Digital Object Identifier

Multi-User Full Duplex Transceiver Design for mmWave Systems Using Learning-Aided Channel Prediction

K. SATYANARAYANA[1], (Student Member, IEEE), MOHAMMED EL-HAJJAR[1], (SENIOR MEMBER, IEEE), ALAIN MOURAD[2], AND LAJOS HANZO[1], (FELLOW, IEEE)
[1]University of Southampton, Southampton SO17 1BJ, U.K
[2]InterDigital Inc., London EC2A 3QR, U.K.

This work was supported by InterDigital as well as in part by the EPSRC under Project EP/Noo4558/1 and Project EP/PO34284/1, in part by the Royal Society's Global Research Challenges Grant, and in part by the European Research Council's Advanced Fellow Grant QuantCom.

ABSTRACT Millimeter Wave (mmWave) technology coupled with full duplex (FD) communication has the potential of increasing the spectral efficiency. However, the self interference (SI) encountered in the FD mode and the ubiquitous multi-user interference (MI) contaminates the signal. Furthermore, the system performance may also be limited by channel aging that arises because of the time-varying nature of the channel. Therefore, in this paper, we conceive FD hybrid beamforming (HBF) for K-user multiple-input multiple-output (MIMO) aided orthogonal frequency division multiplexing (OFDM) using learning-aided channel prediction. We first derive a joint precoder and combiner design for full duplex K-user MIMO-OFDM interference channels, where we aim for minimizing both the residual SI and the MI, followed by an iterative hybrid decomposition technique developed for OFDM systems. Then, we propose a learning-aided channel prediction technique for systems suffering from channel aging relying on a radial basis neural network, where we show by simulation that upon using sufficient training, learning-assisted channel prediction can faithfully estimate the current channel. Furthermore, we demonstrate by simulations that our proposed joint hybrid precoder and combiner design outperforms the popular eigen beamforming (EBF) technique by about 5 dB for a 128×32-element MIMO aided OFDM system having 32 sub-carriers.

INDEX TERMS Millimeter Wave, Hybrid Precoding, Full Duplex, MIMO, Machine Learning, Beamforming.

NOMENCLATURE

AA	Antenna Array
ANN	Artificial Neural Network
AoA	Angle-of-Arrival
AoD	Angle-of-Departure
BER	Bit Error Rate
BF	Beamforming
BS	Base Station
CSI	Channel State Information
EBF	Eigen Beamforming
FD	Full Duplex
FDD	Frequency Division Duplex
HBF	HBF
HD	Half Duplex
INR	Interference-to-Noise Ratio
MIMO	Multiple-Input Multiple-Output
mmWave	Millimeter Wave
MI	Multi-User Interference
MS	Mobile Station
MSE	Mean Square Error
MU	Multi-User
NMSE	Normalized Mean Squared Error
OF	Objective Function
OFDM	Orthogonal Frequency-Division Multiplexing
RF	Radio Frequency
SI	Self Interference
SIR	Signal-to-Interference Ratio
SNR	Signal-to-Noise Ratio
SU	Single User
SVD	Singular Value Decomposition

TPC	Transmit Precoder
ULA	Uniform Linear Array

I. INTRODUCTION

THE rapid proliferation of wireless devices has led to escalating tele-traffic. Given that the dearth of spectral resources at sub-6 GHz frequencies limits the attainable data rates, harnessing the large bandwidth reserves available at mmWave frequencies appears to be an attractive proposition for accommodating high-data rate transmissions, despite the high path loss [1]. Furthermore, until recently, it has been believed that the bidirectional communication in the same frequency is prevented by the high self interference (SI) that results from its own transmission [2]. However, the pioneering studies reported in [3]–[5] conceived a solution for simulatenous transmission and reception in the same physical resource block. This philosophy of simultaneous transmission and reception is termed as full-duplex (FD) communication. Recently, FD communication has captured the attention of the wireless community, given its potential of doubling the spectral efficiency. Hence, combining the concept of full duplex communication with mmWave technology is a promising solution to attain further high spectral efficiency. It is instructive to note that the SI cancellation at the local receiver is typically carried out by an amalgamation of passive and active methods [3], [6]. The passive methods include antenna isolation, where the SI suppression is carried out by increasing the path loss relying on the antenna separation and the angular position of the antenna array (AA) [5]. In contrast to the passive methods, the active methods rely on the information of the transmitted signal and they are typically carried out by pursuing three different approaches [7], namely radio-frequency (RF) cancellation, antenna cancellation, and digital cancellation. In the RF cancellation, the known transmitted signal is used as a reference signal, where the transmitted signal is reconstructed in the RF and subtracted from the received signal so as to remove the SI. In the antenna cancellation, the SI is suppressed by adding the two opposite phase replicas of the transmitted signal by utilizing multiple transmit and receive antennas [8]. By contrast, digital cancellation is typically carried out along with the RF and antenna cancellation to further improve the suppressing of the SI in the baseband [8], [9]. In other words, the digital cancellation shares the burden of the imperfections in the analog cancellation. However, the advantages obtained by employing RF-and-digital cancellation to suppress the SI may be circumscribed because of the undesired distortions at in the face of high non-linearities in the I/Q imbalance, power amplifier, and phase noise [2]. The residual SI is particularly high in mmWave systems because of the increased imperfections of the mmWave power amplifiers having a high gain times bandwidth product. In literature, most of the contributions on the FD has been focused on MIMO relaying systems with SI cancellation [10]–[14], while the related upper-layer work includes dynamic resource allocation conceived for FD systems [15]. Upon considering the limitations of the input circuitry, Day *et al.* derived both the upper and lowers bounds of the achievable sum rate for FD MIMO [16]. Furthermore, Everett *et al.* [17] provided the performance analysis for FD infrastructure nodes in the context of passive SI cancellation. More recently, we proposed a HBF design for a single user mmWave link [18], where we demonstrated the suppression of the residual SI upto 30 dB.

Although mmWave frequencies have abundant bandwidth resources for supporting large numbers of users, they suffer from high propagation losses due to atmospheric absorption, foliage attenuation and rain induced fading, which typically results in low signal-to-noise ratio (SNR) at the receiver. Hence, to mitigate the propagation losses high-gain directional transmission has to be employed. In mmWave communications, typically a hybrid architecture is employed, where both analog and digital signal processing aided SI-cancellation is carried out for reducing both the cost and hardware complexity [19]. To elaborate further, in hybrid systems, the signals are digitally precoded using a transmit precoder (TPC) in the baseband and then phase shifted in the RF using analog phase shifters before they are finally transmitted. Additionally, it is instructive to note that beamforming plays an important role in mitigating the SI, where the received signal is concentrated in a particular direction, thereby reducing the amount of SI it experiences in contrast to the omni-directional transmission. In the state-of-the-art, two HBF designs are reported, namely fully-connected and sub-array-connected architectures [20], [21].

In this paper, we focus our attention on full duplex HBF for K-user frequency selective interference channels. Most of the recent work in frequency selective HBF is focused on codebook design. Alkhateeb and Heath [22] developed codebooks for spatial multiplexing in wideband channels, followed by a hybrid precoding technique relying on popular Gram-Schmidt orthonormalization. Yu *et al.* [23] conceived the hybrid precoder design for frequency selective channels by formulating the problem as a norm distance between the optimal fully-digital precoder and the hybrid combination. In [24], a heuristic hybrid precoding design was derived for SU-MIMO OFDM systems which maximizes the achievable rate of the system. Kong *et al.* [25] discussed an iterative hybrid precoding technique, which aims for maximizing the sum-rate of all users. However, the effectiveness of the proposed algorithm is heavily reliant on the type of the frequency scheduling adopted. Furthermore, the proposed algorithm is only suitable when the total transmit power constraint is considered, whilst in practice, the precoders are designed by considering the more strict per sub-carrier transmit power. Additionally, Sohrabi and Yu [26] proposed a heuristic hybrid precoding design for a single-user mmWave system. However, this design works only for the setting when the digital precoder obeys the orthonormal property, which is typically not the case in face of multi-user interference.

On the other hand, the above-mentioned techniques rely on the idealized simplifying assumption of having perfect channel state information (CSI). However, this assumption is un-

TABLE 1: Summary of our design contrasting with state-of-the-art, where ✓and the blank correspond to with and without contributions, respectively .

Themes of Contributions	This Paper	[18]-2018	[13]-2017	[7]-2016	[10]-2011
Full Duplex	✓	✓	✓	✓	✓
Multi-User	✓				
Multi-Carrier	✓				
Beamforming	✓	✓		✓	
MI Considerations	✓				
Machine Learning	✓				
Channel Prediction	✓				
Multi-Carrier HBF	✓				

realistic [27]. This is because, typically in the FDD systems, the channel is estimated at the receiver in the downlink and fedback to the transmitter in the uplink, where the transmitter leverages the CSI to enhance the system performance [28]. However, the CSI acquired at the transmitter is imperfect owing to the estimation errors introduced by the non-linear components in the hardware and channel estimation per se. These errors in the CSI would be exacerbated at the transmitter when the ubiquitous phenomenon of channel aging is also considered [27]. To elaborate, the channel aging is a phenomenon that mainly arises because of the mobility of the users relative to the base station (BS). As a result, the channel information arriving at the transmitter would become outdated [27]. It should be noted that channel aging may also arise because of the phase noise due to local oscillators, which causes Doppler shift [29].

To counter the effects of channel aging that arises due to Doppler shift in frequencies, an extensive body of work is focused on channel prediction [27], [30]–[32]. However, when the channel is subjected to high Doppler spreads, these techniques do not accurately predict the channel.

More recently, artificial neural network (ANN)-aided wireless transmission has gained substantial attention owing to its accurate predictions, which is often superior to conventional methods dispensing with learning [33]. The state-of-the-art designs of the localization literature are focused on learning [34]–[36]. More particularly, learning based approaches used in localization have been shown to be more effective in terms of minimizing the location error. Wang *et al.* [37] validated by experimental demonstration that a CSI-based fingerprint relying on the deep learning approach achieves better performance than that of its counterparts. An experimental study of indoor localization conducted by Chen *et al.* [38] also attributed its superior performance to learning, where the authors invoke a convolutional neural network assisted learning scheme.

Inspired by this, in this paper we resort to a neural network based approach for channel prediction. In other words, we predict the channel using pilot-assisted channel estimate, which will become outdated by the time they reach the transmitter for beamforming or precoding, relying on a neural network-aided learning technique. More particularity, we employ a radial basis neural network for channel prediction, where the transmitter employs HBF relying on the channel predicted from the pilot-assisted channel estimates acquired at the receiver and signaled to the transmitter. Historically, radial basis functions were introduced for the purpose of exact function interpolation [39]. In other words, a linear combination of radial basis functions is capable of reproducing or interpolating a function with maximum fidelity. Motivated by this, in this treatise, we invoke a radial basis neural network. The construction and the design of the network for channel prediction is elucidated later in this paper.

Against this background, to best of our knowledge, this is the first paper studying the FD HBF designed for K-user MIMO-OFDM interference channels operating in mmWave communications using learning-aided channel estimation. Table 1 contrasts our design to state-of-the-art. More explicitly, our contributions in this paper are summarized as follows. Our specific contributions in this paper are as follows.

1) We derive a precoder and combiner design for K-user frequency selective interference channels which minimizes both the SI and multi-user interference (MI) in mmWave systems using beamforming. In this design, we aim for preserving the signal dimension, while mitigating the interference.
2) We develop an iterative matrix decomposition for hybrid precoding aided OFDM systems, where the digital TPC weights are employed in the OFDM scheme's frequency-domain, while the analog RF beamformer weights are applied to the time-domain signal.
3) To counteract the effects of channel aging, we propose a learning-assisted channel prediction relying on a radial basis function neural network, where we show by simulation that upon involving sufficient training neural network-aided channel prediction can faithfully reproduce the current channel.
4) We empirically show that depending on the Doppler spread, the neural network used for channel predictions has to be periodically retrained. In other words, we demonstrate by simulations that the overhead involved in the CSI feedback for faithful reproduction of the actual channel at the transmitter is low for lower Doppler spreads.
5) We show by our simulations that the proposed HBF design outperforms the classic EBF by more than 5 dB

at sum rate of 40 bps/Hz.

