbf{s}_{k,u} \in \mathbb{C}^{N_s \times 1}$. Instead of a fully-digital design which would require a dedicated RF chain per antenna element, both the precoder and combiner comprise separate digital and analog processing blocks. This approach allows for the use of reduced digital blocks with only a few RF chains, which are complemented by the analog blocks, that can be supported solely on networks of phase shifters and switches. Since the analog precoder (combiner) is located after (before) the IFFT (FFT) blocks, it is shared between the different subcarriers, as in [24,25]. Regarding the analog precoder and combiner, which are represented by matrices $\mathbf{F}_{RF} \in \mathbb{C}^{N_{tx} \times N_{RF}^{tx}}$ and $\mathbf{W}_{RF,u} \in \mathbb{C}^{N_{rx} \times N_{RF}^{rx}}$ with $u = 1, \dots, N_u$, it is assumed that $N_u N_s \leq N_{RF}^{tx} \leq N_{tx}$ and $N_s \leq N_{RF}^{rx} \leq N_{rx}$, where N_{RF}^{tx} and N_{RF}^{rx} are the number of

RF chains at the BS and each user, respectively. The received signal model at subcarrier k after the combiner can be written as

$$\mathbf{Y}_{k,u} = \sqrt{\rho_u} \mathbf{W}_{\text{BB},k,u}^H \mathbf{W}_{\text{RF},k,u}^H \mathbf{H}_{k,u} \mathbf{F}_{\text{RF},k,u} \mathbf{F}_{\text{BB},k} \mathbf{s}_k + \mathbf{W}_{\text{BB},k,u}^H \mathbf{W}_{\text{RF},k,u}^H \mathbf{n}_{k,u}, \quad (1)$$

where $\mathbf{H}_{k,u} \in \mathbb{C}^{N_{rx} \times N_{tx}}$ is the frequency domain channel matrix (assumed to be perfectly known at the transmitter and receiver) between the base station and the u th receiver at subcarrier k . Vector $\mathbf{n}_{k,u} \in \mathbb{C}^{N_{rx} \times 1}$ contains independent zero-mean circularly symmetric Gaussian noise samples with covariance $\sigma_n^2 \mathbf{I}_{N_{rx}}$ and ρ_u denotes the average received power. The digital baseband precoders and combiners are denoted by $\mathbf{F}_{\text{BB},k} \in \mathbb{C}^{N_{RF}^{tx} \times N_u N_s}$ and $\mathbf{W}_{\text{BB},k,u} \in \mathbb{C}^{N_{RF}^{rx} \times N_s}$, respectively. Regarding the channel model, it is important to note that even though the mmWave and THz bands share a few commonalities, the THz channel has several peculiarities that distinguish it from the mmWave channel. For example, the very high scattering and diffraction losses in the THz band will typically result in a much sparser channel in the angular domain with fewer multipath components (typically less than 10) [21]. Furthermore, the gap between the line of sight (LOS) and non-line of sight (NLOS) components tends to be very large, making it often LOS-dominant with NLOS-assisted [25]. An additional aspect relies on the much larger bandwidth of THz signals which can suffer performance degradation due to the so-called beam split effect, where the transmission paths squint into different spatial directions depending on the subcarrier frequency [20]. In light of this, in this paper we consider a clustered wideband geometric channel, which is commonly adopted both in the mmWave [15] and THz literature [20,25–28]. However, it should be noted that the hybrid precoding/combining approach proposed in this paper is independent of a specific MIMO channel. In this case, the frequency domain channel matrices can be characterized as

$$\begin{aligned} \mathbf{H}_{k,u} = & \gamma \left(\alpha_u^{\text{LOS}} \mathbf{a}_r(\phi_u^{\text{r,LOS}}, \theta_u^{\text{r,LOS}}) \mathbf{a}_t(\phi_u^{\text{t,LOS}}, \theta_u^{\text{t,LOS}})^H + \right. \\ & \left. + \sum_{i=1}^{N_d} \sum_{l=1}^{N_{ray}} \alpha_{i,l,u} \mathbf{a}_r(\phi_{i,l,u}^{\text{r}}, \theta_{i,l,u}^{\text{r}}) \mathbf{a}_t(\phi_{i,l,u}^{\text{t}}, \theta_{i,l,u}^{\text{t}})^H \right) e^{-j2\pi\tau_{i,u}f_k}, \end{aligned} \quad (2)$$

where N_d denotes the scattering clusters with each cluster i having a time delay of $\tau_{i,u}$ and N_{ray} propagations paths. α_u^{LOS} and $\alpha_{i,l,u}$ are the complex gains of the LOS component and of the l th ray from cluster i . Index u is the user ($u = 1, \dots, N_u$), $f_k = f_c + \frac{B}{F} \left(k - 1 - \frac{F-1}{2} \right)$ ($k = 1, \dots, F$) is the k th subcarrier frequency, B is the bandwidth, f_c is the central frequency and γ is a normalizing factor such that $\mathbb{E} \left[\|\mathbf{H}_{k,u}\|_F^2 \right] = N_{tx} N_{rx}$. Vectors $\mathbf{a}_t(\phi_{i,l,u}^{\text{t}}, \theta_{i,l,u}^{\text{t}})$ and $\mathbf{a}_r(\phi_{i,l,u}^{\text{r}}, \theta_{i,l,u}^{\text{r}})$ represent the transmit and receive antenna array responses at the azimuth and elevation angles of $(\phi_{i,l,u}^{\text{t}}, \theta_{i,l,u}^{\text{t}})$ and $(\phi_{i,l,u}^{\text{r}}, \theta_{i,l,u}^{\text{r}})$, respectively. Vectors $\mathbf{a}_t(\phi_u^{\text{t,LOS}}, \theta_u^{\text{t,LOS}})$ and $\mathbf{a}_r(\phi_u^{\text{r,LOS}}, \theta_u^{\text{r,LOS}})$ have similar meanings but refer to the LOS path angles $(\phi_u^{\text{t,LOS}}, \theta_u^{\text{t,LOS}})$ and $(\phi_u^{\text{r,LOS}}, \theta_u^{\text{r,LOS}})$. By carefully selecting the parameters of the channel model we can make it depict a mmWave or a THz channel. Considering Gaussian signaling, the spectral efficiency achieved by the system for the transmission to MS- u in subcarrier k is [28]

$$\mathbf{R}_{k,u} = \log_2 \left| \mathbf{I}_{N_{RF}^{rx}} + \mathbf{R}_u^{-1} \mathbf{W}_{BB,k,u}^H \mathbf{W}_{RF,u}^H \mathbf{H}_{k,u} \mathbf{F}_{RF} \mathbf{F}_{BB,k,u} \times \mathbf{F}_{BB,k,u}^H \mathbf{F}_{RF}^H \mathbf{H}_{k,u}^H \mathbf{W}_{RF,u} \mathbf{W}_{BB,k,u} \right|, \quad (3)$$

where $\mathbf{R}_{k,u}$ is the covariance matrix of the total inter-user interference plus noise at MS- u , which is characterized by

$$\mathbf{R}_{k,u} = \mathbf{W}_{BB,k,u}^H \mathbf{W}_{RF,u}^H \left(\mathbf{H}_{k,u} \sum_{j \neq u}^{N_u} \mathbf{F}_{RF} \mathbf{F}_{BB,k,j} \mathbf{F}_{BB,k,j}^H \mathbf{F}_{RF}^H \mathbf{H}_{k,u}^H + \sigma^2 \mathbf{I}_{N_{RF}^{rx}} \right) \mathbf{W}_{RF,u} \mathbf{W}_{BB,k,u}. \quad (4)$$

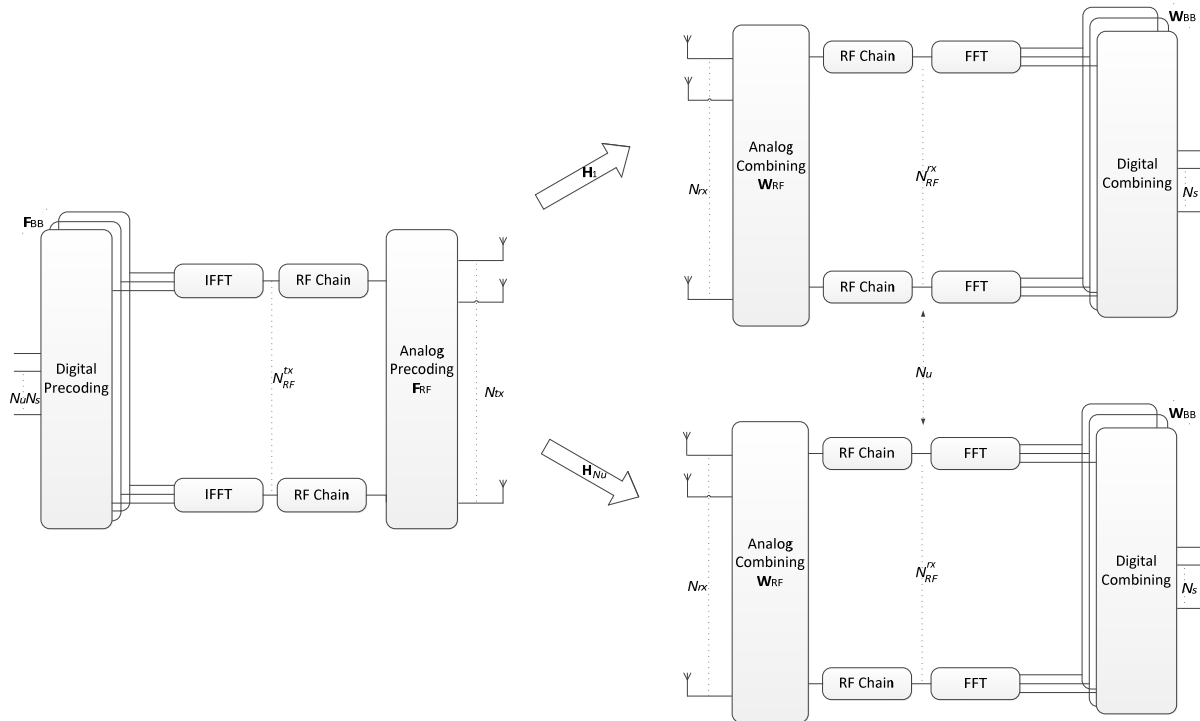


Figure 1. A multiuser OFDM mmWave/THz MIMO system with hybrid precoding.

3. Proposed Hybrid Design Algorithm

In this section, we will introduce the algorithm for the hybrid precoding problem and show how it can be adapted to different architectures. Although we will focus on the precoder design, a similar approach can be adopted for the combiner. However, since our design assumes that inter-user interference suppression is applied at the transmitter, only single-user detection is required at the receiver and therefore the algorithm reduces to the one described in [10].

3.1. Main Algorithm

Although there are several problem formulations for the hybrid design proposed in the literature, one of the most effective relies on the minimization of the Frobenius norm of the difference between the fully-digital precoder and the hybrid precoder [22,29–31]. In this paper we follow this matrix approximation-based approach. First, we compute fully-digital precoders, which we assume to be designed so as to enforce zero inter-user interference, using, for example, the block-diagonalization (BD) approach described in [32]. Using the BD procedure, we obtain one different digital precoder matrix for each subcarrier, \mathbf{F}_{opt_k} (with $k = 1, \dots, F$, and $\mathbf{F}_{opt_k} = [\mathbf{F}_{opt_{k,1}}, \dots, \mathbf{F}_{opt_{k,N_u}}]$), which satisfy $\mathbf{H}_{k,u} \mathbf{F}_{k,u} = \mathbf{0}$ for all $u' \neq u$ ($u, u' = 1, \dots, N_u$), thus guaranteeing no inter-user interference.

