

Hybrid Iterative Space-Time Equalization for Multi-User mmW Massive MIMO Systems

Roberto Magueta, Daniel Castanheira, Adão Silva, *Member, IEEE*,
Rui Dinis, *Senior Member, IEEE*, and Atílio Gameiro

Abstract—The combination of millimeter wave (mmW) with massive MIMO is a promising approach to achieve the multi Gb/s required by future wireless systems. Fully digital architectures are not feasible due to hardware limitations, and thus, the design of signal processing techniques for hybrid analog-digital architectures is of paramount importance. In this paper, we propose a new hybrid iterative block space-time receiver structure for multiuser mmW massive MIMO systems. We consider low-complexity user terminals employing analog-only random precoding and a single RF chain. At the base station, a hybrid analog-digital equalizer/detector is designed to efficiently remove the multiuser interference. The analog and digital parts of the equalizer are jointly optimized using as a metric the average bit-error-rate. The specificities of the analog domain impose several constraints in the joint optimization. To efficiently handle these constraints, the analog part is selected from a dictionary based on the array response vectors. We also propose a simple, yet an accurate semi-analytical approach for obtaining the performance of the proposed hybrid receiver structure. The results show that the performance of the hybrid iterative equalizer is close to the fully digital counterpart after only a few iterations. Moreover, it clearly outperforms the linear receivers recently considered for hybrid mmW massive MIMO architectures.

Index Terms—Massive MIMO, millimeter wave communications, iterative block equalization, hybrid analog/digital architectures.

I. INTRODUCTION

MILLIMETER wave (mmW) communications and massive MIMO (mMIMO) have been considered as two of the key enabling technologies needed to meet the quality of service requirements for future wireless communication [1]. The global bandwidth shortage facing wireless carriers has motivated the exploration of the underutilized mmW frequency spectrum for future broadband cellular communication networks [2]. Therefore, several tens of GHz could become

Manuscript received June 23, 2016; revised September 29, 2016 and November 15, 2016; accepted November 20, 2016. Date of publication November 24, 2016; date of current version February 14, 2017. This work is funded by National Funds through FCT—Fundação para a Ciência e a Tecnologia under the project PURE-5GNET (UID/EEA/50008/2013). The work of D. Castanheira was supported by FCT under the postdoctoral grant SFRH/BPD/95375/2013. The associate editor coordinating the review of this paper and approving it for publication was X. Gao.

R. Magueta, D. Castanheira, A. Silva, and A. Gameiro are with the Departamento de Eletrónica, Telecomunicações e Informática, Instituto de Telecomunicações, University of Aveiro, 3810-193 Aveiro, Portugal (e-mail: rlm@av.it.pt; dcastanheira@av.it.pt; asilva@av.it.pt; amg@ua.pt).

R. Dinis is with the Faculdade de Ciências e Tecnologia, Instituto de Telecomunicações, Universidade Nova de Lisboa, Caparica 2829-516, Portugal (e-mail: rdinis@fct.unl.pt).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCOMM.2016.2632732

available for future wireless systems, offering multi Gb/s. Together with the access to more bandwidth, the deployment of a large number of antennas has been also considered an enabling technology for meeting the ever increasing demand of higher data rates in future wireless networks [3]. Massive MIMO can scale up MIMO by orders of magnitude compared to the conventional MIMO approaches [4]. A survey on mMIMO, also known as large-scale MIMO, including channel modeling, applications scenarios and physical and networking techniques was provided in [5]. The use of MIMO with mmW is very attractive, since it allows packing more antennas in the same volume due to the smaller wavelength compared to the current communication systems [6], and hence, the terminals can be equipped with a large number of antennas.

MmW massive MIMO may exploit new and efficient spatial processing techniques such as beamforming/precoding and spatial multiplexing at the transmitter and/or receiver sides [7]. The system design for these techniques should follow different approaches than the ones adopted for lower frequencies counterparts, mainly due to the hardware limitations [8]. In fact, the high cost and power consumption of some mmW mixed-signal components, make it difficult to have a fully dedicated radio frequency (RF) chain for each antenna [9] as in conventional MIMO systems [10]. Another important issue is that the mmW propagation characteristics are quite different from those at lower frequencies since the channels do not have so rich multipath propagation effects [11], [12], and this should be taken into account in the designing of beamforming techniques. A simple and immediate approach to overcome the limitation on the number of RFs chains is to make beamforming only in the analog domain by using phase shifters [8]. Some analog beamforming schemes have been considered in [14], and specifically designed for mmW systems in [15] and [16]. The works in [15] and [16] considered beam steering solutions relying on phase shifters that are used to optimally orient the response of arrays in space, which are based on statistical channel knowledge. Nevertheless, the performance of the pure analog signal processing approach are limited by the availability of only quantized phase shifters and the constraints on the amplitudes of these phase shifters, and thus analog beamforming is usually limited to single-stream transmission [17].

To overcome these limitations, hybrid analog/digital architectures, where some signal processing is done at the digital level and some left to the analog domain, have been discussed in [13], [17], and [18]. Beamforming and/or combin-

ing/equalization schemes for single-user systems have been proposed for hybrid architectures in [19]–[23]. In [19], it was proposed a joint design of transmit-receive mixed analog/digital beamformers that aims at maximizing the received average signal-to-noise-ratio (SNR), for single carrier (SC) systems with linear frequency domain equalization (FDE). A simple precoding scheme, only based on the knowledge of partial channel information at both terminals, in the form of angles of arrival (AoA) and departure (AoD), was proposed in [20]. Ayach *et al.* [21] designed a hybrid spatially sparse precoding and combining approach for mmW massive MIMO systems. The spatial structure of mmW channels was exploited to formulate the single-user multi-stream precoding/combining scheme as a sparse reconstruction problem. A digitally assisted analog beamforming technique for mmW systems was considered in [22], where a digital beamsteering system using coarsely quantized signals assists the analog beamformer. In [23], a turbo-like beamforming was proposed to jointly compute the transmit and receive analog beamforming coefficients, but the digital processing part was not taken into account. The previous approaches considered a fully connected hybrid architecture, where each RF chain is connected to all receive/transmit antennas. Beamforming solutions for subconnected hybrid architecture, where each RF chain is connected to only a subset of antennas were presented in [24]. Codebook design approaches for the analog beamforming structure and for single stream transmissions were addressed in [25]. Recently, some beamforming approaches for multiuser systems have been also proposed for hybrid architectures [26], [27]. The work in [26] focused on the optimization of a downlink hybrid analog/digital beamforming to minimize the transmit power subject to signal to interference plus noise ratio (SINR) constraints. In that paper, all processing is done at the transmitter side by assuming perfect knowledge of the channel state information (CSI) and single antenna user terminals (UTs). A downlink hybrid beamforming scheme based on a weighted sum mean square error (WSMSE) optimization problem was proposed for massive MIMO systems [27]. It was shown that the performance degradation is severe when the number of RF chains is smaller than half of the number of antennas. Hybrid approaches explicitly designed for mmW communications were considered in [28] and [29]. Li *et al.* [28] designed an uplink receive beamforming that handles the multiuser interference at both analog and digital stages. However, only single antenna UTs were assumed. A limited feedback hybrid two-stage precoding/combining algorithm was proposed in [29] for the downlink. In the first stage transmit and receive analog beamforming are jointly computed for each base-station (BS) – user terminal pair assuming interference free links. In the second stage the interference is explicitly mitigated, in the digital domain, by employing a conventional linear zero forcing (ZF) precoding.

Nonlinear equalizers were considered in past to efficiently separate the spatial streams and/or remove the multiuser interference in the current MIMO based networks [30]. Iterative block decision feedback equalization (IB-DFE) approach are one of the most promising nonlinear equalization

schemes [30]. IB-DFE was originally proposed in [31], and it can be regarded as a low complexity turbo equalizer implemented in the frequency-domain that does not require the channel decoder output in the feedback loop. In the last years, it has been extended to a wide range of scenarios, ranging from diversity scenarios, conventional MIMO, cooperative MIMO systems, among many others [32]–[37]. The IB-DFE principles can be used in mmW massive MIMO context to efficiently remove the multiuser interference.

However, as discussed, mmW massive MIMO brings new major challenges that prevent a direct plug and play of the iterative detection based solutions developed for conventional MIMO systems. Therefore, in this manuscript, we design an efficient hybrid iterative block space-time equalizer for mmW massive MIMO systems. To the best of our knowledge iterative block detection, designed for mmW massive MIMO systems, has not been addressed in the literature. We consider low-complexity UTs employing a single RF chain and analog-only random precoding, assuming that the UTs do not have knowledge of the CSI. Before the analog precoder, a time encoder/precoder is employed to ensure that the transmit signal and consequently the noise plus interference, at the receiver side, are Gaussian distributed (which simplifies the receiver optimization) and increase the inherent diversity of the mmW massive MIMO system. The designed hybrid equalizer is optimized by using as a metric the average bit-error-rate (BER), which is shown to be equivalent to minimize the mean square error (MSE) of each encoded time slot. The specificities of the analog domain impose several constraints in the joint optimization. To efficiently handle these constraints the analog part is selected from a dictionary based on the array response vectors similar to the approach discussed in [21]. We also design a fully digital iterative space-time receiver, which provides a lower bound performance for the proposed hybrid approach. Semi-analytic results are presented to validate the Gaussian assumption in the derivation of the proposed hybrid iterative equalizer. We also show that the performance of the hybrid receiver tends to the performance of the digital one and the single-user scenario as the number of iterations increases.

The remainder of the paper is organized as follows: section II describes the hybrid multiuser mmW massive MIMO systems model. Section III, starts by briefly describing the iterative space-time receiver structure. Then, the fully digital equalizer is presented and finally the proposed hybrid space-time equalizer is derived in detail. Section IV presents the main performance results and the conclusions will be drawn in section V.

Notations: Boldface capital letters denote matrices and boldface lowercase letters denote column vectors. The operations $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$ and $tr(\cdot)$ represent the transpose, the Hermitian transpose, the conjugate and the trace of a matrix. The operator $\text{sign}(a)$ represents the sign of real number a . For a complex number c , $\text{sign}(c) = \text{sign}(\Re(c)) + j\text{sign}(\Im(c))$, where $\Re(c)$ ($\Im(c)$) represents the real part of c (imaginary part of c). The operator $\text{sign}(\cdot)$ is applied element-wise to matrices. Consider a vector \mathbf{a} and a matrix \mathbf{A} , then $\text{diag}(\mathbf{a})$ and $\text{diag}(\mathbf{A})$ correspond to a diagonal matrix with diagonal entries equal

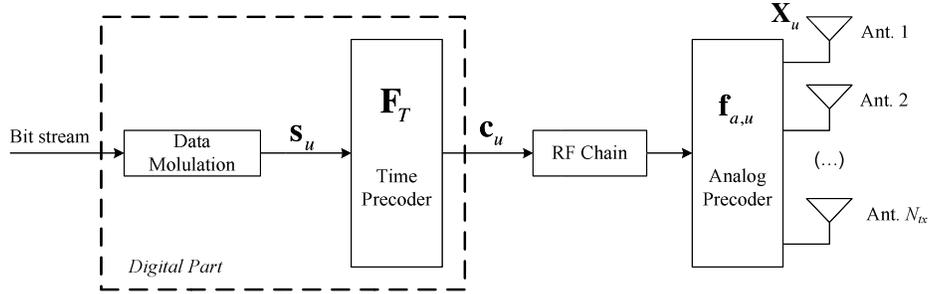


Fig. 1. User terminal u block diagram.

to vector \mathbf{a} and a diagonal matrix with entries equal to the diagonal entries of the matrix \mathbf{A} , respectively. $\mathbf{A}(j, l)$ denotes the element at row j and column l of the matrix \mathbf{A} . \mathbf{I}_N is the identity matrix with size $N \times N$.