The rest of the paper is organized as follows. In Sec. II, we detail both the system and channel models considered. The derived precoder and combiner design is detailed in Sec. III, while the iterative hybrid decomposition is developed in Sec. IV. Finally, our simulation results and conclusions are presented in Sec. VI and Sec. VIII, respectively.

Notations: We use $\mathbf{A}$ for matrices, $\mathbf{a}$ for vectors, and $\mathbf{I}$ for identity matrix, while we use H for the Hermitian transpose. Finally, $\mathcal{U}$, and $\mathcal{CN}$ represent the uniform distribution and the complex-valued normal distribution respectively.

II. SYSTEM MODEL

In this paper, we consider a K-user interference channel, where the users are communicating over a mmWave channel. To counter act the channel-induced dispersion, each user employs OFDM based transmission [40]. Fig. 1 depicts the K-user interference channel, where each user communicates with its intended receiver. As an example, in Fig. 1, Nodes 1 and 2 operate in full duplex mode, while all other nodes operate in half duplex mode[1]. To elaborate further, User 1 at Node 1 is transmitting to its receiver at Node 2 over the wireless channel $\mathbf{H}_{11}$, while User 2 at Node 2 is transmitting to its receiver at Node 1 over the channel $\mathbf{H}_{22}$. Likewise, the user at Node 3 is transmitting to its respective receiver over the channel characterized by the matrix $\mathbf{H}_{33}$ and so forth, while the channel matrix $\mathbf{H}_{ij}, i \neq j$ represents the interfering channel from transmitter i to the receiver j. For example, in Fig. 1, $\mathbf{H}_{31}$ is the interfering channel between transmitter 3 and receiver 1.

In this paper, we aim for designing the hybrid precoders that minimize the MI for users operating in the half duplex mode and both the MI as well as SI for the users in the FD mode. A similar approach can be used when all the users are in full duplex mode by considering the respective interference channels. It is instructive to note that the users at Nodes 1 and 2 of Fig. 1 transmit and receive simultaneously, as shown in Fig. 2. It can be seen from Fig. 2 that both nodes are equipped with N_t transmit and N_r receive antennas. Furthermore, the architecture of the transmitter and receiver of each node rely on the fully-connected architecture as shown in Fig. 2, where the signals in the baseband are digitally precoded and then phase shifted in the analog RF domain before transmission.

Furthermore, the transmit and local receive AAs at each node may be positioned at different angles with a separation D between the arrays as shown in Fig. 3. The placement of the AAs also plays an important role in mitigating the SI. The SI can be reduced by increasing the separation D between the arrays, whilst also relying on the appropriate angle θ. Intuitively, it can be seen that when $D = 0$ and $\theta = 0^\circ$, the SI would be significantly high, as the transmitter and receiver share the same array. On the other hand, the users at other

[1]The design applies to any configuration, where any node can be FD or HD. In this paper, we used this example for simplifying explanation.

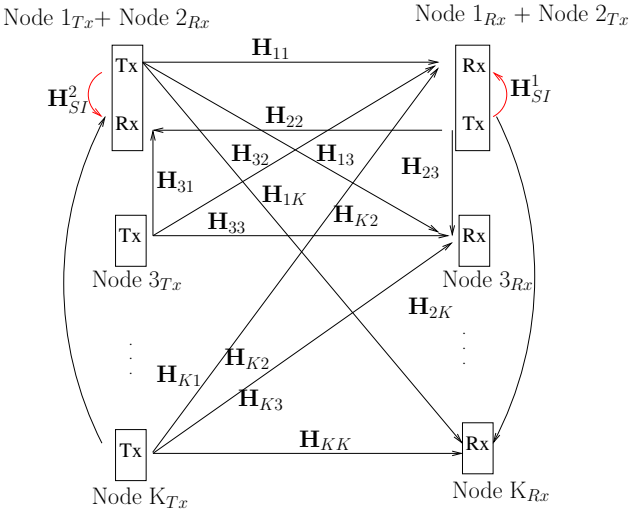


FIGURE 1: *K-user interference channel, where User 1 and 2 rely on FD communication, while the remaining users operate in HD mode.*

nodes rely on half-duplex communication using the fully-connected design. The users' signals before transmission are precoded using a digital TPC matrix $\mathbf{F}_{\text{BB}}$ of size $N_t^{\text{RF}} \times N_s$ in the baseband and then they are phase-shifted using the analog beamformer matrix $\mathbf{F}_{\text{RF}}$ of size $N_t \times N_t^{\text{RF}}$. At the receiver, the signals are processed using the beamformer matrix $\mathbf{W}_{\text{RF}}$ of size $N_r \times N_r^{\text{RF}}$ in the analog RF domain, followed by the digital combiner matrix $\mathbf{W}_{\text{BB}}$ of size $N_r^{\text{RF}} \times N_s$, where N_t^{RF} and N_r^{RF} are the number of RF chains at the transmitter and receiver, respectively. Then, the received signal vector $\mathbf{y}$ for the n^{th} sub-carrier after combining at the i^{th} receiver node is given by[2]

$$\begin{aligned}\mathbf{y}^{(i)}[n] = &\underbrace{\mathbf{W}_{\text{BB}}^{H^{(i)}}[n]\mathbf{W}_{\text{RF}}^{H^{(i)}}\mathbf{H}_{\text{ii}}[n]\mathbf{F}_{\text{RF}}^{(i)}\mathbf{F}_{\text{BB}}^{(i)}[n]\mathbf{s}^{(i)}[n]}_{\text{desired signal}} + \\ &\underbrace{\mathbf{W}_{\text{BB}}^{H^{(i)}}[n]\mathbf{W}_{\text{RF}}^{H^{(i)}}\mathbf{H}_{\text{SI}}^{(i)}[n]\mathbf{F}_{\text{RF}_{\text{SI}}}^{(i)}\mathbf{F}_{\text{BB}_{\text{SI}}}^{(i)}[n]\mathbf{s}^{(i)}[n]}_{\text{self interference (SI)}} \\ &+ \underbrace{\sum_{j=1}^{K}\mathbf{W}_{\text{BB}}^{H^{(i)}}[n]\mathbf{W}_{\text{RF}}^{H^{(i)}}[n]\mathbf{H}_{ji}[n]\mathbf{F}_{\text{RF}}^{(j)}\mathbf{F}_{\text{BB}}^{(j)}[n]\mathbf{s}^{(j)}[n]}_{\text{multi-user interference (MI)}} \\ &+ \underbrace{\mathbf{W}_{\text{BB}}[n]^{H^{(i)}}\mathbf{W}_{\text{RF}}^{H^{(i)}}\mathbf{n}^{(i)}}_{\text{noise}}, \quad \forall j \neq i, \end{aligned} \tag{1}$$

where $\mathbf{n}$ is the noise vector of dimension $N_r^{(i)}$, whose elements obey the distribution $\mathcal{CN}(0, \sigma^2)$, $\mathbf{s}^{(i)}$ is the signal vector of dimension $N_s^{(i)}$, $\mathbf{H}_{\text{ji}}$ is the mmWave channel matrix of size $N_r \times N_t$ spanning from the j^{th} node transmitter to the i^{th} node receiver so that $\mathbb{E}[\|\mathbf{H}_{\text{ji}}\|_F^2] = N_t N_r$, which is given

[2]With slight abuse of notation, we represent the matrices $\mathbf{F}_{\text{RF}_{\text{SI}}}^{(i)}\mathbf{F}_{\text{BB}_{\text{SI}}}^{(i)}[n]$ as the precoder matrices of the transmitter causing SI at the node i.

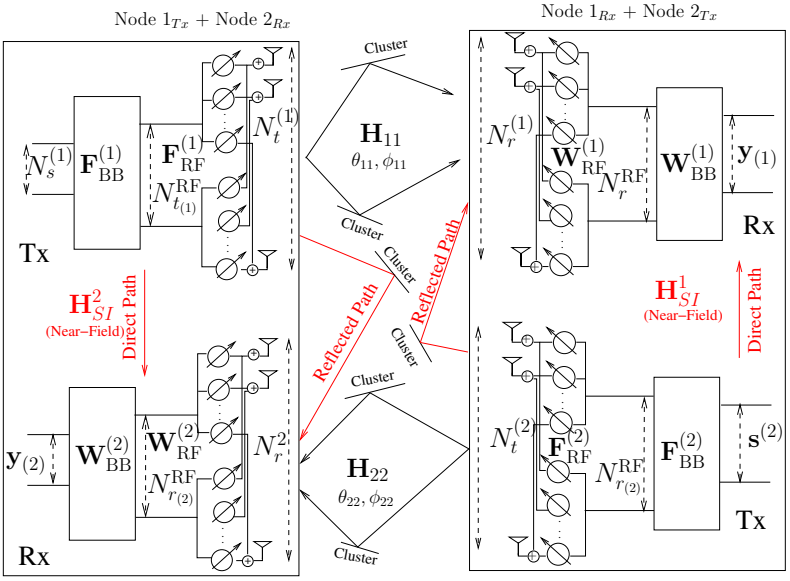


FIGURE 2: *Full duplex scenario using HBF, where the constituents (Tx/Rx) of each node are fully-connected.*

by [41]

$$\mathbf{H}_{ji}[n] = \sqrt{\frac{N_r^{(i)} N_t^{(j)}}{N_c N_{\text{ray}}}} \times \tag{2}$$

$$\sum_{n_c=1}^{N_c} \sum_{n_{\text{p}}=1}^{N_{\text{p}}} \alpha_{ji}^{(n_{\text{p}},N_c)} \mathbf{a}_r(\phi_{ji}^{n_{\text{p}},n_c}) \mathbf{a}_t^T(\theta_{ji}^{n_{\text{p}},n_c}) e^{-\hat{j}2\pi(n_c-1)n/N}, \tag{3}$$

where $\hat{j} = \sqrt{-1}$, N is the number of sub-carriers, while the distribution of $\alpha_{n_c}^{n_{\text{p}}}$ is $\sim \mathcal{CN}(0,1)$. For a uniform linear array (ULA) having N_r and N_t antenna elements the response vectors $\mathbf{a}_r$ and $\mathbf{a}_t$ are expressed as:

$$\mathbf{a}_r(\phi_r) = \frac{1}{N_r}[1\ e^{\hat{j}\frac{2\pi}{\lambda}d\cos(\phi_r)} \dots e^{\hat{j}\frac{2\pi}{\lambda}(N_r-1)d\cos(\phi_r)}]^T, \tag{4}$$

$$\mathbf{a}_t(\theta_t) = \frac{1}{N_t}[1\ e^{\hat{j}\frac{2\pi}{\lambda}d\cos(\phi_t)} \dots e^{\hat{j}\frac{2\pi}{\lambda}(N_t-1)d\cos(\phi_t)}]^T. \tag{5}$$

where θ_t and ϕ_r, N_c and N_{p} are the angles of departure (AOD) and arrival (AOA) as well as the number of clusters and rays, respectively.