Using these digital precoder matrices, we design the hybrid digital/analog precoder by solving a matrix approximation problem formulated as

$$\min_{\mathbf{F}_{RF}, \mathbf{F}_{BB_k}} \sum_{k=1}^F \left\| \mathbf{F}_{opt_k} - \mathbf{F}_{RF} \mathbf{F}_{BB_k} \right\|_F^2 \quad (5)$$

$$\text{subject to } \mathbf{F}_{RF} \in \mathcal{C}_{N_{tx} \times N_{RF}^{tx}} \quad (6)$$

$$\left\| \mathbf{F}_{RF} \mathbf{F}_{BB_k} \right\|_F^2 = N_u N_s \quad (7)$$

where Equation (7) enforces the transmitter's total power constraint and $\mathcal{C}_{N_{tx} \times N_{RF}^{tx}}$ is the set of feasible analog precoding matrices, which is defined according to the adopted RF architecture (it will be formally defined for several different architectures in the next subsection). Matrix \mathbf{F}_{opt_k} denotes the fully-digital precoder. Even if \mathbf{F}_{opt_k} is selected in order to cancel all interference between users, the hybrid design resulting as a solution of Equations (5)–(7) will correspond to an approximation and, as such, residual inter-user interference will remain. To avoid the performance degradation that will result from this, an additional constraint can be added to the problem formulation, namely

$$\sum_{\substack{u'=1 \\ u' \neq u}}^{N_u} \mathbf{H}_{k,u'} \mathbf{F}_{RF} \mathbf{F}_{BB_{k,u}} = 0, \quad k=1, \dots, F, \quad u=1, \dots, N_u \quad (8)$$

where $\mathbf{F}_{BB_{k,u}} = \mathbf{F}_{BB_k} [:(u-1)N_s + 1 : uN_s]$. This restriction is equivalent to enforcing $\mathbf{F}_{RF} \mathbf{F}_{BB_{k,u}}$ to lie in the null space of $\bar{\mathbf{H}}_{k,u} \in \mathbb{C}^{(N_u-1)N_{rx} \times N_{tx}}$ ($\bar{\mathbf{H}}_{k,u}$ is a matrix corresponding to \mathbf{H}_k with the N_{rx} lines of user u removed) which we denote as $\mathcal{N}(\bar{\mathbf{H}}_{k,u})$. The overall optimization problem can be then expressed as

$$\min_{\mathbf{F}_{RF}, \mathbf{F}_{BB_k}} \sum_{k=1}^F \left\| \mathbf{F}_{opt_k} - \mathbf{F}_{RF} \mathbf{F}_{BB_k} \right\|_F^2 \quad (9)$$

$$\text{subject to } \mathbf{F}_{RF} \in \mathcal{C}_{N_{tx} \times N_{RF}^{tx}} \quad (10)$$

$$\left\| \mathbf{F}_{RF} \mathbf{F}_{BB_k} \right\|_F^2 = N_u N_s \quad (11)$$

$$\mathbf{F}_{RF} \mathbf{F}_{BB_{k,u}} \in \mathcal{N}(\bar{\mathbf{H}}_{k,u}), \quad k=1, \dots, F, \quad u=1, \dots, N_u. \quad (12)$$

To derive a hybrid precoder/design algorithm that can cope with the different RF architectures, we can integrate the RF constraint directly into the objective function of the optimization problem. This can be accomplished through the addition of an auxiliary variable, \mathbf{R} , combined with the use of the indicator function. The indicator function for a generic set \mathcal{A} is defined as $I_{\mathcal{A}}(\mathbf{x})$, returning 0 if $\mathbf{x} \in \mathcal{A}$ and $+\infty$ otherwise. A similar approach can be adopted for integrating the other constraints, Equations (11) and (12), also into the objective function. The optimization problem can then be rewritten as

$$\begin{aligned} \min_{\mathbf{F}_{RF}, \mathbf{F}_{BB_k}, \mathbf{R}, \mathbf{B}_k, \mathbf{F}_{approx_{k,u}}} \sum_{k=1}^F \left\| \mathbf{F}_{opt_k} - \mathbf{F}_{RF} \mathbf{F}_{BB_k} \right\|_F^2 + I_{\mathcal{C}_{N_{tx} \times N_{RF}^{tx}}}(\mathbf{R}) + \\ + \sum_{k=1}^F I_{\left\| \cdot \right\|_F^2 = N_u N_s}(\mathbf{B}_k) + \sum_{k=1}^F \sum_{u=1}^{N_u} I_{\mathcal{N}(\bar{\mathbf{H}}_{k,u})}(\mathbf{F}_{approx_{k,u}}) \end{aligned} \quad (13)$$

$$\text{subject to } \mathbf{R} = \mathbf{F}_{\text{RF}} \quad (14)$$

$$\mathbf{B}_k = \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}_k} \quad (15)$$

$$\mathbf{F}_{\text{aprox}_{k,\mu}} = \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}_k}, \quad (16)$$

where $\mathbf{F}_{\text{aprox}_k} = [\mathbf{F}_{\text{aprox}_{k,1}}, \dots, \mathbf{F}_{\text{aprox}_{k,N_u}}]$. The augmented Lagrangian function (ALF) for Equations (13)–(16) can be written as

$$\begin{aligned} L_{\rho,\eta,\mu}(\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}, \mathbf{R}, \mathbf{B}, \mathbf{F}_{\text{aprox}}, \mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Gamma}) = & \sum_{k=1}^F \left\| \mathbf{F}_{\text{opt}_k} - \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}_k} \right\|_F^2 + I_{C_{N_{\text{tx}} \times N_{\text{RF}}}}(\mathbf{R}) + \\ & + \sum_{k=1}^K I_{\|\cdot\|_F^2 = N_u N_s}(\mathbf{B}_k) + \sum_{k=1}^F \sum_{u=1}^{N_u} I_{\mathcal{A}(\bar{\mathbf{h}}_{k,\mu})}(\mathbf{F}_{\text{aprox}_{k,\mu}}) + 2 \operatorname{Re}\{\operatorname{tr}(\mathbf{\Lambda}^H (\mathbf{F}_{\text{RF}} - \mathbf{R}) + \\ & + \operatorname{tr}(\mathbf{\Psi}_k^H \sum_{k=1}^F (-\mathbf{B}_k + \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}_k})) + \operatorname{tr}(\mathbf{\Gamma}_k^H \sum_{k=1}^F (-\mathbf{F}_{\text{aprox}_k} + \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}_k}))\} + \rho \|\mathbf{F}_{\text{RF}} - \mathbf{R}\|_F^2 + \\ & + \eta \sum_{k=1}^F \left\| -\mathbf{B}_k + \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}_k} \right\|_F^2 + \mu \sum_{k=1}^F \left\| -\mathbf{F}_{\text{aprox}_k} + \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}_k} \right\|_F^2, \end{aligned} \quad (17)$$

where $\mathbf{\Lambda} \in \mathbb{C}^{N_{\text{tx}} \times N_{\text{RF}}}$, $\mathbf{\Psi} \in \mathbb{C}^{N_{\text{tx}} \times N_{\text{RF}}}$ and $\mathbf{\Gamma} \in \mathbb{C}^{N_{\text{tx}} \times N_{\text{RF}}}$ are dual variables and ρ , η , μ are penalty parameters. After some straightforward algebraic manipulation and working with scaled dual variables $\mathbf{U} = \mathbf{\Lambda} / \rho$, $\mathbf{W}_k = \mathbf{\Psi} / \eta$ and $\mathbf{Z}_k = \mathbf{\Gamma}_k / \mu$ we can rewrite the ALF as

$$\begin{aligned} L_{\rho,\eta,\mu}(\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}, \mathbf{R}, \mathbf{B}, \mathbf{F}_{\text{aprox}}, \mathbf{U}, \mathbf{W}, \mathbf{Z}) = & \sum_{k=1}^F \left\| \mathbf{F}_{\text{opt}_k} - \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}_k} \right\|_F^2 + I_{C_{N_{\text{tx}} \times N_{\text{RF}}}}(\mathbf{R}) + \\ & + \sum_{k=1}^F I_{\|\cdot\|_F^2 = N_u N_s}(\mathbf{B}_k) + \sum_{k=1}^F \sum_{u=1}^{N_u} I_{\mathcal{A}(\bar{\mathbf{h}}_{k,\mu})}(\mathbf{F}_{\text{aprox}_{k,\mu}}) + \rho \|\mathbf{F}_{\text{RF}} - \mathbf{R} + \mathbf{U}\|_F^2 - \rho \|\mathbf{U}\|_F^2 + \\ & + \eta \sum_{k=1}^F \left\| -\mathbf{B}_k + \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}_k} + \mathbf{W}_k \right\|_F^2 - \eta \sum_{k=1}^F \|\mathbf{W}_k\|_F^2 + \\ & + \mu \sum_{k=1}^F \left\| -\mathbf{F}_{\text{aprox}_k} + \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}_k} + \mathbf{Z}_k \right\|_F^2 - \mu \sum_{k=1}^F \|\mathbf{Z}_k\|_F^2. \end{aligned} \quad (18)$$

In the following, we apply ADMM [22] as a heuristic for solving problem Formulation (13)–(16). To accomplish this, we can apply the gradient ascent to the dual problem involving the ALF, which allows us to obtain an iterative precoding algorithm comprising the following sequence of steps. We start with the minimization of the ALF over \mathbf{F}_{RF} for iteration $t+1$ defined as

$$\mathbf{F}_{\text{RF}}^{(t+1)} = \min_{\mathbf{F}_{\text{RF}}} L_{\rho,\eta,\mu}(\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}^{(t)}, \mathbf{R}^{(t)}, \mathbf{B}^{(t)}, \mathbf{F}_{\text{aprox}}^{(t)}, \mathbf{U}^{(t)}, \mathbf{W}^{(t)}, \mathbf{Z}^{(t)}), \quad (19)$$

which can be obtained from

$$\nabla_{\mathbf{F}_{\text{RF}}} L_{\rho,\eta,\mu}(\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}^{(t)}, \mathbf{R}^{(t)}, \mathbf{B}^{(t)}, \mathbf{F}_{\text{aprox}}^{(t)}, \mathbf{U}^{(t)}, \mathbf{W}^{(t)}, \mathbf{Z}^{(t)}) = 0 \quad (20)$$

leading to the closed form expression

$$\mathbf{F}_{\text{RF}}^{(t+1)} = \left[\sum_{k=1}^F [\mathbf{F}_{\text{opt}_k} + \eta(\mathbf{B}_k^{(t)} - \mathbf{W}_k^{(t)}) + \mu(\mathbf{F}_{\text{aprox}_k}^{(t)} - \mathbf{Z}_k^{(t)})] \times \mathbf{F}_{\text{BB}_k}^{(t)H} + \rho(\mathbf{R}^{(t)} - \mathbf{U}^{(t)}) \right] \times \\ \times [(1 + \eta + \mu) \sum_{k=1}^F \mathbf{F}_{\text{BB}_k}^{(t)} \mathbf{F}_{\text{BB}_k}^{(t)H} + \rho \mathbf{I}_{N_{\text{RF}}^{2t}}]^{-1}. \quad (21)$$

After obtaining the expression for $\mathbf{F}_{\text{RF}}^{(t+1)}$, $\mathbf{F}_{\text{BB}}^{(t+1)}$ can be found by following the same methodology. In this case the minimization is expressed as

$$\mathbf{F}_{\text{BB}}^{(t+1)} = \min_{\mathbf{F}_{\text{BB}}} L_{\rho, \eta, \mu}(\mathbf{F}_{\text{RF}}^{(t)}, \mathbf{F}_{\text{BB}}, \mathbf{R}^{(t)}, \mathbf{B}^{(t)}, \mathbf{F}_{\text{aprox}}^{(t)}, \mathbf{U}^{(t)}, \mathbf{W}^{(t)}, \mathbf{Z}^{(t)}), \quad (22)$$

which can be obtained from

$$\nabla_{\mathbf{F}_{\text{BB}}^H} L_{\rho, \eta, \mu}(\mathbf{F}_{\text{RF}}^{(t)}, \mathbf{F}_{\text{BB}}, \mathbf{R}^{(t)}, \mathbf{B}^{(t)}, \mathbf{F}_{\text{aprox}}^{(t)}, \mathbf{U}^{(t)}, \mathbf{W}^{(t)}, \mathbf{Z}^{(t)}) = 0 \quad (23)$$

and leads to the closed form expression

$$\mathbf{F}_{\text{BB}_k}^{(t+1)} = (1 + \eta + \mu) \left(\mathbf{F}_{\text{RF}}^{(t+1)H} \mathbf{F}_{\text{RF}}^{(t+1)} \right)^{-1} \mathbf{F}_{\text{RF}}^{(t+1)H} \cdot \\ \cdot \left(\mathbf{F}_{\text{opt}_k} + \eta(\mathbf{B}_k^{(t)} - \mathbf{W}_k^{(t)}) + \mu(\mathbf{F}_{\text{aprox}_k}^{(t)} - \mathbf{Z}_k^{(t)}) \right), k = 1, \dots, F. \quad (24)$$