II. SYSTEM CHARACTERIZATION

In this section, we present the mmW massive MIMO signal definition and describe both the user terminals and receiver signals.

A. System Model

Consider a multiuser mmW system with U users, each with N_{tx} transmit antennas and a base station with N_{rx} receive antennas, where each user sends one data stream to the receiver, per time-slot. We consider a block fading channel, i.e. the channel remains constant during a block, with size T , but it varies independently between blocks. For a given block, the received signal is given by

$$\mathbf{Y} = \sum_{u=1}^U \mathbf{H}_u \mathbf{X}_u + \mathbf{N} = \mathbf{H}\mathbf{X} + \mathbf{N}, \quad (1)$$

where $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_T] \in \mathbb{C}^{N_{rx} \times T}$ denotes the received signal matrix, $\mathbf{X}_u = [\mathbf{x}_{u,1}, \dots, \mathbf{x}_{u,T}] \in \mathbb{C}^{N_{tx} \times T}$ the transmitted signal of the u th user, $\mathbf{X} = [\mathbf{X}_1^H, \dots, \mathbf{X}_U^H]^H \in \mathbb{C}^{N_{tx} \times T}$ is the concatenation of the transmitted signals of all U users and $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_T] \in \mathbb{C}^{N_{rx} \times T}$ is the zero mean Gaussian noise with variance σ_n^2 . $\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_U] \in \mathbb{C}^{N_{rx} \times N_{tx} U}$ is the concatenated channel matrix, where $\mathbf{H}_u \in \mathbb{C}^{N_{rx} \times N_{tx}}$ denotes the channel matrix of the u th user, which follows the clustered sparse mmW channel model discussed in [21]. The channel of the u th user, i.e. \mathbf{H}_u , is the sum of the contribution of N_{cl} clusters, each of which contribute N_{ray} propagation paths to the channel matrix \mathbf{H}_u . The channel matrix of the u th user may be expressed as

$$\mathbf{H}_u = \mathbf{A}_{rx,u} \mathbf{\Lambda}_u \mathbf{A}_{tx,u}^H \quad (2)$$

where $\mathbf{\Lambda}_u$ is a diagonal matrix, with entries (j, l) that correspond to the paths gains of the l th ray in the j th scattering cluster. $\mathbf{A}_{tx,u} = [\mathbf{a}_{tx,u}(\theta_{1,1}^{tx,u}), \dots, \mathbf{a}_{tx,u}(\theta_{N_{cl}, N_{ray}}^{tx,u})]$, $\mathbf{A}_{rx,u} = [\mathbf{a}_{rx,u}(\theta_{1,1}^{rx,u}), \dots, \mathbf{a}_{rx,u}(\theta_{N_{cl}, N_{ray}}^{rx,u})]$ are the matrix of array response vectors at the transmitter and receiver, corresponding to user u , whereas $\theta_{j,l}^{rx,u}$ and $\theta_{j,l}^{tx,u}$ are the azimuth angles of arrival and departure, respectively. The channel path gains and

the angles are generated according to the random distributions discussed in [21], such that $\mathbb{E}[\|\mathbf{H}_u\|_F^2] = N_{rx} N_{tx}$.

B. User Terminal Model Description

We consider an analog only architecture for each user terminal, as shown in Fig 1. User u transmits only one data stream per time slot. The analog domain of user u at the instant t is modeled mathematically by the precoder matrix $\mathbf{f}_{a,u,t} \in \mathbb{C}^{N_{tx}}$. Due to hardware constraints, the analog part is implemented using a matrix of analog phase shifters, which force all elements of vector $\mathbf{f}_{a,u,t}$ to have equal norm ($|\mathbf{f}_{a,u,t}(l)|^2 = N_{tx}^{-1}$). We assume that the user terminals have no access to CSI simplifying the overall system design. As such the analog precoder vector of the u th user is generated randomly accordingly to

$$\mathbf{f}_{a,u,t} = [e^{j2\pi\phi_n^{u,t}}]_{1 \leq n \leq N_{tx}, 1 \leq t \leq T}, \quad (3)$$

where $\phi_n^{u,t}$, $n \in \{1, \dots, N_{tx}\}$, $t \in \{1, \dots, T\}$ and $u \in \{1, \dots, U\}$ are i.i.d uniform random variables with support $\phi_n^{u,t} \in [0, 1]$.

The transmitter total power constraint is $\|\mathbf{X}_u\|_F^2 = T$. The transmit signal $\mathbf{X}_u = [\mathbf{x}_{u,1}, \dots, \mathbf{x}_{u,T}]$ is constructed by using a space-time block code (STBC), where a DFT performs the time and the analog precoder the space encoding. The STBC operation may be mathematically expressed by

$$\mathbf{x}_{u,t} = \mathbf{f}_{a,u,t} \mathbf{c}_{t,u}, \quad (4)$$

$$\mathbf{c}_u^T = \mathbf{s}_u^T \mathbf{F}_T, \quad (5)$$

where $\mathbf{F}_T \in \mathbb{C}^{T \times T}$ denotes a T -point DFT matrix, $\mathbf{c}_u = [c_{t,u}]_{1 \leq t \leq T}$ the time encoded version of the data symbols $\mathbf{s}_u = [s_{t,u}]_{1 \leq t \leq T} \in \mathbb{C}^T$, $t = 1, \dots, T$ denotes the time index, $s_{t,u}$, $t \in \{1, \dots, T\}$ denotes a complex data symbol chosen from a QAM constellation and with $\mathbb{E}[|s_{t,u}|^2] = \sigma_u^2$, where $\sum_{u=1}^U \sigma_u^2 = U$. For the sake of simplicity and, without loss of generality, in this work we consider only QPSK constellations.

C. Receiver Model Description

At the receiver, we consider a hybrid iterative block space-time decoder, as shown in Fig. 2. The received signal is firstly processed through the analog phase shifters, modeled by the matrix $\mathbf{W}_{a,t} \in \mathbb{C}^{N_{rx}^{RF} \times N_{rx}}$, then follows the baseband processing, composed of N_{rx}^{RF} processing chains. All elements of the matrix $\mathbf{W}_{a,t}$ must have equal norm ($|\mathbf{W}_{a,t}(i, l)|^2 = N_{rx}^{-1}$).

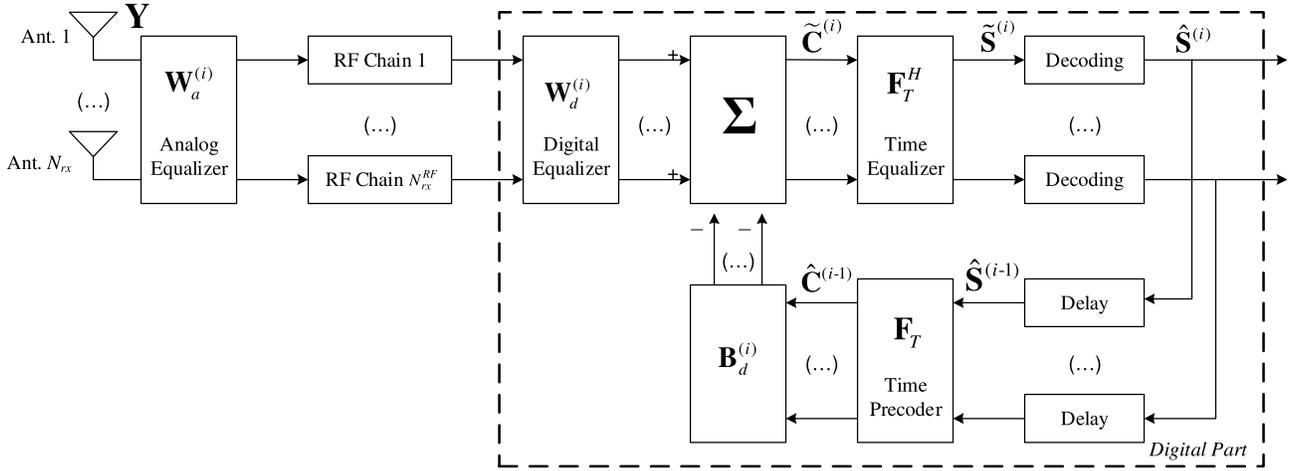


Fig. 2. Receiver block diagram.

Specifically, the baseband processing includes a digital feedback closed-loop comprising a forward and a feedback path. For the forward path the signal first passes through a linear filter $\mathbf{W}_{d,t} \in \mathbb{C}^{U \times N_{rx}^{RF}}$, then follows time decoding and demodulation. In the feedback path, the data recovered in the forward path is first modulated and time encoded, then it passes through the feedback matrix $\mathbf{B}_{d,t} \in \mathbb{C}^{U \times U}$. The time encoding of the data symbols follows (5) and its decoding obeys

$$\tilde{\mathbf{S}} = \tilde{\mathbf{C}} \mathbf{F}_T^H. \quad (6)$$

More specifically to obtain a soft estimate of the transmitted symbols we apply the IDFT transform to the codeword matrix $\tilde{\mathbf{C}} = [\tilde{\mathbf{c}}_1, \dots, \tilde{\mathbf{c}}_T]$, as in (6), where $\tilde{\mathbf{c}}_t = [\tilde{c}_{u,t}]_{1 \leq u \leq U}$. The feedback and feedforward paths are combined by subtracting the signal output of the feedback path from the filtered received signal $\mathbf{W}_{d,t} \mathbf{W}_{a,t} \mathbf{y}_t$. The proposed receiver structure can be referred to as a hybrid iterative space-time feedback equalizer (HIST-DFE). The main difference from the conventional iterative block decision feedback based equalizer is the analog front-end of phase shifters. The constraint that all elements of the analog matrix must have the same norm makes the design of the hybrid iterative equalizer harder than for the conventional fully digital one. The analog and digital forward matrices and the digital feedback matrix are designed in the following sections.

III. ITERATIVE SPACE-TIME RECEIVER DESIGN

In this section, we derive the proposed hybrid iterative space-time receiver structure for multiuser mmW massive MIMO based systems discussed in the previous section. We start by designing the fully digital receiver, that can serve as lower bound for the hybrid one and then a detailed formulation of the proposed iterative approach is presented. In this manuscript, we assume a decoupled joint transmitter-receiver optimization problem and focus on the design of the hybrid equalizer. Note that the joint transmitter-receiver design is a very hard task as discussed in [21] due to the coupled nature of the transmitter precoder and receiver equalizer.

