IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 October 2022, accepted 20 October 2022, date of publication 25 October 2022, date of current version 31 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3216834

RESEARCH ARTICLE

Channel Estimation and Prediction in a Pilot-Less Massive MIMO TDD Using Non-Coherent DMPSK

MANUEL J. LOPEZ-MORALES, (Member, IEEE), AND ANA GARCIA-ARMADA, (Senior Member, IEEE)

Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Spain

Corresponding author: Manuel J. Lopez-Morales (mjlopez@tsc.uc3m.es)

This work was supported in part by the European Union Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie European Training Network (ETN) TeamUp5G under Grant 813391, and in part by the Spanish National Project IRENE-EARTH (MINECO/AEI/FEDER/UE) under Grant PID2020-115323RB-C33.

ABSTRACT A novel time division duplex massive MIMO technique is proposed based on performing a pilot-less channel estimation in the uplink (UL) utilizing reconstructed differentially encoded data. Spatial multiplexing and differentially encoded data is applied both in the UL and in the downlink (DL). In this system, a reference signal is the first one of the differentially encoded streams in the UL and DL, and the pilots for data estimation are avoided while maintaining spatial multiplexing capabilities. To improve the channel estimation we propose to use a linear Wiener filter and we also propose different symbols placing strategies in an OFDM grid. We also propose a detection improvement of the UL data utilizing the predicted channels. We perform an analysis of the MSE of the blind channel estimation using the differentially encoded data and analyze the symbol-error-rate for both the UL and the DL when channel aging is considered. The analysis is corroborated via numerical results and the proposed scheme is shown to outperform its pilot-based counterpart.

INDEX TERMS Non-coherent, TDD massive MIMO, differential modulation, spatial multiplexing, blind channel estimation.

I. INTRODUCTION

In the last years, wireless networks have seen a tremendous development in order to cope with the technical requirements of several emerging innovative applications, such as 4K/8K video streaming, virtual/augmented reality (VR/AR), cloud gaming, real-time remote control, autonomous driving, and others. The fifth generation of mobile communications (5G) is already under deployment to cope with some of these needs [1], but its capabilities will not be sufficient for the most demanding applications. For this reason, beyond 5G and 6G novel techniques are currently under development to cope with the ever increasing demands of future wireless networks [2]. Higher data rates, lower latency, higher reliability, greater number of devices, lower energy consumption, greater network intelligence and others are some of the requirements foreseen for future networks. Apart from these requirements, another goal is to make the wireless networks work with more stringent conditions in scenarios such as those with fast varying channels (both in time and/or frequency) and/or low signal-to-noise ratios (SNR) [2].

One of the most interesting technologies during the last years for the advancement of wireless communications has been massive multiple-input-multiple-output (MIMO), a technology where a base station (BS) is equipped with a large number of antennas [3]. In massive MIMO, time division duplex (TDD) is typically used as the strategy to organize the uplink (UL) and downlink (DL) usage of the channel resources [4]. In TDD, each user's equipment sends its reference signals (ideally orthogonal), which are typically referred to as pilots, multiplexed with data in the uplink wireless resources. Then, the channel is estimated using the signals received in these resources to obtain the so-called channel state information (CSI). With the obtained CSI, the

The associate editor coordinating the review of this manuscript and approving it for publication was Walid Al-Hussaibi.

uplink data are coherently demodulated. Additionally, using the same CSI and making use of the reciprocity property of the channel [5], the downlink data are precoded to spatially separate the users. The TDD massive MIMO approach can be applied due to the channel reciprocity that exists between the uplink and the downlink streams when the channel variations are controlled by properly designing the UL and DL time slots' duration. We ensure the reciprocity in this manuscript by assuming soft channel variations which are present in most of the wireless propagation channels and assume transceivers calibration, which is needed to ensure full channel reciprocity [6], [7].