The next steps consist of the minimization over \mathbf{R} and \mathbf{B}_k . The minimization of Equation (18) with respect to \mathbf{R} and \mathbf{B}_k can be written as

$$\mathbf{R}^{(t+1)} = \min_{\mathbf{R}} \left\{ \mathcal{I}_{\mathcal{C}_{N_{\text{Tx}} \times N_{\text{RF}}}}(\mathbf{R}) + \rho \left\| \mathbf{F}_{\text{RF}}^{(t+1)} - \mathbf{R} + \mathbf{U}^{(t)} \right\|_F^2 \right\} \\ = \Pi_{\mathcal{C}_{N_{\text{Tx}} \times N_{\text{RF}}}} \left(\mathbf{F}_{\text{RF}}^{(t+1)} + \mathbf{U}^{(t)} \right), \quad (25)$$

and

$$\mathbf{B}_k^{(t+1)} = \min_{\mathbf{B}_k} \left\{ \mathcal{I}_{\|\cdot\|_F^2 = N_u N_s}(\mathbf{B}_k) + \eta \left\| \mathbf{F}_{\text{RF}}^{(t+1)} \mathbf{F}_{\text{BB}_k}^{(t+1)} - \mathbf{B}_k + \mathbf{W}^{(t)} \right\|_F^2 \right\} \\ = \Pi_{\|\cdot\|_F^2 = N_u N_s} \left(\mathbf{F}_{\text{RF}}^{(t+1)} \mathbf{F}_{\text{BB}_k}^{(t+1)} + \mathbf{W}^{(t)} \right), k = 1, \dots, F, \quad (26)$$

where $\Pi_{\mathcal{C}_{a \times b}}(\cdot)$ and $\Pi_{\|\cdot\|_F^2 = N_u N_s}(\cdot)$ denote the projection onto set $\mathcal{C}_{a \times b}$ and onto the set of matrices whose squared Frobenius norm is $N_u N_s$, respectively. While the former projection depends on the adopted analog architecture and will be explained in the next subsection, the second projection is simply computed as

$$\mathbf{B}_k^{(t+1)} = \frac{\left(\mathbf{F}_{\text{RF}}^{(t+1)} \mathbf{F}_{\text{BB}_k}^{(t+1)} + \mathbf{W}^{(t)} \right) \sqrt{N_u N_s}}{\left\| \mathbf{F}_{\text{RF}}^{(t+1)} \mathbf{F}_{\text{BB}_k}^{(t+1)} + \mathbf{W}^{(t)} \right\|_F^2}. \quad (27)$$

The minimization of (18) with respect of $\mathbf{F}_{\text{aprox}_{k,\mu}}$ can be written as

$$\mathbf{F}_{\text{aprox}_{k,\mu}}^{(t+1)} = \min_{\mathbf{F}_{\text{aprox}_{k,\mu}}} \left\{ \mathcal{I}_{\mathcal{A}(\overline{\mathbf{H}}_{k,\mu})}(\mathbf{F}_{\text{aprox}_{k,\mu}}) + \mu \left\| \mathbf{F}_{\text{RF}}^{(t+1)} \mathbf{F}_{\text{BB}_{k,(u-1)N_s+1:uN_s}}^{(t+1)} - \mathbf{F}_{\text{aprox}_{k,\mu}} + \mathbf{Z}_{k,(u-1)N_s+1:uN_s}^{(t)} \right\|_F^2 \right\}$$

$$= \Pi_{\mathcal{N}(\bar{\mathbf{H}}_{k,u})} \left(\mathbf{F}_{\text{RF}}^{(t+1)} \mathbf{F}_{\text{BB}_{k,(u-1)N_s+1:uN_s}}^{(t+1)} + \mathbf{Z}_{k,(u-1)N_s+1:uN_s}^{(t)} \right), \quad (28)$$

which also involves a projection, $\Pi_{\mathcal{N}(\bar{\mathbf{H}}_{k,u})}(\cdot)$, but in this case onto the null-space of $\bar{\mathbf{H}}_{k,u}$. Let us use \mathbf{A} to denote $\mathbf{A} = \mathbf{F}_{\text{RF}}^{(t+1)} \mathbf{F}_{\text{BB}_{k,u}}^{(t+1)} + \mathbf{Z}_{k,(u-1)N_s+1:uN_s}^{(t)}$. The procedure to compute the projection of matrix \mathbf{A} onto the null-space of $\bar{\mathbf{H}}_{k,u}$ can be formulated as another optimization problem, which can be expressed as

$$\min \sum_{i=1}^{N_s} \|\mathbf{A}_{:,i} - \mathbf{X}_{:,i}\|_F^2 \quad (29)$$

$$\text{subject to } \bar{\mathbf{H}}_{k,u} \mathbf{X}_{:,i} = 0. \quad (30)$$

The general solution for this problem is presented in [29] corresponding to

$$\mathbf{X}_{:,i} = \left(\mathbf{I}_{N_{\text{rx}}} - \bar{\mathbf{H}}_{k,u}^H \left(\bar{\mathbf{H}}_{k,u} \bar{\mathbf{H}}_{k,u}^H \right)^{-1} \bar{\mathbf{H}}_{k,u} \right) \mathbf{A}_{:,i}, \quad i = 1, \dots, N_s. \quad (31)$$

Reordering the column vectors in the original matrix form results in the final expression which can be rewritten as

$$\begin{aligned} \mathbf{X} &= \left(\mathbf{I}_{N_{\text{rx}}} - \bar{\mathbf{H}}_{k,u}^H \left(\bar{\mathbf{H}}_{k,u} \bar{\mathbf{H}}_{k,u}^H \right)^{-1} \bar{\mathbf{H}}_{k,u} \right) \mathbf{A} \\ &= \left(\mathbf{I}_{N_{\text{rx}}} - \bar{\mathbf{V}}_{k,u}^{(1)} \left(\bar{\mathbf{V}}_{k,u}^{(1)} \right)^H \right) \mathbf{A}. \end{aligned} \quad (32)$$

In this expression, $\bar{\mathbf{V}}_{k,u}^{(1)}$ denotes the matrix containing the right singular vectors corresponding to the non-zero singular values associated with the singular value decomposition (SVD) given by $\bar{\mathbf{H}}_{k,u} = \bar{\mathbf{U}}_{k,u} \bar{\mathbf{\Lambda}}_{k,u} \left[\bar{\mathbf{V}}_{k,u}^{(1)} \bar{\mathbf{V}}_{k,u}^{(0)} \right]^H$. Therefore, to compute matrix \mathbf{X} , one can perform a single value decomposition of $\bar{\mathbf{H}}_{k,u}$ and then use this to remove the projection of \mathbf{A} onto the row space of $\bar{\mathbf{H}}_{k,u}$. Finally, the expressions for the update of dual variables \mathbf{U} , \mathbf{W} and \mathbf{Z} are given by

$$\mathbf{U}^{(t+1)} = \mathbf{U}^{(t)} + \mathbf{F}_{\text{RF}}^{(t+1)} - \mathbf{R}^{(t+1)}, \quad (33)$$

$$\mathbf{W}_k^{(t+1)} = \mathbf{W}_k^{(t)} + \mathbf{F}_{\text{RF}}^{(t+1)} \mathbf{F}_{\text{BB}_k}^{(t+1)} - \mathbf{B}_k^{(t+1)}, \quad (34)$$

$$\mathbf{Z}_k^{(t+1)} = \mathbf{Z}_k^{(t)} + \mathbf{F}_{\text{RF}}^{(t+1)} \mathbf{F}_{\text{BB}_k}^{(t+1)} - \mathbf{F}_{\text{aprox}_k}^{(t+1)}. \quad (35)$$

Appropriate values for the penalty parameters can be obtained in a heuristic manner by performing numerical simulations. Regarding the initialization and termination of the algorithm, the same approach described in [10] can be adopted. The whole algorithm is summarized in Table 1. In this table, Q denotes the maximum number of iterations. The projection operation is the only step specific to the implemented architecture, as will be explained in the next subsection.

Table 1. General Iterative Hybrid Design Algorithm.

1:	Input: $\mathbf{F}_{\text{opt}_k}^{(0)}, \mathbf{F}_{\text{RF}}^{(0)}, \mathbf{F}_{\text{BB}_k}^{(0)}, \mathbf{R}^{(0)}, \mathbf{B}_k^{(0)}, \mathbf{F}_{\text{approx}_{k,u}}^{(0)}, \rho, Q$
2:	for $t = 0, 1, \dots, Q-1$ do
3:	Compute $\mathbf{F}_{\text{RF}}^{(t+1)}$ using (21).
4:	Compute $\mathbf{F}_{\text{BB}_k}^{(t+1)}$ using (24), for all $k = 1, \dots, F$.
5:	Compute $\mathbf{R}^{(t+1)}$ using (25).
6:	Compute $\mathbf{B}_k^{(t+1)}$ using (26), for all $k = 1, \dots, F$.
7:	Compute $\mathbf{F}_{\text{approx}_{k,u}}^{(t+1)}$ using (28), for all $k = 1, \dots, F$ and $u = 1, \dots, N_u$.
8:	Update $\mathbf{U}^{(t+1)}$ using (33).
9:	Update $\mathbf{W}_k^{(t+1)}$ using (34), for all $k = 1, \dots, F$.
10:	Update $\mathbf{Z}_k^{(t+1)}$ using (35), for all $k = 1, \dots, F$.
11:	end for.
12:	$\hat{\mathbf{F}}_{\text{RF}} \leftarrow \mathbf{R}^{(Q)}$.
13:	$\hat{\mathbf{F}}_{\text{BB}_k} \leftarrow \left(\hat{\mathbf{F}}_{\text{RF}}^H \hat{\mathbf{F}}_{\text{RF}} \right)^{-1} \hat{\mathbf{F}}_{\text{RF}}^H \mathbf{F}_{\text{approx}_k}^{(Q)}$, for all $k = 1, \dots, F$.
14:	$\hat{\mathbf{F}}_{\text{BB}_k} \leftarrow \sqrt{N_u N_s} \left\ \hat{\mathbf{F}}_{\text{BB}_k}^H \hat{\mathbf{F}}_{\text{BB}_k} \right\ _F^{-1} \hat{\mathbf{F}}_{\text{BB}_k}$.
15:	Output: $\hat{\mathbf{F}}_{\text{RF}}, \hat{\mathbf{F}}_{\text{BB}}$.

3.2. Analog RF Precoder/Combiner Structure

The projection required for obtaining matrix \mathbf{R} in step 5 of the precoding algorithm has to be implemented according to the specific analog beamformer [6,20,33–37]. This makes the proposed scheme very generic, allowing it to be easily adapted to different RF architectures. In the following, we will consider a broad range of architectures that can be adopted for the RF precoder for achieving reduced complexity and power consumption implementations. We will consider FC, AoSA and DAoSA structures as illustrated in Figure 2, where we assume single phase shifters (SPS). Besides SPS, we will also consider other alternative implementations for these structures, as illustrated in Figure 3 for AoSA. The different solutions either rely on selectors, switches, inverters or phase shifters, or combinations of these. The overall analog structure is defined as a combination of one of the architectures in Figure 2 with either SPS or one of the alternatives illustrated in Figure 3.

(1) Unquantized Phase Shifters (UPS)

In the first case, we consider the use of infinite resolution phase shifters which, while being ideal, are often used as a reference benchmark. For this architecture the RF constraint set is given by

$$\mathcal{C}_{a \times b} = \left\{ \mathbf{X} \in \mathbb{C}^{a \times b} : |X_{i,j}| = 1 \right\} \quad (36)$$

and the corresponding projection can be performed simply using

$$\mathbf{R}^{(t+1)} = \left(\mathbf{F}_{\text{RF}}^{(t+1)} + \mathbf{W}^{(t)} \right) \oslash \left| \mathbf{F}_{\text{RF}}^{(t+1)} + \mathbf{W}^{(t)} \right|, \quad (37)$$

where \oslash denotes the Hadamard (i.e., element-wise) division.

(2) Quantized Phase Shifters (QPS)

The second case considers a more realistic scenario, in which phase shifters can be digitally controlled with N_b bits. These devices allow the selection of 2^{N_b} different quantized phases and the RF constraint set becomes

$$\mathcal{C}_{a \times b} = \left\{ \mathbf{X} \in \mathbb{C}^{a \times b} : X_{i,j} = e^{2\pi k i / 2^{N_b}}, k=0, \dots, 2^{N_b} - 1 \right\}. \quad (38)$$

The implementation of the projection in line 5 of Table 1 can be obtained as the following element-wise quantization

$$R_{i,j}^{(t+1)} = e^{k_{0, \dots, 2^{N_b}-1} \left\{ \text{angle} \left(F_{RF,j}^{(t+1)} + W_{i,j}^{(t)} \right) - 2\pi k / 2^{N_b} \right\}}, i = 1, \dots, N_{tx}, j = 1, \dots, N_{RF}^{tx}. \quad (39)$$

Phase shifters are typically one of the best solutions for analog processing blocks but also have a higher implementation cost and power consumption, especially when they have high resolutions.