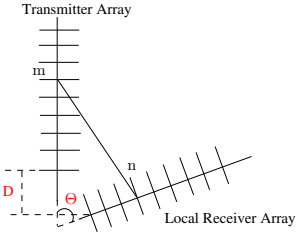


FIGURE 3: *Geometrical representation of the transmit and receive AAs of a node.*

Furthermore, $\mathbf{H}_{\text{SI}}^{(i)}$ is the near-field SI channel. Typically, when the Tx and Rx are geographically separated by a

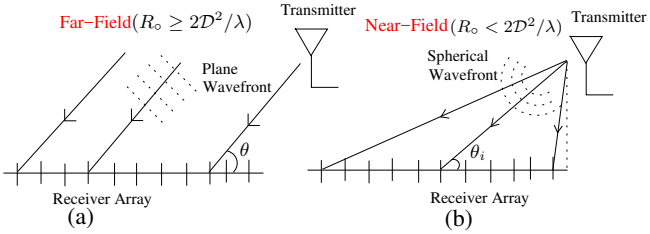


FIGURE 4: *Depiction of far-field model and near-field model. (a) In the far-field model ($R_o \geq 2\mathcal{D}^2/\lambda$), the signal is assumed to behave as a planar wave. (b) The signal is assumed to behave as a spherical wavefront ($R_o < 2\mathcal{D}^2/\lambda$).*

distance of $R_o \geq 2\mathcal{D}^2/\lambda$ as shown in Fig. 4 (a), where $\mathcal{D}$ is the measure of the antenna aperture, the underlying assumption is that the signal ray strikes the AA as a planar wave [42], where all the rays experience the same propagation losses. However, in the FD mode the transmit and local receive AAs are closely spaced, which violates the planar wave assumption. Hence, a more realistic spherical wavefront is considered, as shown in Fig. 4(b). Under the spherical wavefront assumption the SI channel coefficients are given by [7], [43]

$$[\mathbf{H}_{\text{LOS}}^{(i)}]_{pq}[n] = \frac{\rho}{r_{pq}} \exp(-j2\pi \frac{r_{pq}}{\lambda}), \tag{6}$$

where ρ is the normalization factor introduced for ensuring $\mathbb{E}[\|\mathbf{H}_{\text{SI}}^{(i)}\|_F^2] = N_t N_r$ and r_{pq} is the distance between the p^{th} element of the transmitter and q^{th} element of the receiver [7], which is given by (8). Thus the SI channel by accounting the reflected path is expressed as

$$\mathbf{H}_{\text{SI}}^{(i)}][n] = \sqrt{\frac{\kappa}{\kappa+1}} \mathbf{H}_{\text{LOS}}^{(i)}][n] + \sqrt{\frac{1}{\kappa+1}} \mathbf{H}_r[n], \tag{7}$$

where $\mathbf{H}_r$ follows the statistical channel model given in (2) and κ is the Rician factor.

It is important to note that while the $\mathbf{H}_{\text{LOS}}^{(i)}[n]$ component of $\mathbf{H}_{\text{SI}}^{(i)}[n]$ is independent of any fading and can be estimated without any overhead, the channel matrices $\mathbf{H}_r[n]$ and $\mathbf{H}_{ji}[n]$ have to be estimated periodically by pilots. Furthermore, given the large number of antennas, the overhead involved in pilot-assisted channel estimation would typically be high and scales with the number of antennas [44]–[47]. However, exploiting the sparsity of the mmWave channel is capable of reducing the overhead required for channel estimation [48]

In the next section, we discuss the precoder and combiner design proposed that mitigates both the SI and the MI.

III. PROPOSED PRECODER AND COMBINER DESIGN

In this section, we discuss the precoder and combiner design proposed for the system model of (1) which reduces both the SI and the MI. To design the hybrid precoders of (1), we first derive the fully-digital optimal precoder and combiner. After obtaining the fully-digital solution, we decompose the fully-digital precoder into analog RF beamformer and digital TPC matrices.

To arrive at the fully-digital solution, we set $\mathbf{F}^{(i)} = \mathbf{F}_{\text{RF}}^{(i)}\mathbf{F}_{\text{BB}}^{(i)}$ and $\mathbf{W}^{H^{(i)}} = \mathbf{W}_{\text{BB}}^{H^{(i)}}\mathbf{W}_{\text{RF}}^{H^{(i)}}$. Then, the system model in (1) reduces to

$$\mathbf{y}^{(i)}[n] = \underbrace{\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n]\mathbf{s}^{(i)}[n]}_{\text{desired signal}} + \underbrace{\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{\text{SI}}^{(i)}[n]\mathbf{F}_{\text{SI}}^{(i)}[n]\mathbf{s}^{(i)}[n]}_{\text{SI}} + \underbrace{\sum_{j=1}^{K}\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ji}[n]\mathbf{F}^{(j)}[n]\mathbf{s}^{(j)}[n]}_{\text{MI}} + \underbrace{\mathbf{W}^{H^{(i)}}[n]\mathbf{n}^{(i)}}_{\text{noise}}, \forall j \neq i. \tag{9,10}$$

The second and third terms of (10) represent the SI and the MI, respectively. Then, the total interference plus noise ($\mathbf{TI}$) at the receiver Node i is

$$\mathbf{TI}^{(i)}[n] = \mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{\text{SI}}^{(i)}[n]\mathbf{F}^{(i)}[n]\mathbf{s}^{(i)}[n] + \tag{11}$$

$$\sum_{j=1}^{K}\mathbf{W}^{H^{(j)}}[n]\mathbf{H}_{ji}[n]\mathbf{F}^{(j)}\mathbf{s}^{(j)}[n] + \mathbf{W}^{H^{(i)}}[n]\mathbf{n}^{(i)}[n]. \tag{12}$$

Upon denoting the co-variance matrix of the interference plus noise power at the receiver Node i by $\mathbf{Q}^{(i)}[n]$, we have

$$\mathbf{Q}^{(i)}[n] = \text{Tr}\left(\mathbf{W}^{H^{(i)}[n]}\mathbf{R}^{(i)}[n]\mathbf{W}^{(i)}[n]\right), \tag{13}$$

with $\mathbf{R}^{(i)}[n]$ represented by

$$\mathbf{R}^{(i)}[n] = \underbrace{P_{\text{SI}}^{(i)}\mathbf{H}_{\text{SI}}^{(i)}[n]\mathbf{F}_{\text{SI}}^{(i)}[n](\mathbf{H}_{\text{SI}}^{(i)}[n]\mathbf{F}_{\text{SI}}^{(i)}[n])^H}_{\text{SI}} + \underbrace{\sum_{j=1}^{K}P^{(j)}\mathbf{H}_{ji}[n]\mathbf{F}^{(j)}[n]\left(\mathbf{H}_{ji}[n]\mathbf{F}^{(j)}[n]\right)^H}_{\text{MI}} + \sigma^2\mathbf{I}_n, \tag{14}$$

where $P_{\text{SI}}^{(i)}$ is the SI power, $P^{(j)}$ is the power of the j^{th} user, and σ^2 is the noise power. Our aim is to evaluate $\mathbf{W}$ and $\mathbf{F}$ so that the interference power $\mathbf{Q}^{(i)}$ at the receiver is minimized while also preserving the signal dimensionality, i.e. rank$\left(\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n]\right) = N_s^{(i)}$.

Proposition 1: For the optimization problem of

$$\min_{\mathbf{W}^{(i)}[n]} \text{Tr}\left(\mathbf{W}^{H^{(i)}}[n]\mathbf{R}^{(i)}[n]\mathbf{W}^{(i)[n]}\right) \tag{15}$$

$$s.t.\ \mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n] = \alpha\mathbf{I}_{N_s},$$

where $\mathbf{R}^{(i)}[n]$ is a positive definite matrix ($\mathbf{R}^{(i)}[n] \succ 0$), the optimal solution is given by

$$\mathbf{W}_{\text{opt}}^{(i)}[n] = \alpha\mathbf{R}^{(i)^{-1}}[n]\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n] * \left(\left(\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n]\right)^H\mathbf{R}^{(i)^{-1}}[n]\left(\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n]\right)\right)^{-1} \tag{16,17}$$

and α is the normalization factor given by

$$\alpha = \frac{1}{\sqrt{\text{Tr}\left(\mathbf{W}^{H(i)}[n]_{\text{opt}}\mathbf{W}_{\text{opt}}^{(i)}[n]\right)}}. \tag{18}$$

Proof: See Appendix A.

We can see from (16) that $\mathbf{W}_{\text{opt}}^{(i)}$ is dependent on the matrix $\mathbf{F}^{(i)}$, while our aim is to jointly evaluate $\mathbf{W}$ and $\mathbf{F}$. Therefore, we resort to an alternating design, where we initialize $\mathbf{F}^{(i)}$ to the right singular vector of the channel matrix $\mathbf{H}_{ii}$ to obtain $\mathbf{W}_{\text{opt}}^{(i)}$.

We then proceed to design the precoder matrix $\mathbf{F}^{(i)}$ by considering the interference imposed by the transmitter of Node i upon the local receiver of the same node. More explicitly, we aim for specifically designing the precoder to minimize the interference inflicted upon the unintended receivers due to its own transmission.

Let us consider the total interference imposed by the transmitter j on the unintended receivers, which is given by

$$\mathbf{J}^{(j)}[n] = \text{Tr}\left(\mathbf{F}^{H^{(j)}}[n]\mathbf{S}^{(j)}[n]\mathbf{F}^{(j)}[n]\right), \tag{19}$$

$$r_{mn} = \sqrt{\left(\frac{D}{\tan(\Theta)} + (n-1)\frac{\lambda}{2}\right)^2 + \left(\frac{D}{\sin(\Theta)} + (m-1)\frac{\lambda}{2}\right)^2 - 2\left(\frac{D}{\tan(\Theta)} + (n-1)\frac{\lambda}{2}\right)\left(\frac{D}{\sin(\Theta)} + (m-1)\frac{\lambda}{2}\right)\cos(\Theta)}, \tag{8}$$

where D is the separation between the transmit and receive AAs, while Θ is the angular orientation of the transmit AA with respect to the receive AA, as shown in Fig. 3.

where $\mathbf{S}^{(j)}[n]$ is given by[3]

$$\mathbf{S}^{(j)}[n] = (P_{\text{SI}}^{(j)}\mathbf{W}^{H^{(j)}}[n]\mathbf{H}_{\text{SI}}^{(j)}[n])^H(\mathbf{W}^{H^{(j)}}[n]\mathbf{H}_{\text{SI}}^{(j)}[n]) + \sum_{i=1}^{K} P^{(i)}\left(\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ji}[n]\right)^H\left(\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ji}[n]\right) + \mathbf{I}, \forall i \neq j. \tag{20}$$

Similar to (14), our objective here is to minimize the interference power $\mathbf{S}^{(j)}$ caused by the transmitter.

Thus, the associated constrained optimization problem is given by

$$\min_{\mathbf{F}^{(j)}[n]} \text{Tr}\left(\mathbf{F}^{H^{(j)}}[n]\mathbf{S}^{(j)}[n]\mathbf{F}^{(j)}[n]\right) \tag{21}$$
$$s.t.\ \mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{jj}[n]\mathbf{F}^{(j)}[n] = \beta\mathbf{I}_{N_s}.$$

By invoking Proposition 1, the optimal solution is expressed as

$$\mathbf{F}_{\text{opt}}^{(j)}[n] = \beta\mathbf{S}^{(j)^{-1}}[n]\mathbf{H}_{jj}^H[n]\mathbf{W}^{H^{(j)}}[n]* \tag{22}$$
$$\left(\left(\mathbf{W}^{H^{(j)}}[n]\mathbf{H}_{jj}[n]\mathbf{S}^{(j)^{-1}}[n]\right)\left(\mathbf{W}^{H^{(j)}}[n]\mathbf{H}_{jj}[n]\right)^H\right)^{-1} \tag{23}$$

and β is a normalization factor given as

$$\beta = \frac{1}{\sqrt{\text{Tr}\left(\mathbf{F}_{\text{opt}}^{H(j)}[n]\mathbf{F}_{\text{opt}}^{(j)}[n]\right)}}. \tag{24}$$

Algorithm 1 HBF Design Proposed for FD System

1: Let $\mathbf{H}_{ii} = \mathbf{U}_{ii}\Sigma_{ii}\mathbf{V}_{ii}$
2: Set $\mathbf{F}^{(i)}$ to $\mathbf{V}_{ii}$
3: Compute $\mathbf{R}^{(i)}, i = 1, 2$
4: Compute $\mathbf{W}^{(i)} = \mathbf{R}^{(i)^{-1}}\mathbf{H}_{ii}\mathbf{F}^{(i)}\left((\mathbf{H}_{ii}\mathbf{F}^{(i)})^H\mathbf{R}^{(i)^{-1}}(\mathbf{H}_{ii}\mathbf{F}^{(i)})\right)^{-1}$
5: Then normalize $\mathbf{W}^{(i)}$ as $\frac{\mathbf{W}^{(i)}}{\sqrt{\text{Tr}(\mathbf{W}^{H(i)}\mathbf{W}^{(i)})}}$
6: Compute $\mathbf{S}^{(i)}$
7: Obtain $\mathbf{F}^{(i)} = \mathbf{S}^{(i)^{-1}}\mathbf{H}_{jj}^H\mathbf{W}^{H(i)}\left(\mathbf{W}^{H(i)}\mathbf{H}_{ii}\mathbf{S}^{(i)^{-1}}(\mathbf{W}^{H(i)}\mathbf{H}_{ii})^H\right)^{-1}$
8: Then normalize $\mathbf{F}^{(i)}$ as $\frac{\mathbf{F}^{(i)}}{\sqrt{\text{Tr}(\mathbf{F}^{H(i)}\mathbf{F}^{(i)})}}$
9: Continue from Step 2 until convergence of (15) and (21) is reached.