In the previously mentioned scenarios of interest, such as those with high mobility and/or low SNR, the use of accurate enough CSI is problematic. This comes from the fact that too many pilots may be needed [8], [9] either to effectively track channel variations or to average channel estimations to increase the estimation quality. This results, inevitably, in a detriment in the efficiency of the communication link since more resources will be used for channel estimation and less for useful data transmission [10]. Moreover, this problem aggravates in scenarios with many users since each of them needs to transmit orthogonal (or at least quasi-orthogonal) pilots to obtain their respective CSI with the minimum possible interference. Noteworthily, apart from the pilots included in the uplink for CSI estimation, some additional pilot symbols must be added in the data stream of the downlink of each user, as shown in [11] and references therein. This is necessary to perform the coherent demodulation due to a couple of reasons: (1) the channel estimation performed in the uplink is always imperfect due to the presence of noise in the estimation, (2) the employed precoding and filtering matrices do not always completely compensate the channel effects and they may not be designed with that goal, as it happens for example with maximum ratio combining/transmission (MRC/MRT), and (3) an imperfect channel is compensated with respect to the estimated channel due to channel aging (CA) effects. All these three channel effects result in a phase and amplitude shift of the received symbols, which requires demodulation pilots to compensate it. Therefore, some downlink resources, which are often called demodulation pilots or downlink pilots, (even though potentially shared among several users) must be used for channel estimation and thus an additional processing is needed in the downlink. The overhead due to uplink and downlink pilots may become excessive in the above mentioned circumstances, causing an intolerable decrease of the actual data throughput. Some blind and semi-blind channel estimation techniques permit to avoid or reduce the use of pilots [12], [13], [14], [15], [16], but suffer from the need of very large coherence time intervals to perform the estimation, and thus reprehend their use in fast varying channels. Also, they need high SNR values and their detection complexity is very large. Other approaches perform channel prediction to improve the channel estimation [17], [18], [19], but they cannot either cope with extremely fast time varying channels. For this purpose, in this work we combine channel estimation with non-coherently detected data and prediction utilizing the approach proposed in [8], which solves many of the previously mentioned limitations of the state-of-the-art works.

A technique that allows to receive data without the use of CSI, while benefiting from the use of many antennas at the BS and also reducing the complexity of the receivers, is called non-coherent (NC) massive single-input-mutiple-output (SIMO) [9], [20]. As it was shown in [8], the NC detection based on differential M-ary phase shift keying (DMPSK) [20] can be used to not only detect the transmitted data without having CSI knwoledge but also to estimate the channel response. The process for this novel channel estimation is to first detect the NC data, then reconstruct the NC symbol sent from the transmitter and use this reconstructed symbol to estimate the channel. Its main advantage is that it does not need large coherence time intervals to perform the channel estimation (contrary to other techniques [12], [13], [14]).

It is worth noting that the literature of NC massive MIMO is very focused on the uplink and not much has been proposed for the downlink [21]. Then, a natural question arises: can we leverage the channel estimation method proposed in [8] to design a whole pilot-less TDD (uplink and downlink) communication system? In [8], the authors proposed a differentially data-aided channel estimation approach based on reconstructing differentially encoded and detected data. It was shown to be useful to increase the data efficiency of an OFDM grid in an uplink connection. The MSE of the channel estimation was analyzed considering erroneously reconstructed differential data and without considering CA effects.

To answer the previous question, in this paper, we propose and analyze an innovative approach for TDD massive MIMO in which both the uplink and the downlink are based on applying differentially encoded data following a NC massive SIMO scheme. When including spatial multiplexing by estimating the channel utilizing differentially reconstructed encoded data as done in [8]. In this manuscript we consider CA effects when analyzing the MSE, which are critical in fast time varying channels. In a first TDD time slot, the users reference signals must be placed in orthogonal resources in the uplink. The subsequent downlink and uplink data streams will be spatially multiplexed using the initially estimated channel in the first uplink slot to design the precoder or receiver filter, respectively. To be able to cope with time varying channels, the channel estimation will be updated by reconstructing the NC data obtained in each uplink slot of each TDD slot and using it to estimate the channel, similarly to how it was proposed in [8]. The channel estimation is improved by utilizing a pth linear Wiener filter that acts as a channel predictor, which to the best knowledge of the authors has never been applied to perform channel prediction in non-coherent massive MIMO systems. The users are spatially separated in both the uplink and the downlink through the precoder and their data is modulated in DMPSK. The main

reason behind this is that the DMPSK modulation is very robust against phase and amplitude residual channel effects. These come from the fact that the channel estimation is always imperfect, and they worsen in fast varying or low SNR channels. On the contrary, the quadrature-amplitude modulation (QAM) greatly suffers from phase and amplitude shifts, which makes it not suitable for the proposed pilot-less TDD massive MIMO.