A. Description of Iterative Receiver

A block diagram of the proposed HIST-DFE based receiver is presented in Fig. 2. At the i th iteration the received signal at the t th time slot is given by

$$\tilde{\mathbf{c}}_t^{(i)} = \mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} \mathbf{y}_t - \mathbf{B}_{d,t}^{(i)} \hat{\mathbf{c}}_t^{(i-1)}, \quad (7)$$

where $\mathbf{W}_{a,t}^{(i)} \in \mathbb{C}^{N_{rx}^{RF} \times N_{rx}}$ denotes the analog part of the feedforward matrix, $\mathbf{W}_{d,t}^{(i)} \in \mathbb{C}^{U \times N_{rx}^{RF}}$ the digital part of the feedforward matrix and $\mathbf{B}_{d,t}^{(i)} \in \mathbb{C}^{U \times U}$ the feedback matrix. The matrix $\hat{\mathbf{C}}^{(i-1)} = [\hat{\mathbf{c}}_1^{(i-1)}, \dots, \hat{\mathbf{c}}_T^{(i-1)}] \in \mathbb{C}^{U \times T}$ is the DFT of the detector output $\hat{\mathbf{S}}^{(i-1)}$

$$\hat{\mathbf{C}}^{(i-1)} = \hat{\mathbf{S}}^{(i-1)} \mathbf{F}_T, \quad (8)$$

where $\hat{\mathbf{C}}^{(i)}$ is the hard estimate of the transmitted codeword \mathbf{C} and $\hat{\mathbf{S}}^{(i)}$ the hard decision associated with the QPSK data symbols $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_t, \dots, \mathbf{s}_T]$ with $\mathbf{s}_t = [s_{t,1}, \dots, s_{t,U}]^T$, at iteration i

$$\hat{\mathbf{S}}^{(i)} = \text{sign}(\tilde{\mathbf{S}}^{(i)}). \quad (9)$$

From the central limit theorem the entries of vector $\mathbf{c}_t = [c_{u,t}]_{1 \leq u \leq U}$, $t \in \{1, \dots, T\}$ (the output of the DFT at user u) are Gaussian distributed, then as the input-output relationship between variables \mathbf{c}_t and $\hat{\mathbf{c}}_t^{(i)}$, $t \in \{1, \dots, T\}$ is memoryless, by the Bussgang theorem [38] follows

$$\hat{\mathbf{c}}_t^{(i)} = \Psi^{(i)} \mathbf{c}_t + \hat{\mathbf{e}}_t^{(i)}, \quad t \in \{1, \dots, T\}, \quad (10)$$

where $\Psi^{(i)}$ is a diagonal matrix given by

$$\Psi^{(i)} = \text{diag}(\psi_1^{(i)}, \dots, \psi_u^{(i)}, \dots, \psi_U^{(i)}), \quad (11)$$

$$\psi_u^{(i)} = \frac{\mathbb{E}[\hat{\mathbf{c}}_t^{(i)}(u) \mathbf{c}_t^*(u)]}{\mathbb{E}[|\mathbf{c}_t(u)|^2]}, \quad u \in \{1, \dots, U\}, \quad (12)$$

and $\hat{\mathbf{e}}_t^{(i)}$ is a zero mean error vector uncorrelated with \mathbf{c}_t , $t \in \{1, \dots, T\}$, with

$$\mathbb{E}[\hat{\mathbf{e}}_t^{(i)} \hat{\mathbf{e}}_t^{(i)H}] = \left(\mathbf{I}_{N_s} - |\Psi^{(i)}|^2 \right) \sigma_u^2. \quad (13)$$

Let us consider $\mathbf{H}_t = [\mathbf{H}_1 \mathbf{f}_{a,1,t}, \dots, \mathbf{H}_U \mathbf{f}_{a,U,t}] \in \mathbb{C}^{N_{rx} \times U}$ and define the vector $\boldsymbol{\Gamma}^{(i)} = T^{-1} \sum_{t=1}^T (\mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} \mathbf{H}_t)$, that corresponds to the equivalent overall channel from signal \mathbf{s}_t to $\tilde{\mathbf{s}}_t^{(i)}$. If there is no intersymbol interference (ISI) and channel noise at the output of the IDFT, then $\tilde{\mathbf{s}}_t^{(i)} = \boldsymbol{\Gamma}^{(i)} \mathbf{s}_t$ [34]. On the other hand with ISI plus channel noise

$$\tilde{\mathbf{s}}_t^{(i)} = \boldsymbol{\Gamma}^{(i)} \mathbf{s}_t + \tilde{\mathbf{e}}_t^{(i)}, \quad (14)$$

where $\tilde{\mathbf{e}}_t^{(i)} = \tilde{\mathbf{s}}_t^{(i)} - \boldsymbol{\Gamma}^{(i)} \mathbf{s}_t$ denotes an overall error that includes both the channel noise and the residual ISI. From (14), it follows that

$$\tilde{\mathbf{S}}^{(i)} = \boldsymbol{\Gamma}^{(i)} \mathbf{S} + \tilde{\mathbf{E}}^{(i)}, \quad (15)$$

where $\mathbf{S}^{(i)} = [\mathbf{s}_1^{(i)}, \dots, \mathbf{s}_T^{(i)}]$, $\tilde{\mathbf{S}}^{(i)} = [\tilde{\mathbf{s}}_1^{(i)}, \dots, \tilde{\mathbf{s}}_T^{(i)}]$ and $\tilde{\mathbf{E}}^{(i)} = [\tilde{\mathbf{e}}_1^{(i)}, \dots, \tilde{\mathbf{e}}_T^{(i)}]$. By applying a DFT to (15), in agreement with (6), we have

$$\tilde{\mathbf{c}}_t^{(i)} = \boldsymbol{\Gamma}^{(i)} \mathbf{c}_t + \tilde{\boldsymbol{\epsilon}}_t^{(i)}, \quad (16)$$

where

$$\tilde{\boldsymbol{\epsilon}}_t^{(i)} = \tilde{\mathbf{c}}_t^{(i)} - \boldsymbol{\Gamma}^{(i)} \mathbf{c}_t = \tilde{\mathbf{E}}^{(i)} \mathbf{f}_t$$

Therefore, from (7) and (10), it follows that

$$\begin{aligned} \tilde{\mathbf{c}}_t^{(i)} &= \mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} (\mathbf{H}_t \mathbf{c}_t + \mathbf{n}_t) - \mathbf{B}_{d,t}^{(i)} \left(\boldsymbol{\Psi}^{(i-1)} \mathbf{c}_t + \hat{\boldsymbol{\epsilon}}_t^{(i-1)} \right) \\ &= \boldsymbol{\Gamma}^{(i)} \mathbf{c}_t - \boldsymbol{\Gamma}^{(i)} \mathbf{c}_t + \mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} \mathbf{H}_t \mathbf{c}_t \\ &\quad - \mathbf{B}_{d,t}^{(i)} \boldsymbol{\Psi}^{(i-1)} \mathbf{c}_t - \mathbf{B}_{d,t}^{(i)} \hat{\boldsymbol{\epsilon}}_t^{(i-1)} + \mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} \mathbf{n}_t, \end{aligned} \quad (17)$$

and, consequently, the error term has three parts 1) the residual ISI, 2) the error stemming from the erroneous estimate made by $\hat{\mathbf{c}}_t$ of the signal \mathbf{c}_t and 3) the part corresponding to the channel noise:

$$\begin{aligned} \tilde{\boldsymbol{\epsilon}}_t^{(i)} &= \underbrace{(\mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} \mathbf{H}_t - \boldsymbol{\Gamma}^{(i)} - \mathbf{B}_{d,t}^{(i)} \boldsymbol{\Psi}^{(i-1)}) \mathbf{c}_t}_{\text{Residual ISI}} \\ &\quad - \underbrace{\mathbf{B}_{d,t}^{(i)} \hat{\boldsymbol{\epsilon}}_t^{(i-1)}}_{\text{Estimate } \hat{\mathbf{c}}_t^{(i)}} + \underbrace{\mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} \mathbf{n}_t}_{\text{Channel Noise}}. \end{aligned} \quad (18)$$

Let $\mathbf{W}_{ad,t}^{(i)} = \mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)}$, then $\boldsymbol{\Gamma}^{(i)} = T^{-1} \sum_{t=1}^T \mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t$ and (18) simplifies to

$$\begin{aligned} \tilde{\boldsymbol{\epsilon}}_t^{(i)} &= \left(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \boldsymbol{\Gamma}^{(i)} - \mathbf{B}_{d,t}^{(i)} \boldsymbol{\Psi}^{(i-1)} \right) \mathbf{c}_t \\ &\quad - \mathbf{B}_{d,t}^{(i)} \hat{\boldsymbol{\epsilon}}_t^{(i-1)} + \mathbf{W}_{ad,t}^{(i)} \mathbf{n}_t. \end{aligned} \quad (19)$$

Without loss of generality let us condition $\boldsymbol{\Gamma}^{(i)}$ to be equal to the identity matrix, then

$$\begin{aligned} \tilde{\boldsymbol{\epsilon}}_t^{(i)} &= \tilde{\mathbf{c}}_t^{(i)} - \mathbf{c}_t \\ &= \left(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U - \mathbf{B}_{d,t}^{(i)} \boldsymbol{\Psi}^{(i-1)} \right) \mathbf{c}_t - \mathbf{B}_{d,t}^{(i)} \hat{\boldsymbol{\epsilon}}_t^{(i-1)} \\ &\quad + \mathbf{W}_{ad,t}^{(i)} \mathbf{n}_t \end{aligned} \quad (20)$$

If we further assume that error $\boldsymbol{\epsilon}_t^{(i)}$ is complex Gaussian distributed, then for a QPSK constellation with Gray mapping the average BER is given by

$$\text{BER} = \frac{1}{UT} \sum_{u=1}^U \sum_{t=1}^T \mathcal{Q} \left(\sqrt{\left(\text{MSE}_{t,u}^{(i)} \right)^{-1}} \right), \quad (21)$$

where $\mathcal{Q}(\cdot)$ denotes the Q-function, $\text{MSE}_{t,u}^{(i)}$ is the mean square error on samples $\tilde{c}_{t,u}^{(i)}$, with $\tilde{\mathbf{c}}_t^{(i)} = [(\tilde{c}_{t,1}^{(i)})^H, \dots, (\tilde{c}_{t,U}^{(i)})^H]^H$, at iteration i . The accuracy of the Gaussian assumption will be validated by numerical simulations, in the numerical results section. For low values of x the function $\mathcal{Q}(x^{-1})$ is convex, then from the Jensen inequality (21) is lower bounded by

$$\text{BER} \geq \mathcal{Q} \left(\sqrt{\left(\frac{1}{UT} \sum_{t=1}^T \text{MSE}_t^{(i)} \right)^{-1}} \right), \quad (22)$$

where $\text{MSE}_t^{(i)} = \sum_{u=1}^U \text{MSE}_{t,u}^{(i)}$ is the MSE at time slot t and is given by

$$\text{MSE}_t^{(i)} = \mathbb{E}[\|\tilde{\mathbf{c}}_t^{(i)} - \mathbf{c}_t\|^2] = \mathbb{E}[\|\tilde{\boldsymbol{\epsilon}}_t^{(i)}\|^2]. \quad (23)$$

The lower bound (22) is valid for the high signal to noise ratio regime, where $\text{MSE}_{t,u}^{(i)}, t \in \{1, \dots, T\}, u \in \{1, \dots, U\}$ tends to zero. If $\text{MSE}_{t,u}^{(i)} = \text{MSE}_{t',u'}^{(i)}, \forall t, t' \in \{1, \dots, T\}, u, u' \in \{1, \dots, U\}$, then the BER expression given by (21) matches with the corresponding lower bound (22).

