This is a postprint version of the following published document:

Vega Delgado, A., Sánchez- Fernández, M., Llorca J. y Tulino A. (2018). Feasible Transmission Strategies for Downlink MIMO in Sparse Millimeter Wave Channels. *IEEE Communications Magazine*, 56(7), pp. 49-55.

DOI: https://doi.org/10.1109/MCOM.2018.1701022

Feasible Transmission Strategies for Downlink MIMO in Sparse Millimeter Wave Channels

Adrián Vega Delgado, Matilde Sánchez-Fernández, *Senior Member, IEEE*, Jaime Llorca, *Senior Member, IEEE*, Antonia Tulino, *Fellow, IEEE*

Abstract

Addressing the disruptive capacity requirements of 5G networks calls for a thorough exploration of multiple technological solutions. Two promising approaches are i) the use of multiple-input multiple-output (MIMO) technologies that enable multiplicative capacity gains, and ii) the exploration of new frequency bands to enlarge available bandwidth. While millimeter spectrum, one of the main bands under study, poses significant challenges due to its cumbersome propagation characteristics (particularly severe path-loss and channel sparsity), its ten-fold frequency increase favors the deployment of reduced-size large antenna arrays for massive MIMO. However, the high cost and power consumption of its required signal mixers and analog-to-digital converters precludes mmWave beamforming to be performed entirely at baseband using digital precoders. A possible cost-effective alternative is the hybrid precoding transceiver architecture, which combines digital and analog precoders. In this paper, we exploit the sparse nature of the channel to unveil an advantage in the design of the hybrid precoder. Specifically, by reformulating the hybrid precoder design as a matrix factorization problem, and adopting an atomic norm minimization approach, we propose a new hybrid precoding algorithm that takes advantage of the sparse nature of the mmWave channel to approach the performance of the optimal fully-digital precoder. Simulation results confirm that the proposed algorithm can approach the performance achieved by unconstrained digital beamforming solutions.

Index Terms

mmWave, massive multiple-input multiple-output (MIMO), sparsity, precoding, RF chain, compressed sensing, atomic norm.

I. Introduction

The fifth generation of mobile communication networks (5G) will require a multi-fold increase in overall system capacity. Recent studies project that, by 2020, 5G networks will support $\times 1000$ larger capacity than current Long Term Evolution (LTE) networks [1]. In order to meet this demanding requirement, a number of physical layer technologies have been proposed, including massive multiple–input multiple–output (MIMO), carrier aggregation, advanced channel coding, and interference coordination. However, the saturated use of the spectrum in current cellular networks makes the improvements achieved by the above-referenced technologies insufficient to meet the $\times 1000$ increase in capacity. Thus, exploring alternative spectrum bands that are either underutilized or not yet utilized for mobile communications has become critical for 5G.

According to the GSM Association, 5G will make use of three key frequency ranges to provide widespread coverage and support all the requirements for the next generation of mobile communications [2]. The first range encompasses the frequencies below 1 GHz, whose aim is to provide widespread coverage in urban, suburban, and rural areas, as well as to support low data-rate Internet of Things (IoT) services. The second band goes from 1 to 6 GHz and it is expected to support emergent 5G services

This work has been partly funded by the Spanish Government through projects MIMOTEX (TEC2014-61776-EXP), CIES (RTC-2015-4213-7) and TERESA-ADA (TEC2017-90093-C3-2-R) (MINECO/AEI/FEDER, UE).

A. Vega Delgado and M. Sánchez-Fernández are with the Signal Theory & Communications Department, Universidad Carlos III de Madrid, Spain. E-mail: {avega, mati}@tsc.uc3m.es

J. Llorca is with Nokia Bell Labs, Holmdel, NJ 07733, USA. E-mail: jaime.llorca@nokia-bell-labs.com.

A. M. Tulino is with the DIETI Department, University of Napoli, Federico II, 80138 Napoli, Italy and with Nokia Bell Labs, Holmdel, NJ 07733, USA. E-mail: a.tulino@nokia-bell-labs.com

such as intelligent transport systems (e.g., autonomous vehicles). Finally, the spectrum above 6 GHz, and specifically the millimeter wave (mmWave) band (30 GHz to 300 GHz) [3], [4], will be used to provide ultra–high broadband services (e.g., augmented reality).

Millimeter wave band usage for 5G presents many challenges, some of which are discussed next. First, this band experiences significant free–space path loss due to the ten-fold increase in carrier frequency. Fortunately, this increase in carrier frequency implies a decrease in wavelength, which facilitates the deployment of large antenna arrays in small areas that can take advantage of the well-known multiplexing and diversity gain of MIMO to mitigate path loss. A second important challenge relates to mmWave hardware design, a subject of significant research since the 1970s. While the first mmWave system implementations were based on gallium arsenide (GaAs), the improvement that CMOS technology brought to traditional microwave systems has recently led researchers and industry practitioners to develop CMOS subsystems for mmWave systems. Nonetheless, mmWave system design based on CMOS technology is still under ongoing investigation, as mmWave CMOS is not easily extrapolated from microwave CMOS [5]. A last challenge to take into consideration is the cost of mmWave hardware (low noise amplifiers, power amplifiers, antennas, etc.), mainly due to its reduced size and low power consumption requirements [3], which creates the need for the design of low complexity MIMO communication solutions in these bands. With MIMO becoming a key technology for 5G due to its improved spectral efficiency and diversity gains, it is important to address the specific challenges of implementing MIMO, or even massive MIMO, in the mmWave band.

We focus on a multiuser MIMO (MU-MIMO) downlink (DL) scenario. Under conventional microwave propagation, MIMO precoding can be easily implemented digitally at baseband, and requires dedicated Radio Frequency (RF) hardware, i.e., an RF chain, for each antenna element. However, RF hardware for mmWave wavelengths is more expensive and power consuming than for conventional microwave wavelengths, which makes a fully-digital (FD) precoding implementation in mmWave non-viable [3]. Hybrid (HB) precoding divides the precoding implementation into digital baseband processing, bandpass modulation and analog processing, allowing a potential reduction on the number of RF chains and thus enabling a viable precoding architecture for mmWave systems. In this paper, by reformulating the hybrid precoder design as a matrix factorization problem, and adopting an atomic norm minimization approach, we propose a new hybrid precoding algorithm that takes advantage of the sparse nature of the mmWave channel and that it is able to closely approach the performance of the optimal fully-digital precoder.

The paper is organized as follows: in Section II, propagation in mmWave channels is discussed; in Section III, different feasible transmission strategies are explored with a focus on propagation in sparse channels; Section IV explores different hybrid precoder alternatives in the literature, together with our proposed design approach; finally, concluding remarks are given in Section V.

II. Propagation in mmWave Sparse channels

Signal propagation in the mmWave band faces important challenges that include severe free–space path loss, atmospheric absorption, rain and foliage losses, and specular scattering effects. In fact, already at 30 GHz, free–space path loss can reach 102 dB over a distance of 100 m. In addition, the frequencies around 60 GHz and 180 GHz exhibit attenuation due to atmospheric absorption above 10 dB/Km. Hence, mmWave propagation faces severe path loss that can be worsen by rain and obstacles such as buildings or trees in outdoor environments, or walls and furniture in indoor settings. This often leads to channel scenarios with a reduced number of propagation paths, characterized by low angular spreads both in elevation and azimuth [4], and delay-spreads below 10 ns in outdoor urban environments [3], [6], and below 20 ns in indoor environments [7].