[3]The term $\mathbf{I}$ in (20) ensures the matrix $\mathbf{S}^{(j)}[n]$ is invertible.

This process is iterative and it is continued until the convergence of (15) and (21) is reached. The related pseudo code is presented in Algorithm 1. With slight abuse of notation we have dropped the notation $[n]$.

The exposition of Algorithm 1 is as follows: initialize the digital precoder to the first N_s columns of the singular matrix of the channel. Then compute the matrix $\mathbf{R}^{(i)}$ using (14). Having computed the matrix $\mathbf{R}^{(i)}$, obtain the combiner matrix $\mathbf{W}^{(i)}$ using (16) and ensure it is normalized. Similarly, compute the matrix $\mathbf{S}^{(i)}$ using (20). Then obtain the precoder matrix $\mathbf{F}^{(i)}$ using (22) and normalize it for ensuring that the total transmit power is constant. Iterate this process until the convergence of (13) and (19) is achieved.

Convergence: We characterize the convergence of the objective functions (OFs) in (15) and (21). We note that the value of the OF computed in every iteration decreases and converges to its local optimum in a few iterations. Here, we provide the proof of convergence for the proposed algorithm. The SI plus noise power at the receiver of node i is given by (25). Similarly, the SI plus noise power at node j is given by (27).

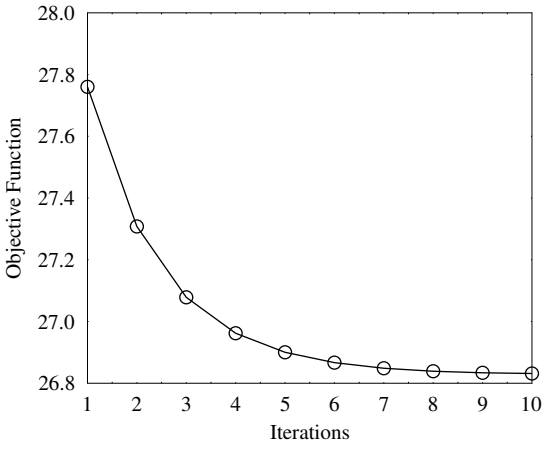


FIGURE 5: *Convergence of the proposed design.*

We observe that the Equations (25) and (27) are identical. Therefore, the matrix $\mathbf{W}$ which is designed for minimizing (25) also minimizes (27) and similarly the matrix $\mathbf{F}$ designed for minimizing (27) also minimizes (25).

Fig. 5 characterizes the convergence of our precoder and combiner design. It is evident from the figure that the OF (25) is minimized in each iteration and it converges to a locally optimal solution in as few as 8-9 iterations.

$$\mathbf{Q}^{(i)} = \mathrm{Tr}\left(\mathbf{W}^{H^{(i)}}\mathbf{R}^{(i)}\mathbf{W}^{(i)}\right) \tag{25}$$

$$= \mathrm{Tr}\left(\mathbf{W}^{H^{(i)}}\left(\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}^{(j)}(\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}^{(j)})^{H} + \sum_{j=1}^{K} P^{(j)}\mathbf{H}_{ji}[n]\mathbf{F}^{(j)}[n]\left(\mathbf{H}_{ji}[n]\mathbf{F}^{(j)}[n]\right)^{H} + \mathbf{I}_n\right)\mathbf{W}^{(i)}\right) \tag{26}$$

$$\mathbf{J}^{(j)} = \mathrm{Tr}\left(\mathbf{F}^{H^{(j)}}\mathbf{S}^{(j)}\mathbf{F}^{(j)}\right), \tag{27}$$

$$= \mathrm{Tr}\left(\mathbf{F}^{H^{(j)}}\left((\mathbf{W}^{H^{(j)}}\mathbf{H}_{\mathrm{SI}})^{H}(\mathbf{W}^{H^{(j)}}\mathbf{H}_{\mathrm{SI}}) + \sum_{i=1}^{K} P^{(i)}\left(\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ji}[n]\right)^{H}\left(\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ji}[n]\right) + \mathbf{I}_n\right)\mathbf{F}^{(j)}\right) \tag{28}$$

Note that in this design we assumed perfect CSI. However, considering the practical limitations on the knowledge of having perfect CSI, we proposed channel prediction in Section V of the paper. Therefore, the proposed precoder/combiner design is applied on the predicted channel.

In the next section, we present the decomposition of the fully-digital solution into analog RF beamformer and digital TPC.

IV. ITERATIVE HYBRID DECOMPOSITION

In the previous section, we have derived a fully-digital solution. In this section, we aim for designing the hybrid precoder matrices that closely approximates the obtained solution. In OFDM systems the analog RF beamformer is conveniently applied in the time-domain, while the digital baseband TPC is employed in the frequency-domain. Therefore, to decompose the fully-digital solution obtained into the analog and digital precoder matrices, we pursue a similar approach to [41], where the objective is to minimize the Frobenius norm between the fully-digital and hybrid precoders. This can be formulated as [41]

$$\min_{\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}^{(i)}[n]} \sum_{n=1}^{N} \|\mathbf{F}_{\mathrm{opt}}^{(i)}[n] - \mathbf{F}_{\mathrm{RF}_k}^{(i)}\mathbf{F}_{\mathrm{BB}_k}^{(i)}[n]\|_F^2, \tag{29}$$

$$s.t. \begin{cases} |\mathbf{F}_{\mathrm{RF}}(a,b)| = 1 \\ \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[n]\|_{\mathrm{F}}^2 = N_s \end{cases}.$$

In order to solve the above optimization problem, we first fix the matrix $\mathbf{F}_{\mathrm{RF}}^{(i)}$ to $\angle\mathbf{F}_{\mathrm{opt}}^{(i)}$ and then obtain a locally optimal solution $\mathbf{F}_{\mathrm{BB}}^{(i)}$, which is then utilized for obtaining the unconstrained $\mathbf{F}_{\mathrm{RF}}^{(i)}$ in the next step.

Thus, the optimization problem of (29) can be interpreted as a two-stage problem. Furthermore, since the baseband TPC is obtained separately for each sub-carrier, we can get rid of the summation in (29), while optimizing the baseband TPC $\mathbf{F}_{\mathrm{BB}}[n]$. Therefore, for k^{th} iteration we have

$$\mathbf{F}_{\mathrm{BB}_{k+1}}^{(i)}[n] \triangleq \min_{\mathbf{F}_{\mathrm{BB}}^{(i)}[n]} \|\mathbf{F}_{\mathrm{opt}}^{(i)}[n] - \mathbf{F}_{\mathrm{RF}_k}^{(i)}\mathbf{F}_{\mathrm{BB}_k}^{(i)}[n]\|_F^2, \tag{30}$$

$$\mathbf{F}_{\mathrm{RF}_{k+1}}^{(i)} \triangleq \sum_{n=1}^{N} \min_{\mathbf{F}_{\mathrm{RF}}^{(i)}} \|\mathbf{F}_{\mathrm{opt}}^{(i)}[n] - \mathbf{F}_{\mathrm{RF}_k}^{(i)}\mathbf{F}_{\mathrm{BB}_k}^{(i)}[n]\|_F^2. \tag{31}$$

By invoking the classic least-squares solution, we have

$$\mathbf{F}_{\mathrm{BB}_{k+1}}^{(i)}[n] = \left(\mathbf{F}_{\mathrm{RF}_k}^{H^{(i)}}\mathbf{F}_{\mathrm{RF}_k}^{(i)}\right)^{-1}\mathbf{F}_{\mathrm{RF}_k}^{H^{(i)}}\mathbf{F}_{\mathrm{opt}}^{(i)}[n], \tag{32}$$

(33)

$$\mathbf{F}_{\mathrm{RF}_{k+1}}^{(i)} = \left(\sum_{n=1}^{N}\mathbf{F}_{\mathrm{opt}}^{(i)}[n]\mathbf{F}_{\mathrm{BB}_{k+1}}^{H(i)}[n]\right) \times \tag{34}$$

$$\left(\sum_{n=1}^{N}\mathbf{F}_{\mathrm{BB}_{k+1}}^{(i)}[n]\mathbf{F}_{\mathrm{BB}_{k+1}}^{H(i)}[n]\right)^{-1}. \tag{35}$$

Note that the analog beamforming matrix $\mathbf{F}_{\mathrm{RF}_{k+1}}^{(j)}$ derived in (35) is unconstrained. Therefore, we invoke Proposition 2 [49] to compute the constrained $\mathbf{F}_{\mathrm{RF}_{k+1}}^{(j)}$, where each entity of the matrix has a constant modulus.

Proposition 2 [49]: Let us assume that $\mathbf{A} \in \mathbb{C}^{N_t \times N_s}$ is expressed as $\mathbf{A}(a,b) = |\mathbf{A}(a,b)|e^{j\angle\mathbf{A}(a,b)}, \forall a,b$. Furthermore, let $\mathcal{S} = \{\mathbf{B} \in \mathbb{C}^{N_t \times N_s} \mid \ |\mathbf{B}(a,b)| = 1/\sqrt{N_t}, \forall a,b\}$ and

$$\mathbf{U}' = f(\mathbf{A}) \triangleq \arg\min_{\mathbf{U}\in\mathcal{S}} \|\mathbf{A} - \mathbf{U}\|_{\mathrm{F}}^2, \tag{36}$$

Then, $\mathbf{U}' = \frac{1}{\sqrt{N_t}}e^{\angle\mathbf{A}(a,b)}$.

Proof: We refer readers to [49].

Thus, we have $\mathbf{F}_{\mathrm{RF}_{k+1}}^{(i)} = f(uv)$, where $u = \left(\sum_{n=1}^{N}\mathbf{F}_{\mathrm{opt}}^{(i)}[n]\mathbf{F}_{\mathrm{BB}_{k+1}}^{H(i)}[n]\right)$ and $v = \left(\sum_{n=1}^{N}\mathbf{F}_{\mathrm{BB}_{k+1}}^{(i)}[n]\mathbf{F}_{\mathrm{BB}_{k+1}}^{H(i)}[n]\right)^{-1}$. In other words, the magnitude of each entry in $\mathbf{F}_{\mathrm{RF}_{k+1}}^{(i)}$ is set to $1/\sqrt{N_t^{(i)}}$ while the phase component is set to $\angle\mathbf{F}_{\mathrm{RF}_{k+1}}^{(i)}$. A similar procedure is followed to obtain $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}[n]$.

The pseudo-code of the hybrid decomposition for developed for FD OFDM systems is presented in Algorithm 2.