The motivation of this work is to propose a TDD massive MIMO approach that can maintain a large data efficiency in fast varying channels and/or low SNR values. This application is of interest for future wireless communication systems, such as the foreseen 6G, which aims for scenarios with users moving at a very high speed and battery powered or energy efficient devices [2]. This application is useful for any frequency range where TDD massive MIMO may be applied, but it is particularly interesting in the sub-6GHz range where the channel characteristics are more suitable to achieve a high degree of spatial multiplexing (please note the spectrum range can be seen in [22, Fig. 1]).

Please note that [23] is a preliminary conference version of this work, which has been greatly extended and improved in this manuscript. Here, we improve the analysis of the channel error effects and the spatial multiplexing of the DL, include spatial multiplexing in the UL, consider an orthogonal frequency division multiplexing (OFDM) waveform [24] in the performance analysis, with channels correlated in time and introduce a channel prediction approach that improves the blind channel estimation, proposed in [8]. The contributions of this manuscript are:

- A novel pilot-less TDD massive MIMO scheme is proposed. It is based on estimating the channel utilizing reconstructed NC data. Its main advantage is that it does not need long coherence times and can thus cope with fast varying channels. To improve the channel estimation, the use of a linear Wiener filter for prediction is proposed. Also, different symbol placing strategies over and OFDM grid are proposed.
- An analysis of the performance of the proposed pilot-less TDD scheme is provided for fast time varying channels. First, the mean squared error (MSE) of the channel estimation is given. Second, the system performance in terms of BER is analyzed for both the UL and the DL. Last, a qualitative analysis of the channel prediction is given.
- Some numerical results are provided to corroborate the previous analysis. Besides this, the proposed pilot-less scheme is compared with a classical pilot-symbol-assisted-modulation (PSAM) system to show that the proposed approach is a better option in fast varying channels.

The remainder of this paper is organized as follows. Section II introduces the System Model for the TDD coherent massive MIMO. Section III explains the proposed pilot-less scheme, with a symbol placing structure and a Wiener channel predictor to improve the performance of the system. In Section IV, we analyze the mean square error (MSE) of the channel estimation using NC data with time variability and characterize the theoretical expression of the symbol-error-rate (SER) of the precoded uplink and downlink data, where we also give a qualitative analysis of the Wiener channel predictor. Numerical results in Sec. V confirm our derivations and show a performance comparison between the proposed scheme and the coherent counterpart. Some conclusions and future work are pointed out in Sec. VI.

Notation: matrices, vectors and scalar quantities are denoted by boldface uppercase, boldface lowercase, and normal letters, respectively. $[\mathbf{A}]_{m,n}$ denotes the element in the m-th row and n-th column of $\mathbf{A}$, where a subindex $\forall$ indexes all the elements in that dimension. $[\mathbf{a}]_n$ represents the n-th element of vector $\mathbf{a}$. The superscripts $(\cdot)^H$, $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^{-1}$ and $*$ denote Hermitian, complex conjugate, matrix transpose, matrix inverse and convolution, respectively. $\mathbb{E}\{\cdot\}$ represents the expected value. $\mathcal{CN}(0,\sigma^2)$ represents the circularly-symmetric and zero-mean complex normal distribution with variance σ^2, $U(a,b)$ represents the uniform distribution between a and b and $VG(\mu,\alpha,\beta,\lambda)$ refers to the variance gamma distribution where μ is the location, α and β are the asymmetry parameters and λ is the shape parameter. $\Re\{(\cdot)\}$ and $\Im\{(\cdot)\}$ refer, respectively, to the real and imaginary parts of a complex number. $\|x\|^2$ denotes the euclidean norm of x. $\mathbf{1}$ and $\mathbf{0}$ indicate a column vector of all ones and all zeros, respectively. $Q(\cdot)$ is the Q-function and $\mathbf{I}_R$ the identity matrix of size $(R \times R)$. $\angle(\cdot)$ is the angle function. δ^n is the Kronecker delta. erfc() refers to the complementary error function.

II. SYSTEM MODEL FOR TDD COHERENT MASSIVE MIMO

In this section we explain the system model for a TDD coherent massive MIMO in which the channel is estimated in the UL typically using pilots, and then this estimation is used for filtering (in the UL) and precoding (in the DL) of multiuser data by spatial multiplexing. This system is useful since it is the baseline from which our pilot-less scheme is proposed.

We consider a massive MIMO base-station (BS), with R antennas and U single-antenna users. An OFDM signal composed of K subcarriers is used for transmission in UL and DL, with a cyclic prefix length (L_{CP}), which is designed to be long enough to absorb the effects of the multipath channel. At each receiving side, after removing the cyclic prefix and performing a fast Fourier Transform (FFT), it is possible to process each subcarrier independently.