These propagation characteristics lend themselves to a sparse channel model, characterized by a few discrete propagation paths between transmitter and receiver, as described in [8]. A mathematical model for the narrowband multiuser MIMO channel, with M antennas at the base station (BS), K single-antenna users, and L and L' scatterers placed in directions $\mathbf{k}_l$ and $\mathbf{k}_{l'}$ at the transmitter and receiver, respectively, can be described as:

$$\mathbf{H} = \sum_{l=1}^{L} \sum_{l'=1}^{L'} s\left(\mathbf{k}'_{l'}, \mathbf{k}_l\right) \mathbf{u}_{l'} \mathbf{t}_l^{\dagger} = \mathbf{U}\mathbf{S}\mathbf{T}^{\dagger}, \tag{1}$$

where $[\cdot]^{\dagger}$ denotes the Hermitian operator, $s\left(\mathbf{k}', \mathbf{k}\right)$ is a scattering function that defines the energy of the plane wave transmitted in direction $\mathbf{k}$ and received in direction $\mathbf{k}'$, and $\mathbf{t}_l^{\dagger}$ and $\mathbf{u}_{l'}$ are the channel beamformers at the transmitter and receiver, respectively, which depend on the location of the active scatterers and on the geometry and/or position of the transmit and receive antennas. Under a linear array structure assumption, the channel beamformers at the transmitter have the following structure: $\mathbf{t}_l^{\dagger} = e^{j2\pi\gamma_l}\left[1 \; e^{-j2\pi f_l} e^{-j2\pi 2 f_l} \ldots e^{-j2\pi(M-1)f_l}\right]$, where both the frequency f_l and the phase γ_l are normalized taking values between 0 and 1 and are solely defined by the direction (azimuth and/or elevation) in which the L scatterers seen by the transmitter are placed.

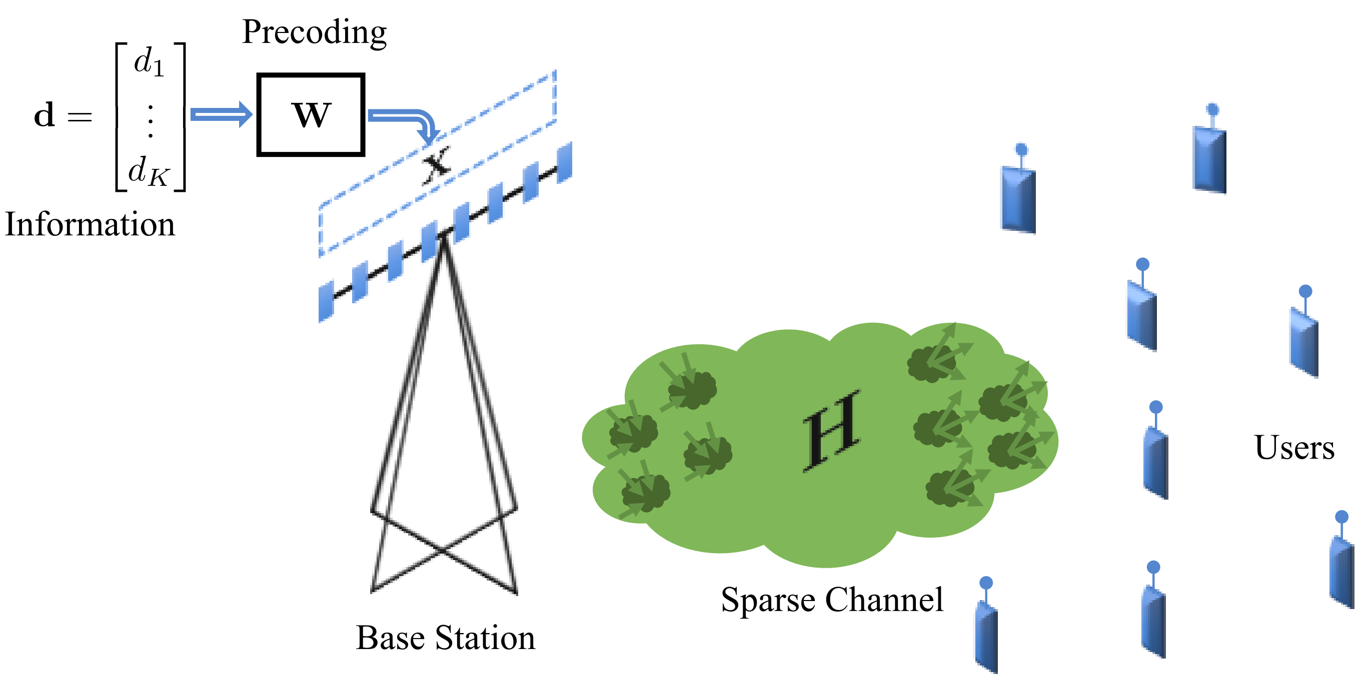

Fig. 1. Downlink multiuser system model. Information vector $\mathbf{d}$, composed of K threads, is precoded with matrix $\mathbf{W}$, and the resulting signal $\mathbf{x}$ is sent though channel $\mathbf{H}$. All users receive their corresponding signal and no further processing other than scaling may be done.

III. Feasible transmission strategies for the downlink

We consider a mmWave downlink transmission scenario, as depicted in Fig. 1, where a base station with M antennas serves K single-antenna users. In this setting, under the assumption that users do not cooperate, the key challenge is to design a suitable transmission strategy. Indeed, the best solution to overcome the co-channel interference, and to eventually provide each user with reasonable performance, is to pre-process the signal intended for each user at the BS with a proper transmission strategy.

Many transmission strategies have been studied in the literature. The DL MU-MIMO sum-rate, defined as the maximum aggregation of all the users' data rates, is achieved via dirty paper coding (DPC), a highly complex transmission strategy for which practical codes that approach the capacity limit are still unknown. On the other hand, linear precoding can achieve the same multiplexing gain as DPC, with a certain offset with respect to the sum-rate performance [9]. Thus, a linear precoder would be able to transmit as many data threads as DPC, while requiring a much less complex transmission implementation.

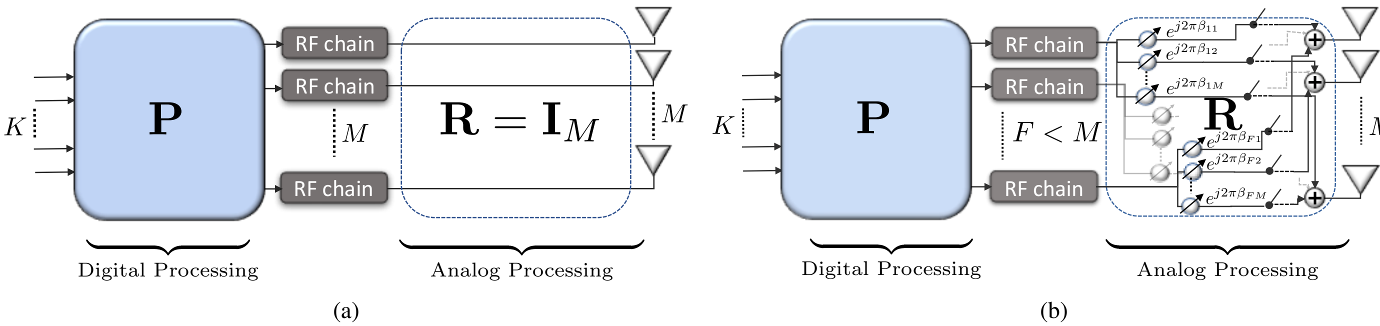

Fig. 2. Block diagram of the processing to be implemented at the BS: (a) Digital and analog processing for a FD implementation of the transmission strategy; (b) Digital and analog processing for a HB implementation of the transmission strategy.