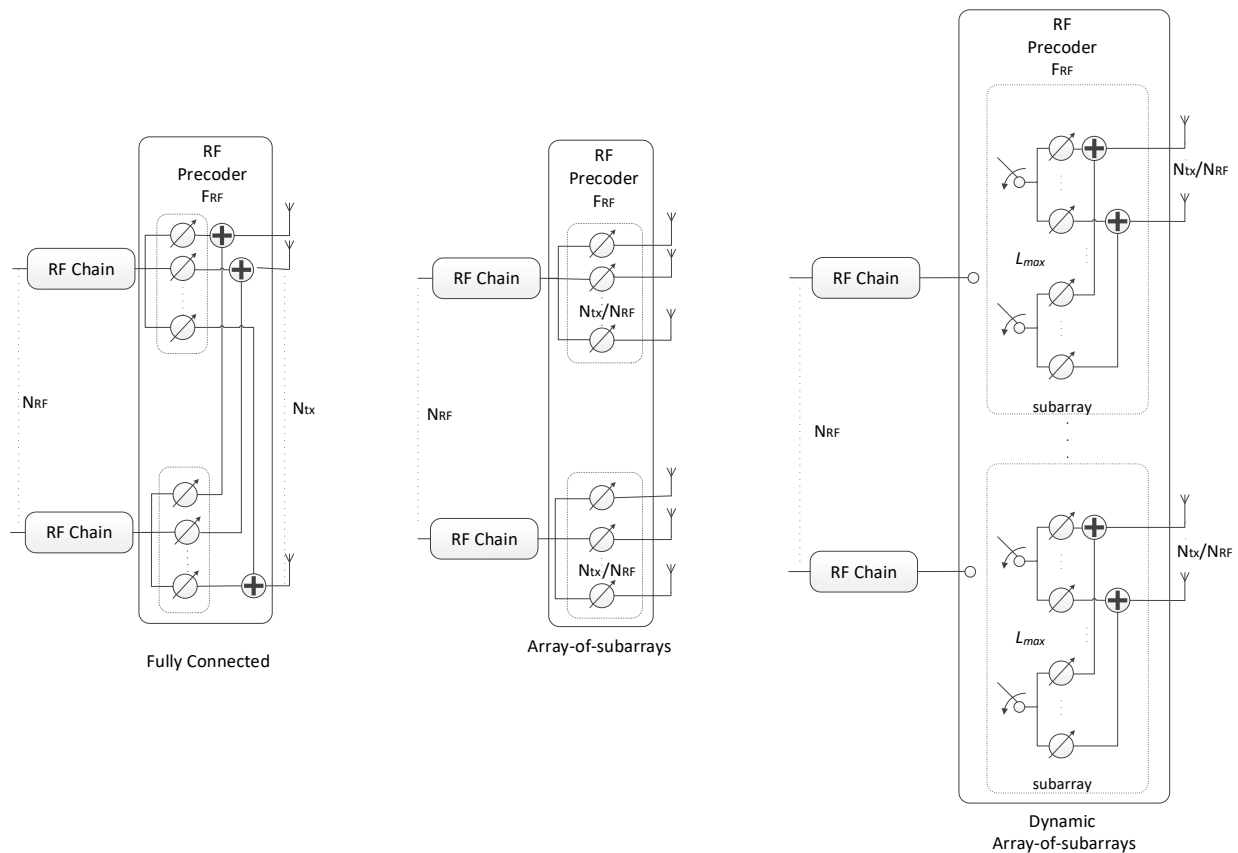


Figure 2. Different precoder architectures for a mmWave/THz MIMO system based on phase shifters.

(3) Double Phase Shifters (DPS)

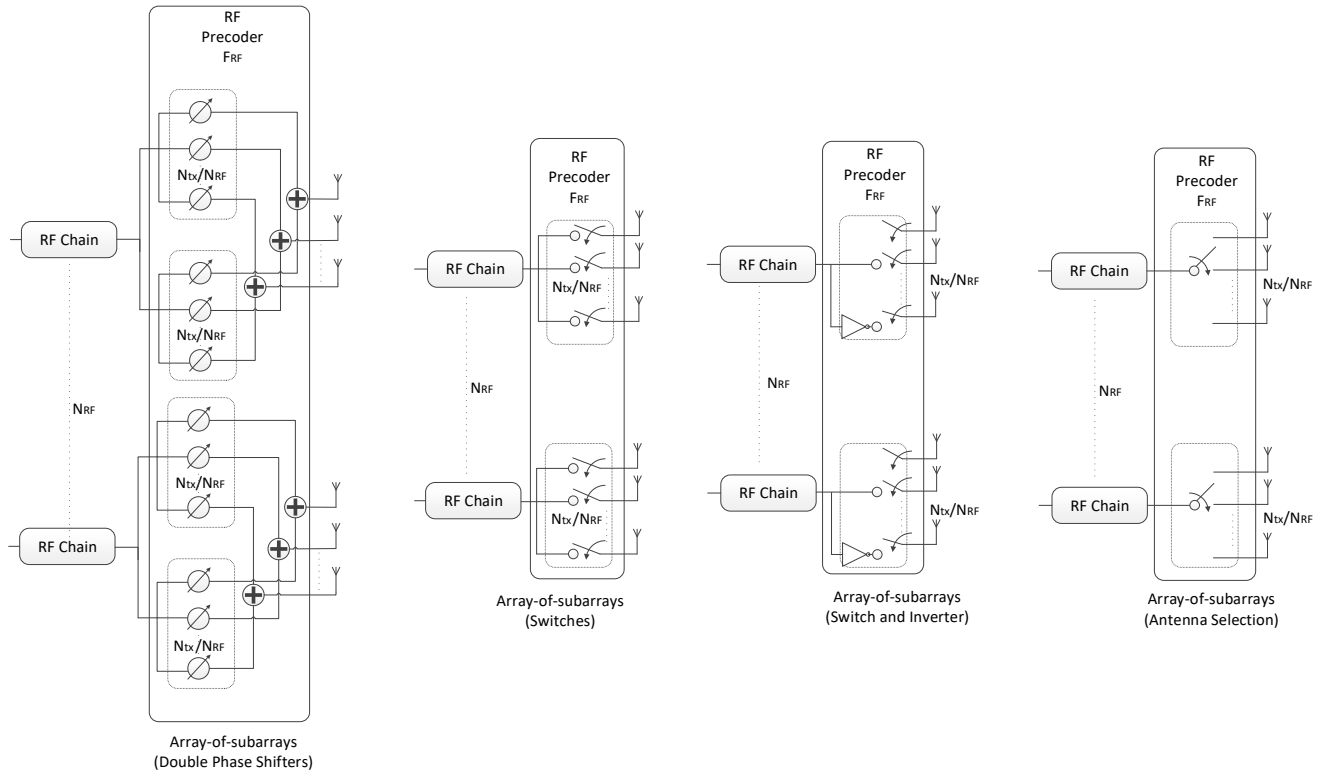


Figure 3. Alternative implementations to single phase shifters based on array-of-subarrays for a mmWave/THz MIMO system.

Another appealing architecture relies on the use of double phase shifters (DPS) since these remove the constant modulus restriction on the elements of \mathbf{F}_{RF} , following the idea in [37]. The main difference between SPS and DPS structures relies on the number of phase shifters in use to compose each connection from an RF chain to a connected antenna element, which in this case is doubled. Even though it increases the implementation complexity and power consumption, this solution can increase the spectral efficiency and approach the performance of the fully-digital one [37]. In this case, the projection can be implemented element-wise simply as

$$R_{i,j}^{(t+1)} = \left(F_{\text{RF},i,j}^{(t+1)} + W_{i,j}^{(t)} \right) - e^{i \cdot \text{angle}\left(F_{\text{RF},i,j}^{(t+1)} + W_{i,j}^{(t)} \right)} \times \max\left(0, \left| F_{\text{RF},i,j}^{(t+1)} + W_{i,j}^{(t)} \right| - 2 \right) \quad (40)$$

Similarly to other architectures, DPS can be used not only in the fully-connected approach but also in the AoSA and DAoSA cases, replacing the constant modulus setting operation.

(4) Switches (Swi)

In an architecture based on switches, each of the variable phase shifters can be replaced by a switch which typically consumes less power [34]. This simplification results in a network of switches connecting each RF chain to the antennas. The RF constraint set can be represented as

$$\mathcal{C}_{a \times b} = \left\{ \mathbf{X} \in \mathbb{R}^{a \times b} : X_{i,j} = 0 \text{ or } X_{i,j} = 1 \right\} \quad (41)$$

(set of matrices having solely '0's or '1's as elements) and the projection can be implemented element-wise as

$$R_{i,j}^{(t+1)} = 1/2 + 1/2 \cdot \text{sign}\left(2 \text{Re}\left[F_{\text{RF},j}^{(t+1)} + W_{i,j}^{(t)}\right] - 1\right). \quad (42)$$

(5) Switches and Inverters (SI)

Assuming that $N_b = 1$, then each variable phase shifter of the QPS architecture can be replaced by a pair of switched lines, including also an inverter. The corresponding constraint set can be reduced to

$$C_{a \times b} = \left\{ \mathbf{X} \in \mathbb{R}^{a \times b} : X_{i,j} = \pm 1 \right\} \quad (43)$$

and the implementation of the projection simplifies to

$$R_{i,j}^{(t+1)} = \text{sign}\left(\text{Re}\left[F_{\text{RF},j}^{(t+1)} + W_{i,j}^{(t)}\right]\right). \quad (44)$$

(6) Antenna Selection (AS)

The simplest scenario that we can consider corresponds to an architecture, where each RF chain can be only connected to a single antenna (and vice-versa). Antenna selection is a low-cost low-complexity alternative where only a specified subset of antennas are active at any given time [34]. The RF constraint set will comprise a matrix with only one non-zero element per column and per row, i.e.,

$$C_{a \times b} = \left\{ \mathbf{X} \in \mathbb{R}^{a \times b} : X_{i,j} = 0 \text{ or } X_{i,j} = 1, \|X_{i,:}\|_0 = 1, \|X_{:,j}\|_0 = 1 \right\}. \quad (45)$$

In the definition, $\|\cdot\|_0$ represents the cardinality of a vector (sometimes referred to as the ℓ_0 -norm in the literature). Defining $\mathbf{X} = \mathbf{F}_{\text{RF}}^{(t+1)} + \mathbf{W}^{(t)}$, the projection can be approximately implemented by setting all the elements in \mathbf{X} as 0 except for $X_{t_j,j} = 1$, where t_j is the row position with the highest real component in column j :

$$t_j = \arg \max_{i=1, \dots, N_{\text{tx}}} \left\{ \text{Real}\left[X_{i,j}\right] \right\}. \quad (46)$$

The computation of t_j is performed for all columns $j = 1, \dots, N_{\text{RF}}^{\text{tx}}$, sorted by descending order in terms of highest real components. It should be noted that during this operation, the same row cannot be repeated.

(7) Array-of-Subarrays (AoSAs)

Within the context of UM-MIMO, one of the most appealing architectures for keeping the complexity acceptable relies on the use of AoSA, where each RF chain is only connected to one or more subsets of antennas (subarrays). Denoting the number of subarrays as n_{SA} , which is typically set as $n_{\text{SA}} = N_{\text{RF}}$, and the size of each subarray as

$N_{\text{tx}}^{\text{SA}}$, then we have $N_{\text{tx}}^{\text{SA}} = \frac{N_{\text{tx}}}{n_{\text{SA}}} = \frac{N_{\text{tx}}}{N_{\text{RF}}}$. To limit the complexity of the architecture, each RF

chain can connect to a maximum of L_{max} consecutive subarrays. In this case, the RF constraint set comprises matrices where each column has a maximum of L_{max} blocks of $N_{\text{tx}}^{\text{SA}}$ constant modulus elements, with all the remaining elements being zero. Defining

$\mathbf{X} = \mathbf{F}_{\text{RF}}^{(t+1)} + \mathbf{W}^{(t)}$, the projection can be implemented by setting all the elements in \mathbf{X} as 0 except for the subblocks in each column j which fulfill

$$\left\| \mathbf{X}_{\left\{ \left((j-1)N_{\text{tx}}^{\text{SA}} + (i-1)N_{\text{tx}}^{\text{SA}} + 1 : (j-1)N_{\text{tx}}^{\text{SA}} + iN_{\text{tx}}^{\text{SA}} \right) \bmod N_{\text{tx}} \right\} + 1, j} \right\|_1 > \frac{N_{\text{tx}}^{\text{SA}}}{2} \quad (47)$$

with $i = 1, \dots, L_{\max}$ and $j = 1, \dots, N_{RF}$. In this case, the corresponding elements of \mathbf{R} are set as $R_{i,j}^{(t+1)} = (X_{i,j}) / |X_{i,j}|$, assuming UPS in these connections. Clearly, the phase shifters can be replaced by any of the other alternatives presented previously.