From the Bussgang [38] theorem, $\hat{\boldsymbol{\epsilon}}_t^{(i)}$ is a zero mean error vector uncorrelated with \mathbf{c}_t , as previously mentioned, then from (20) the average error ($\tilde{\boldsymbol{\epsilon}}_t^{(i)}$) power is given by

$$\begin{aligned} \text{MSE}_t^{(i)} &= \mathbb{E}[\|\tilde{\boldsymbol{\epsilon}}_t^{(i)}\|^2] \\ &= \left\| \mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \boldsymbol{\Gamma}_t^{(i)} - \mathbf{B}_{d,t}^{(i)} \boldsymbol{\Psi}^{(i-1)} \right\|_F^2 \sigma_u^2 \\ &\quad + \left\| \mathbf{B}_{d,t}^{(i)} (\mathbf{I}_U - |\boldsymbol{\Psi}^{(i-1)}|^2)^{1/2} \right\|_F^2 \sigma_u^2 + \left\| \mathbf{W}_{ad,t}^{(i)} \right\|_F^2 \sigma_n^2. \end{aligned} \quad (24)$$

B. Design of Digital Iterative Space-Time Receiver

Firstly, we design the fully digital iterative space-time receiver based on the IB-DFE principles, i.e., the analog part is not considered. The performance of this approach can be regarded as a lower bound for the proposed iterative hybrid space-time equalizer designed in the next section. This section is included for completeness since the proposed hybrid iterative space-time receiver will be compared against the digital iterative space-time receiver and as will be shown in the next sections the design of the hybrid receiver relies on the digital counterpart, see (33).

To design the equalizer, we use as a metric the average BER. Accordingly, to (21)-(23), the minimization of the average BER is equivalent to the minimization of the MSE

$$\begin{aligned} &\left((\mathbf{W}_{ad,t}^{(i)})_{opt}, (\mathbf{B}_{d,t}^{(i)})_{opt} \right) \\ &= \arg \min \text{MSE}_t^{(i)} \\ &\text{s.t. } \sum_{t=1}^T \text{diag}(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t) = T \mathbf{I}_U. \end{aligned} \quad (25)$$

In this case, the number of receiver RF chains is equal to the number of receiver antennas, and thus we only have a digital linear feedforward filter referred as $\mathbf{W}_{ad,t}$ and a feedback filter $\mathbf{B}_{d,t}$. The solution to the optimization problem (25) is

(see Appendix A),

$$(\mathbf{W}_{ad,t}^{(i)})_{opt} = \boldsymbol{\Omega} \left(\mathbf{R}_t^{(i-1)} \right)^{-1} \mathbf{H}_t^H, \quad (26)$$

$$(\mathbf{B}_{d,t}^{(i)})_{opt} = \left((\mathbf{W}_{ad,t}^{(i)})_{opt} \mathbf{H}_t - \mathbf{I}_U \right) \left(\boldsymbol{\Psi}^{(i-1)} \right)^H, \quad (27)$$

$$\boldsymbol{\Omega} = T \left(\sum_{t=1}^T \text{diag} \left(\left(\mathbf{R}_t^{(i-1)} \right)^{-1} \mathbf{H}_t^H \mathbf{H}_t \right) \right)^{-1}, \quad (28)$$

$$\mathbf{R}_t^{(i-1)} = \mathbf{H}_t^H \mathbf{H}_t (\mathbf{I}_U - |\boldsymbol{\Psi}^{(i-1)}|^2) + \sigma_n^2 \sigma_u^{-2} \mathbf{I}_U. \quad (29)$$

C. Design of Hybrid Iterative Space-Time Receiver

In this section, we design the proposed iterative hybrid space-time receiver. Clearly, the previous optimization problem of (25) does not take into account the analog domain constraints. Let us denote by \mathcal{W}_a the set of feasible RF equalizers, i.e. the set of $N_{rx} \times N_{rx}^{RF}$ matrices with constant-magnitude entries, then the reformulated optimization problem for the hybrid iterative equalizer is as follows

$$\begin{aligned} & \left((\mathbf{W}_{a,t}^{(i)})_{opt}, (\mathbf{W}_{d,t}^{(i)})_{opt}, (\mathbf{B}_{d,t}^{(i)})_{opt} \right) \\ & = \arg \min \text{MSE}_t^{(i)} \\ & \text{s.t.} \quad \sum_{t=1}^T \text{diag}(\mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} \mathbf{H}_t) = T \mathbf{I}_U \\ & \quad \mathbf{W}_{a,t}^{(i)} \in \mathcal{W}_a. \end{aligned} \quad (30)$$

Due to the digital nature of the feedback equalizer $\mathbf{B}_{d,t}^{(i)}$ and since the new constraint does not impose any restriction on this matrix, the feedback equalizer for the hybrid iterative equalizer is similar to the fully digital iterative equalizer discussed in the previous section, and thus given by

$$(\mathbf{B}_{d,t}^{(i)})_{opt} = \left((\mathbf{W}_{d,t}^{(i)})_{opt} (\mathbf{W}_{a,t}^{(i)})_{opt} \mathbf{H}_t - \mathbf{I}_U \right) \left(\boldsymbol{\Psi}^{(i-1)} \right)^H. \quad (31)$$

From (24) and (31), the MSE expression simplifies to, see Appendix B

$$\begin{aligned} \text{MSE}_t^{(i)} & = \left\| (\mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U) (\mathbf{I}_U - |\boldsymbol{\Psi}^{(i-1)}|^2)^{1/2} \right\|_F^2 \sigma_u^2 \\ & \quad + \left\| \mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} \right\|_F^2 \sigma_n^2. \end{aligned} \quad (32)$$

As can be seen, in Appendix C, the MSE expression (32) is equal (up to a constant) to

$$\overline{\text{MSE}}_t^{(i)} = \left\| \left(\mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} - (\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt} \right) \left(\tilde{\mathbf{R}}_t^{(i-1)} \right)^{1/2} \right\|_F^2, \quad (33)$$

$$\tilde{\mathbf{R}}_t^{(i-1)} = \mathbf{H}_t (\mathbf{I}_U - |\boldsymbol{\Psi}^{(i-1)}|^2) \mathbf{H}_t^H + \sigma_n^2 \sigma_u^{-2} \mathbf{I}_U, \quad (34)$$

$$(\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt} = (\mathbf{I}_U - |\boldsymbol{\Psi}^{(i-1)}|^2) \boldsymbol{\Omega}^{-1} (\mathbf{W}_{ad,t}^{(i)})_{opt}, \quad (35)$$

where $(\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt}$ and $\tilde{\mathbf{R}}_t^{(i-1)}$ denote a non-normalized version of the optimum fully digital feedforward matrix and the correlation of the ISI plus channel noise. Therefore, the

optimization problem (30) simplifies to

$$\begin{aligned} & \left((\mathbf{W}_{a,t}^{(i)})_{opt}, (\mathbf{W}_{d,t}^{(i)})_{opt} \right) \\ & = \arg \min \overline{\text{MSE}}_t^{(i)} \\ & \text{s.t.} \quad \sum_{t=1}^T \text{diag}(\mathbf{W}_{d,t}^{(i)} \mathbf{W}_{a,t}^{(i)} \mathbf{H}_t) = T \mathbf{I}_U \\ & \quad \mathbf{W}_{a,t}^{(i)} \in \mathcal{W}_a. \end{aligned} \quad (36)$$

Due to the non-convex nature of the feasible set \mathcal{W}_a , an analytical solution to the problem (36) is difficult to obtain, if not impossible. Nevertheless, we find an approximate solution to problem (36) by assuming that the matrix $\mathbf{W}_{a,t}^{(i)}$ is a N_{rx}^{RF} sparse linear combination of vectors $\mathbf{a}_{rx,u}(\theta_{j,l}^{rx,u})$ or equivalently a N_{rx}^{RF} sparse linear combination of the columns of matrix $\mathbf{A}_{rx} = [\mathbf{A}_{rx,1}, \dots, \mathbf{A}_{rx,U}]$. We may say that $\mathbf{W}_{a,t}^{(i)}$ has a N_{rx}^{RF} term representation over the dictionary \mathbf{A}_{rx} . Equivalently, we may think of \mathbf{A}_{rx} as the analog feedforward equalizer and that only N_{rx}^{RF} outputs of the equalized signal $\mathbf{A}_{rx} \mathbf{Y}$ are connected/processed by the digital part of the feedforward equalizer. Therefore, the optimization problem (36) can be approximated as follows

$$\begin{aligned} & (\ddot{\mathbf{W}}_{d,t}^{(i)})_{opt} \\ & = \arg \min \left\| \left(\ddot{\mathbf{W}}_{d,t}^{(i)} \mathbf{A}_{rx}^H - (\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt} \right) \left(\tilde{\mathbf{R}}_t^{(i-1)} \right)^{1/2} \right\|_F^2 \\ & \text{s.t.} \quad \sum_{t=1}^T \text{diag}(\ddot{\mathbf{W}}_{d,t}^{(i)} \mathbf{A}_{rx}^H \mathbf{H}_t) = T \mathbf{I}_U \\ & \quad \left\| \text{diag}((\ddot{\mathbf{W}}_{d,t}^{(i)})^H \ddot{\mathbf{W}}_{d,t}^{(i)}) \right\|_0 = N_{rx}^{RF}, \end{aligned} \quad (37)$$

where $\left\| \text{diag}((\ddot{\mathbf{W}}_{d,t}^{(i)})^H \ddot{\mathbf{W}}_{d,t}^{(i)}) \right\|_0 = N_{rx}^{RF}$ represents the sparsity constraint and enforces that only N_{rx}^{RF} columns of matrix $\ddot{\mathbf{W}}_{d,t}^{(i)}$ are non-zero. The optimum digital feedforward matrix $(\mathbf{W}_{d,t}^{(i)})_{opt}$ is obtained from the solution $(\ddot{\mathbf{W}}_{d,t}^{(i)})_{opt}$, of optimization problem (37), by removing the zero columns and the optimum analog feedforward matrix $(\mathbf{W}_{a,t}^{(i)})_{opt}$ is obtained from \mathbf{A}_{rx}^H by selecting the rows corresponding to the non-zero columns of $\ddot{\mathbf{W}}_{d,t}^{(i)}$.

Let us consider optimization problem (37), without the sparsity constraint (the sparsity constraint is enforced next)

$$\begin{aligned} & (\ddot{\mathbf{W}}_{d,t}^{(i)})_{opt} \\ & = \arg \min \left\| \left(\ddot{\mathbf{W}}_{d,t}^{(i)} \mathbf{A}_{rx}^H - (\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt} \right) \left(\tilde{\mathbf{R}}_t^{(i-1)} \right)^{1/2} \right\|_F^2 \\ & \text{s.t.} \quad \sum_{t=1}^T \text{diag}(\ddot{\mathbf{W}}_{d,t}^{(i)} \mathbf{A}_{rx}^H \mathbf{H}_t) = T \mathbf{I}_U. \end{aligned} \quad (38)$$

Using (46) to rewrite the constraint of optimization problem (38) follows that the associated Lagrangian

is [39], [40]

$$\begin{aligned} \mathcal{L}(\mu_u, \ddot{\mathbf{W}}_{d,t'}^{(i)}) &= \left\| \left(\ddot{\mathbf{W}}_{d,t'}^{(i)} \mathbf{A}_{rx}^H - (\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt} \right) \left(\tilde{\mathbf{R}}_t^{(i-1)} \right)^{1/2} \right\|_F^2 \\ &+ \sum_{u=1}^U \mu_u \text{tr} \left(\ddot{\mathbf{W}}_{d,t'}^{(i)} \mathbf{A}_{rx}^H \mathbf{H}_t \mathbf{e}_u \mathbf{e}_u^H \right) \\ &+ \sum_{u=1}^U \mu_u \left(\sum_{t'=1, t' \neq t}^T \text{tr} \left(\ddot{\mathbf{W}}_{d,t'}^{(i)} \mathbf{A}_{rx}^H \mathbf{H}_t \mathbf{e}_u \mathbf{e}_u^H \right) - T \right), \quad (39) \end{aligned}$$

where μ_u , $u \in \{1, \dots, U\}$ are the Lagrange multipliers. Let $\mathbf{U}_d = \text{diag}(\mu_1, \dots, \mu_U) = \sum_{u=1}^U \mu_u \mathbf{e}_u \mathbf{e}_u^H$, then by taking the derivate in relation to $\ddot{\mathbf{W}}_{d,t}^{(i)}$ we obtain the optimality condition [39], [40]

$$\frac{\partial \mathcal{L}(\mu_u, \ddot{\mathbf{W}}_{d,t}^{(i)})}{\partial \left(\ddot{\mathbf{W}}_{d,t}^{(i)} \right)^*} = \mathbf{W}_{res,t}^{(i)} \mathbf{A}_{rx} = \mathbf{0}, \quad (40)$$

where (40) represents the orthogonality principle subjacent to MSE estimator (37) and $\mathbf{W}_{res,t}^{(i)}$ is the residue matrix that is given by

$$\mathbf{W}_{res,t}^{(i)} = \left(\ddot{\mathbf{W}}_{d,t}^{(i)} \mathbf{A}_{rx}^H - (\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt} \right) \left(\tilde{\mathbf{R}}_t^{(i-1)} \right) + \mathbf{U}_d \mathbf{H}_t^H. \quad (41)$$