All transmission strategies rely on a certain level of channel knowledge at the BS. The best scenario is characterized by full channel knowledge at the transmitter, also referred to as Channel State Information (CSI) at the transmitter (CSIT). Other more realistic assumptions provide the BS with partial information of the channel (PCSI). Examples include quantized information, delayed information, or even long-term variation (statistics). These scenarios differ in performance, with CSI outperforming PCSI; and also in complexity, with CSI being more complex to acquire than PCSI [10]. For the rest of this work, we focus on linear precoders with CSIT.

A. *Linear transmission strategies*

Linear precoding strategies provide implementation advantages compared to non-linear techniques, allowing a trade-off between complexity and performance. A linear precoder implements a transformation of K information threads into M transmitted symbols, which is linearly modeled in its low-pass equivalent by an $M \times K$ precoding matrix $\mathbf{W}$, whose columns define the transformation applied to each user signal and transmitted through the M antennas, referred to as the BS beamformers (see Fig. 1).

When designing a linear precoder, an optimization criteria should be determined and applied according to the level of channel knowledge at the transmitter. Focusing on full CSI, where the BS has full access to the channel matrix, the linear precoder may be chosen to optimize i) the sum-rate, ii) the signal to interference-plus-noise ratio (SINR), or iii) the mean squared error (MSE), as well as to eliminate the inter-user interference (block–diagonalization) or the spatial interference (zero-forcing, or conjugate beamforming).[1]

B. *Building linear precoders*

Any linear transmission strategy has two distinctive components. One is the digital processing, which is the baseband transformation of the discrete information symbols. And the second is the band-pass modulation and analog processing, which is the transformation of the discrete symbols into analog signals in band-pass (see Fig. 2). From the steps mentioned above (i.e., digital processing, band-pass modulation, and analog processing), the digital and analog processing steps can be modeled via the factorization of the precoding matrix $\mathbf{W}$ as the product of an $F \times K$ baseband processing matrix $\mathbf{P}$ (with no predefined structure beyond the one eventually needed for transmit power allocation purposes), and an $M \times F$ analog processing matrix $\mathbf{R}$ referred to as the RF matrix. F is a new design parameter that represents the number of RF chains that implement the band-pass modulation. The RF matrix may also model a set of switches, mixers, and analog phase shifters, and therefore needs to have a concrete structure, as shown in Fig. 2. Switches are implemented in the RF matrix by setting to 0 (disconnected) or 1 (connected) those elements

[1] It should be noted that when the users have a single antenna, both block–diagonalization and zero-forcing beamforming provide the same performance.

of the RF matrix that link each of the RF chains to each antenna. Furthermore, a phase shift between the output of the i-th RF chain and the m-th antenna sets the correspondent element of the RF matrix to $e^{j2\pi\beta_{im}}$.

The choice of F and of the analog/digital processing split determine different hardware/software precoder implementations.If $F = M$, we have as many RF chains as antennas. In this scenario, the RF matrix is a square matrix, and it is typically forced to be an identity matrix, given that any phase shift and mixer in a square matrix structure can be implemented in a digital manner and moved to the baseband matrix (see Fig. 2(a)). In this case, there is no analog processing, only band-pass modulation, which is always necessary. The hardware/software implementation is then said to be *fully-digital* (FD). Alternatively, we may decide to reduce the number of RF chains making $F < M$. In this scenario, the tall RF matrix may have a non-diagonal structure, allowing a certain number of mixers and phase shifters to be active in order to implement the analog processing (see Fig. 2(b)). In this case, the transmitter implements digital and analog processing together with the band-pass modulation. This hardware/software implementation is said to be *hybrid*.

From the above consideration, it should be clear that any hybrid implementation of a precoder boils down to the factorization of the precoder matrix in terms of the baseband matrix $\mathbf{P}$ and the RF matrix $\mathbf{R}$.

In the following, we discuss the most adequate hardware/software implementation for the transmission strategy: fully-digital or hybrid. The answer is obviously not universal and depends on several factors, such as target performance, hardware cost for analog processing and band-pass modulation, and space constraints for the deployment of the RF chains, among others. Most of the precoders in the literature for DL MU-MIMO provide a structure for the precoding matrix that, apparently, and without digging into any suitable potential decomposition, would only match the FD implementation. In this case, an hybrid implementation with $F < M$ would not be feasible without compromising performance unless we are able to leverage a suitable internal decomposition of the precoder matrix that matches the hybrid implementation. Assuming performance is not an issue, either because there is an internal decomposition of the precoder matrix that allows a hybrid implementation, or because the performance loss can be compensated with other hardware implementation potential benefits (e.g., reduced cost, reduced complexity), there is still scope for discussion on when it is suitable to reduce the number of RF chains compared to the number of antennas at the BS. In any system, reducing the number of RF chains would force analog processing, with the increase in cost that this hardware implementation would incur. However, since the number of RF chains are being reduced, there is a trade-off between the additional cost due to analog processing and the cost reduction due to the smaller number of RF chains. In this scenario, there is typically a net reduction of the hardware cost (RF chains are more expensive than the analog processing hardware required).

A remaining issue to address when deciding between a FD or HB transmitter is the role of the baseband processing. Typical system parametrization in MU-MIMO, and especially in mmWave channels, assumes $M >> K$, i.e., the number of antennas at the BS should be larger than the number of information threads K to be transmitted. In this case, we could still face two scenarios: $F \geq K$ or $F < K$. In the first scenario, the baseband processing is performed by a tall or square matrix, ensuring a level of redundancy that helps overcoming the channel impairments and allows for an accurate estimate of the transmitted thread. However, the second scenario is not as straightforward since the baseband processing matrix would be fat and therefore the thread information is being compressed before the analog processing step. In this scenario, we should ensure that the analog processing step overcomes this compressing loss, which in many cases may not be possible due to the spatial multiplexing gain reduction [10].

C. Unveiling an Hybrid precoding structure in the Fully-Digital implementation

While existing literature provides a number of criteria for the design of hybrid precoders, in the following we focus on the minimum MSE (MMSE) criterion. Due to the hardness of directly solving the associated constrained optimization problem, our approach is to find the hybrid implementation that best approximates the unconstrained MMSE precoder for a given number of RF chains F. To this end, a convenient mathematical structure for the unconstrained MMSE precoder can be obtained by resorting to the dowlink/uplink (DL/UL) duality property [11].[2] Indeed, using DL/UL duality, the MMSE precoder admits the following expression:

$$\begin{aligned}\mathbf{W}^{\mathrm{MMSE}} &= \mathbf{H}^{\dagger}\left(\mathbf{C}\mathbf{H}\mathbf{H}^{\dagger} + \sigma_z^2\mathbf{I}\right)^{-1}\mathbf{C}\mathbf{B}^{-1} \\ &= \mathbf{T}\mathbf{S}^{\dagger}\mathbf{U}^{\dagger}\left(\mathbf{C}\mathbf{H}\mathbf{H}^{\dagger} + \sigma_z^2\mathbf{I}\right)^{-1}\mathbf{C}\mathbf{B}^{-1} \\ &= \mathbf{T}\mathbf{P}^{\mathrm{MMSE}}, \end{aligned} \tag{2}$$

where the diagonal matrix $\mathbf{C}$ defines the DL power allocation, σ_z^2 is the DL noise power, and $\mathbf{B}$ is a scaling matrix for normalization purposes.