In the UL, each user u transmits a symbol $x_u^{n,k}$, containing either data or a pilot, placed in an orthogonal subcarrier k at time instant n. The propagation channel between the user u and the BS at time instant n and subcarrier k is represented by $\mathbf{h}_u^{n,k}$ of size $(R \times 1)$ where $\mathbf{h}_u^{n,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_R)$ is modeling a Rayleigh distributed channel, as defined in [25]. Following [19], [26], the channel coefficients suffer from time variability and an autoregressive model approximates the temporally correlated fading channel coefficients of

subcarrier k at time instant n as

$$\mathbf{h}_u^{n',k} = \alpha_d \mathbf{h}_u^{n,k} + \mathbf{w}_{u,h}^{n',k}, \tag{1}$$

where n' refers to a time instant in the future with respect to n ($d = |n' - n|$ time difference in OFDM symbols), $\alpha_d = J_0\left(2\pi d f_D \left(\frac{K+L_{CP}}{K\Delta f}\right)\right) < 1$ is the temporal correlation parameter, Δf is the subcarrier spacing in Hertz, $J_0(\cdot)$ denotes the zero-th order Bessel function of the first kind and f_D represents the maximum Doppler shift experienced by the transmitted signal, also in Hertz. Also, we assume that $\mathbf{w}_{u,h}^{n',k}$ is uncorrelated with $\mathbf{h}_u^{n',k}$ since it represents a random channel change component and is modeled [27] as a Gaussian random process with i.i.d. entries and distribution $\mathbf{w}_{u,h}^{n',k} \sim \mathcal{CN}\left(\mathbf{0}, (1-\alpha_d^2)\mathbf{I}_R\right)$. Therefore, the distribution of the channel accounting for channel time variations is $\mathbf{h}_u^{n',k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_R)$.

The received signal in the UL for each user u at the k-th subcarrier and n-th time instant is given by

$$\mathbf{q}_{u,ul}^{n,k} = \mathbf{h}_u^{n,k} x_{u,ul}^{n,k} + \boldsymbol{\nu}_b^{n,k}, \tag{2}$$

where $\boldsymbol{\nu}_b^{n,k}$ $(R \times 1)$ denotes the additive white Gaussian noise (AWGN) in the UL (therefore introduced by the base station) with all the elements distributed according to $[\boldsymbol{\nu}_{n,b}^k]_r \sim \mathcal{CN}(0, \sigma_b^2)$. In the BS, $\mathbf{h}_u^{n,k}$ is estimated from $\mathbf{y}_{u,ul}^{n,k}$. In case $x_{u,ul}^{n,k}$ is a known pilot, the channel for each user u is estimated following pilot-symbol-assisted modulation (PSAM) [28] and denoted as $\hat{\mathbf{h}}_u^{n,k}$, which is normalized for analysis purposes. The estimated channels of all the users are stacked as $\hat{\mathbf{H}}_{ul}^{n,k} = [\hat{\mathbf{h}}_1^{n,k}, \cdots, \hat{\mathbf{h}}_U^{n,k}]$ and the actual channels as $\mathbf{H}_{ul}^{n,k} = [\mathbf{h}_1^{n,k}, \cdots, \mathbf{h}_u^{n,k}]$ both of size $(R \times U)$. Representing the symbols of all the users stacked in $\mathbf{x}_{ul}^{n,k} = [x_{1,ul}^{n,k}, \cdots, x_{U,ul}^{n,k}]^T$ of size $(U \times 1)$ for each time instant n and subcarrier k, we can define the received signal from all users in the UL as

$$\mathbf{q}_{ul}^{n,k} = \mathbf{H}_{ul}^{n,k} \mathbf{x}_{ul}^{n,k} + \boldsymbol{\nu}_b^{n,k}, \tag{3}$$

for which we apply a filtering matrix $\mathbf{B}_{ul}^{n,k}$ in the BS in reception as

$$\mathbf{y}_{ul}^{n,k} = \mathbf{B}_{ul}^{n,k} \mathbf{q}_{ul}^{n,k} = \mathbf{B}_{ul}^{n,k} \mathbf{H}_{ul}^{n,k} \mathbf{x}_{ul}^{n,k} + \mathbf{B}_{ul}^{n,k} \boldsymbol{\nu}_b^{n,k}, \tag{4}$$