(8) Dynamic Array-of-Subarrays (DAoSAs)

As a variation of the previous AoSA architecture, we also consider an implementation where each subarray can be connected to a maximum of L_{\max} RF chains (which can be non-adjacent). In this case, the constraint set comprises matrices where each $N_{tx}^{SA} \times N_{RF}$ component submatrix contains a maximum of L_{\max} columns with constant modulus elements. The rest of the matrix contains only zeros. In this case, starting with $\mathbf{X} = 0$, the projection can be obtained by selecting the L_{\max} columns of

$$\left\| \mathbf{X}_{\lfloor (j-1)N_{tx}^{SA} + (i-1)N_{tx}^{SA} + 1 : (j-1)N_{tx}^{SA} + iN_{tx}^{SA} \rfloor} \right\|_1 > \frac{N_{tx}^{SA}}{2} \quad (48)$$

where $j = 1, \dots, n_{SA}$ with the largest ℓ_1 -norm and setting the corresponding elements of \mathbf{R} as $R_{i,j}^{(t+1)} = (X_{i,j}) / |X_{i,j}|$, assuming the use of UPS. Care must be taken to guarantee that at least one sub-block will be active in every column of \mathbf{R} . Similarly to the AoSA, the phase shifters can be replaced by any of the other presented alternatives.

3.3. Complexity

In the proposed algorithm, the $\mathbf{F}_{RF}^{(t+1)}$ and $\mathbf{F}_{BB}^{(t+1)}$ updates (steps 3 and 4 in Table 1) are defined using closed-form expressions that encompass several matrix multiplications, sums and an $N_{RF} \times N_{RF}$ matrix inverse (with an assumed complexity order of $\mathcal{O}(N_{RF}^3)$). These steps require a complexity order of $\mathcal{O}(QN_u N_s N_{RF} N_{tx} + F^{-1} Q N_{RF}^2 N_{tx})$ and $\mathcal{O}(QN_u N_s N_{RF} N_{tx} + Q N_{RF}^2 N_{tx})$, respectively.

The $\mathbf{R}^{(t+1)}$ update (step 5) involves simple element-wise division (assuming UPS) with $\mathcal{O}(Q N_{RF} N_{tx})$ while variable $\mathbf{B}_k^{(t+1)}$ (step 6) comprises a Frobenius norm computation with $\mathcal{O}(Q N_u N_s N_{RF} N_{tx})$. Step 7, the $\mathbf{F}_{\text{approx},u}^{(t+1)}$ update, has a complexity order of $\mathcal{O}(Q N_{tx}^2 N_u N_s + N_u^3 N_{tx} N_{rx}^2 + N_u^4 N_{rx}^3)$, whereas the dual variables updates (steps 8–10) have a complexity of $\mathcal{O}(Q N_u N_s N_{RF} N_{tx} + Q N_{RF}^2 N_{tx})$. Therefore, keeping only the dominant terms, the overall complexity order for the proposed algorithm is $\mathcal{O}(Q(N_{tx}^2 N_u N_s + N_{RF}^2 N_{tx}) + N_u^3 N_{tx} N_{rx}^2 + N_u^4 N_{rx}^3)$. Table 2 presents the total complexity order of the proposed method and compares it against other existing low-complexity alternatives, namely AM – Based [15], LASSO – Based Alt-Min (SPS and DPS) [14] and element-by-element (EBE) [20] algorithms. Taking into account that in UM-MIMO, N_{tx} will tend to be very large, it means the algorithms with higher complexity will typically be EBE and the one proposed in this paper due to the terms $\mathcal{O}(Q N_{tx}^2)$ and $\mathcal{O}(Q N_{tx}^2 N_u N_s)$. It is important to note, however, that while the computational complexity of these two design methods may be higher, both algorithms can be applied to simple AoSA/DAoSA architectures. In particular, the proposed approach directly supports structures with lower practical implementation complexity (and are more energy-efficient) such as those based on switches. Furthermore, in a single-user scenario, the interference cancellation step of the proposed algorithm is unnecessary, and the complexity reduces to $\mathcal{O}(Q(N_u N_s N_{RF} N_{tx} + N_{RF}^2 N_{tx}))$. Regarding the other algorithms, they have similar

complexities. However, the AM-based algorithm is designed for single stream scenarios whereas the others consider multiuser multi-stream scenarios.

Table 2. Overall Complexity of Different Hybrid Precoding Algorithms (per subcarrier).

AM—Based	
Operation	Complexity Order
Overall [15]	$\mathcal{O}\left(Q\left(N_u N_s N_{RF} N_{tx} + N_{RF}^2 N_u N_s + F^{-1} N_{RF}^3\right) + F^{-1} N_{RF}^2 N_{tx} + N_u^3 N_s^3\right)$
LASSO—Based Alt-Min (SPS)	
Operation	Complexity Order
Overall [14]	$\mathcal{O}\left(Q\left(N_u N_s N_{RF} N_{tx} + N_{RF}^2 N_u N_s + F^{-1} N_{RF}^3\right) + N_u^2 N_s N_{RF} N_{tx} + N_u^4 N_s^3\right)$
ADMM	
Operation	Complexity Order
Overall [10]	$\mathcal{O}\left(Q\left(N_s N_{RF} N_{tx} + N_{RF}^2 N_{tx}\right)\right)$
EBE	
Operation	Complexity Order
Overall [20]	$\mathcal{O}\left(Q N_{tx}^2\right)$
Proposed	
Operation	Complexity Order
\mathbf{F}_{RF}	$\mathcal{O}\left(Q N_u N_s N_{RF} N_{tx} + F^{-1} Q N_{RF}^2 N_{tx}\right)$
\mathbf{F}_{BB}	$\mathcal{O}\left(Q N_u N_s N_{RF} N_{tx} + Q N_{RF}^2 N_{tx}\right)$
\mathbf{R}	$\mathcal{O}\left(Q N_{RF} N_{tx}\right)$
\mathbf{B}	$\mathcal{O}\left(Q N_u N_s N_{RF} N_{tx}\right)$
$\mathbf{F}_{\text{approx}}$	$\mathcal{O}\left(Q N_{tx}^2 N_u N_s + N_u^3 N_{tx} N_{rx}^2 + N_u^4 N_{rx}^3\right)$
$\mathbf{U}, \mathbf{W}, \mathbf{Z}$	$\mathcal{O}\left(Q N_u N_s N_{RF} N_{tx} + Q N_{RF}^2 N_{tx}\right)$
Overall	$\mathcal{O}\left(Q\left(N_{tx}^2 N_u N_s + N_{RF}^2 N_{tx}\right) + N_u^3 N_{tx} N_{rx}^2 + N_u^4 N_{rx}^3\right)$

4. Numerical Results

In this section, the performance of the proposed algorithm will be evaluated and compared against other existing alternatives from the literature, considering multiuser MIMO systems. We consider that both the transmitter and receivers are equipped with uniform planar arrays (UPAs) with $\sqrt{N_{tx}} \times \sqrt{N_{tx}}$ antenna elements at the transmitter and $\sqrt{N_{rx}} \times \sqrt{N_{rx}}$ at the receiver. The respective array response vectors are given by

$$\mathbf{a}_{t/r}(\phi_{i,l,u}^{t/r}, \theta_{i,l,u}^{t/r}) = \frac{1}{\sqrt{N_{tx/rx}}} \times \left[1, \dots, e^{j \frac{2\pi}{\lambda} d (p \sin \phi_{i,l,u}^{t/r} \sin \theta_{i,l,u}^{t/r} + q \cos \theta_{i,l,u}^{t/r})}, \dots, e^{j \frac{2\pi}{\lambda} d ((\sqrt{N_{tx/rx}} - 1) \sin \phi_{i,l,u}^{t/r} \sin \theta_{i,l,u}^{t/r} + (\sqrt{N_{tx/rx}} - 1) \cos \theta_{i,l,u}^{t/r})} \right]^T, \quad (49)$$

where $p, q = 0, \dots, \sqrt{N_{tx/rx}} - 1$ are the antenna indices, λ is the signal wavelength and d is the inter-element spacing, which we assume to be $d = \lambda/2$. We consider a sparse channel with limited scattering where $N_{ray} = 4$ and $N_d = 6$. The angles of departure and arrival were selected according to a Gaussian distribution whose means are uniformly distributed in $[0, 2\pi]$ and whose angular spreads are 10 degrees. While we include a few results for a NLOS channel, which is often considered a possible scenario in mmWave communications [14,15,25,31], we also present results for a channel with a LOS component which is more realistic, especially in the THz band. In the scenarios with a LOS component, a ratio of $E[|\alpha_u^{LOS}|^2] / \sum_{i=1}^{N_d} \sum_{l=1}^{N_{ray}} E[|\alpha_{i,l,u}|^2] = 10$ is assumed (in this case we are admitting very weak NLOS paths compared to LOS which is typical in the THz band [27]). A fully-digital combiner was considered at each receiver and all simulation results were computed with 5000 independent Monte Carlo runs.

4.1. Fully-Connected Structures

First, we evaluate the performance assuming a fully-connected structure. Simulation results for a scenario where a base station with $N_{tx} = 100$ antennas transmits a single data stream ($N_s = 1$) to $N_u = 4$ users with $N_{rx} = 4$ antennas are shown in Figure 4 for $F = 1$ and Figure 5 for $F = 64$. The number of RF chains in the transmitter (N_{RF}^{tx}) is equal to $N_u N_s$. Besides our proposed precoder, several alternative precoding schemes are compared against the fully-digital solution, namely the LASSO-Based Alt-Min, the AM-Based and ADMM-Based precoding [10,14,15].

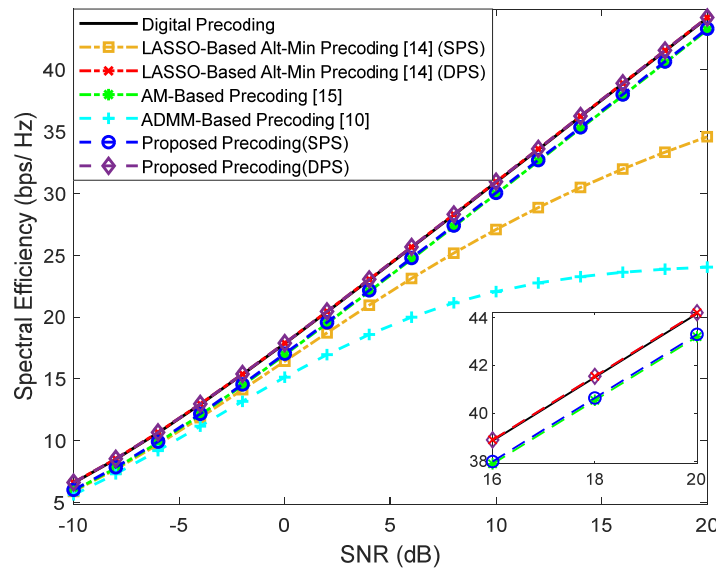


Figure 4. Spectral efficiency versus SNR achieved by different methods with $N_u = 4$, $N_s = 1$, $N_{RF}^{tx} = 4$, $F = 1$, $N_{tx} = 100$ and $N_{rx} = 4$ (only NLOS).

It can be observed that when $F = 1$, only the LASSO-Based Alt-Min with SPS and the ADMM-Based precoder from [10] (which does not remove the inter-user interference) lie far from fully-digital precoder. All the others achieve near optimum results and, in fact, can even match them when adopting DPS (proposed approach and LASSO-based Alt-Min). As explained in Section II, whereas for $F = 1$ we have F_{BB} and F_{RF} designed for

that specific carrier, when $F = 64$, \mathbf{F}_{RF} has to be common to all subcarriers. While this reduces the implementation complexity, it also results in a more demanding restriction that makes the approximation of $\mathbf{F}_{\text{opt}_k}$ (Problems (5)–(7)) to become worse. Additionally, when this approximation worsens, there can also be increased interference between users. Therefore, it can be observed in the results of Figure 5, that the gap between the fully-digital precoder and all the different hybrid algorithms is substantially wider. Still, the proposed precoder manages to achieve the best results.