Given the knowledge of the channel beamformer matrix $\mathbf{T}$, or at least of the L frequencies associated with each channel beamformer,[3] then the factorization form of the MMSE precoder in Eq. (2) immediately reveals how to obtain its hybrid implementation.

Specifically, if F is unconstrained, then i) mapping the factorization form in Eq. (2) to the hybrid decomposition in terms of the RF matrix $\mathbf{R}$ and the baseband processing matrix $\mathbf{P}$ described in Section III-B, and ii) recalling that each channel beamformer $\mathbf{t}_l$ is exclusively implementing a phase shift, it follows that the optimal RF matrix is equal to the channel beamformer matrix, and consequently the optimal F is equal to L. Hence, in a sparse channel with a small number of scatterers at the transmitter ($L << M$), the MMSE precoder can exhibit a hybrid structure with a number of RF chains much smaller than the number of antennas. Such a desirable hybrid structure leads to a cost/complexity reduction with respect to a classical FD implementation, while maintaining the same performance.

On the other hand, if F is constrained to be smaller than L (e.g., for cost/complexity reduction), the hybrid precoder structure can be obtained by finding the F channel beamformers in $\mathbf{T}$ that best approximate the MMSE precoder. This is in general an NP-hard problem for which existing approaches are described in Section IV-A.

We remark that the procedure described above to identify a hybrid implementation of the MMSE precoder can be applied to other CSI-based precoders, such as the zero–forcing and the conjugate beamforming precoders.

IV. Hybrid Precoder design alternatives

As stated in the previous section, any hybrid implementation of a precoder $\mathbf{W}$ is a matrix factorization problem in terms of the baseband matrix $\mathbf{P}$ and the RF matrix $\mathbf{R}$. This factorization problem can also be seen as expressing each BS beamformer $\mathbf{w}_k$ to be a linear (convex) combination of the columns of the RF matrix. Recalling that the columns of the RF matrix are defined by a proper subset of the columns of the channel beamformer matrix, such factorization problem can in general be very complex due to two main reasons. First, the channel beamformer matrix is in general not known at the transmitter and hence needs to be estimated. Second, in order to reduce cost/complexity, F can be chosen to be smaller than L, which, in general, results in an NP-hard problem (even for known $\mathbf{T}$).

A. Existing Approaches

One of the first approaches to solve this factorization problem is given in [8] for a single–user MIMO scenario. In this case, the reference precoder maximizes the mutual information and it is given by the right-hand side eigenvectors of the channel matrix. To make the search of the RF phases tractable, they assume that the channel beamformer matrix, which contains the frequencies that characterize the position

[2]The DL/UL duality property states that the DL achievable region in terms of normalized linear MSE is the same as the region of an equivalent uplink problem, which is obtained by switching the role of transmitter and receiver.

[3]According to the definition of the $\mathbf{t}_l$ beamformers in Sec. II, two parameters fully characterize the l-th beamformer: the normalized frequency f_l and the normalized phase γ_l. However, the effect on $\mathbf{T}$ of the multiplicative term $e^{2\pi\gamma_l}$ of each of the $\mathbf{t}_l$ can be further factorized as a diagonal matrix that contains all these terms, and can be included into the digital processing matrix of the hybrid implementation

of the scatterers, is known, and they choose the "best" F vectors within this matrix by means of an orthogonal matching pursuit (OMP) algorithm. This work is further extended to the multi-user scenario in [10] by optimizing, in two stages, the RF beamformers to maximize each user desired power, and the baseband beams to overcome the remaining multiuser interference. In this case, again, the search for the phases is not over all the phase space, but it is still brute force search over a discrete predefined phase dictionary whose cardinality depends on the discrete set of phases that the phase shifters can implement. Other approaches in the literature use zero-forcing as the reference precoder, and again, to avoid the search over all possible phases of the RF matrix, the channel beamformer matrix is assumed to be known, and set equal to the RF matrix without any further search.

An extension of the OMP approach is proposed in [12] for a single user scenario and the maximization of the mutual information. In this approach, the authors find the RF matrix phases via gradient descend over a space of $M \times F$ unit radius circles on the complex plane. Each point found in one circle provides the correspondent phase.

Finally, another proposal for solving the hybrid precoder factorization problem [13] is also based on an OMP-based search on extended dictionaries rather than the one based on the channel beamformer matrix knowledge. Some examples of the proposed dictionaries are a set of eigenvectors that can only be implemented in analog processing by adding gain controllers to the corresponding phase shifters, a set of discrete Fourier transform beamformers, or a set of discrete cosine transform beamformers.

It should be noted that all the approaches described so far are based on different searches over a discrete dictionary. Furthermore, there are other approaches in the literature, not specifically proposed for Hybrid Precoding design, that could also be applied for finding the channel beamformer matrix. This is the case of the Nonparametric Iterative Adaptive Approach (IAA) used for source location in [14]. Here, again, the search of the frequency set that characterizes the channel beamformer matrix is done over a large dictionary defined by sampling the frequency space.

B. Hybrid factorization based on an Atomic Norm continuous alphabet search

Differently from previous works, our approach does not assume any knowledge of the frequencies $f_1, f_2, \dots, f_L$ that define the channel beamformer matrix $\mathbf{T}$. In the following, we assume that the columns of the RF matrix $\mathbf{R}$ belong to the analog processing chain set $\mathcal{R} = \left\{ \begin{bmatrix} 1 & e^{j2\pi f} e^{j2\pi 2f} \dots e^{j2\pi(M-1)f} \end{bmatrix}^{\top} \right\}$, with f being the continuous normalized frequency. We refer to the elements of the set $\mathcal{R}$ as *atoms*. Then, the factorization problem whose solution provides the desired hybrid implementation, can be posed as follows: find F atoms and a matrix $\mathbf{P}$, such that when using the F atoms as the columns of the RF matrix $\mathbf{R}$, the product of both matrices best approximates the MMSE precoding matrix.

Recall that for unconstrained F, the atoms found by solving the factorization problem are equal to the columns of the channel beamformer matrix $\mathbf{T}$. Hence, under uncostrainted F, one can just focus on estimating the channel beamformer matrix. The case of fixed F, which, as stated earlier, requires solving an NP-Hard problem, is out of the scope of this paper. Therefore, the problem becomes: find, for each user k, the set of atoms whose linear combination best approximates the k-th column of the MMSE precoder:

$$\mathbf{w}_k^{\mathrm{AN}} = \arg\min_{\mathbf{w}_k} \left\{ \frac{1}{2} \|\mathbf{w}_k^{\mathrm{MMSE}} - \mathbf{w}_k\|^2 + \tau \|\mathbf{w}_k\|_{\mathcal{R}} \right\}. \tag{3}$$

where $\mathbf{w}_k^{\mathrm{MMSE}}$ denots the k-th column of the MMSE precoder (i.e., the MMSE BS beamformer) and $\mathbf{w}_k^{\mathrm{AN}}$ its hybrid implementation, referred to as AN beamformer.