Remark: It is important to emphasize that the proposed transceiver design relies on the knowledge of the CSI at the transmitter, which is estimated at the receiver using pilots and fedback to the transmitter. However, owing to the mobility of the users, the CSI relayed to the transmitter in the uplink would become outdated, as mentioned in Sec. I. In other words, the channel estimate acquired from the receiver would no longer be accurate for the current channel conditions. Furthermore, the multiple local oscillators used at the transmitter would further contribute the Doppler shift, hence increasing the inaccuracies in the CSI at the transmitter. In other

Algorithm 2 Hybrid Decomposition for FD OFDM system

1: Invoke algorithm 1 to obtain $\mathbf{F}_{\text{opt}}^{(i)}$
2: Set $\mathbf{F}_{\text{RF}_k}^{(i)} = \angle \mathbf{F}_{\text{opt}}^{(i)}$
3: Compute $\mathbf{F}_{\text{BB}_{k+1}}^{(i)}[n] \leftarrow \left(\mathbf{F}_{\text{RF}_k}^{H(i)} \mathbf{F}_{\text{RF}_k}^{(i)}\right)^{-1} \mathbf{F}_{\text{RF}_k}^{H(i)} \mathbf{F}_{\text{opt}}^{(i)}[n]$
4: Compute $\mathbf{F}_{\text{RF}_{k+1}}^{(i)} \leftarrow$
$\left(\sum_{n=1}^{N} \mathbf{F}_{\text{opt}}^{(i)}[n] \mathbf{F}_{\text{BB}_{k+1}}^{H(i)}[n]\right) \left(\sum_{n=1}^{N} \mathbf{F}_{\text{BB}_{k+1}}^{(i)}[n] \mathbf{F}_{\text{BB}_{k+1}}^{H(i)}[n]\right)^{-1}$
5: Then, set $\mathbf{F}_{\text{RF}_{k+1}}^{(i)} = 1/\sqrt{N_t^{(i)}} \angle \mathbf{F}_{\text{RF}_{k+1}}^{(i)}$
6: Continue from Step 3 until convergence.

words, regardless of how sophisticated channel estimation is used at the receiver, the CSI would become outdated at the transmitter. We also note that the loss in the accuracy of the CSI is also dependent on the mobility (speed) of the users. Therefore, it is of crucial importance to predict the channel at the transmitter from the CSI estimate acquired at the receiver.

In the next section, we propose a learning-aided channel prediction, where the current channel estimate is used to predict the future channel information *at the transmitter* for HBF in order to enhance the system's performance by employing the algorithm proposed in Sec. III.

V. LEARNING-AIDED CHANNEL PREDICTION

Typically, in the FDD systems, the channel is estimated at the receiver in the downlink and fedback to the transmitter in the uplink, where the transmitter leverages the CSI to enhance the system performance. However, the CSI acquired at the transmitter is imperfect owing to the estimation errors introduced by the pilot-assisted CSI estimate as well as by the non-linear components in the hardware. The errors in the CSI at the transmitter would be exacerbated when the ubiquitous phenomenon of channel aging is considered [27]. To elaborate, the channel aging is a phenomenon that arises due to the time-varying nature of the channel. As a result the channel information arriving at the transmitter would become outdated. In this scenario, any signal processing performed at the transmitter relying on the outdated CSI would result in performance losses. Therefore, to circumvent the problem of channel aging, it is pertinent to predict the channel from the current estimates — where in this paper we resort to a learning approach. In other words, we use the pilot-based channel estimate at the receiver, which would become outdated by the time it reaches the transmitting terminal, to predict the channel estimate at the transmitter. Then the transmitter employs the proposed HBF using the predicted channel estimate. More explicitly, in this section, we commence our discussion with some preliminaries on learning methods, namely on radial basis neural networks and then followed by the discussion of channel prediction, where we invoke on radial basis neural network.

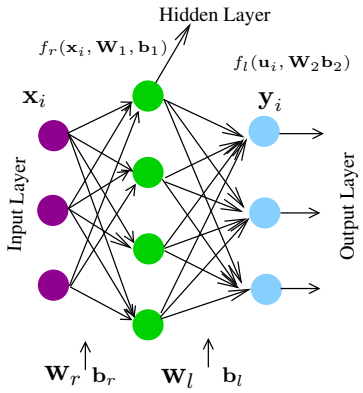


FIGURE 6: *A typical radial basis neural network.*

A. NEURAL NETWORK PRELIMINARIES

Artificial Neural Networks (ANN) inspired by the structural and functional aspects of biological neural networks constitute one of the main pillars of machine learning [50]. A typical ANN consists of interconnected layers of elements which process the input information in parallel to arrive at a decision. This process is akin to the human nervous system. Hence, the elements in the layers are termed as neurons. A neural network having three layers is shown in Fig. 6, where the first and last layers are the input and output layers, respectively, while the layer between them is referred to as the hidden layer. Furthermore, a network is said to be a deep neural network, if the number of hidden layers is more than one [33]. Fig. 6 shows a neural network having one hidden layer. To elaborate further, each layer of the ANN has a weight matrix $\mathbf{W}$ and a bias vector $\mathbf{b}$ that determines the performance of the network. More explicitly, in our example of Fig. 6, $\mathbf{W}_r, \mathbf{W}_l$ are the weight matrices, while $\mathbf{b}_r, \mathbf{b}_l$ are the bias vectors, and each neuron in the ANN is activated by a function $f(.)$ called score (or activation) function. The choice of the activation function used defines the type of the neural network. We note that there is no constraint on the type of activation function other than the differentiability of the function [39], [50]. The choice of the activation function depends on the type of the problem. In this paper, we choose a radial basis function as the activation function because of its ability to infer and reproduce any function, which is appealing for our channel prediction scenario. Therefore, this neural network is referred to as a radial basis neural network. The learning strategy in an ANN is a twofold process, which consists of a training phase and a testing phase. In the training phase, a known set of input samples and its corresponding output samples are used for computing the weight vectors of the network. This is reminiscent of supervised learning, where the weight matrices and bias vectors for the network are designed under supervision. On the other hand, in the testing phase, the network applies the weights obtained in the training phase to the real-time data to evaluate and predict the outcome.

B. CHANNEL PREDICTION

Let us we now focus our attention on the channel prediction. In learning-aided channel prediction design, the training is performed to design the weights using the outdated channel estimate and the actual channel. A pictorial illustration of outdated, of current and of predicted channel estimates is shown in Fig. 7.

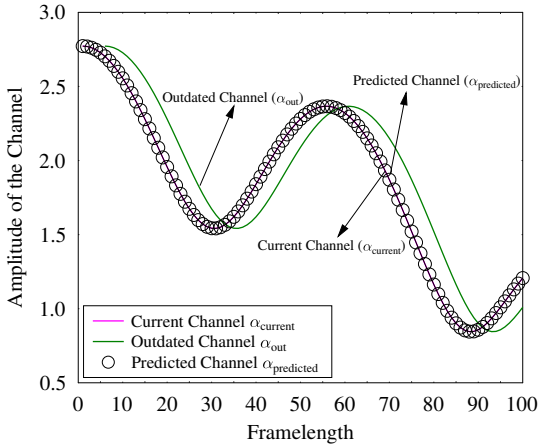


FIGURE 7: *Pictorial representation of outdated, current and predicted channel estimates.*

More explicitly, the outdated channel samples and the corresponding actual (current) channel samples are used as the input and output of the network, respectively. In other words, the outdated channel is fed to the input layer, whose initial weights are randomly assigned from the distribution $\mathcal{N}(0,1)$ to obtain scores for the hidden layer, where the radial basis function $f_r(.)$ is employed. The score of neuron, v in the hidden layer of the radial basis function is given by [39]

$$o_v = f_r(\|\mathbf{w}_r - \mathbf{x}_i\| \, \mathrm{b}_r), \tag{37}$$

where $\mathbf{w}_r, \mathbf{x}_i$ and b_v are the weight vector, the input and the bias, respectively, while $f(.)$ is the radial basis function expressed as [39]

$$f_r(n) = e^{-n^2}. \tag{38}$$

The score computed using (37) serves as the input to its succeeding layer, where a linear function is applied on the score obtained, which is expressed as

$$\mathbf{y}_i = \mathbf{W}_l \mathbf{o}_r + \mathbf{b}_l. \tag{39}$$

The score $\mathbf{y}_i$ obtained using (39) in the output layer is the predicted channel, which is then compared to the known actual channel to compute the error. The error between the predicted and the actual channel is evaluated using a loss function. The loss function used in this design is the mean square error given by [39]

$$L = \frac{1}{\mathcal{S}} \sum_{i=1}^{\mathcal{S}} \left\| \mathbf{y}_i - \mathbf{y}_i^t \right\| + \rho_1 \|\mathbf{W}_r\|_2^2 + \rho_2 \|\mathbf{W}_l\|_2^2, \tag{40}$$

where $\mathcal{S}$ is the total number of training samples, $\mathbf{y}^t$ is the current channel, while ρ_1, ρ_2 are the regularization factors used to avoid over-fitting [39]. Having obtained the loss function, the weights are then designed to minimize the loss function using back-propagation. In back-propagation, the gradient of the loss function with respect to weights and biases is computed for each layer. A more detailed discussion on back-propagation is presented by Chauvin *et al.* in [51]. This process is continued until convergence is reached. In our design, $\mathbf{x}_i$ is the outdated channel vector α_{out} and $\mathbf{y}_i$ is the predicted channel vector α_{predict}. More explicitly, α_{out} is the channel learned at the receiver during pilot transmission which would become outdated at the transmitter because of the time-varying nature of the channel.

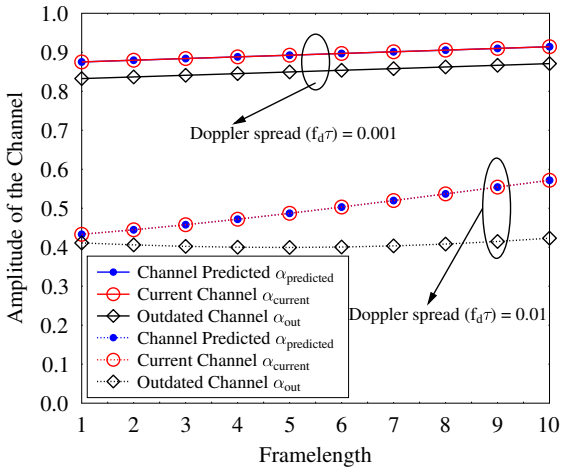
Let us now again consider the channel presented in (2) in the face of channel aging, where the AoA and AoD change slowly when compared to the small-scale fading coefficient α, which is the instantaneous CSI. More explicitly, α evolves with time according to the Jakes' model, whose autocorrelation is given by the zero order Bessel function of the first kind [52]. The correlation coefficient is given by

$$\zeta = J_\circ(2\pi f_d \tau),$$

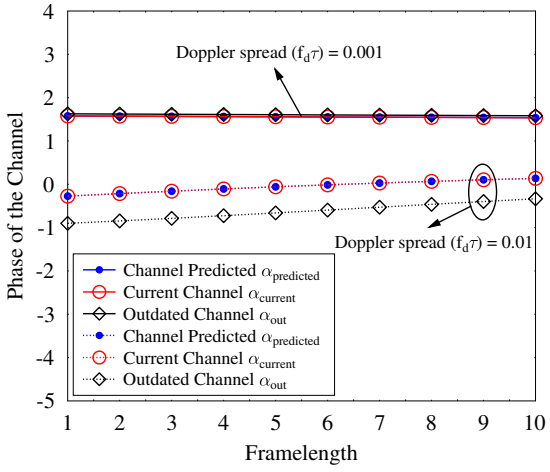
where f_d is the maximum Doppler frequency and τ is the sample time.

It is important to emphasize that the pilot-assisted channel estimate at the receiver which is outdated because of the mobility of the users is $\alpha_{\text{out}} = \alpha(\tau - 1)$, while the current channel estimate is $\alpha_{\text{current}} = \alpha(\tau)$. Note that during the training phase of the learning α_{out} and α_{current} are used as the training samples to design the weights of the neural network as discussed in the previous section. Having designed the training weights, the transmitter predicts the current channel estimate α_{predict} from the outdated channel estimate α_{out} in real time.