which can be designed to satisfy several criteria, and we focus on either the maximum ratio combining (MRC) as $\mathbf{B}_{mr,ul}^{n,k} = (\hat{\mathbf{H}}_{ul}^{n,k})^H$, the zero-forcing (ZF) as $\mathbf{B}_{zf,ul}^{n,k} = (\hat{\mathbf{H}}_{ul}^{n,k})^H (\hat{\mathbf{H}}_{ul}^{n,k} (\hat{\mathbf{H}}_{ul}^{n,k})^H)^{-1}$, or the minimum mean squared error (MMSE) as $\mathbf{B}_{mm,ul}^{n,k} = (\hat{\mathbf{H}}_{ul}^{n,k})^H (\hat{\mathbf{H}}_{ul}^{n,k} (\hat{\mathbf{H}}_{ul}^{n,k})^H + \sigma_b^{-2})^{-1}$.

Similarly in the DL, the symbols of all the users are stacked in $\mathbf{x}_{dl}^{n,k}$ of size $(U \times 1)$ and the downlink channel is defined as $\mathbf{H}_{dl}^{n,k} = (\mathbf{H}_{ul}^{n,k})^T$ of size $(U \times R)$ both for each time instant n and subcarrier k. The DL symbols are precoded before transmission using a precoding matrix defined simlarly as for the UL, but utilizing the DL channel matrix instead of the UL one. Thus, the DL received signal is

$$\mathbf{y}_{dl}^{n,k} = \mathbf{H}_{dl}^{n,k} \mathbf{B}_{dl}^{n,k} \mathbf{x}_{dl}^{n,k} + \boldsymbol{\nu}_{dl}^{n,k}, \tag{5}$$

where $\boldsymbol{\nu}_{dl}^{n,k}$ is a vector $(U \times 1)$ where each element is $\nu_u^{n,k}$ and is the noise at each user u distributed as $\nu_u^{n,k} \sim \mathcal{CN}(0, \sigma_u^2)$.

The signal of each user can be separated for the DL by extending (5) to separate the matrix between the desired user and the rest of the users and selecting the value of user u as

$$[\mathbf{y}_{dl}^{n,k}]_u = (\mathbf{h}_u^{n,k})^T \mathbf{b}_{u,dl}^{n,k} x_{u,dl}^{n,k} + \sum_{u' \neq u} (\mathbf{h}_u^{n,k})^T \mathbf{b}_{u',dl}^{n,k} x_{u',dl}^{n,k} + \boldsymbol{\nu}_{dl}^{n,k}, \tag{6}$$

where $\mathbf{b}_{u,dl}^{n,k} = [\mathbf{B}_{dl}^{n,k}]_{\forall,u}$ and $x_{u,dl}^{n,k} = [\mathbf{x}_{dl}^{n,k}]_u$. It could similarly be done for the UL as

$$[\mathbf{y}_{ul}^{n,k}]_u = (\mathbf{b}_{u,ul}^{n,k})^T \mathbf{h}_u^{n,k} x_{u,ul}^{n,k} + \sum_{u' \neq u} (\mathbf{b}_{u',ul}^{n,k})^T \mathbf{h}_u^{n,k} x_{u',ul}^{n,k} + (\mathbf{b}_{u,ul}^{n,k})^T \boldsymbol{\nu}_{ul}^{n,k}. \tag{7}$$

For the coherent demodulation, a coherence time of at least four OFDM symbols as shown in Fig. 3 is needed. Thus, due to the reciprocity of the channel and for coherent demodulation, we should ensure that $\mathbf{h}_u^{n_{ul},k} \approx (\mathbf{h}_u^{n_{dl},k})^T$ so long as $|n_u - n_d| < n_c = 4$. In the previous equation, $n_c = 0.15 f_D^{-1} \Delta f$ [26] denotes the coherence time in OFDM symbol units according to the model in (1).

III. PROPOSED PILOT-LESS TDD MASSIVE MIMO

In this section we explain how the proposed pilot-less TDD massive MIMO system works. Our proposal is based on applying differential encoding with DMPSK for both the UL and DL data and estimating the channel in the UL utilizing differentially reconstructed encoded data, avoiding the use of pilot symbols. We first explain how the differentially encoded data is constructed for both the uplink and downlink streams. Secondly, we indicate how the channel can be estimated in the uplink utilizing NC data. Thirdly, we give insights on how the symbols can be placed in the OFDM grid. Last, we explain the Wiener channel predictor to improve the channel estimation.