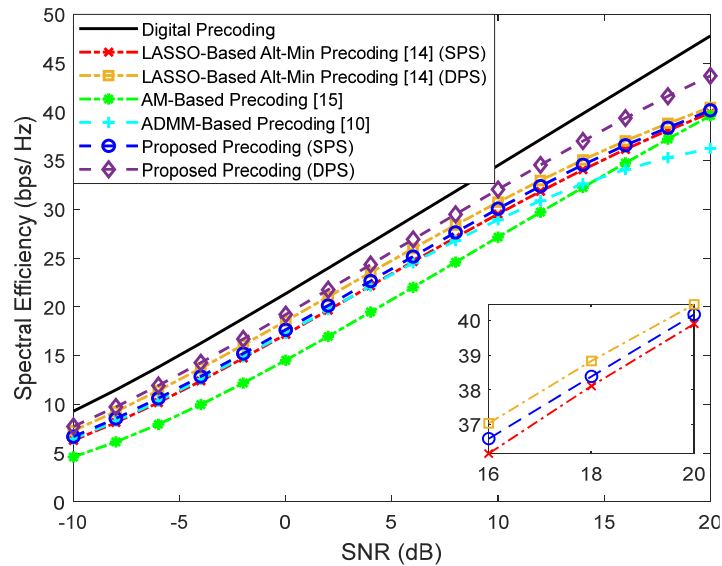


Figure 5. Spectral efficiency versus SNR achieved by different methods with $N_u = 4$, $N_s = 1$, $N_{\text{RF}}^{\text{tx}} = 8$, $F = 64$, $N_{\text{tx}} = 100$ and $N_{\text{rx}} = 4$ (only NLOS).

Given the performances of the different approaches, it is important to keep in mind that the AM-based precoding algorithm has the lowest performance in wideband but also one of the lowest computational complexities (see Table 2 of Section 3.3). In general, the proposed precoding algorithm is the one that can achieve better results at the cost of some additional computational complexity. Later on, we will address strategies based on lower complexity architectures that will allow for reducing the power consumption associated with its complexity.

In Figure 6, we consider a scenario where the BS employs a larger array with $N_{\text{tx}} = 256$ antennas to transmit $N_s = 2$ simultaneous streams to each user, where $N_u = 2$. To better fit this scenario to a typical communication in the THz band, we consider the existence of a LOS component, a center frequency of $f_c = 300\text{GHz}$ and a bandwidth of $B = 15\text{GHz}$ (it is important to note that the beam split effect is also considered in the channel model). The AM precoder from [15] requires a single stream per user and thus was not included in the figure. It is important to note that in this scenario the use of $F = 64$ with only 4 RF chains results in a more demanding restriction that makes the approximation to $\mathbf{F}_{\text{opt}_k}$ (Equation (5)) more difficult, thus widening the gap between all the schemes and the fully-digital curve. Still, the LASSO-based Alt-Min precoding schemes present a performance substantially lower when compared to the proposed approaches. Furthermore, the best performance is achieved with the use of double phase shifters, as expected. Once again, in comparing the curves of the proposed precoder against the ADMM-based precoder from [10], the advantage of adopting an interference-cancellation-based design over a simple matrix approximation one is clear.

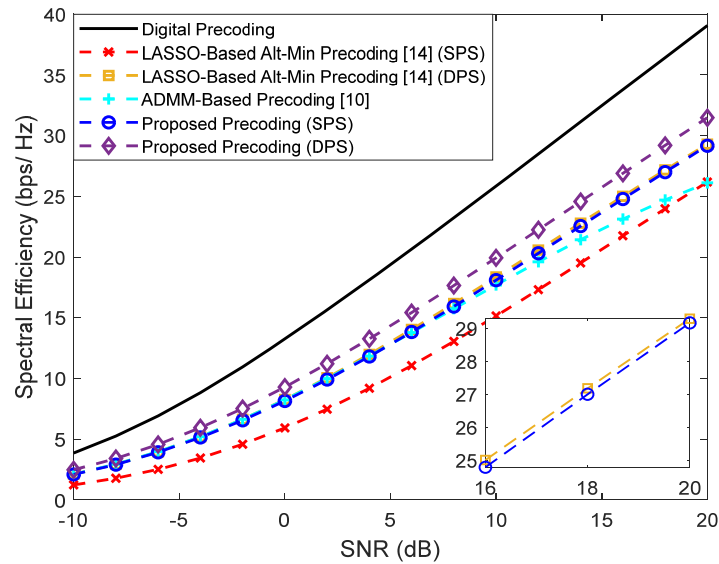


Figure 6. Spectral efficiency versus SNR achieved by different methods with $N_u = 2$, $N_s = 2$, $N_{RF}^{tx} = 4$, $F = 64$, $N_{tx} = 256$ and $N_{rx} = 4$ (with LOS component).

4.2. Reduced Complexity Architectures

Next, we will focus on the adoption of different reduced complexity architectures according to the typologies presented in Section 3.2. The objective is to evaluate the performance degradation when simpler architectures are adopted.

Figure 7 considers a scenario in which we have more than one data stream ($N_s = 2$) being sent from the BS to each user ($N_u = 4$) in a system with $N_{RF}^{tx} = N_u N_s$, $F = 1$, $N_{tx} = 256$ and $N_{rx} = 4$. We considered the same penalty parameters configuration: $\rho = 0.05$, $\mu = 1$ and $\eta = \rho$. This figure is placed in a perspective of simplifying the implementation of the analog precoder but keeping a fully-connected structure. We can observe that the versions based on DPS and single UPS achieve the best results, as expected. Considering the more realistic QPS versions, the results can worsen but it is visible that it is not necessary to use high resolution phase shifters since, with only 3 bits resolution, the results are already very close to the UPS curve. It can also be observed that the simplest of the architectures, AS, results in the worst performance but the SE improves when the antenna selectors are replaced by a network of switches, or, even better, if branches with inverters are also included.

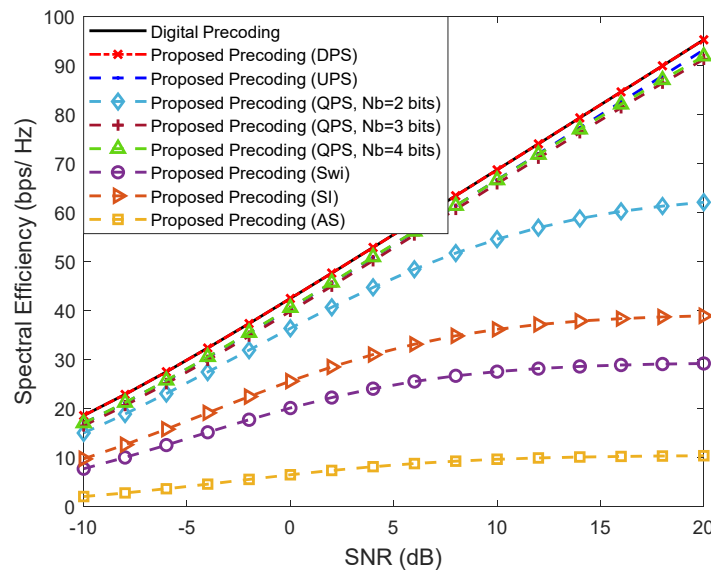


Figure 7. Spectral efficiency versus SNR achieved by the proposed precoder using different fully-connected architectures for $N_u = 4$, $N_s = 2$, $N_{RF}^x = 8$, $F = 1$, $N_{tx} = 256$ and $N_{rx} = 4$ (only NLOS).

In Figure 8, we intend to simplify the implementation even further with the adoption of AoSAs. In this case we considered that the maximum number of subarrays that can be connected to an RF chain (L_{max}) is only one. This imposes a very demanding restriction on matrix \mathbf{F}_{RF} since most of it will be filled with zeros, thus substantially deteriorating the approximation to \mathbf{F}_{opt_k} (Equation (5)). The scenario is the same in Figure 7 but considers the existence of a LOS component with a few weak NLOS paths.

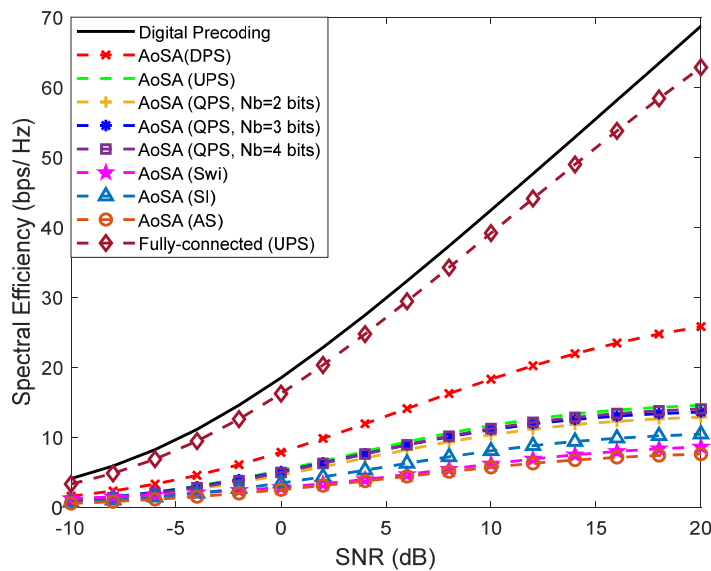


Figure 8. Spectral efficiency versus SNR achieved by the proposed precoder using different AoSA architectures with $L_{max} = 1$, $N_u = 4$, $N_s = 2$, $N_{RF}^x = 8$, $F = 1$, $N_{tx} = 256$ and $N_{rx} = 4$ (with LOS component).

In fact, hereafter, the existence of an LOS component is assumed for the remaining figures of the paper in order to fit the AoSA/DAoSA results to a more typical scenario in

the THz band. We can observe that for AoSA structures, the degradation of the SE is notorious, since all candidate versions present worse results when compared to the corresponding fully-connected design and are all far from the fully-digital solution. To reduce the large performance loss due to the adoption of a simple AoSA architecture, we can allow the dynamic connection of more subarrays to each RF chain by adopting a DAoSA structure, as introduced in Section 3.2.

In Figure 9, we study the effect of increasing the maximum number of subarrays that can be connected to an RF chain (L_{max}) in the performance of these schemes. Each subarray has a size of 32 antennas (n_i). Curves assuming the use of SPS as well as of DPS are included. It can be observed that the increase in the number of connections to subarrays, L_{max} , has a dramatic effect on the performance, resulting in a huge improvement by simply going from $L_{max} = 1$ to $L_{max} = 2$. Increasing further to $L_{max} = 4$, the results become close to the fully-connected case, showing that the DAoSA can be a very appealing approach for balancing spectral efficiency with hardware complexity and power consumption. Combining the increase in L_{max} with the adoption of DPS can also improve the results but the gains become less pronounced for $L_{max} > 1$. It is important to note that the penalty parameters can be fine-tuned for different system configurations.

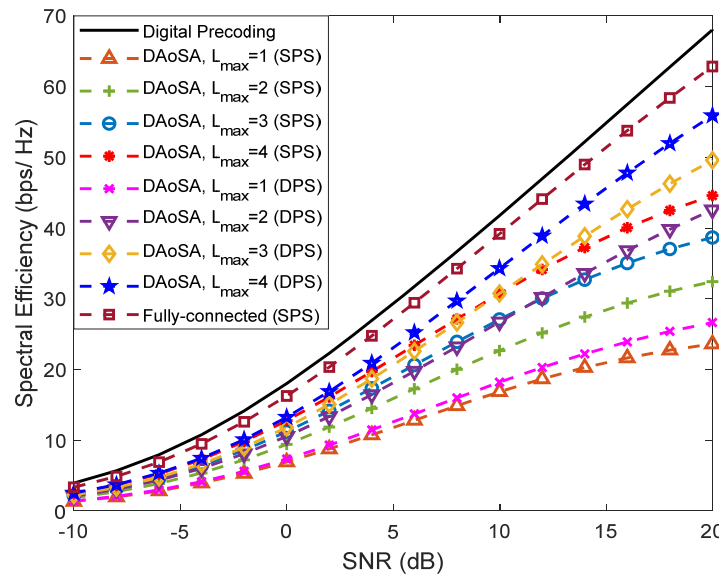


Figure 9. Spectral efficiency versus SNR achieved by the proposed precoder considering an architecture based on DAoSAs and the variation of the maximum number of subarrays that can be connected to an RF chain (L_{max}) for $N_u = 4$, $N_s = 2$, $N_{RF}^{tx} = 8$, $F = 1$, $N_{rx} = 256$ and $N_{rx} = 4$ (with LOS component).