The objective of the above optimization is to obtain the closest vector to the MMSE BS beamformer, forcing the optimized vector $\mathbf{w}_k$ to be a sparse linear combination of some of the elements in $\mathcal{R}$ by means of an atomic norm (AN) $\|\cdot\|_{\mathcal{R}}$ penalty. The regularization parameter τ should be carefully chosen to provide the right balance between how far we want to be from the reference precoder matrix and how strong we want to make the structured approximation of the precoder beam. Higher values of τ allow more error between the precoder beam and our structured approximation, thus giving more importance to the AN penalization. In this scenario, we leverage the structured approximation, but the precoder beam obtained does not have to be close to the reference precoder. Smaller values of τ incur less error, forcing the approximated precoder beam to be closer to the reference beam. However, in this scenario the required structure for the precoder beam may not be achieved.

The optimization problem in (3) turns into a semidefinite optimization problem [15] that can be easily solved via semidefinite programming (SDP). From the solution to (3), which provides the best AN-based approximation to each MMSE BS beamformer, it remains to find the atoms in the dictionary $\mathcal{R}$ (i.e., the columns of the RF matrix) whose linear combination generates the correspondent AN beamformer. According to [15, Corollary 1], this can be done by: i) computing the vector error between the MMSE beamformer and the AN beamformer, which coincides with the unique solution to the dual problem of (3); ii) evaluating the inner product, $\rho_{k,f}$, between the aforementioned error and a generic atom defined by frequency f; and iii) identifying the values of f where the absolute value of the inner product equals the regularization parameter τ in (3). The number of atoms matching this condition will determine the number of RF chains required in the hybrid implementation of the precoder. Finally, the digital processing matrix $\mathbf{P}$ is obtained from the K AN beamformers and the RF matrix obtained by finding the set of frequencies associated with each AN beamformer.

C. Results

To show the feasibility of hybrid precoder implementations and their performance, we present a scenario with $M = 40$ antennas at the BS, $K = 5$ users to be served, and $L = L' = 5$ scatterers. Both users and scatterers are randomly placed and the BS has a linear array structure. Our reference precoder matrix is set to be the MMSE precoder given by (2), which maximizes each user SINR and minimizes their MSE.

Given the above setting, Fig. 3 shows the reference frequencies that define each of the channel beamformers, i.e. the frequencies to be found to characterize the RF matrix, represented by blue vertical lines. Furthermore, following the procedure described in Sec. III-C, each of the colored lines represent the frequencies found for each user. It is observed that AN finds exactly $F = 5$ frequencies, which are the same set of frequencies for all users, allowing the same set of analog processing chains to generate the hybrid beamformers.

Next, Fig. 4 compares the performance of existing algorithms such as OMP [8] and IAA [14], with our proposed method, AN. Specifically, we show how close the different hybrid factorizations (AN, IAA, and OMP) are to the MMSE (FD precoder) in terms of the normalized squared error norm. It is worth noticing that AN and IAA do not assume any number of RF chains a priori. Nevertheless, they are able to find the $L = 5$ frequencies that characterize the channel beamformer matrix, which leads to $F = 5$ RF chains. For the OMP, instead, we consider two strategies: restricting the number of RF chains to match the number of RF chains selected by AN, or leaving the number of RF chains as a free parameter, and restricting the error with respect to the MMSE precoder to be the same as the AN error. We can see in Fig. 4 that if we match the number of RF chains of OMP and AN, OMP is the worst algorithm in terms of error, followed by IAA, and then AN. Furthermore, if we fix the OMP error to be the same as the AN error, we see that the OMP requires $F = 8$ RF chains to match the AN error.

V. Conclusions

In this paper, we address one of the main challenges presented by mmWave systems via the design of a hybrid precoder in a DL MIMO scenario with single-antenna users. This precoder takes advantage of the sparse nature of mmWave channels to reduce the number or RF chains, therefore scaling down the total hardware cost, and yielding a feasible transmission strategy in this millimeter bands. We described some of the state-of-the-art algorithms for hybrid precoding, such as OMP and IAA, and present a novel algorithm based on AN by reformulating the hybrid precoder design as a matrix factorization problem and adopting an AN minimization approach. The AN–based precoder allows reducing hardware complexity

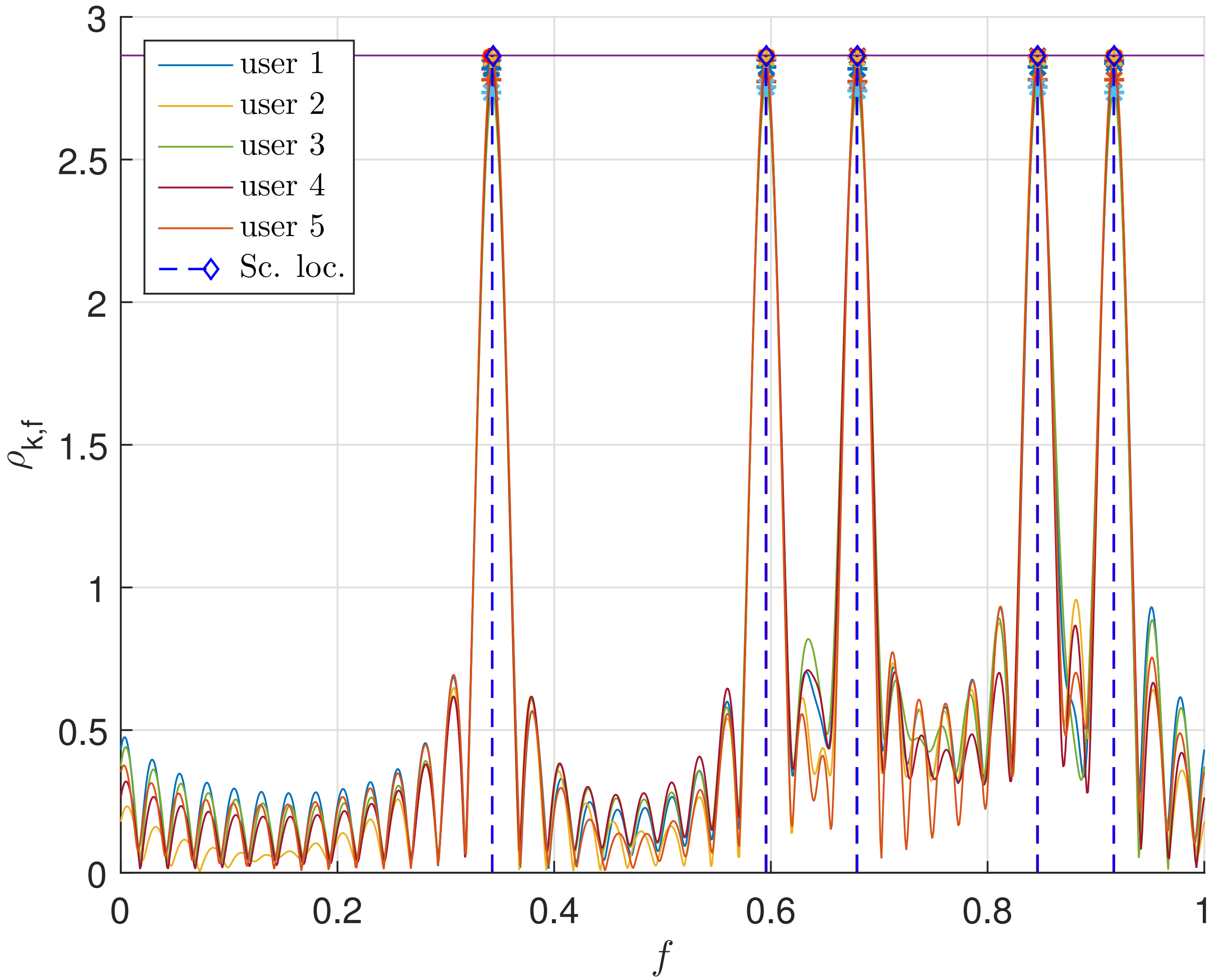

Fig. 3. Set of normalized frequencies found by AN and defining the analog processing implemented for each of the RF chain vectors. The dashed vertical blue lines (labelled Sc. loc.) represent the set of $L = 5$ frequencies that characterize the columns of the channel beamforming matrix. The colored lines are the frequencies found by AN for each of the $K = 5$ users.