A. DIFFERENTIAL UPLINK AND DOWNLINK

We differentiate between the first TDD slot, where no spatial multiplexing is done since a reference orthogonal signal is needed for each user and the rest of the TDD slots, where the NC data can be spatially multiplexed in the UL and the DL.

We propose a system in which each UL and DL data symbol of each user $s_u^{n,k}$ is differentially encoded as

$$x_u^{n,k} = x_u^{n-1,k} s_u^{n,k} \quad 1 \leq n \leq N_u, \tag{8}$$

where N denotes the length of the differentially encoded data stream which only needs one reference symbol $s_u^{0,k}$ known at both ends of the communication link, and $s_u^{n,k}$ belongs to a DMPSK constellation for both the UL and the DL. Because the information is only encoded in the phase of unit modulus symbols, we know that

$$(x_u^{n,k})^* x_u^{n-1,k} = s_u^{n,k} \quad \text{for} \quad 1 \leq n \leq N_u. \tag{9}$$

The reception is performed via differential detection of two consecutive received signals (4) for the UL and (5) for

the DL in time, as shown in [29],

$$z_u^{n,k} = \left([\mathbf{y}^{n-1,k}]_u\right)^H [\mathbf{y}^{n,k}]_u, \tag{10}$$

and the transmitted symbol for each user is estimated in the BS (for the uplink) and in each user terminal (for the downlink) according to [20] as

$$\hat{s}_u^{n,k} = \arg\min_{s_u^{n,k}} \left\{ \left| s_u^{n,k} - z_u^{n,k} \right|, \; s_u^{n,k} \in \mathfrak{M} \right\}, \tag{11}$$

where $\mathfrak{M}$ indicates the DMPSK constellation set, of size M, either for the uplink (M_{UL}) or downlink (M_{DL}).

B. CHANNEL ESTIMATION IN THE UPLINK

Once the decision is taken on the UL NC data symbols, the channel can be estimated following [8, Eq. (21)]. First, we recompose the differential symbols from the detected ones using (11) as

$$\hat{x}_{u,ul}^{n,k} = \hat{x}_{u,ul}^{n-1,k} \hat{s}_{u,ul}^{n,k}, \quad \text{for} \quad 1 \le n \le N_u, \tag{12}$$

where $\hat{x}_{u,ul}^{0,k}$ is known at the receiver side, as explained above. Then, the LS criterion is applied to estimate the channel as

$$\hat{\mathbf{h}}_u^{n,k} = \frac{\mathbf{q}_{u,ul}^{n,k}}{\hat{x}_{u,ul}^{n,k}} = \mathbf{h}_u^{n,k} \cdot \frac{x_{u,ul}^{n,k}}{\hat{x}_{u,ul}^{n,k}} + \frac{\boldsymbol{\nu}_b^{n,k}}{\hat{x}_{u,ul}^{n,k}}, \tag{13}$$

where there is no noise enhancement since $|\hat{x}_{u,ul}^{n,k}| = 1$. A maximum length measured in OFDM symbols in the uplink differential data stream of each user is defined by N_u so that the probability that an erroneous detection ($\hat{x}_{u,ul}^{n,k} \neq x_{u,ul}^{n,k}$) is reduced to ensure a proper performance, as defined in [8]. With this estimated channel, the DL is precoded following (5) in the same time slot and the UL is filtered in the UL of the next time slot.

It is worth noting that before estimating the channel, it is necessary to separate the users by filtering them with $\mathbf{B}_{ul}^{n,k}$ as shown in (4). This process should spatially separate the users and thus, we can detect the data for each of them independently in the BS by differentially detecting the data for each user symbol with (10) and then perform the detection with (11). The estimated channel, which is obtained applying (13), in a certain TDD slot is used to construct the filtering matrix for the UL $\mathbf{B}_{ul}^{n,k}$ for the next UL symbols $\mathbf{q}_{ul}^{n,k}$ and the precoding matrix for the DL $\mathbf{B}_{dl}^{n,k}$ for the next DL symbols. By looking at Fig. 1 we can better understand this idea. Every time the channel is estimated, we may store this estimation for a combined processing with subsequent channel estimations in the UL.

C. SYMBOL PLACING AND STRUCTURE IN THE OFDM GRID

The proposed pilot-less TDD process starts with a first TDD slot in which the first uplink symbol is the reference $\hat{x}_{u,ul}^{0,k}$, so for this symbol, the channel estimation follows the PSAM approach [28]. The reference symbol of each user must be placed in orthogonal resources, and we propose to place them in orthogonal subcarriers for the first OFDM symbol. For the next UL symbols, either in the first TDD slot or on the rest, the channel estimation is performed following (13), without any kind of reference signals.