One of the objectives of adopting these low-complexity solutions is to reduce overall power consumption. Based on [20], we can calculate the total power consumption of each precoding scheme using

$$P_C = P_{BB} N_{BB} + (P_{DAC} + P_{OS} + P_M) N_{RF}^{tx} + P_{PA} N_{tx} + P_{PC} N_{tx} + P_{PS} N_{PS} + P_{SWI} N_{SWI} + P_{tx} \quad (50)$$

where P_{BB} is the power of the baseband block (with $N_{BB} = 1$), P_{DAC} is the power of a DAC, P_{OS} is the power of an oscillator, P_M is the power of a mixer, P_{PA} is the power of a power amplifier, P_{PC} is the power of a power combiner, P_{PS} is the power of a phase shifter, P_{SWI} is the power of a switch and P_{tx} denotes the transmit power. The N_x variable represents the number of elements of each device used in the precoder configuration.

Based on the values provided in [20,38] for the power consumption of individual devices in the 300 GHz band, we adopt the following values: $P_{BB} = 200$ mW, $P_{DAC} = 110$ mW, $P_{OS} = 4$ mW, $P_M = 22$ mW, $P_{PA} = 60$ mW, $P_{PC} = 6.6$ mW, $P_{SWI} = 24$ mW and $P_T = 100$ mW.

Regarding the phase shifters, we assume values of $P_{PS} = 10, 20, 40$ and 100 mW for 1, 2, 3 and 4 quantization bits. Considering the same configuration scenario as Figures 7–9 with $N_u = 4, N_s = 2, N_{RF}^{tx} = N_u N_s, F = 1$ and $N_{tx} = 256$, we provide the values of power consumption for different precoder configurations in Table 3.

Table 3. Power Consumption for Different Implementations of the Proposed Precoder for $N_u = 4, N_s = 2, N_{RF}^{tx} = 8, F = 1$ and $N_{tx} = 256$.

	Precoder	Estimated Power Consumption [W]
Fully-Connected	DPS	428.04
	UPS	223.24
	QPS ($N_b = 2$)	59.4
	QPS ($N_b = 3$)	100.36
	SWI	67.59
	SI	38.92
DAoSA SPS	$L_{max} = 1$	28.87
	$L_{max} = 2$	39.30
	$L_{max} = 3$	49.73
	$L_{max} = 4$	60.17
DAoSA DPS	$L_{max} = 1$	39.11
	$L_{max} = 2$	59.78
	$L_{max} = 3$	80.45
	$L_{max} = 4$	101.13

For the fully-connected structure with UPS, we assumed that $P_{PS} = 100$ mW, which corresponds to quantized phase shifters with $N_b = 4$ bits [38]. For the remaining phase-shifter-based precoder structures, we assumed that $P_{PS} = 40$ mW, which corresponds to quantized phase shifters with $N_b = 3$ bits, since with only 3 bits resolution the results are already very close to the UPS curve (see Figure 7). As can be seen from this table, the use of architectures based on DAoSAs allows us to reduce considerably the amount of power that is consumed by the precoder. In fact, we can reduce the amount of consumed power up to 55% if we consider a precoder scheme based on DAoSA with DPS and $L_{max} = 4$ versus an FC structure precoder based on UPS, with only a small performance penalty (Figure 9). This reduction increases to 73% if the DPS structure is replaced by an SPS one.

In the particular case of architectures based on quantized phase shifters, we observed that by decreasing the number of quantization bits, it is possible to substantially reduce the power consumption without excessively compromising the complexity (as seen in Figure 7). This conclusion is corroborated by [20,38], since the architectures based on low resolution QPS, AoSAs and DAoSAs present superior energy efficiency when compared to the fully-connected structure with UPS.

In Figures 10 and 11, we provide a comparison between our proposed precoder and the EBE precoder from [20], considering an architecture based on DAoSAs (with SPS) and a scenario configuration similar to Figure 9, i.e., with $N_s = 2, N_{RF}^{tx} = 8, F = 1, N_{tx} = 256$ and $N_{rx} = 4$. These figures present various curves where the maximum number of subarrays that can be connected to an RF chain, L_{max} , is changed. Figure 10 refers to an SU scenario ($N_u = 1$) whereas Figure 11 corresponds to an MU scenario with $N_u = 4$. In the SU case, the proposed precoder achieves results very close to the fully-digital precoder, even with only $L_{max} = 2$. Compared to the proposed algorithm, EBE shows a wider gap even though it has a smaller complexity (as presented in Table 2 of Section 3.3).

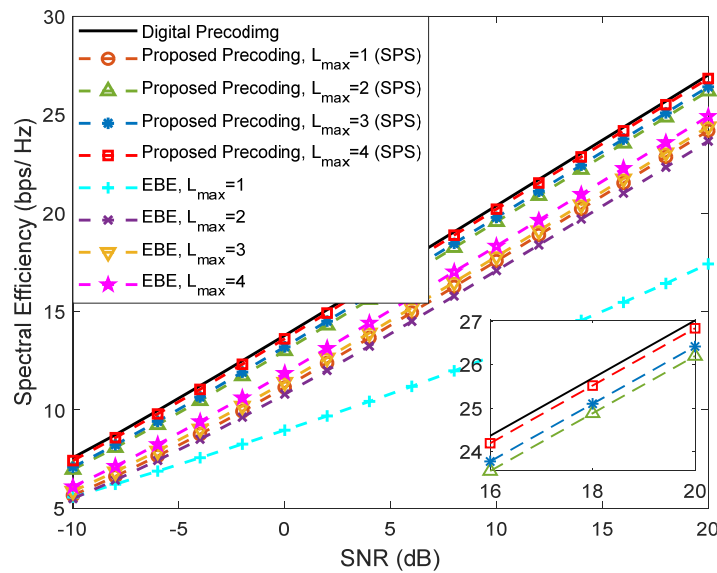


Figure 10. Spectral efficiency versus SNR achieved by the proposed precoder and by the EBE algorithm considering an architecture based on DAoSAs and the variation of the maximum number of subarrays that can be connected to an RF chain (L_{max}) for $N_u=1$, $N_s=2$, $N_{RF}^{tx}=8$, $F=1$, $N_{tx}=256$ and $N_{rx}=4$ (with LOS component).

When we increase the number of users from $N_u=1$ to $N_u=4$, we can clearly observe that the EBE algorithm suffers a substantial degradation compared to the proposed solution which can be explained by the lack of inter-user interference cancellation (it was not specifically designed for MU scenarios).

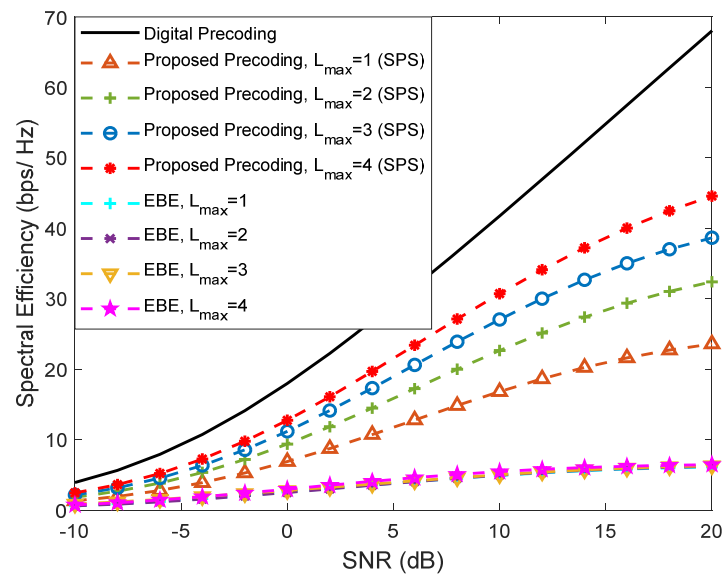


Figure 11. Spectral efficiency versus SNR achieved by the proposed precoder and by the EBE algorithm considering an architecture based on DAoSAs and the variation of the maximum number of subarrays that can be connected to an RF chain (L_{max}) for a mmWave/THz system with $N_u=4$, $N_s=2$, $N_{RF}^{tx}=8$, $F=1$, $N_{tx}=256$ and $N_{rx}=4$ (with LOS component).

Even though a sub-6 GHz system often adopts fully-digital processing [39], where each antenna element has a dedicated RF chain, it is possible to apply the proposed hybrid design

algorithm to a sub-6 GHz channel since it is independent of a specific MIMO channel (as are the other alternative algorithms that we used as benchmarks and which are targeted at solving the matrix approximation problem). To exemplify, Figure 12 presents the simulated results obtained for the same scenario of Figure 4 but considering an ideal uncorrelated channel which approximates a rich scattering environment that is typical in sub-6 GHz bands. It can be observed that the proposed approach displays similar behavior to the ones in the upper-bands channel, showing that it can also be used for this particular type of channel (even though it may require a higher number of RF chains to achieve a good approximation of the fully-digital solution in some scenarios, due to the channel not being sparse, as noted in [40]). It is important to highlight that even though the proposed approach can be applied to other channels, the algorithm was designed with the aim of dealing with architectures with a very large number of antennas and with large hardware constraints, making it especially interesting for mmWave and THz.

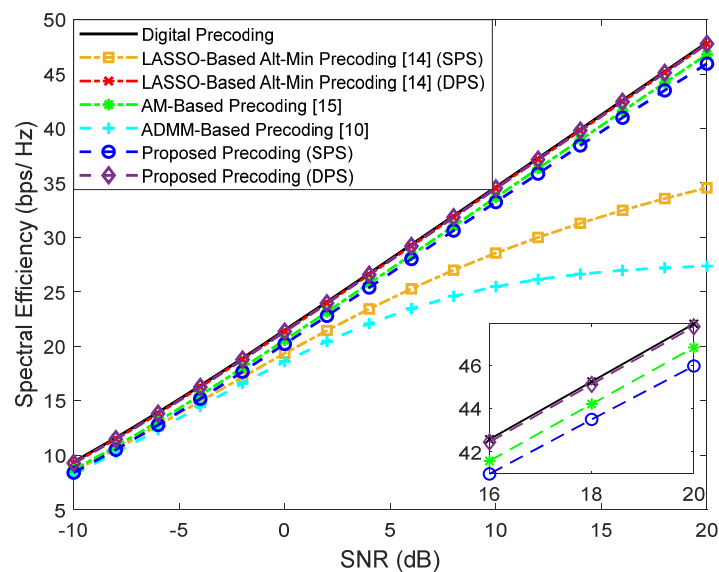


Figure 12. Spectral efficiency versus SNR achieved by different methods for a mmWave/THz MIMO-OFDM system with $N_u = 4$, $N_s = 1$, $N_{RF}^x = 4$, $F = 1$, $N_{tx} = 100$ and $N_{rx} = 4$ considering an uncorrelated channel.

While we have shown how the proposed approach can deal with several relevant types of analog precoders/combiners, it is important to note that there are other alternative structures that have been recently proposed in the literature. For example, some authors have considered precoding paradigms based on time-delayer structures for THz systems [27,41]. One of the most notorious is the Delay Phase Precoding (DPP), which consists in the use of a Time-Delay (TD) network between the RF chains and the traditional phase shifters network in order to convert phase-controlled analog precoding into delay-phase-controlled analog precoding. The main advantage related to this type of precoding is that the time delays in the TD network are carefully designed to generate frequency-dependent beams which are aligned with the spatial directions over the whole bandwidth [41]. While we do not address the adoption of time-delay structures in this paper, it should be possible to derive a projection algorithm that simultaneously takes into account the constraints imposed in both analog-precoding steps: time-delay networks and frequency-independent phase shifters.

5. Conclusions

In this paper, we proposed an iterative algorithm for hybrid precoding design which is suitable for multiuser UM-MIMO systems operating in mmWave and THz bands. The adopted approach replaces unfeasible fully-digital precoders/combiners relying on a dedicated RF chain per antenna with a hybrid architecture comprising low dimension digital blocks with only a few RF chains which are complemented by analog blocks supported solely on networks of phase shifters and switches. The proposed hybrid design algorithm is based on the approximation of the fully-digital approach whose problem formulation is split into a sequence of smaller subproblems with closed-form solutions and can work with a broad range of configuration of antennas, RF chains and data streams. The separability of the design process allows for the adaptability of the algorithm to different architectures, making it suitable for implementation with low-complexity AoSA and DAoSA structures, which are particularly relevant for the deployment of UM-MIMO in hardware-constrained THz systems. It was shown that good trade-offs between SE and hardware implementation complexity can in fact be achieved by the proposed algorithm for several different architectures. Numerical results showed that the use of architectures based on DAoSAs allows us to reduce considerably the amount of power that is consumed at the precoder. In fact, in a reference scenario, we showed that it was possible to reduce the amount of consumed power up to 55% if we consider a precoder scheme based on DAoSAs with DPS and $L_{\max} = 4$ versus an FC structure based on UPS, with only a small performance penalty. This reduction increases to 73% if the DPS structure is replaced by an SPS one.