by requiring less RF chains than the state-of-the-art precoders, for a fixed error in terms of distance to the optimal fully-digital precoder. Thus, mmWave spectrum can be exploited in a feasible way via the use of MIMO and hybrid precoding. Furthermore, AN is revealed as a promising algorithm to leverage the sparse nature of the mmWave channel in order to reduce the number or RF chains.

References

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What Will 5G Be?," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1065–1082, June 2014.

[2] "5G Spectrum," Public Policy Position, GSM Asociation, 2016. https://www.gsma.com/spectrum/5g-spectrum-policy-position/. Accessed January 2018.

[3] Z. Pi and F. Khan, "An introduction to Millimeter-Wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, pp. 101–107, June 2011.

[4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[5] A. A. Abidi, "CMOS microwave and millimeter-wave ICs: The historical background," *IEEE International Symposium on Radio-Frequency Integration Technology (RFIT)*, Aug. 2014.

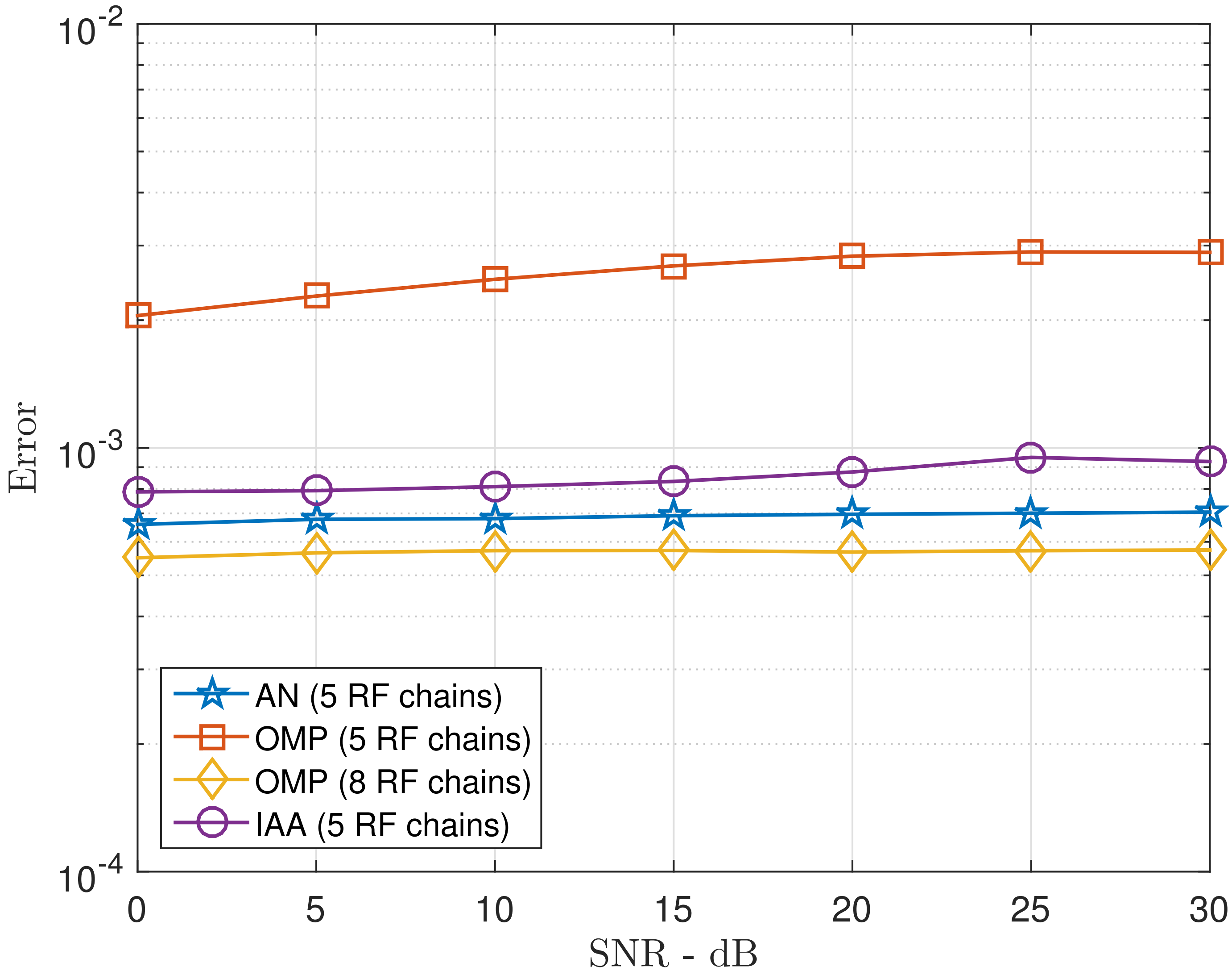

Fig. 4. Error in terms of squared distance between the MMSE (FD) precoder beams and the hybrid precoder vectors.

[6] G. R. MacCartney, M. K. Samimi, and T. S. Rappaport, “Exploiting directionality for millimeter-wave wireless system improvement,” in *2015 IEEE International Conference on Communications (ICC)*, pp. 2416–2422, June 2015.

[7] B. Ai, K. Guan, R. He, J. Li, G. Li, D. He, Z. Zhong, and K. M. S. Huq, “On indoor millimeter wave massive MIMO channels: Measurement and simulation,” *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 1678–1690, July 2017.

[8] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 13, pp. 1499–1513, March 2014.

[9] J. Lee and N. Jindal, “High SNR analysis for MIMO broadcast channels: Dirty paper coding versus linear precoding,” *IEEE Transactions on Information Theory*, vol. 53, pp. 4787–4792, Dec 2007.

[10] A. Alkhateeb, G. Leus, and R. W. H. Jr., “Limited Feedback Hybrid Precoding for Multi-User Millimeter Wave Systems,” *IEEE Transactions on Wireless Communications*, vol. 14, Nov. 2015.

[11] S. Shi, M. Schubert, and H. Boche, “Downlink MMSE Transceiver Optimization for Multiuser MIMO Systems: MMSE Balancing,” *IEEE Transactions on Signal Processing*, vol. 56, pp. 3702–3712, Aug 2008.

[12] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, “Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 485–500, April 2016.

[13] M. Kim and Y. H. Lee, “MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels,” *IEEE Transactions on Vehicular Technology*, vol. 64, pp. 2714–2720, June 2015.