By looking at Fig. 1, the first TDD slot has the orthogonal reference signals $s_u^{0,k}$ in the first OFDM symbol which belongs to the first UL slot. A reference symbol is introduced every N_u symbols, which is much larger that the TDD slot. Please note the channel can be directly estimated not only for each time instant in the UL and for each subcarrier, so the interpolation typical in the PSAM is not needed. The configuration shown in Fig. 2 shows that the proposed scheme can work for a coherence time of 3 OFDM symbols, which is not attainable in the coherent counterpart which needs at least 4 OFDM symbols, as shown in Fig. 3. Besides this, the efficiency of the coherent scheme is 66% of the OFDM grid, while close to 100% for the proposed pilot-less scheme.

D. CHANNEL PREDICTION USING A LINEAR WIENER PREDICTOR

To improve the channel estimation process, and taking into account that a lot of knowledge of the channel is available thanks to the channel estimation based on the use of differentially encoded data, we propose to perform a channel prediction utilizing a pth linear Wiener predictor as it was done in [19]. We briefly explain the equations to define the Wiener predictor used in this paper, but for further details the reader is referred to [19] and references therein.

We define the optimal pth linear Wiener for the autoregressive model of (1) with $d = 1$ as

$$\mathbf{V} = \alpha_1 [\boldsymbol{\omega}_{\alpha_1}^p \otimes \mathbf{I}_R] \mathbf{T}_{\alpha_1}^p, \quad \text{of size} \;\; R \times R(p+1), \tag{14}$$

where $[\boldsymbol{\omega}_{\alpha_1}^p \otimes \mathbf{I}_R]$ has size $R \times R(p+1)$ and is defined

$$[\boldsymbol{\omega}_{\alpha_1}^p \otimes \mathbf{I}_R] = [\mathbf{I}_R, \alpha_1 \mathbf{I}_R, \alpha_1^2 \mathbf{I}_R, \cdots, \alpha_1^p \mathbf{I}_R], \tag{15}$$

and defining $\mathbf{L}_{\hat{\mathbf{h}}}^{n-n'} = \left(\alpha_1^{|n-n'|} + \delta^{n-n'} \sigma_b^2\right) \mathbf{I}_R$, we have

$$\mathbf{T}_{\alpha_1}^p = \begin{pmatrix} \mathbf{L}_{\hat{\mathbf{h}}}^0 & \mathbf{L}_{\hat{\mathbf{h}}}^1 & \cdots & \mathbf{L}_{\hat{\mathbf{h}}}^p \\ \mathbf{L}_{\hat{\mathbf{h}}}^1 & \mathbf{L}_{\hat{\mathbf{h}}}^0 & \cdots & \mathbf{L}_{\hat{\mathbf{h}}}^{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{L}_{\hat{\mathbf{h}}}^p & \mathbf{L}_{\hat{\mathbf{h}}}^{p-1} & \cdots & \mathbf{L}_{\hat{\mathbf{h}}}^0 \end{pmatrix}^{-1} \tag{16}$$

which has a size $R(p+1) \times R(p+1)$. And the predicted channel $\tilde{\mathbf{h}}_u^{n+1,k}$ is defined as

$$\tilde{\mathbf{h}}_u^{n+1,k} = \mathbf{V} \left[(\hat{\mathbf{h}}_u^{n,k})^H, (\hat{\mathbf{h}}_u^{n-1,k})^H, \cdots, (\hat{\mathbf{h}}_u^{n-p,k})^H \right]^H. \tag{17}$$

It is worth noting that older observations have a lower impact on the prediction than newer observations. In case the noisy channel observation $\hat{\mathbf{h}}_u^{n,k}$ is not available, due to the fact that the resource at time instant n is not dedicated for channel estimation, the channel observation required to predict the channel in time instant $n+1$ would be the channel predicted at time instant n ($\hat{\mathbf{h}}_u^{n+1,k} = \tilde{\mathbf{h}}_u^{n,k}$). This means that