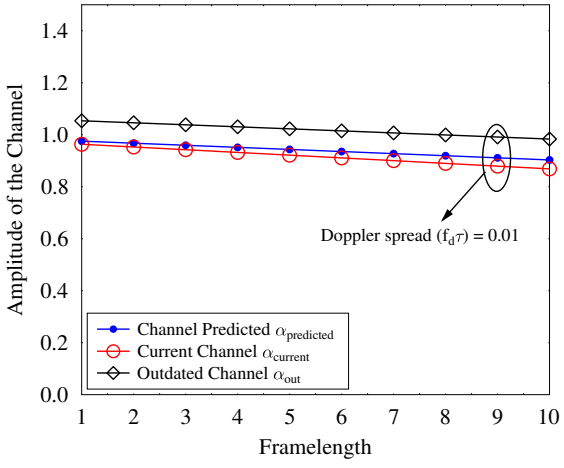
Fig. 8 shows the channel response in for Doppler spreads of 0.001 and 0.01. More explicitly, Fig. 8(a) and Fig. 8(b) characterizes the amplitude and phase of the channel response, where 200 training samples are used to design the training weights of the neural network. It can be seen from Figures 8(a) and 8(b) that the learning-aided channel prediction has faithfully reproduced the current channel from the outdated channel both for a Doppler spread as high as 0.01 and for Doppler spread as low as 0.001. It is important to emphasize that the fidelity of the learning assisted channel prediction relies on the number of training samples used to design the training weights. This is evident from Fig. 8(c), where the predicted channel is a bit off the mark from the



(a) *Amplitude of the CSI with 200 training samples.*

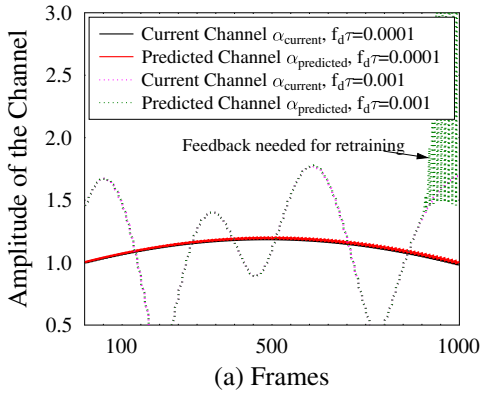


(b) *Phase of the CSI with 200 training samples.*



(c) *Amplitude of the CSI with 100 training samples.*

FIGURE 8: *Amplitude and phase of the outdated channel, of the current channel and of the predicted channel.*



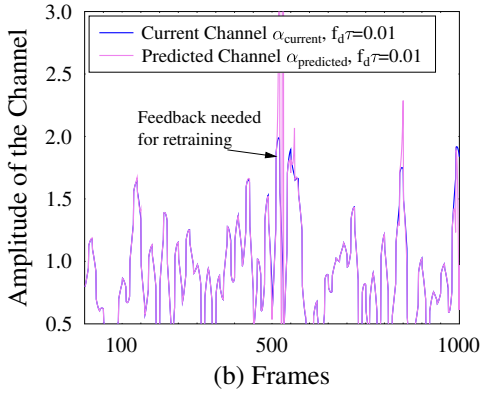
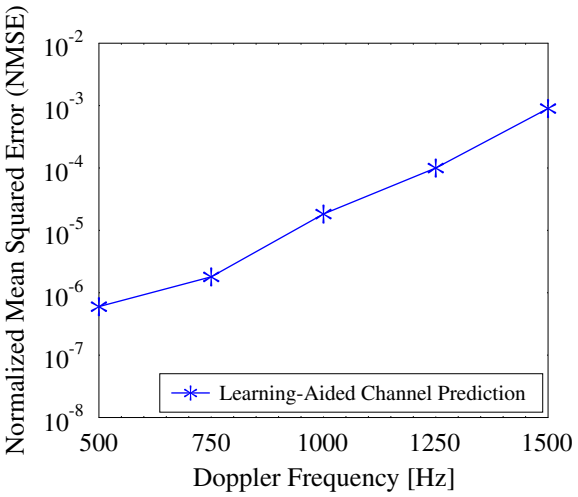


FIGURE 9: *Amplitude of the predicted channel and the current channel for different Doppler spread values.*

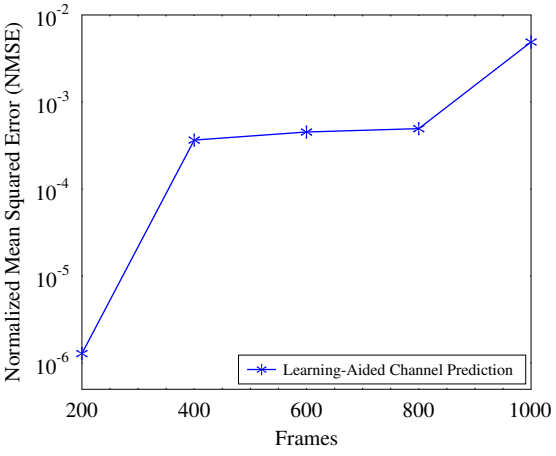
current channel, when only 100 training samples are used[4].

Fig. 9 shows the amplitude of the predicted channel and the current channel for different Doppler spread values. It is evident from the Fig. 9(a) that for a normalized Doppler spread as low as 0.0001, the neural network weights evaluated during training at the start of the frame are capable of predicting the channel even after 1000 frames. In other words, the CSI feedback from the receiver, which is estimated using pilots, is not needed at the transmitter. However, it can be also seen from Fig. 9(a) that as the Doppler spread is increased to 0.001, the weights of the neural network evaluated during training at the transmitter becomes unable to faithfully reproduce the channel after 800 frames. Therefore, feedback from the receiver becomes necessary for the neural network retraining after around 800 frames of transmission. Observe from Fig. 9(b) that as the Doppler spread is increased to 0.01, the number of frames transmitted using the same weights decreases. In other words, more frequent feedback from the receiver becomes necessary. Therefore, this implies that the overhead required for retraining the neural network depends on the Doppler spread.

[4]The number of training samples required for faithful channel prediction would also depend on the Doppler spread.



(a) Normalized mean squared error between the predicted channel and the accurate channel at different Doppler frequencies.



(b) Normalized mean squared error between the predicted channel and the accurate channel at different number of frames for a Doppler spread of 0.001.

FIGURE 10: Normalized mean squared error between the predicted channel and the accurate channel.

Fig. 10 shows the normalized mean squared error (NMSE) between the predicted channel and the current channel. More explicitly, Fig. 10(a) shows the NMSE at different Doppler frequencies. It can be noted from Fig. 10(a) that the NMSE is substantially lower between the predicted and current channel. In other words, the ANN is capable of reproducing the current channel with high accuracy. Although the NMSE increases with the Doppler frequency, it still is on the order of $\mathcal{O}(10^{-3})$. On the other hand, Fig. 10(b) shows the NMSE between the predicted and current channels when transmitting at different number of frames between a pair of training events. It can be seen from Fig. 10(b) that NMSE becomes insignificant for a low number of frames, while it is on the order of $\mathcal{O}(10^{-3})$ for say 800 frames. More explicitly, the physical significance of this is that the training weights designed before the transmission of the first frame can be applied without retraining for upto 800-1000 frame durations depending on the required NMSE.

VI. SIMULATION RESULTS

In this section, we present our simulations characterizing the sum rate of the system relying on the proposed hybrid precoder design. We performed Monte Carlo simulations for studying the performance difference between the proposed and the EBF designs. Our simulation parameters are summarized in Table 2. The angles of arrivals and departures obey Laplacian distribution with an angular spread of 5°, with the mean angles uniformly distributed between $(0, 60^\circ)$. In these simulations, we have used the interference-to-noise ratio (INR) INR_{SI} as a measure of SI level, where we aim for reducing the SI to the level of the noise floor, while the signal-to-interference ratio (SIR) SIR_{MI} is used as a measure of MI levels. Note that in all our simulation results characterizing the sum-rate performance, the channel employed is the predicted channel for all users. Furthermore, the channel described in (2) is employed, which is first estimated using pilots at the receiver and then fedback to the transmitter. Exploiting the sparsity of the channel by following the approach in [53], the pilot overhead can be minimized. The effective sum rate of the system when taking into account the overhead is $f_e = (1 - f_p)f_r$, where f_p is the pilot percentage, f_r is the capacity and f_e is the effective rate. Upon the reception of the pilot-assisted channel estimate, which becomes outdated because of the user's mobility, the transmitter predicts the channel by invoking the ANN weights described in Sec. V.

TABLE 2: System parameters.

Parameters	*Values*
N (sub-carriers)	32
N_t	128
N_r	32
N_t^{RF}	2
N_r^{RF}	2
N_s	2
θ	Laplacian distributed
ϕ	Laplacian distributed

Fig. 11 studies the BER performance of the system for Doppler spreads of 0.01 and 0.005. More particularly, it characterizes the BER performance of the channel with prediction, of the current channel and of the outdated channel. Furthermore, in Fig. 11 we study the BER performance using channel predicted with both 100 and 200 training samples for designing the training weights of the learning network. In this setting, we have set the SI and MI to 3 dB. It can be seen from Fig. 11 that for a Doppler spread as high as 0.01 and 0.005, the BER of the system relying on channel prediction is close to that of the system using 100 training samples, while it performs similarly to the current channel,

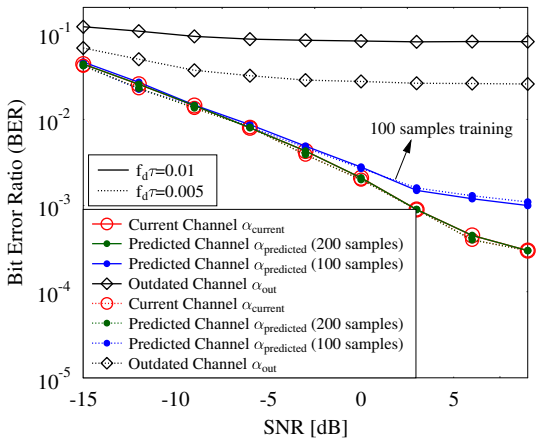


FIGURE 11: *Characterizing the BER performance for Doppler spreads of 0.01 and 0.005 with predicted channel, outdated channel and current channel. In this simulation, BER performance is studied using predicted channel with both 100 and 200 training samples for designing the weights of the neural network. Furthermore, in this setting SI and MI is set to 3 dB.*

when the number of training samples is increased to 200. On the other hand, the outdated channel produces an error floor for both Doppler spreads.

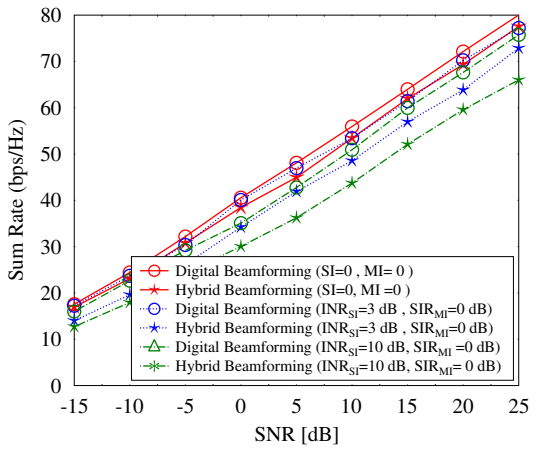


FIGURE 12: *Characterizing the sum rate performance of the proposed design having digital solution and of the proposed design having hybrid solution.*

Fig. 12 shows the sum rate performance of the proposed design relying on both the digital and on the hybrid beamforming solutions. It can be seen from Fig. 12 that the hybrid beamforming solution performs closely to the digital solution, provided that the interference is low. However, when the INR_{SI} is set to 3 dB, there is around 2.5 dB loss at the sum rate of 50 bps/Hz, while it is around 4 dB when the INR_{SI} is set to as high as 10 dB.