From the definition of matrices $\tilde{\mathbf{R}}_t^{(i-1)}$, $\mathbf{R}_t^{(i-1)}$, $(\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt}$ and $(\mathbf{W}_{ad,t}^{(i)})_{opt}$ (see (34), (29), (35) and (26), respectively), equation (41) simplifies to

$$\mathbf{W}_{res,t}^{(i)} = \ddot{\mathbf{W}}_{d,t}^{(i)} \mathbf{A}_{rx}^H \tilde{\mathbf{R}}_t^{(i-1)} - \mathbf{\Omega}_d \mathbf{H}_t^H, \quad (42)$$

where $\mathbf{\Omega}_d = \mathbf{I}_U - |\Psi^{(i-1)}|^2 + \mathbf{U}_d$ denotes a redefined Lagrangian multipliers matrix, that must be selected so that the constraint of the optimization problem (38) is respected.

To enforce the sparsity constraint, the best columns of the dictionary \mathbf{A}_{rx} are selected using an iterative greedy method. At each iteration the column of \mathbf{A}_{rx} that is most correlated with the actual value of the residue $\mathbf{W}_{res,t}^{(i)}$ is selected. It should be emphasized that for a mmW channel the number of paths is small and then the complexity of this selection procedure is small. In the first iteration, the residue is set to the trivial value $\mathbf{W}_{res,t}^{(i)} = -(\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt} \left(\tilde{\mathbf{R}}_t^{(i-1)} \right)$. Then, after identifying a set of columns of the matrix \mathbf{A}_{rx} (one column per iteration) to form the analog feedforward equalizer matrix $(\mathbf{W}_{a,t}^{(i)})_{opt}$ we obtain the optimum digital feedforward equalizer matrix $(\mathbf{W}_{d,t}^{(i)})_{opt}$ using the orthogonality condition (40), as follows.

Let $\mathbf{W}_{d,t}^{(i)}$ and $(\mathbf{W}_{a,t}^{(i)})_{opt}$ denote the digital and analog parts of the feedforward matrix, restricted to the selected indices, i.e. $\ddot{\mathbf{W}}_{d,t'}^{(i)} = [\mathbf{W}_{d,t'}^{(i)}, \mathbf{0}]$, if the selected indices were the first. Then, from (42) the orthogonality condition simplifies to

$$\left(\mathbf{W}_{d,t}^{(i)} (\mathbf{W}_{a,t}^{(i)})_{opt} \tilde{\mathbf{R}}_t^{(i-1)} - \mathbf{\Omega}_d \mathbf{H}_t^H \right) (\mathbf{W}_{a,t}^{(i)})_{opt}^H = \mathbf{0}. \quad (43)$$

The solution to equation (43) is

$$(\mathbf{W}_{d,t}^{(i)})_{opt} = \mathbf{\Omega}_d \left((\mathbf{W}_{a,t}^{(i)})_{opt} \mathbf{H}_t \right)^H \left(\mathbf{R}_{d,t}^{(i-1)} \right)^{-1}, \quad (44)$$

Algorithm 1 The Proposed HIST-DFE Algorithm for Multi-User mmWave Massive MIMO Systems

Input: $(\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt}$
1: $(\mathbf{W}_{a,t}^{(i)})_{opt} = \mathbf{Empty Matrix}$
2: $\mathbf{W}_{res,t}^{(i)} = -(\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt} \tilde{\mathbf{R}}_t^{(i-1)}$
3:for $r \leq N_r^{RF}$ **do**
4: $k = \arg \max_{l=1, \dots, N_{cl} N_{ray}} \left(\mathbf{A}_{rx}^H (\mathbf{W}_{res,t}^{(i)})^H \mathbf{W}_{res,t}^{(i)} \mathbf{A}_{rx} \right)_{l,l}$
5: $(\mathbf{W}_{a,t}^{(i)})_{opt} = [((\mathbf{W}_{a,t}^{(i)})_{opt})^H | (\mathbf{A}_{rx}^{(k)})]^H$
6: $(\mathbf{W}_{d,t}^{(i)})_{opt} = \mathbf{\Omega}_d \left((\mathbf{W}_{a,t}^{(i)})_{opt} \mathbf{H}_t \right)^H \left(\mathbf{R}_{d,t}^{(i-1)} \right)^{-1}$
7: $\mathbf{W}_{res,t}^{(i)} = ((\mathbf{W}_{d,t}^{(i)})_{opt} (\mathbf{W}_{a,t}^{(i)})_{opt} - (\overline{\mathbf{W}}_{ad,t}^{(i)})_{opt}) \tilde{\mathbf{R}}_t^{(i-1)} + \mathbf{U}_d (\mathbf{H}_t)^H$
8:end for
9:return $(\mathbf{W}_{a,t}^{(i)})_{opt}, (\mathbf{W}_{d,t}^{(i)})_{opt}$

where $\mathbf{R}_{d,t}^{(i-1)} = (\mathbf{W}_{a,t}^{(i)})_{opt} \tilde{\mathbf{R}}_t^{(i-1)} (\mathbf{W}_{a,t}^{(i)})_{opt}^H$ and to respect the constraint of problem (37) $\mathbf{\Omega}_d$ is given by

$$\begin{aligned} \mathbf{\Omega}_d &= T \left(\sum_{t=1}^T \text{diag} \left(\left((\mathbf{W}_{a,t}^{(i)})_{opt} \mathbf{H}_t \right)^H \left(\mathbf{R}_{d,t}^{(i-1)} \right)^{-1} \right. \right. \\ &\quad \left. \left. \times (\mathbf{W}_{a,t}^{(i)})_{opt} \mathbf{H}_t \right) \right)^{-1}, \quad (45) \end{aligned}$$

After obtaining the optimum value of the digital feedforward matrix $\mathbf{W}_{d,t}^{(i)}$ the residue matrix (41) is updated. The previous steps iterate on the updated residue value to obtain the N_r^{RF} index set to index the dictionary \mathbf{A}_{rx} . The pseudo-code for the proposed algorithm is given in Algorithm 1.

IV. PERFORMANCE RESULTS

In this section, we access the performance of the proposed hybrid iterative receiver structure. We consider, for each user, a clustered narrowband channel model [41] with $N_{cl} = 8$ clusters, each with $N_{ray} = 4$ rays, with Laplacian distributed azimuth angles of arrival and departure as in [29]. The average power of all N_{cl} clusters is the same and the angle spread at both the transmitter and receiver is set to 8° degrees. Although the hybrid receiver structure developed in the paper can be applied to arbitrary antenna arrays, we use uniform linear arrays (ULAs) to obtain the results. The antenna element spacing is assumed to be half-wavelength and the carrier frequency is set to 72 GHz, [41]. The channel remains constant during a block, with size T , and takes independent values between blocks. We assume perfect synchronization and channel knowledge at the receiver side. In the following we present both semi-analytic results, by evaluating numerically the BER expression (21) and the simulated average BER. All results were obtained by considering QPSK modulation. The semi-analytic results are presented to validate the Gaussian assumption employed in the derivation of the proposed hybrid iterative equalizer.

We present results for three different scenarios. The scenario 1 can be seen as the baseline scenario, and the other two scenarios are a scaling of the first. For the scenario 1, the parameters are $N_{rx} = 16$, $N_{tx} = 8$, $U = 4$, $N_{rx}^{RF} = 4$.

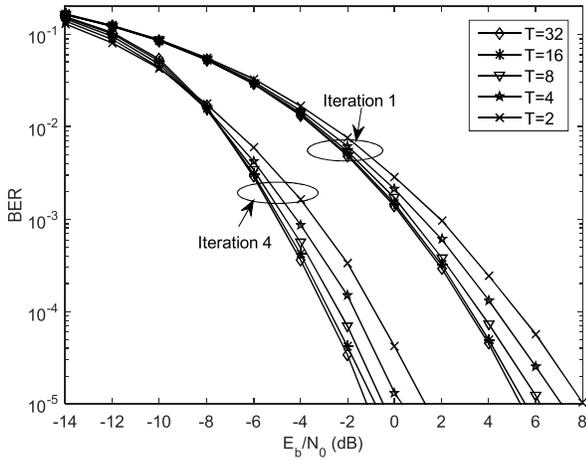


Fig. 3. Performance of the proposed hybrid equalizer as a function of the block size T , for scenario 2.

For scenario 2 the parameters are $N_{rx} = 32$, $N_{tx} = 16$, $U = 8$, $N_{rx}^{RF} = 8$, i.e. all parameters of scenario 2 are two times higher than the ones of scenario 1. Finally, for scenario 3 the parameters are $N_{rx} = 64$, $N_{tx} = 16$, $U = 16$, $N_{rx}^{RF} = 16$. For these three scenarios, we present results for iteration 1, 2 and 4 of the derived digital and hybrid iterative space-time equalizers, which are referred as *digital* and *hybrid*, in the following. To compute the analog and digital parts of the proposed hybrid equalizer we consider that the analog precoder of the u th user is generated randomly accordingly to (3). We also compare the results with the two-stage (TS) approach discussed in [29], extended here from the downlink to the uplink. In the first stage a selfish beamforming is performed to maximize the received power for each UT-BS link assuming free interference links and in the second one a conventional linear MMSE multi-user equalizer is employed to mitigate the interference. For the TS approach we consider beamsteering codebooks [29] with cardinality $32(B_{RF}^{MS} = B_{RF}^{BS} = 5$ bits) both at the mobile stations and BS. This approach is denominated by TS approach in the following.

The performance metric considered is the BER, which is presented as a function of the E_b/N_0 , with E_b denoting the average bit energy and N_0 denoting the one-sided noise power spectral density. We consider $\sigma_1^2 = \dots = \sigma_U^2 = 1$ and then the average E_b/N_0 is identical for all users $u \in \{1, \dots, U\}$ and is given by $E_b/N_0 = \sigma_u^2 / (2\sigma_n^2) = \sigma_n^{-2} / 2$.