[14] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, “Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, pp. 425–443, Jan 2010.

[15] B. N. Bhaskar, G. Tang, and B. Recht, “Atomic Norm Denoising With Applications to Line Spectral Estimation,” *IEEE Transactions on Signal Processing*, vol. 61, pp. 5987–5999, Dec 2013.

Adrián Vega Delgado received the B.Sc. degree and the M.Sc. degree in telecommunication engineering from Universidad Carlos III de Madrid in 2015 and 2017 respectively. He received the research M.Sc. degree in Multimedia and Communications from Universidad Carlos III de Madrid in 2017 and now he holds a research support position at Universidad Carlos III de Madrid. His current research interests are multiple-input-multiple-output techniques and the application of compressed sensing techniques to communications.

Matilde Sánchez-Fernández (SM‘14) received the M.Sc. and Ph.D. degrees in telecommunications engineering from Universidad Politécnica de Madrid, Spain, in 1996 and 2001, respectively. In 2000, she joined Universidad Carlos III de Madrid, where she has been an Associate Professor since 2009. Previously, she was a Telecommunication Engineer with Telefónica. Her current research interests are signal processing for wireless communications, multiple-input multiple-output techniques, channel modeling in wireless communications and game theory and machine learning techniques applied to communications.

Jaime Llorca received the M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Maryland, College Park, in 2003 and 2008, respectively. He joined Bell Labs, New Jersey, in 2010, where he is currently a Senior Research Scientist. His research interests lie in the area of network algorithms, network optimization, distributed control, and network information theory, with special focus on communication networks, distributed cloud networking, and content distribution.

Antonia M. Tulino received the Ph.D. degree in Electrical Engineering from Seconda Universitá degli Studi di Napoli, Italy, in 1999. In 2009, she joined Bell Labs and since 2002, she is a member of the Faculty of the Universitá degli Studi di Napoli Federico II. In 2013, she became IEEE Fellow, and in 2009, she received the Stephen O.Rice Prize in the Field of Communications Theory. Her research interests lie in the area of communication systems, signal processing, information theory and random matrix theory.

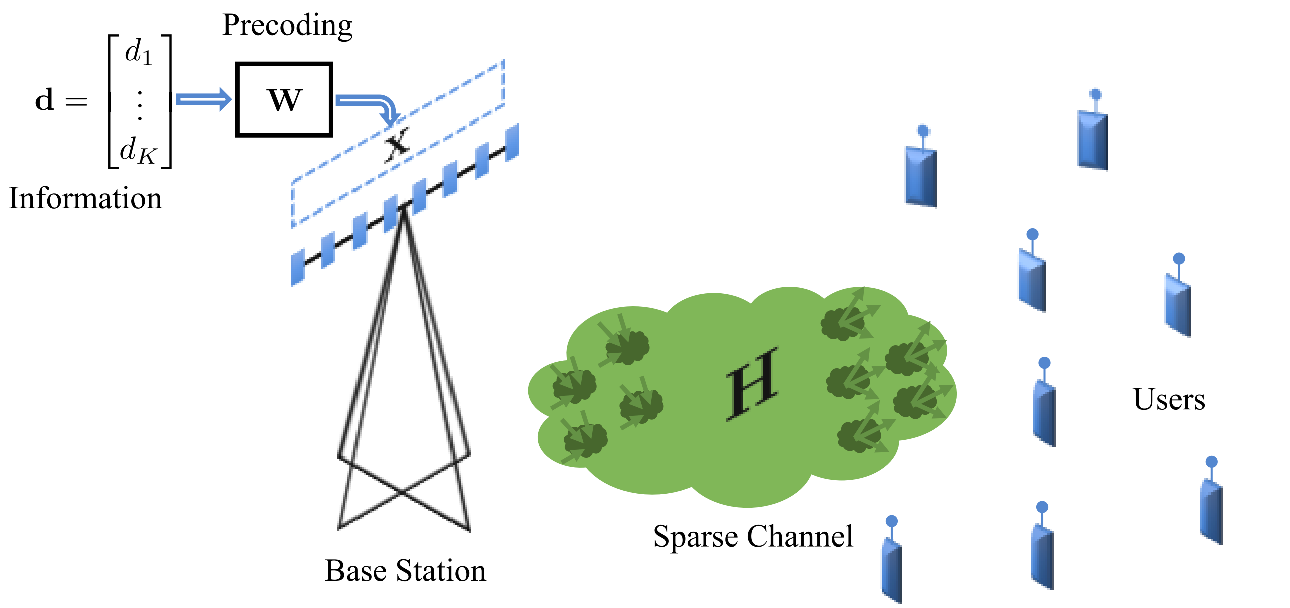

Fig. 1. Downlink multiuser system model. Information vector $\mathbf{d}$, composed of K threads, is precoded with matrix $\mathbf{W}$, and the resulting signal $\mathbf{x}$ is sent though channel $\mathbf{H}$. All users receive their corresponding signal and no further processing other than scaling may be done.

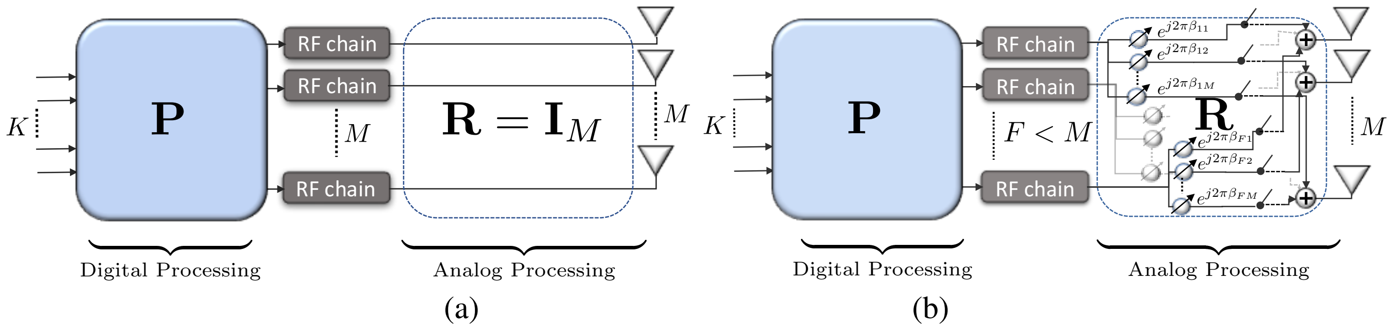

Fig. 2. Block diagram of the processing to be implemented at the BS: (a) Digital and analog processing for a FD implementation of the transmission strategy; (b) Digital and analog processing for a HB implementation of the transmission strategy.