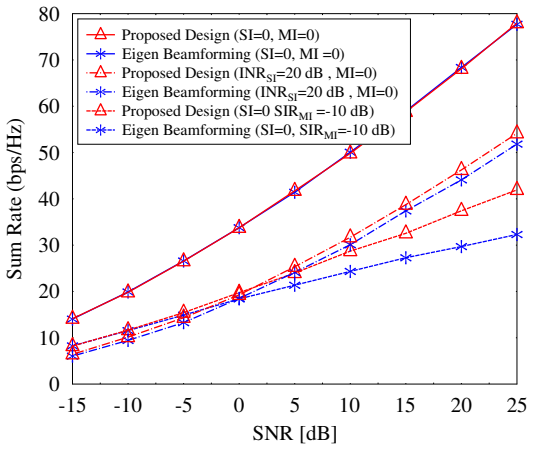


FIGURE 13: *Characterizing the sum rate performance of our FD HBF design and of the EBF for different interference configurations. The parameters in Table 2 are used for simulations. In this simulation, channel with prediction is used.*

Fig. 13 shows the sum rate performance of the proposed design and of the EBF design. The simulations are carried out for three interference configurations, namely in the presence of i) no interference; ii) multi-user interference; iii) self-interference. It can be seen in Fig. 13 that when there is no MI or SI, the performances of the proposed and of the EBF designs are similar. This is because, in the absence of interference, our proposed hybrid precoder design reduces to the EBF solution, which is optimal in this setting. However, when the SI-to-noise ratio (INR_{SI}) is 20 dB, the performance gap between the two designs becomes clear. Furthermore, when the signal-to-MI ratio (SIR_{MI}) is -10 dB, the proposed design outperforms the EBF design significantly, and the performance gap between the two designs is as high as 10 dB.

To study the sum rate of the system by accounting for the pilot overhead involved in the channel estimation at the receiver, Fig. 14 shows the sum rate of the design parameterized by the pilot percentage. We note from the Fig. 14 that there is a loss of around 1.5 bits when the pilot percentage f_p is set to 5%, while exploiting the sparsity of the mmWave channel [53]. The same holds for the rest of the simulations in different configurations.

Fig. 15 shows the sum rate performance of both the proposed design and of the EBF for a given INR_{SI} of 20 dB. It is evident from Fig. 15 that the proposed technique outperforms the design relying on EBF by about 4 dB for SIR_{MI} =-5 dB. Moreover, when the SIR_{MI} is set to -10 dB, the EBF degrades the performance. By contrast, the performance of

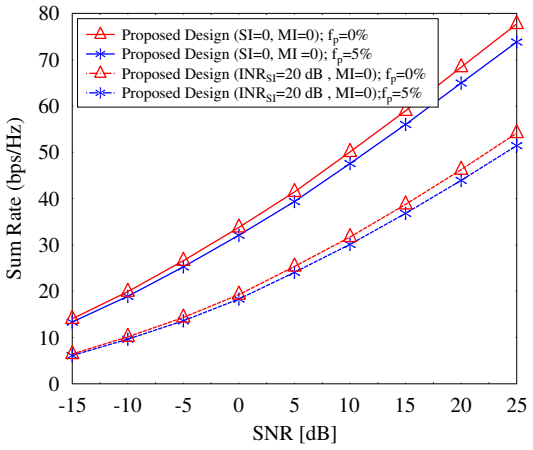


FIGURE 14: *Characterizing the sum rate performance of our FD HBF design when pilot overhead is considered. The parameters in Table 2 are used for simulations. In this simulation, channel with prediction is used.*

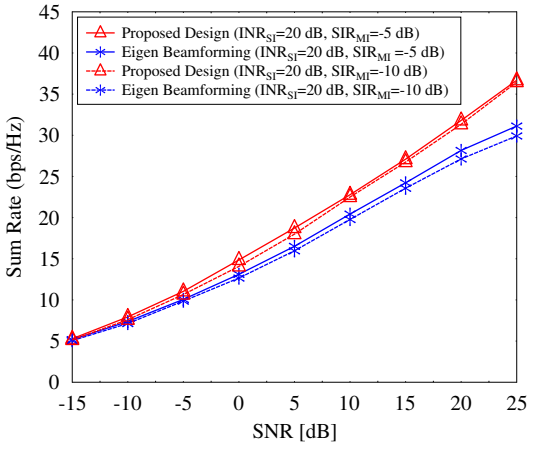


FIGURE 15: *Characterizing the sum rate performance of our FD HBF design and of the EBF for a given SI power with different MI levels. The parameters in Table 2 are used for simulations. In this simulation, channel with prediction is used.*

our proposed technique does not degrade as much as the EBF design and has a similar sum rate to that at SIR_{MI} =-5 dB.

To further understand the behavior of our design, Fig. 16 plots the sum rate of our design in comparison to that of the EBF for a fixed SIR_{MI} of -10 dB with varying INR_{SI}. It can be seen from Fig. 16 that our proposed design performs about 5 dB better than the EBF design at a sum rate of 30 bps/Hz. It is also worth observing the trajectory of the EBF, indicating that slope of the sum rate starts to reduce as it approaches

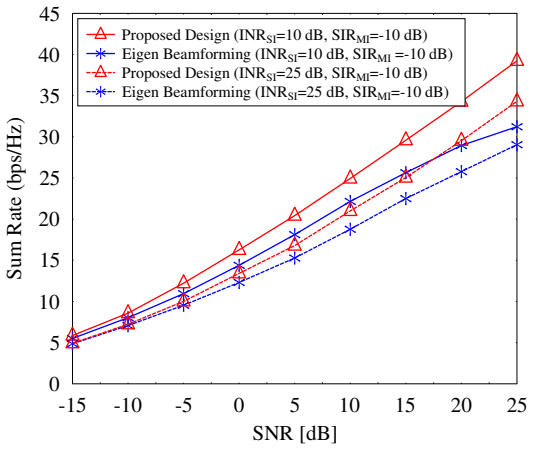


FIGURE 16: *Characterizing the sum rate performance of our FD HBF design and of the EBF for a given MI power with different SI levels. The parameters in Table 2 are used for simulations. In this simulation, channel with prediction is used.*

high SNRs for both INR_{SI}=10 dB and 25 dB. It is instructive to note that despite having high INR_{SI} of 25 dB, our proposed design outperforms the EBF by a significant margin.

VII. COMPLEXITY

In this section, we discuss the complexity of the proposed design. More particularly, we present the complexity in terms of the number of computations involved in both the digital precoder/combiner design and the channel prediction. Having obtained the co-variance matrix $\mathbf{R}$ in (14), the number of complex multiplications required to obtain the digital precoder and combiner matrix for $N_t = N_r = N_s = N$ would be $\mathcal{O}(N^3)$. More explicitly, for a different number of transmit and receiver antennas, the total number of computations required is shown in Table 3.

TABLE 3: Computations required for digital solution.

Design	*Computations*
Proposed Digital Solution	$N_r^2 N_t + 3N_r N_t N_s + N_s N_r^2 + 2N_r N_s^2$
SVD	$N_r^2 N_t + N_t^2 N_r + N_r^3$

On the other hand, the complexity of the radial basis neural network used for channel prediction involves two phases: offline and online. During the offline phase, which is the training phase, the weights of the neural network are designed and stored in memory. If the neural network has n neurons, then the number of computations required to train the weights is $\mathcal{O}(n^4) + \mathcal{O}(n^3)$. We note that the design of weights is carried out only once and stored in the memory during the training phase. Furthermore, the weights are designed at the BS, which may be equipped with high computational resources for channel prediction. However,

during the online stage, the weights are applied to generate the output, where the number of computations is $\mathcal{O}(n^3)$.

For example, in our simulations, where we have set $N_t = 128, N_r = 32, N_s = 2$, the number of complex multiplications required for our propose design is $157,952$, while for SVD it is $688,128$. Furthermore, in our neural network setting, we have set the number of hidden layers to 1 and the number of neurons in the hidden layer to 250.

VIII. CONCLUSIONS

In this paper, we first proposed a joint precoder and combiner design for full duplex K-user MIMO-OFDM interference channels, where we aimed for minimizing both the residual SI and the MI, followed by an iterative hybrid decomposition technique for OFDM systems. Then, for the same system, we proposed a learning-aided channel prediction relying on a radial basis neural network, where we have shown by simulation results that given sufficient training, learning-assisted channel prediction is capable of faithfully reproducing the current channel. Furthermore, we showed by simulations that our proposed joint hybrid precoder and combiner design outperforms the popular eigen-beamforming technique by about 5 dB for a 128×32.

APPENDIX A
PROOF OF PROPOSITION 1

Proof: We aim to solve the problem (15) by invoking the Lagrangian function expressed as (41). With slight abuse of notation we drop the frequency index [n]. It should be noted that the solution obtained is for a sub-carrier. Similar approach should be followed for all the sub-carriers.

$$\mathcal{L}(\mathbf{W}^{(i)}, z) = \left(\mathbf{W}^{H^{(i)}}\mathbf{R}^{(i)}\mathbf{W}^{(i)}\right) + z\left(\mathbf{W}^{H^{(i)}}\mathbf{H}_{ii}\mathbf{F}^{(i)} - \mathbf{I}_{N_s}\right) \tag{41}$$

Then, the Lagrangian equations are given as

$$\nabla_{\mathbf{W}_{\text{opt}}^{H(i)}}\mathcal{L} = 0 \tag{42}$$

$$z^*\left(\mathbf{W}_{\text{opt}}^{H(i)}\mathbf{H}_{ii}\mathbf{F}^{(i)} - \alpha\mathbf{I}_{N_s}\right) = 0 \tag{43}$$

Explicitly (42) can be written as,

$$\nabla_{\mathbf{W}_{\text{opt}}^{H(i)}}\text{Tr}\left(\mathbf{W}_{\text{opt}}^{H(i)}\mathbf{R}^{(i)}\mathbf{W}_{\text{opt}}^{(i)}\right) + z^*\nabla_{\mathbf{W}_{\text{opt}}^{H(i)}}\left(\mathbf{W}_{\text{opt}}^{H^{(i)}}\mathbf{H}_{ii}\mathbf{F}^{(i)} - \mathbf{I}_{N_s}\right) = 0 \tag{44}$$

where ∇ is the gradient operator and z is the Lagrangian coefficient.

By applying the derivative, we achieve

$$\mathbf{R}^{(i)}\mathbf{W}_{\text{opt}}^{(i)} + z\mathbf{H}_{ii}\mathbf{F}^{(i)} = 0 \tag{45}$$

$$\mathbf{W}_{\text{opt}}^{(i)} = -\mathbf{R}^{(i)^{-1}}\mathbf{H}_{ii}\mathbf{F}^{(i)}z. \tag{46}$$

Upon substituting $\mathbf{W}_{\text{opt}}^{(i)}$ in (43), we get

$$\left(-\mathbf{R}^{(i)^{-1}}\mathbf{H}_{ii}\mathbf{F}^{(i)}z\right)^H\mathbf{H}_{ii}\mathbf{F}^{(i)} = \alpha\mathbf{I}_{N_s} \tag{47}$$

$$z = -\alpha\left((\mathbf{H}_{ii}\mathbf{F}^{(i)})^H\mathbf{R}^{(i)^{-1}}\mathbf{H}_{ii}\mathbf{F}^{(i)}\right)^{-1}. \tag{48}$$

Hence,

$$\mathbf{W}_{\text{opt}}^{(i)} = \alpha\mathbf{R}^{(i)^{-1}}\mathbf{H}_{ii}\mathbf{F}^{(i)}\left((\mathbf{H}_{ii}\mathbf{F}^{(i)})^H\mathbf{R}^{(i)^{-1}}(\mathbf{H}_{ii}\mathbf{F}^{(i)})\right)^{-1}\square \tag{49}$$

REFERENCES

[1] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," IEEE Commun. Surveys Tuts., vol. 20, no. 2, pp. 870–913, Secondquarter 2018.