Let us start the analysis of the proposed method by verifying the impact of the parameter T . In Fig. 3 we present BER results for scenario 2 and different values of the parameter T . These results were obtained for iterations 1 and 4. From Fig. 3 we verify that the BER performance improves with T for medium-to-high E_b/N_0 regimes, since the diversity of the mmW massive MIMO system increases with T and the accuracy of the Gaussian assumption is higher. Comparing for example the results for $T = 32$ and $T = 4$, a performance penalty of approximately 1 dB can be observed, at a target BER of 10^{-4} , for both iterations 1 and 4. We can also see that the performance obtained for $T = 16$ and $T = 32$ is

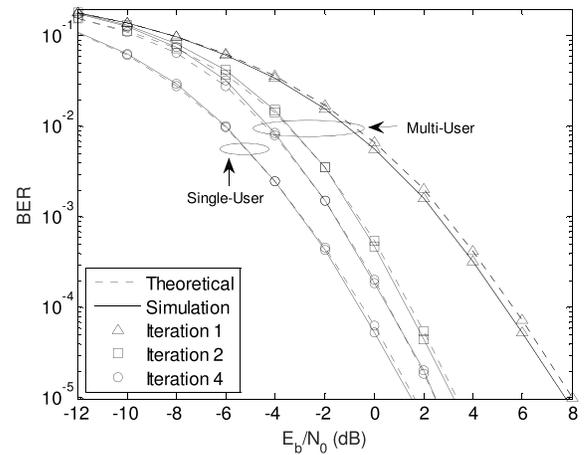


Fig. 4. Performance of the proposed hybrid equalizer for scenario 1, semi-analytic BER approximation.

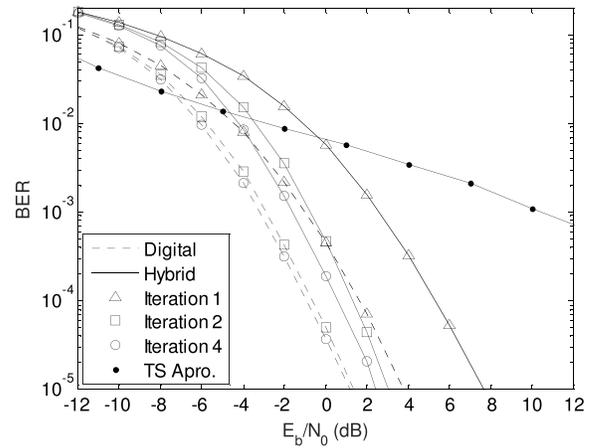


Fig. 5. Performance of the proposed hybrid equalizer for scenario 1, and full-digital equalizer.

approximately the same for both iterations. Hereinafter all results are obtained for $T = 32$.

Now let us analyze the average BER performance of the proposed hybrid equalizer over a predefined E_b/N_0 range. Firstly, consider scenario 1, where $N_{rx} = 16$, $N_{tx} = 8$, $U = N_{rx}^{RF} = 4$. In Fig. 4 we present results for the proposed hybrid iterative equalizer. In this figure, we also present the semi-analytic curves and the single-user bound, where only one user is active. This figure demonstrates that the Gaussian approximation made in the derivation of the proposed algorithm is very accurate, for all iterations and for the whole E_b/N_0 range. Furthermore, it shows that for the fourth iteration the proposed hybrid approach almost achieves the single-user bound (penalty less than 1 dB). In Fig. 5 we present results for the digital and hybrid equalizers and for comparison purposes, we also plot the BER curve for the TS approach. From Fig. 5, we can see that the gaps between the digital and hybrid equalizers, at a target BER of 10^{-3} , are 4, 2 and 1 dB for iterations 1, 2 and 4, respectively. From these results, we verify that the gaps from the 1st to the 2nd iteration are much higher than from the 2nd iteration to the 4th.

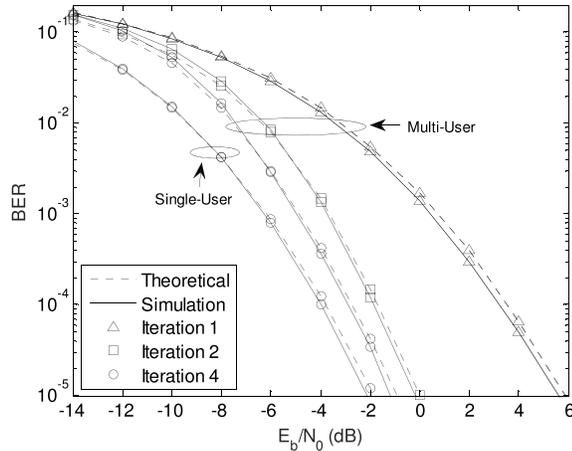


Fig. 6. Performance of the proposed hybrid equalizer for scenario 2, semi-analytic BER approximation.

This larger gap is mainly due to the removal of the residual inter symbol and multi-user interferences which enables the added benefit of a larger diversity. From the 2nd to the 4th iteration there is also a benefit from residual interference removal, but the gains are smaller since most of the interference was already removed in the 2nd iteration. Furthermore, the proposed hybrid equalizer is quite close to the digital counterpart for the 4th iteration. Therefore, the dictionary-based approximation made for the analog part of the feedforward matrix is quite accurate. Finally, comparing the proposed hybrid equalizer with the TS approach, we verify that for the low E_b/N_0 regime the TS approach outperforms the proposed hybrid equalizer. However, for medium to high E_b/N_0 (from -12 dB) the performance penalty of the TS quickly increases as the E_b/N_0 increases (we can see a penalty larger than 10 dB at a target BER target of 10^{-3}). This behavior can be explained by the fact that at the low SNR regime the BER is dominated by the noise and the selfish beamforming strategy with digital interference removal achieves good results. For medium to high SNR regime the systems are mainly dominated by the multi-user interference, and our hybrid equalizer achieves better results since the analog and digital parts of the equalizer are jointly optimized to remove the multi-user interference.

Now, let us consider the other two scenarios (scenario 2 and 3), where the parameters are scaled by a factor of 2 and 4, respectively, i.e. larger antenna arrays are considered. The results are presented in Figs. 6-9 and are similar to the ones obtained for scenario 1. Namely, the Gaussian assumption is still valid, the fourth iteration is still within 1 dB from the single user bound and the TS approach is still around 10 dB worse, when compared against iteration 1. One difference, between the first and these two scenarios, is in the better BER performance achieved. Mainly due to the higher diversity and array gain provided by the larger dimension of the antenna arrays, at the receiver side. We can also see that increasing the number of users the gap between the first and the fourth iterations increases and also the performance of the proposed hybrid approach is closer to the full digital one (compare Fig. 5 with 9).

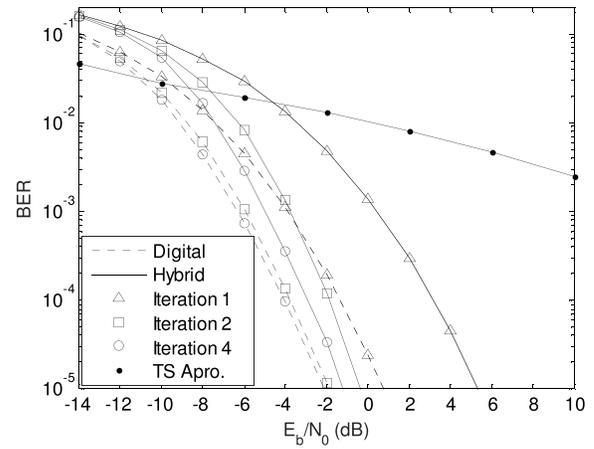


Fig. 7. Performance of the proposed hybrid equalizer for scenario 2, and full-digital equalizer.

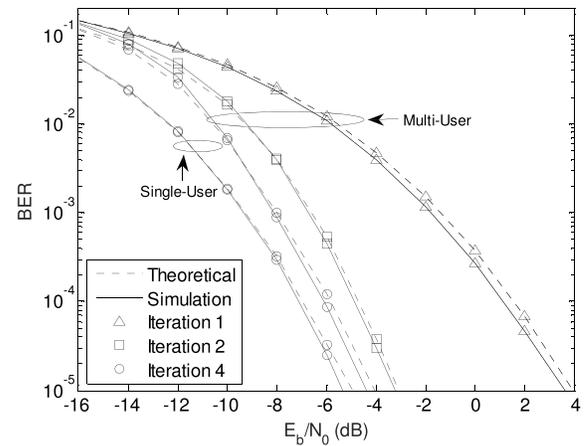


Fig. 8. Performance of the proposed hybrid equalizer for scenario 3, semi-analytic BER approximation.

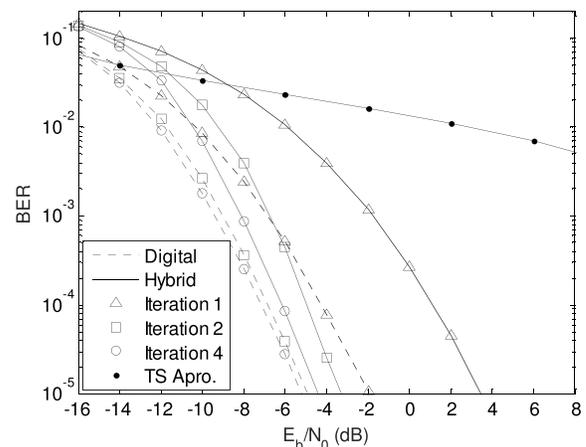


Fig. 9. Performance of the proposed hybrid equalizer for scenario 3, and full-digital equalizer.

This means that our receiver structure is quite efficient to remove the multiuser interference and mitigate the constraints imposed by the hybrid architecture. As the number of users increases, and for the low E_b/N_0 regime, the performance of the TS approach gets closer to the proposed hybrid

equalizer (compare Figs. 5, 7 and 9). These results confirm the performance degradation of the TS approach in scenarios with strong multi-user interference.

It should be also emphasized that the performance of the proposed method is constrained by the fact that the number of streams sent is equal to the number of RF chains. Improved performance would be achieved if the number of streams are lower than the number of RF chains. Indeed, for such a case the gaps between the digital and hybrid equalizer would be even smaller.

In this manuscript we used the average BER as a metric since its minimization ensures some fairness between users. Another metric that is frequently used in the literature is the sum rate (e.g. [27]). To avoid an excessively long manuscript this was not considered in this paper, but we provide here some insights on how the proposed scheme compares against the TS approach when using the sum-rate as metric. For the case with $B_{RF}^{MS} = 0$ bits that corresponds to the scenario where the users have no CSI, the proposed hybrid scheme outperforms the TS approach. By increasing the number of feedback bits B_{RF}^{MS} , the situation reverses, and the TS approach outperforms the proposed scheme, mainly for low Eb/No regime. Such results are expectable since in this manuscript we assume that the CSI is not available at the transmit side and it is well known that schemes that assume CSI at the transmit side have more DoF to achieve higher sum rate.

V. CONCLUSIONS

In this manuscript, we proposed a new hybrid iterative space-time receiver structure for multiuser mmW massive MIMO systems. We considered low-complexity UTs employing analog-only precoding and a single RF chain. At the receiver a hybrid iterative analog-digital equalizer was designed to efficiently remove the multiuser interference. The analog and digital parts of the hybrid iterative equalizer were designed jointly using as a metric the average BER. The analog part was selected from a dictionary, which efficiently models the specific hardware limitation inherent to the analog domain processing in the joint optimization problem. A space-time encoder was used, before the analog precoders, to ensure transmit Gaussian based signals, which allowed to simplify the receiver optimization and to increase the system diversity.

The results have shown that the proposed hybrid iterative space-time receiver is quite efficient to remove the multiuser interference, while allowing a performance close to the digital counterpart with a very few number of iterations, showing that the dictionary based approximation made for the analog part of the feedforward matrix is quite accurate. Moreover, it outperforms the linear TS approach recently considered for hybrid mmW massive MIMO architectures, mainly in the medium to high SNR regime and in scenarios with strong interference. Therefore, we can clearly argue that the proposed receiver structure is quite interesting for practical mmW massive MIMO based systems, where the number of RF chains must be lower than the number of transmit and receive antennas.