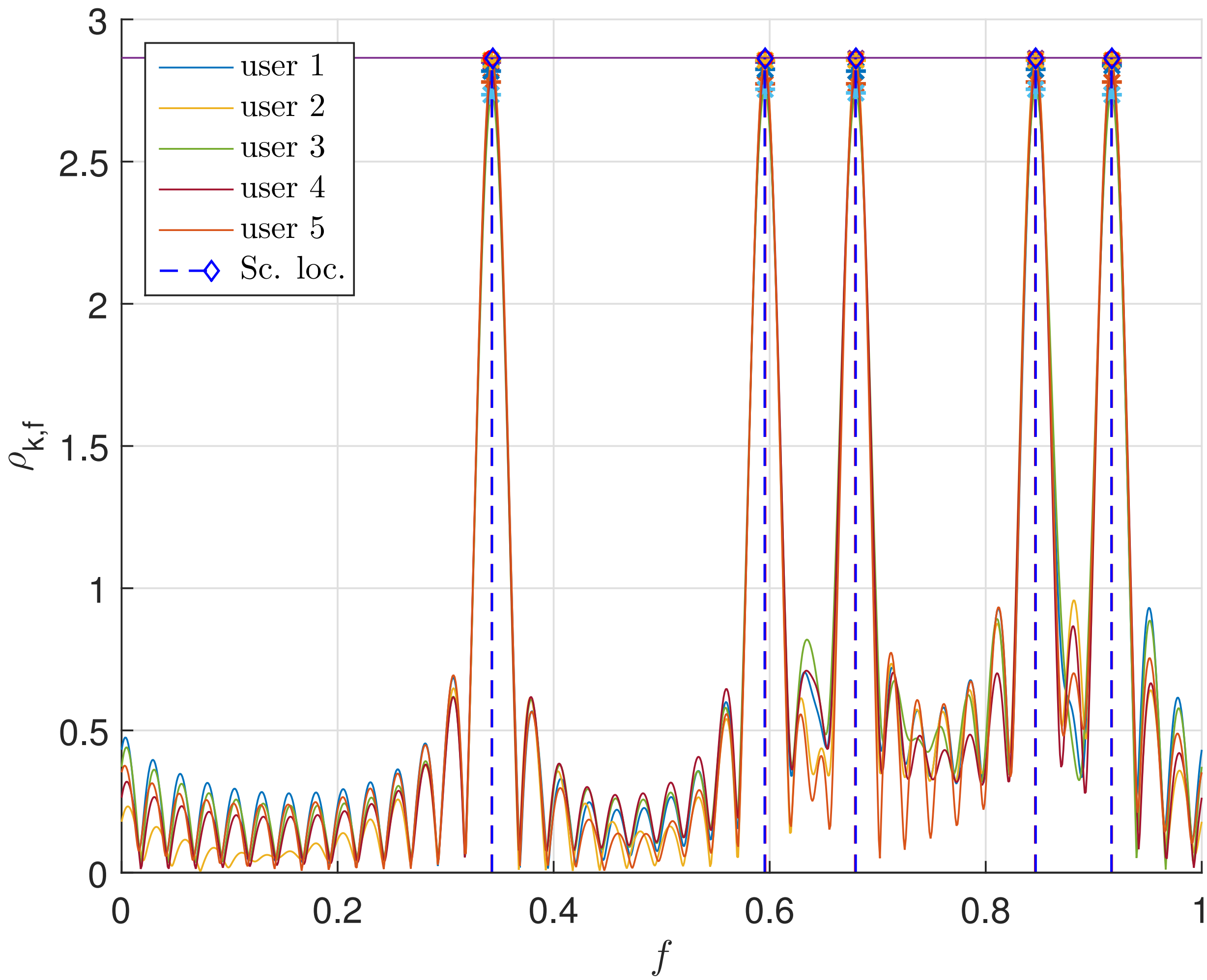

Fig. 3. Set of normalized frequencies found by AN and defining the analog processing implemented for each of the RF chain vectors. The dashed vertical blue lines (labelled Sc. loc.) represent the set of $L = 5$ frequencies that characterize the columns of the channel beamforming matrix. The colored lines are the frequencies found by AN for each of the $K = 5$ users.

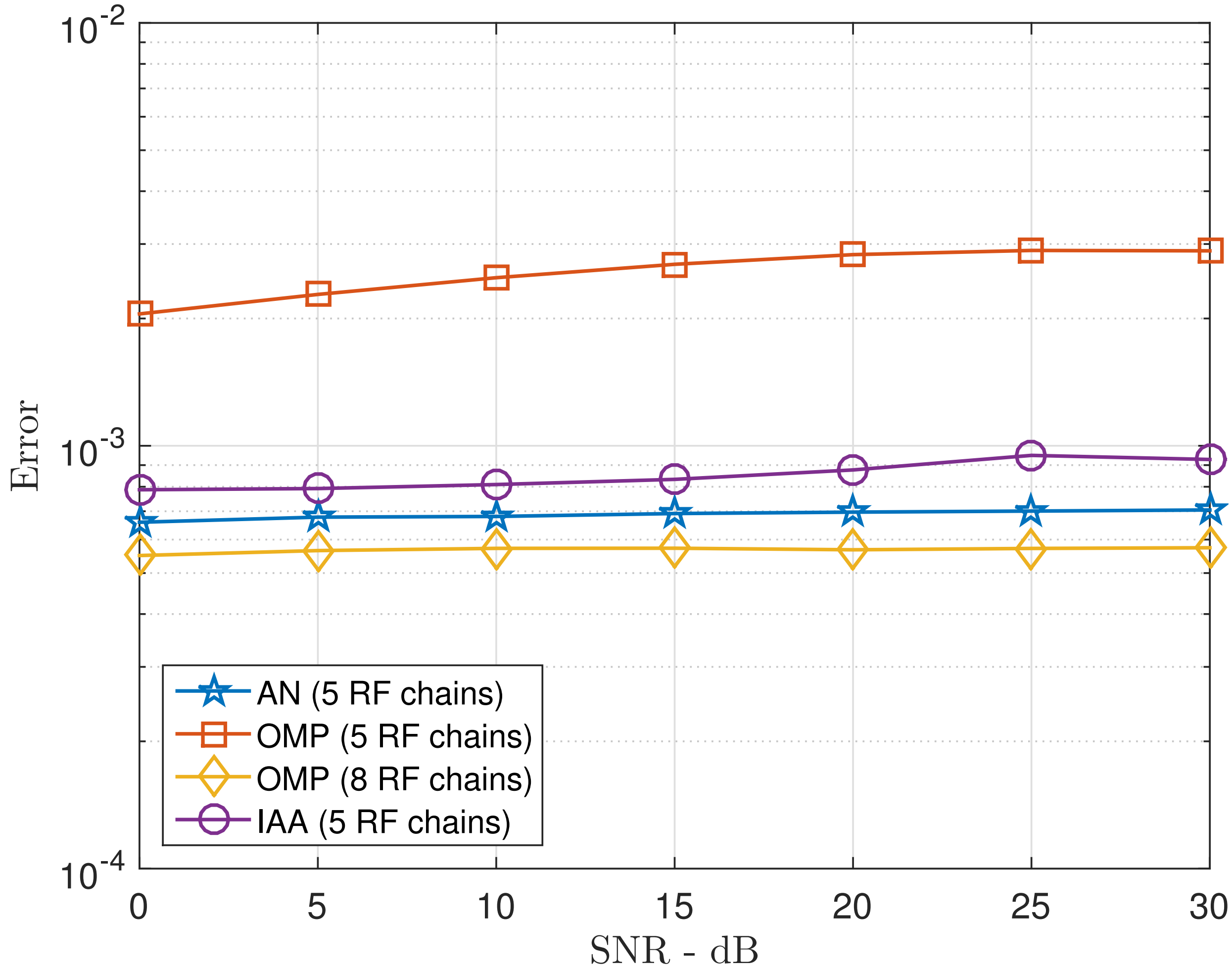

Fig. 4. Error in terms of squared distance between the MMSE (FD) precoder beams and the hybrid precoder vectors.