[2] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," Proc. IEEE, vol. 104, no. 7, pp. 1369–1409, July 2016.

[3] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," ACM SIGCOMM, vol. 43, no. 4, pp. 375–386, 2013.

[4] E. Everett, M. Duarte, C. Dick, and A. Sabharwal, "Empowering full-duplex wireless communication by exploiting directional diversity," in Proc. ASILOMAR, 2011, pp. 2002–2006.

[5] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in Proc. ASILOMAR, 2010, pp. 1558–1562.

[6] A. Sahai, G. Patel, C. Dick, and A. Sabharwal, "On the impact of phase noise on active cancelation in wireless full-duplex," IEEE Trans. Veh. Technol., vol. 62, no. 9, pp. 4494–4510, Nov 2013.

[7] X. Liu, Z. Xiao, L. Bai, J. Choi, P. Xia, and X.-G. Xia, "Beamforming based full-duplex for millimeter-wave communication," Sensors, vol. 16, no. 7, 2016. [Online]. Available: http://www.mdpi.com/1424-8220/16/7/1130

[8] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," IEEE J. Sel. Areas Commun., vol. 32, no. 9, pp. 1637–1652, Sept 2014.

[9] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," IEEE Trans. Wireless Commun., vol. 11, no. 12, pp. 4296–4307, December 2012.

[10] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," IEEE Trans. Signal Process., vol. 59, no. 12, pp. 5983–5993, Dec 2011.

[11] I. Krikidis, H. A. Suraweera, P. J. Smith, and C. Yuen, "Full-duplex relay selection for amplify-and-forward cooperative networks," IEEE Trans. Wireless Commun., vol. 11, no. 12, pp. 4381–4393, December 2012.

[12] L. Li, C. Dong, L. Wang, and L. Hanzo, "Spectral-efficient bidirectional decode-and-forward relaying for full-duplex communication," IEEE Trans. Veh. Technol., vol. 65, no. 9, pp. 7010–7020, Sept 2016.

[13] C. T. Lin, F. S. Tseng, and W. R. Wu, "MMSE transceiver design for full-duplex MIMO relay systems," IEEE Trans. Veh. Technol., vol. 66, no. 8, pp. 6849–6861, Aug 2017.

[14] S. Wang, Y. Liu, W. Zhang, and H. Zhang, "Achievable rates of full-duplex massive MIMO relay systems over Rician fading channels," IEEE Trans. Veh. Technol., vol. 66, no. 11, pp. 9825–9837, Nov 2017.

[15] D. W. K. Ng, E. S. Lo, and R. Schober, "Dynamic resource allocation in MIMO-OFDMA systems with full-duplex and hybrid relaying," IEEE Trans. Commun., vol. 60, no. 5, pp. 1291–1304, May 2012.

[16] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex MIMO relaying: Achievable rates under limited dynamic range," IEEE J. Sel. Areas Commun., vol. 30, no. 8, pp. 1541–1553, September 2012.

[17] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," IEEE Trans. Wireless Commun., vol. 13, no. 2, pp. 680–694, February 2014.

[18] K. Satyanarayana, M. El-Hajjar, P. Kuo, A. Mourad, and L. Hanzo, "Hybrid beamforming design for full-duplex millimeter wave communication," IEEE Trans. Veh. Technol., vol. 68, no. 2, pp. 1394–1404, Feb 2019.

[19] Y. Niu et al., "A survey of millimeter wave (mmwave) communications for 5G: Opportunities and challenges," CoRR, vol. abs/1502.07228, 2015.

[20] S. Han et al., "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," IEEE Commun. Mag., vol. 53, no. 11, pp. 186–194, 2015.
[21] K. Satyanarayana, M. El-Hajjar, P. Kuo, A. Mourad, and L. Hanzo, "Dual-function hybrid beamforming and transmit diversity aided millimeter wave architecture," IEEE Trans. Veh. Technol., vol. 67, no. 3, pp. 2798–2803, March 2018.
[22] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," IEEE Trans. Commun., vol. 64, no. 5, pp. 1801–1818, May 2016.
[23] X. Yu, J. C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," IEEE J. Sel. Topics Sig. Proc., vol. 10, no. 3, pp. 485–500, April 2016.
[24] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmwave MIMO systems," IEEE Trans. Wireless Commun., vol. 16, no. 5, pp. 2907–2920, May 2017.
[25] L. Kong, S. Han, and C. Yang, "Wideband hybrid precoder for massive MIMO systems," in Proc. GlobalSIP, Dec 2015, pp. 305–309.
[26] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmwave OFDM large-scale antenna arrays," IEEE J. Sel. Areas Commun., vol. 35, no. 7, pp. 1432–1443, July 2017.
[27] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," J. of Commun. and Networks, vol. 15, no. 4, pp. 338–351, Aug 2013.
[28] M. Vu and A. Paulraj, "MIMO wireless linear precoding," IEEE Signal Process. Mag., vol. 24, no. 5, pp. 86–105, Sept 2007.
[29] A. K. Papazafeiropoulos, "Impact of general channel aging conditions on the downlink performance of massive MIMO," CoRR, vol. abs/1605.07661, 2016. [Online]. Available: http://arxiv.org/abs/1605.07661
[30] L. Thiele, M. Olbrich, M. Kurras, and B. Matthiesen, "Channel aging effects in CoMP transmission: gains from linear channel prediction," in Proc. ASILOMAR, Nov 2011, pp. 1924–1928.
[31] W. Peng, M. Zou, and T. Jiang, "Channel prediction in time-varying massive MIMO environments," IEEE Access, vol. 5, pp. 23 938–23 946, 2017.
[32] A. Duel-Hallen, "Fading channel prediction for mobile radio adaptive transmission systems," Proc. IEEE, vol. 95, no. 12, pp. 2299–2313, Dec 2007.
[33] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks," CoRR, vol. abs/1710.02913, 2017. [Online]. Available: http://arxiv.org/abs/1710.02913
[34] Z. Wei, Y. Zhao, X. Liu, and Z. Feng, "DoA-LF: A location fingerprint positioning algorithm with millimeter-wave," IEEE Access, vol. 5, pp. 22 678–22 688, 2017.
[35] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," IEEE Internet of Things J., vol. 3, no. 6, pp. 1113–1123, Dec 2016.
[36] Z. E. Khatab, A. Hajihoseini, and S. A. Ghorashi, "A fingerprint method for indoor localization using autoencoder based deep extreme learning machine," IEEE Sensors Letters, vol. 2, no. 1, pp. 1–4, March 2018.
[37] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," IEEE Trans. Veh. Technol., vol. 66, no. 1, pp. 763–776, Jan 2017.
[38] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "Confi: Convolutional neural networks based indoor wi-fi localization using channel state information," IEEE Access, vol. 5, pp. 18 066–18 074, 2017.
[39] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). Berlin, Heidelberg: Springer-Verlag, 2006.
[40] L. Hanzo, M. Münster, B. Choi, and T. Keller, "OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting," May 2003. [Online]. Available: https://eprints.soton.ac.uk/258228/
[41] J. Lee and Y. H. Lee, "AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing," in Proc. ICC, June 2014, pp. 5838–5842.
[42] C. A. Balanis, Antenna Theory: Analysis and Design. Wiley-Interscience, 2005.
[43] J.-S. Jiang and M. A. Ingram, "Spherical-wave model for short-range MIMO," IEEE Trans. Commun., vol. 53, no. 9, pp. 1534–1541, Sept 2005.
[44] X. Xia, Y. Xu, K. Xu, D. Zhang, and W. Ma, "Full-duplex massive MIMO AF relaying with semiblind gain control," IEEE Trans. Veh. Technol., vol. 65, no. 7, pp. 5797–5804, July 2016.
[45] X. Xia, K. Xu, Y. Wang, and Y. Xu, "A 5G-enabling technology: Benefits, feasibility, and limitations of in-band full-duplex mMIMO," IEEE Veh. Technol. Mag., vol. 13, no. 3, pp. 81–90, Sep. 2018.
[46] X. Xia, K. Xu, D. Zhang, Y. Xu, and Y. Wang, "Beam-domain full-duplex massive MIMO: Realizing co-time co-frequency uplink and downlink transmission in the cellular system," IEEE Trans. Veh. Technol., vol. 66, no. 10, pp. 8845–8862, Oct 2017.
[47] X. Xia, D. Zhang, K. Xu, W. Ma, and Y. Xu, "Hardware impairments aware transceiver for full-duplex massive MIMO relaying," IEEE Trans. Signal Process., vol. 63, no. 24, pp. 6565–6580, Dec 2015.
[48] Y. Wu, Y. Gu, and Z. Wang, "Channel estimation for mmwave MIMO with transmitter hardware impairments," IEEE Commun. Lett., vol. 22, no. 2, pp. 320–323, Feb 2018.
[49] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," IEEE J. Sel. Topics Sig. Proc., vol. 10, no. 3, pp. 528–542, April 2016.
[50] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.
[51] Y. Chauvin and D. E. Rumelhart, Eds., Backpropagation: Theory, Architectures, and Applications. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1995.
[52] W. C. Jakes and D. C. Cox, Eds., Microwave Mobile Communications. Wiley-IEEE Press, 1994.
[53] Q. Qin, L. Gui, P. Cheng, and B. Gong, "Time-varying channel estimation for millimeter wave multiuser MIMO systems," IEEE Trans. Veh. Technol., vol. 67, no. 10, pp. 9435–9448, Oct 2018.

K. SATYANARAYANA (www.satyanarayana.xyz) received his B. Tech. degree in Electrical Engineering from Indian Institute of Technology Madras, India, in 2014. During Jul'14-Aug'15, he worked as a research assistant at Indian Institute of Science, Bangalore. Currently, Satya is a research scholar in Wireless Communications at the University of Southampton in liaison with InterDigital Europe, London, UK. His research interests include millimeter wave communications, HBF, with an emphasis on transceiver algorithms for wireless communication systems and multi-functional MIMO.

MOHAMMED EL-HAJJAR is an Associate Professor in the department of Electronics and Computer Science in the University of Southampton. He received his PhD in Wireless Communications from the University of Southampton, UK in 2008. Following the PhD, he joined Imagination Technologies as a design engineer, where he worked on designing and developing Imagination's multi-standard communications platform, which resulted in three patents. He is the recipient of several academic awards and has published a Wiley-IEEE book and in excess of 80 journal and conference papers. Mohammed's research interests include the design of intelligent and energy-efficient transceivers, MIMO, millimeter wave communications, cross-layer optimization for large-scale networks and Radio over fiber network design.

ALAIN A. M. MOURAD holds a PhD degree in Telecommunications from ENST Bretagne in France. He has over 15 years' experience in the wireless networks industry. He is currently leading the research and development of Next Generation Radio Access Networks at InterDigital International Labs (London, Berlin, Seoul). Prior to joining InterDigital, Dr. Mourad was a Principal Engineer at Samsung Electronics R&D (UK) and previously a Senior Engineer at Mitsubishi Electric R&D Centre Europe (France). Throughout his career, Dr. Mourad has been active in the research and standardization of recent communication networks (5G/4G/3G) and broadcasting systems (ATSC 3.0/DVB-NGH/DVB-T2). He has held various leadership roles in the industry, invented over 35 granted patents and several other patent applications, and authored over 50 peer-reviewed publications. He received the Inventor of the Year Award from Samsung Electronics R&D (UK) twice in 2012 and 2013, and in 2016 InterDigital Innovation Award for the "idea, creation, and execution of InterDigital Europe".

LAJOS HANZO (http://www-mobile.ecs.soton.ac.uk) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 100+ PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1600+ research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications please refer to http://www.mobile.ecs.soton.ac.uk Lajos has 30 000+ citations and an H-index of 68.