APPENDIX A

SOLUTION OF THE OPTIMIZATION PROBLEM (25)

Please note that the constraint associated to problem (25) may be equivalently rewritten as

$$\sum_{t=1}^T \mathbf{e}_u^H \mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t \mathbf{e}_u = \sum_{t=1}^T \text{tr} \left(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t \mathbf{e}_u \mathbf{e}_u^H \right) = T, \quad (46)$$

where $\mathbf{e}_u \in \mathbb{C}^{N_s}$ is a unit vector with entry equal to one while all others are zeros, $u = \{1, \dots, U\}$ and $\sum_{t=1}^T \text{diag}(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t) = \sum_{t=1}^T \text{diag} \left(\left(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t \right)^H \right) = T \mathbf{I}_U$. Therefore, the Lagrangian associated to problem (25) is [39], [40]

$$\begin{aligned} \mathcal{L} \left(\mu_u, \mathbf{W}_{ad,t}^{(i)}, \mathbf{B}_{d,t}^{(i)} \right) &= \left\| \mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U - \mathbf{B}_{d,t}^{(i)} \Psi^{(i-1)} \right\|_F^2 \sigma_u^2 \\ &+ \left\| \mathbf{B}_{d,t}^{(i)} (\mathbf{I}_U - |\Psi^{(i-1)}|^2)^{1/2} \right\|_F^2 \sigma_u^2 + \left\| \mathbf{W}_{ad,t}^{(i)} \right\|_F^2 \sigma_n^2 \\ &+ \sum_{u=1}^U \mu_u \text{tr} \left(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t \mathbf{e}_u \mathbf{e}_u^H \right) \\ &+ \sum_{u=1}^U \mu_s \left(\sum_{t'=1, t' \neq t}^T \text{tr} \left(\mathbf{W}_{ad,t'}^{(i)} \mathbf{H}_{t'} \mathbf{e}_u \mathbf{e}_u^H \right) - T \right), \quad (47) \end{aligned}$$

where μ_u , $u \in \{1, \dots, U\}$ are the Lagrange multipliers, and the respective first order optimality condition [39], [40], relative to optimization variable $\mathbf{B}_{d,t}^{(i)}$ is

$$\begin{aligned} \frac{\partial \mathcal{L} \left(\mu_u, \mathbf{W}_{ad,t}^{(i)}, \mathbf{B}_{d,t}^{(i)} \right)}{\partial \left(\mathbf{B}_{d,t}^{(i)} \right)^*} &= \mathbf{B}_{d,t}^{(i)} \sigma_u^2 - \left(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U \right) \left(\Psi^{(i-1)} \right)^H \sigma_u^2 = \mathbf{0}, \quad (48) \end{aligned}$$

whose solution gives the optimum value for $\mathbf{B}_{d,t}^{(i)}$

$$\left(\mathbf{B}_{d,t}^{(i)} \right)_{opt} = \left(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U \right) \left(\Psi^{(i-1)} \right)^H. \quad (49)$$

From (49) the Lagrangian function simplifies to

$$\begin{aligned} \mathcal{L} \left(\mu_u, \mathbf{W}_{ad,t}^{(i)}, \mathbf{B}_{d,t}^{(i)} \right) &= \left\| \left(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U \right) (\mathbf{I}_U - |\Psi^{(i-1)}|^2)^{1/2} \right\|_F^2 \sigma_u^2 \\ &+ \left\| \mathbf{W}_{ad,t}^{(i)} \right\|_F^2 \sigma_n^2 + \sum_{u=1}^U \mu_u \text{tr} \left(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t \mathbf{e}_u \mathbf{e}_u^H \right) \\ &+ \sum_{u=1}^U \mu_u \left(\sum_{t'=1, t' \neq t}^T \text{tr} \left(\mathbf{W}_{ad,t'}^{(i)} \mathbf{H}_{t'} \mathbf{e}_u \mathbf{e}_u^H \right) - T \right). \quad (50) \end{aligned}$$

Now, by taking the derivate in relation to $\mathbf{W}_{ad,t}^{(i)}$

$$\begin{aligned} & \frac{\partial \mathcal{L}(\mu_u, \mathbf{W}_{ad,t}^{(i)}, \mathbf{B}_{d,t}^{(i)})}{\partial (\mathbf{W}_{ad,t}^{(i)})^*} \\ &= (\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U) (\mathbf{I}_U - |\Psi^{(i-1)}|^2) \mathbf{H}_t^H \sigma_u^2 \\ & \quad + \mathbf{W}_{ad,t}^{(i)} \sigma_n^2 + \sum_{u=1}^U \mu_u \mathbf{e}_u \mathbf{e}_u^H \mathbf{H}_t^H, \end{aligned} \quad (51)$$

and setting it equal to zero

$$\begin{aligned} (\mathbf{W}_{ad,t}^{(i)})_{opt} &= \left((\mathbf{I}_U - |\Psi^{(i-1)}|^2) - \frac{1}{\sigma_u^2} \sum_{u=1}^U \mu_u \mathbf{e}_u \mathbf{e}_u^H \right) \mathbf{H}_t^H \\ & \quad \times \left(\mathbf{H}_t (\mathbf{I}_U - |\Psi^{(i-1)}|^2) \mathbf{H}_t^H + \frac{\sigma_n^2}{\sigma_u^2} \mathbf{I}_U \right)^{-1} \end{aligned} \quad (52)$$

Using the matrix inversion lemma (52) simplifies to

$$\begin{aligned} (\mathbf{W}_{ad,t}^{(i)})_{opt} &= \left((\mathbf{I}_U - |\Psi^{(i-1)}|^2) - \frac{1}{\sigma_u^2} \sum_{u=1}^U \mu_u \mathbf{e}_u \mathbf{e}_u^H \right) \\ & \quad \times \left(\mathbf{H}_t^H \mathbf{H}_t (\mathbf{I}_U - |\Psi^{(i-1)}|^2) + \frac{\sigma_n^2}{\sigma_u^2} \mathbf{I}_U \right)^{-1} \mathbf{H}_t^H. \end{aligned} \quad (53)$$

As $\sum_{u=1}^U \mu_u \mathbf{e}_u \mathbf{e}_u^H = \text{diag}(\mu_1, \dots, \mu_U)$ then the Lagrangian multipliers μ_s , may be redefined to $\omega_u = 1 - |\Psi_u^{(i-1)}|^2 - \sigma_u^{-2} \mu_u$, $s \in \{1, \dots, N_s\}$. Let $\mathbf{R}_t^{(i-1)} = \mathbf{H}_t^H \mathbf{H}_t (\mathbf{I}_U - |\Psi^{(i-1)}|^2) + \sigma_n^2 \sigma_u^{-2} \mathbf{I}_U$ and $\mathbf{\Omega} = \text{diag}(\omega_1, \dots, \omega_{N_s})$ then (53) reduces to

$$(\mathbf{W}_{ad,t}^{(i)})_{opt} = \mathbf{\Omega} \left(\mathbf{R}_t^{(i-1)} \right)^{-1} \mathbf{H}_t^H. \quad (54)$$

The optimum feedforward matrix $(\mathbf{W}_{ad,t}^{(i)})_{opt}$ must respect the constraint (24). From (53) and constraint (24) we find that

$$\begin{aligned} & \sum_{t=1}^T \text{tr} \left(\mathbf{\Omega} \left(\mathbf{R}_t^{(i-1)} \right)^{-1} \mathbf{H}_t^H \mathbf{H}_t \mathbf{e}_u \mathbf{e}_u^H \right) \\ &= \sum_{t=1}^T \text{tr} \left(\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t \mathbf{e}_u \mathbf{e}_u^H \right) = T, \quad u \in \{1, \dots, U\}. \end{aligned} \quad (55)$$

From (53), the definition of the trace operator and the definition of the unit vector $\mathbf{e}_u \in \mathbb{C}^U$

$$\begin{aligned} & \sum_{t=1}^T \text{diag} \left(\mathbf{\Omega} \left(\mathbf{R}_t^{(i-1)} \right)^{-1} \mathbf{H}_t^H \mathbf{H}_t \right) \\ &= \mathbf{\Omega} \sum_{t=1}^T \text{diag} \left(\left(\mathbf{R}_t^{(i-1)} \right)^{-1} \mathbf{H}_t^H \mathbf{H}_t \right) = T \mathbf{I}_U, \end{aligned} \quad (56)$$

which reduces to

$$\mathbf{\Omega} = T \left(\sum_{t=1}^T \text{diag} \left(\left(\mathbf{R}_t^{(i-1)} \right)^{-1} \mathbf{H}_t^H \mathbf{H}_t \right) \right)^{-1}. \quad (57)$$

APPENDIX B

DERIVATION OF (32) FROM (24) AND (31)

By substituting (31) in (24)

$$\begin{aligned} \text{MSE}_t^{(i)} &= \left\| (\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U) (\mathbf{I}_U - |\Psi^{(i-1)}|^2) \right\|_F^2 \sigma_u^2 \\ & \quad + \left\| (\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U) (\Psi^{(i-1)})^H (\mathbf{I}_U - |\Psi^{(i-1)}|^2)^{1/2} \right\|_F^2 \sigma_u^2 \\ & \quad + \left\| \mathbf{W}_{ad,t}^{(i)} \right\|_F^2 \sigma_n^2. \end{aligned} \quad (58)$$

As $\Psi^{(i-1)}$ and $\mathbf{I}_U - |\Psi^{(i-1)}|^2$ are diagonal then

$$\begin{aligned} \text{MSE}_t^{(i)} &= \left\| (\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U) (\mathbf{I}_U - |\Psi^{(i-1)}|^2) \right\|_F^2 \sigma_u^2 \\ & \quad + \left\| (\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U) (\mathbf{I}_U - |\Psi^{(i-1)}|^2)^{1/2} (\Psi^{(i-1)})^H \right\|_F^2 \sigma_u^2 \\ & \quad + \left\| \mathbf{W}_{ad,t}^{(i)} \right\|_F^2 \sigma_n^2. \end{aligned} \quad (59)$$

Let $\mathbf{Z} = (\mathbf{W}_{ad,t}^{(i)} \mathbf{H}_t - \mathbf{I}_U) (\mathbf{I}_U - |\Psi^{(i-1)}|^2)^{1/2}$ then from the definition of Frobenius norm

$$\begin{aligned} \text{MSE}_t^{(i)} &= \text{tr} \left(\mathbf{Z} (\mathbf{I}_U - |\Psi^{(i-1)}|^2) \mathbf{Z}^H \right) \sigma_u^2 \\ & \quad + \text{tr} \left(\mathbf{Z} |\Psi^{(i-1)}|^2 \mathbf{Z}^H \right) \sigma_u^2 + \left\| \mathbf{W}_{ad,t}^{(i)} \right\|_F^2 \sigma_n^2 \\ &= \text{tr} \left(\mathbf{Z} \mathbf{Z}^H \right) \sigma_u^2 + \left\| \mathbf{W}_{ad,t}^{(i)} \right\|_F^2 \sigma_n^2. \end{aligned} \quad (60)$$

Now, replacing \mathbf{Z} by its definition we obtain (32).