Author Contributions: Conceptualization, J.P.P. and N.S.; methodology, J.P.P., N.S. and R.D.; software, J.P.P. and N.S.; validation, J.P.P., N.S., M.R. and R.D.; formal analysis, J.P.P., V.V., R.F. and N.S.; investigation, J.P.P., V.V., R.F. and N.S.; resources, N.S. and M.R.; data curation, J.P.P.; writing—original draft preparation, J.P.P. and N.S.; writing—review and editing, J.P.P., V.V., R.F., N.S., M.R., J.S. and R.D.; visualization, J.P.P. and N.S.; supervision, N.S. and M.R.; project administration, N.S., M.R., J.S. and R.D.; funding acquisition, N.S. and R.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the FCT—Fundação para a Ciência e Tecnologia under the grant 2020.05621.BD. The authors also acknowledge the funding provided by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/50008/2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rappaport, T.S.; Sun, S.; Mayzus, R.; Zhao, H.; Azar, Y.; Wang, K.; Wong, G.; Schulz, J.K.; Samimi, M.; Gutierrez, F. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! *IEEE Access* **2013**, *1*, 335–349, <https://doi.org/10.1109/access.2013.2260813>.
2. Uwaechia, A.N.; Mahyuddin, N.M. A Comprehensive Survey on Millimeter Wave Communications for Fifth-Generation Wireless Networks: Feasibility and Challenges. *IEEE Access* **2020**, *8*, 62367–62414, <https://doi.org/10.1109/access.2020.2984204>.
3. Akyildiz, I.F.; Kak, A.; Nie, S. 6G and Beyond: The Future of Wireless Communications Systems. *IEEE Access* **2020**, *8*, 133995–134030, <https://doi.org/10.1109/access.2020.3010896>.
4. Tan, J.; Dai, T. THz Precoding for 6G: Applications, Challenges, Solutions, and Opportunities. *arXiv Prepr.* **2020**, arXiv:2005.10752. Available online: <https://arxiv.org/abs/2005.10752> (accessed on 2 July 2021).
5. Sariaedeen, H.; Alouini, M.-S.; Al-Naffouri, T.Y. An Overview of Signal Processing Techniques for Terahertz Communications. *Proc. IEEE* **2021**, 1–38, <https://doi.org/10.1109/jproc.2021.3100811>.
6. Lin, C.; Li, G.Y.L. Terahertz Communications: An Array-of-Subarrays Solution. *IEEE Commun. Mag.* **2016**, *54*, 124–131, <https://doi.org/10.1109/mcom.2016.1600306cm>.

7. Ahmed, I.; Khammari, H.; Shahid, A.; Musa, A.; Kim, K.S.; De Poorter, E.; Moerman, I. A Survey on Hybrid Beamforming Techniques in 5G: Architecture and System Model Perspectives. *IEEE Commun. Surv. Tutorials* **2018**, *20*, 3060–3097, <https://doi.org/10.1109/comst.2018.2843719>.
8. Chataut, R.; Akl, R. Massive MIMO Systems for 5G and beyond Networks—Overview, Recent Trends, Challenges, and Future Research Direction. *Sensors* **2020**, *20*, 1–35.
9. Sohrabi, F.; Yu, W. Hybrid Digital and Analog Beamforming Design for Large-Scale Antenna Arrays. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 501–513, <https://doi.org/10.1109/jstsp.2016.2520912>.
10. Souto, N.N.; Silva, J.; Pavia, J.; Ribeiro, M. An alternating direction algorithm for hybrid precoding and combining in millimeter wave MIMO systems. *Phys. Commun.* **2019**, *34*, 165–173.
11. Pavia, J.P.; Souto, N.; Ribeiro, M.; Silva, J.; Dinis, R. Hybrid Precoding and Combining Algorithm for Reduced Complexity and Power Consumption Architectures in mmWave Communications. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference: VTC2020-Spring, Antwerp, Belgium, 25–28 May 2020; pp. 1–5, <https://doi.org/10.1109/vtc2020-spring48590.2020.9128553>.
12. Guan, K.; Li, G.; Kuerner, T.; Molisch, A.F.; Peng, B.; He, R.; Hui, B.; Kim, J.; Zhong, Z. On Millimeter Wave and THz Mobile Radio Channel for Smart Rail Mobility. *IEEE Trans. Veh. Technol.* **2016**, *66*, 5658–5674, <https://doi.org/10.1109/tvt.2016.2624504>.
13. Alkhateeb, A.; Mo, J.; Gonzalez-Prelcic, N.; Heath, R. MIMO Precoding and Combining Solutions for Millimeter-Wave Systems. *IEEE Commun. Mag.* **2014**, *52*, 122–131, <https://doi.org/10.1109/mcom.2014.6979963>.
14. Yu, X.; Zhang, J.; Letaief, K.B. Alternating minimization for hybrid precoding in multiuser OFDM mmWave systems. In Proceedings of the 2016 50th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 6–9 November 2016; pp. 281–285, <https://doi.org/10.1109/acssc.2016.7869042>.
15. Yuan, H.; An, J.; Yang, N.; Yang, K.; Duong, T.Q. Low Complexity Hybrid Precoding for Multiuser Millimeter Wave Systems Over Frequency Selective Channels. *IEEE Trans. Veh. Technol.* **2018**, *68*, 983–987, <https://doi.org/10.1109/tvt.2018.2880787>.
16. Liu, F.; Kan, X.; Bai, X.; Du, R.; Liu, H.; Zhang, Y. Hybrid precoding based on adaptive RF-chain-to-antenna connection for millimeter wave MIMO systems. *Phys. Commun.* **2020**, *39*, 100997, <https://doi.org/10.1016/j.phycom.2019.100997>.
17. Xu, K.; Cai, Y.; Zhao, M.; Niu, Y.; Hanzo, L. MIMO-Aided Nonlinear Hybrid Transceiver Design for Multiuser Mmwave Systems Relying on Tomlinson-Harashima Precoding. *IEEE Trans. Veh. Technol.* **2021**, *70*, 6943–6957, <https://doi.org/10.1109/tvt.2021.3087651>.
18. Vizziello, A.; Savazzi, P.; Chowdhury, K.R. A Kalman Based Hybrid Precoding for Multi-User Millimeter Wave MIMO Systems. *IEEE Access* **2018**, *6*, 55712–55722, <https://doi.org/10.1109/access.2018.2872738>.
19. Elmagzoub, H.M. On the MMSE-based multiuser millimeter wave MIMO hybrid precoding design. *Int. J. Commun. Syst.* **2020**, *33*, e4409, <https://doi.org/10.1002/dac.4409>.
20. Yan, L.; Han, C.; Yuan, J. A Dynamic Array-of-Subarrays Architecture and Hybrid Precoding Algorithms for Terahertz Wireless Communications. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 2041–2056, <https://doi.org/10.1109/jsac.2020.3000876>.
21. Busari, S.; Huq, K.; Mumtaz, S.; Rodriguez, J. Terahertz Massive MIMO for Beyond-5G Wireless Communication. In Proceedings of the ICC 2019-2019 IEEE International Conference on Communications (ICC), Shanghai, China, 11–22 March 2019.
22. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.* **2010**, *3*, 1–122, <https://doi.org/10.1561/22000000016>.
23. Wang, Y.; Yin, W.; Zeng, J. Global Convergence of ADMM in Nonconvex Nonsmooth Optimization. *J. Sci. Comput.* **2018**, *78*, 29–63, <https://doi.org/10.1007/s10915-018-0757-z>.
24. Alkhateeb, A.; Heath, R.W. Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems. *IEEE Trans. Commun.* **2016**, *64*, 1801–1818.
25. Yuan, H.; Yang, N.; Yang, K.; Han, C.; An, J. Hybrid Beamforming for Terahertz Multi-Carrier Systems Over Frequency Selective Fading. *IEEE Trans. Commun.* **2020**, *68*, 6186–6199, <https://doi.org/10.1109/tcomm.2020.3008699>.
26. Tarboush, S.; Sameddeen, H.; Chen, H.; Loukil, M.H.; Jemaa, H.; Alouini, M.S.; Al-Naffouri, T.Y. TeraMIMO: A Channel Simulator for Wideband Ultra-Massive MIMO Terahertz Communications. *arXiv Prepr.* **2021**, arXiv:2104.11054. Available online: <https://arxiv.org/abs/2104.11054> (accessed on 2 July 2021).
27. Lin, C.; Li, G.Y.; Wang, L. Subarray-Based Coordinated Beamforming Training for mmWave and Sub-THz Communications. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 2115–2126, <https://doi.org/10.1109/jsac.2017.2720038>.
28. Nguyen, D.D.; Le, L.; Le-Ngoc, T.; Heath, R. Hybrid MMSE Precoding and Combining Designs for mmWave Multiuser Systems. *IEEE Access* **2017**, *5*, 19167–19181.
29. Bertsekas, D. *Nonlinear Programming*; Athena Scientific: Belmont, MA, USA, 2016.
30. El Ayach, O.; Rajagopal, S.; Abu-Surra, S.; Pi, Z.; Jr, R.W.H. Spatially Sparse Precoding in Millimeter Wave MIMO Systems. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 1499–1513, <https://doi.org/10.1109/twc.2014.011714.130846>.
31. Yu, X.; Shen, J.-C.; Zhang, J.; Letaief, K.B. Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 485–500, <https://doi.org/10.1109/jstsp.2016.2523903>.
32. Spencer, Q.; Swindlehurst, A.L.; Haardt, M. Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels. *IEEE Trans. Signal Process.* **2004**, *52*, 461–471, <https://doi.org/10.1109/tsp.2003.821107>.
33. Mendez-Rial, R.; Rusu, C.; Gonzalez-Prelcic, N.; Alkhateeb, A.; Heath, R. Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches? *IEEE Access* **2016**, *4*, 247–267.

34. Lee, J.; Lee, Y.H. AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing. In Proceedings of the 2014 IEEE International Conference on Communications (ICC), Sydney, Australia, 10–14 June 2014; pp. 5838–5842, <https://doi.org/10.1109/icc.2014.6884253>.
35. Payami, S.; Ghorraishi, M.; Dianati, M.; Sellathurai, M. Hybrid Beamforming with a Reduced Number of Phase Shifters for Massive MIMO Systems. *IEEE Trans. Veh. Technol.* **2018**, *67*, 4843–4851, <https://doi.org/10.1109/tvt.2018.2807921>.
36. Tian, M.; Zhang, J.; Zhao, Y.; Yuan, L.; Yang, J.; Gui, G. Switch and Inverter Based Hybrid Precoding Algorithm for mmWave Massive MIMO System: Analysis on Sum-Rate and Energy-Efficiency. *IEEE Access* **2019**, *7*, 49448–49455, <https://doi.org/10.1109/access.2019.2910094>.
37. Yu, X.; Zhang, J.; Letaief, K.B. Doubling Phase Shifters for Efficient Hybrid Precoder Design in Millimeter-Wave Communication Systems. *J. Commun. Inform. Netw.* **2019**, *4*, 51–67.
38. Yan, L.; Han, C.; Yang, N.; Yuan, J. Dynamic-subarray with Quantized- and Fixed-phase Shifters for Terahertz Hybrid Beamforming. In Proceedings of the GLOBECOM 2020-2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6, <https://doi.org/10.1109/globecom42002.2020.9348113>.
39. Li, Z.; Zhang, C.; Lu, I.-T.; Jia, X. Hybrid Precoding Using Out-of-Band Spatial Information for Multi-User Multi-RF-Chain Millimeter Wave Systems. *IEEE Access* **2020**, *8*, 50872–50883, <https://doi.org/10.1109/access.2020.2979712>.
40. Park, S.; Alkhateeb, A.; Jr., R.W.H. Dynamic Subarrays for Hybrid Precoding in Wideband mmWave MIMO Systems. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 2907–2920, <https://doi.org/10.1109/twc.2017.2671869>.
41. Tan, J.; Dai, L. Delay-Phase Precoding for THz Massive MIMO with Beam Split. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Big Island, Hawaii, USA, 9–13 December 2019; pp. 1–6, <https://doi.org/10.1109/globecom38437.2019.9014304>.