APPENDIX C

RELATIONSHIP BETWEEN (32) AND (33)

Starting with (32), which we rewrite here for convenience

$$\text{MSE}_t^{(i)} = \left\| (\mathbf{W} \mathbf{H} - \mathbf{I}) \mathbf{I}_\Psi^{1/2} \right\|_F^2 \sigma_u^2 + \left\| \mathbf{W} \right\|_F^2 \sigma_n^2 \quad (32)$$

Please note that we removed all indices to simplify the notation and $\mathbf{W} = \mathbf{W}_d \mathbf{W}_a$, $\mathbf{I}_\Psi = \mathbf{I} - |\Psi|^2$. By expanding (32), we obtain

$$\begin{aligned} \text{MSE}_t^{(i)} &= \left\| (\mathbf{W} \mathbf{H} - \mathbf{I}) \mathbf{I}_\Psi^{1/2} \right\|_F^2 \sigma_u^2 + \left\| \mathbf{W} \right\|_F^2 \sigma_n^2 \\ &= \text{tr} \left((\mathbf{W} \mathbf{H} - \mathbf{I}) \mathbf{I}_\Psi (\mathbf{W} \mathbf{H} - \mathbf{I})^H \sigma_u^2 + \mathbf{W} \mathbf{W}^H \sigma_n^2 \right) \\ &= \text{tr} \left(\mathbf{W} \left(\sigma_u^2 \mathbf{H} \mathbf{I}_\Psi \mathbf{H}^H + \sigma_n^2 \mathbf{I} \right) \mathbf{W}^H - \mathbf{W} \mathbf{H} \mathbf{I}_\Psi \sigma_u^2 \right) \\ & \quad + \text{tr} \left(-\mathbf{I}_\Psi \mathbf{H}^H \mathbf{W}^H \sigma_u^2 + \mathbf{I}_\Psi \sigma_u^2 \right). \end{aligned} \quad (61)$$

Consider, now the MSE expression (33) which we also rewrite here (with all indices removed) for convenience

$$\overline{\text{MSE}}_t^{(i)} = \left\| (\mathbf{W} - \overline{\mathbf{W}}) \tilde{\mathbf{R}}^{1/2} \right\|_F^2 \quad (33)$$

By expanding (33) and using matrices $(\mathbf{W}_{ad,t}^{(i)})_{opt}$, $\bar{\mathbf{W}}$ and $\tilde{\mathbf{R}}$ definition, equations (26), (34) and (35), respectively

$$\begin{aligned} \overline{\text{MSE}}_t^{(i)} &= \text{tr} \left((\mathbf{W} - \bar{\mathbf{W}}) \tilde{\mathbf{R}} (\mathbf{W} - \bar{\mathbf{W}})^H \right) \\ &= \text{tr} \left(\mathbf{W} \tilde{\mathbf{R}} \mathbf{W}^H - \mathbf{W} \tilde{\mathbf{R}} \bar{\mathbf{W}}^H - \bar{\mathbf{W}} \tilde{\mathbf{R}} \mathbf{W}^H + \bar{\mathbf{W}} \bar{\mathbf{W}}^H \right) \\ &= \text{tr} \left(\mathbf{W} \left(\mathbf{H} \mathbf{I}_\Psi \mathbf{H}^H + \sigma_n^2 \sigma_u^{-2} \mathbf{I} \right) \mathbf{W}^H - \mathbf{W} \mathbf{H} \mathbf{I}_\Psi \right) \\ &\quad + \text{tr} \left(-\mathbf{I}_\Psi \mathbf{H}^H \mathbf{W}^H + \bar{\mathbf{W}} \bar{\mathbf{W}}^H \right), \end{aligned} \quad (62)$$

where $\tilde{\mathbf{R}} \bar{\mathbf{W}}^H = \mathbf{H} \mathbf{I}_\Psi$, from the matrix inversion lemma and matrices $(\mathbf{W}_{ad,t}^{(i)})_{opt}$ and $\tilde{\mathbf{R}}$ definition, equations (26) and (35), respectively.

From the last line of (61) and (62) we verify that

$$\text{MSE}_t^{(i)} - \sigma_u^2 \overline{\text{MSE}}_t^{(i)} = \sigma_u^2 \text{tr} \left(\mathbf{I}_\Psi - \bar{\mathbf{W}} \bar{\mathbf{W}}^H \right), \quad (63)$$

which amount to a constant, independent of the optimization variables of the problem (30) and therefore not affecting the final result of this optimization problem.

REFERENCES

- [1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
- [2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
- [3] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [5] K. Zheng, L. Zhao, J. Mei, B. Shao, W. Xiang, and L. Hanzo, "Survey of large-scale MIMO systems," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1738–1760, 3rd Quart. 2015.
- [6] A. Swindlehursts, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 52–62, Sep. 2014.
- [7] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
- [8] T. Rappaport *et al.*, *Millimeter Wave Wireless Communications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2014.
- [9] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.
- [10] M. Vu and A. Paulraj, "MIMO wireless linear precoding," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 86–105, Sep. 2007.
- [11] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.
- [12] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
- [13] 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.
- [14] V. Venkateswaran and A. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.
- [15] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.
- [16] O. El Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2012, pp. 100–104.
- [17] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.
- [18] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.
- [19] J. Nsenga, A. Bourdoux, and F. Horlin, "Mixed analog/digital beamforming for 60 GHz MIMO frequency selective channels," in *Proc. IEEE ICC*, May 2010, pp. 1–6.
- [20] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2013, pp. 1–5.
- [21] 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.
- [22] A. B. J. Kokkeler and G. J. M. Smit, "Digitally assisted analog beamforming for millimeter-wave communication," in *Proc. ICC-Workshop Beyond—Enabling Technol. Appl.*, Jun. 2015, pp. 1065–1070.
- [23] X. Gao, L. Dai, C. Yuen, and Z. Wang, "Turbo-like beamforming based on Tabu search algorithm for millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5731–5737, Jul. 2015.
- [24] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.
- [25] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.
- [26] J. Geng, Z. Wei, X. Wang, W. Xiang, and D. Yang, "Multiuser hybrid analog/digital beamforming for relatively large-scale antenna arrays," in *Proc. IEEE GLOBECOM Workshop*, Dec. 2013, pp. 123–128.
- [27] T. E. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 4066–4071.
- [28] J. Li, L. Xiao, X. Xu, and S. Zhou, "Robust and low complexity hybrid beamforming for uplink multiuser mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1140–1143, Jun. 2016.
- [29] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6404–6481, Nov. 2015.
- [30] N. Benvenuto, R. Dinis, D. Falconer, and S. Tomasin, "Single carrier modulation with non linear frequency domain equalization: An idea whose time has come—Again," *Proc. IEEE*, vol. 98, no. 1, pp. 69–96, Jan. 2010.
- [31] N. Benvenuto and S. Tomasin, "Block iterative DFE for single carrier modulation," *Electron. Lett.*, vol. 39, no. 19, pp. 1144–1145, Sep. 2002.
- [32] Y.-C. Liang, S. Sun, and C. K. Ho, "Block-iterative generalized decision feedback equalizers for large MIMO systems: Algorithm design and asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2035–2048, Jun. 2006.
- [33] R. Kalbasi, D. Falconer, A. Banihashemi, and R. Dinis, "A comparison of frequency domain block MIMO transmission systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 165–175, Jan. 2009.
- [34] M. Luzzio, R. Dinis, and P. Montezuma, "SC-FDE for offset modulations: An efficient transmission technique for broadband wireless systems," *IEEE Trans. Commun.*, vol. 60, no. 7, pp. 1851–1861, Jul. 2012.
- [35] P. Li and R. C. de Lamare, "Adaptive decision-feedback detection with constellation constraints for MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 853–859, Feb. 2012.
- [36] A. Silva, S. Teodoro, R. Dinis, and A. Gameiro, "Iterative frequency-domain detection for IA-precoded MC-CDMA systems," *IEEE Trans. Commun.*, vol. 62, no. 4, pp. 1240–1248, Apr. 2014.
- [37] D. Castanheira, A. Silva, R. Dinis, and A. Gameiro, "Efficient transmitter and receiver designs for SC-FDMA based heterogeneous networks," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2500–2510, Jul. 2015.
- [38] H. E. Rowe, "Memoryless nonlinearities with Gaussian inputs: Elementary results," *Bell System Tech. J.*, vol. 61, no. 7, pp. 1519–1525, Sep. 1982.
- [39] S. Boyd and V. Lieven, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

- [40] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1997.
- [41] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.



Roberto Magueta received the M.Sc. degree in electronics and telecommunications engineering from the University of Aveiro, Portugal, in 2013. He is currently pursuing the Ph.D. degree. His Ph.D. thesis is on transmitter and receiver designs for future mm-wave and massive MIMO-based wireless systems. He joined the Instituto de Telecomunicações, Aveiro, as a Researcher, where he was involved in the Project RadioVoip-Smart Antenna for Maritime Communications.



Daniel Castanheira received the degree in electronics and telecommunications in 2007 and the Ph.D. degree in telecommunications from the University of Aveiro, Portugal, in 2012. In 2011, he was an Assistant Professor with the Departamento de Eletrónica, Telecomunicações e Informática, University of Aveiro. He is currently a Post-Doctoral Researcher with the Mobile Networks Group, Instituto de Telecomunicações, Portugal, where he is involved in several national and European projects, namely COPWIN, HETCOP, PURE-5GNET, and SWING 2, within the Portuguese National Scientific Foundation, and CODIV, FUTON, and QOSMOS within the European Seventh Framework Program. His research interests lie in signal processing techniques for digital communications, with emphasis on physical layer issues, including channel coding, precoding/equalization, and interference cancellation.



Adão Silva (M'12) received the M.Sc. and Ph.D. degrees in electronics and telecommunications from the University of Aveiro, in 2002 and 2007, respectively. He is currently an Assistant Professor with the Department of Electronics, Telecommunications and Informatics, University of Aveiro, and a Senior Researcher with the Instituto de Telecomunicações. He has been participating in several national and European projects, namely the ASILUM, MATRICE, 4 MORE within the ICT program and the FUTON and CODIV projects with the FP7 ICT. He has led several research projects in the broadband wireless communications area at the national level. His interests include multiuser MIMO, multicarrier based systems, cooperative networks, precoding, and multiuser detection. He has acted as a TPC member of several international conferences.



Rui Dinis (S'96–M'00–SM'14) received the Ph.D. degree from the Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal, in 2001, and the Habilitation degree in telecommunications from the Faculdade de Ciências e Tecnologia (FCT), Universidade Nova de Lisboa (UNL), in 2010. He was a Researcher with the Centro de Análise e Processamento de Sinal, IST, from 1992 to 2005. From 2001 to 2008, he was a Professor with IST. In 2003, he was an Invited Professor with Carleton University, Ottawa, ON, Canada. He was also a Researcher with the Instituto de Sistemas e Robótica from 2005 to 2008. Since 2009, he has been a Researcher with the Instituto de Telecomunicações. He is currently an Associate Professor with FCT, UNL. He has been actively involved in several national and international research projects in the broadband wireless communications area. His research interests include modulation, equalization, channel estimation, and synchronization. He is an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS (Transmission Systems–Frequency-Domain Processing and Equalization) and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was also a Guest Editor of *Physical Communication* Special Issue on Broadband Single-Carrier Transmission Techniques (Elsevier).



Atilio Gameiro received the Licenciatura and Ph.D. degrees from the University of Aveiro in 1985 and 1993, respectively. His industrial experience includes a period of one year with BT Labs and one year with NKT Elektronik. He is currently an Associate Professor with the Department of Electronics and Telecommunications, University of Aveiro, and a Researcher with the Instituto de Telecomunicações, Pólo de Aveiro, where he is the Head of the Group. His main interests lie in signal processing techniques for digital communications and communication protocols, and within this research line he has done work for optical and mobile communications, either at the theoretical and experimental level. His current research activities involve space-time-frequency algorithms for the broadband wireless systems and cross-layer design. He has been involved and has led IT or University of Aveiro participation on more than 20 national and European projects. He has authored over 200 technical papers in international journals and conferences.