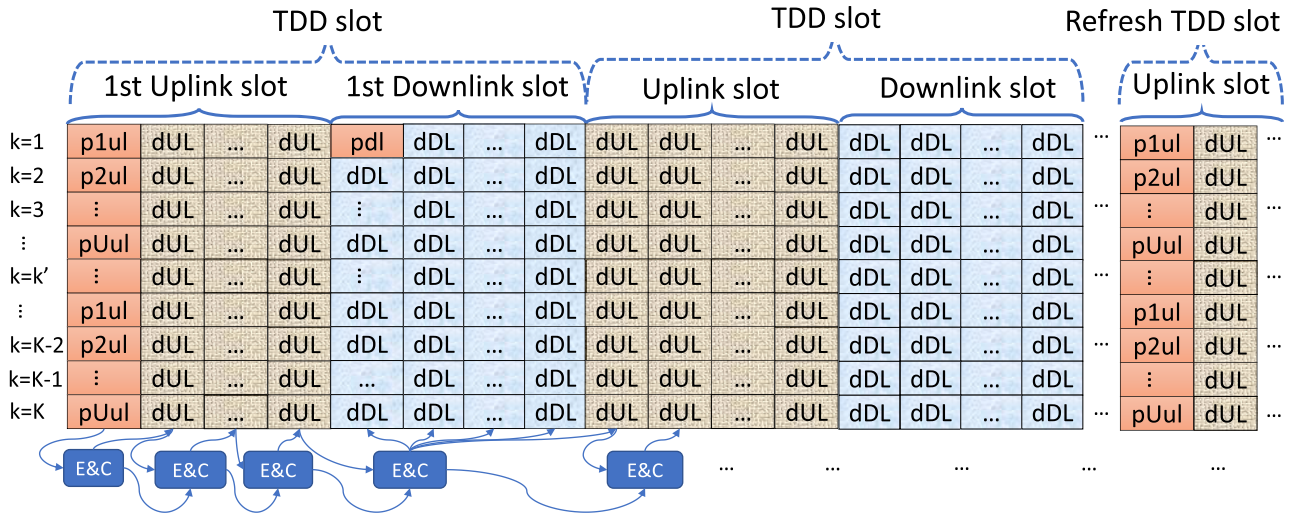

FIGURE 1. OFDM grid schematic showing TDD slots with flexible uplink and downlink slots. E&C boxes refer to channel estimation and compensation and it can be seen that the CSI can be estimated in each UL slot and utilized in the subsequent slots.

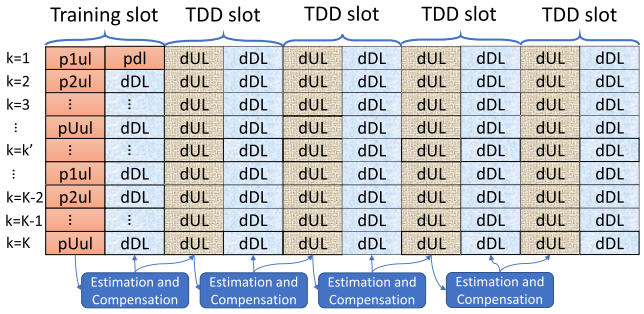

FIGURE 2. OFDM grid schematic showing TDD slots and users' reference symbols placed in orthogonal subcarriers in the first OFDM symbols, for the extreme case of a TDD slot of 2 OFDM symbols (requires a coherence time of at least 3 OFDM symbols).

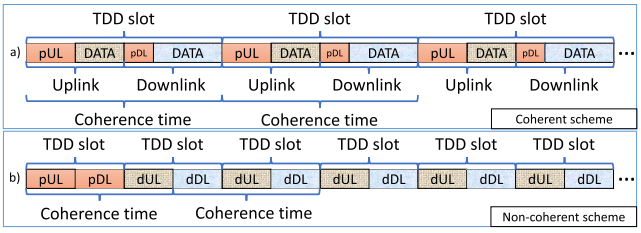

FIGURE 3. TDD frame comparison for coherent (a) and non-coherent (b) for the first subcarrier.

a set of consecutively predicted channels will be used in case no noisy channel observation is available. This can be iterated for a few time instants, but not many since the performance of the predictor would greatly decrease, due to the reliance on too many predictions.

Therefore, and to summarize, we can define the predicted channel at time instant $n+1$ as $\tilde{\mathbf{h}}_u^{n+1,k}$, which would utilize the $p+1$ previously estimated or predicted channels $\tilde{\mathbf{h}}_u^{n,k}, \ldots, \tilde{\mathbf{h}}_u^{n-p,k}$.

IV. ANALYSIS OF THE PROPOSED PILOT-LESS TDD MASSIVE MIMO

Since the proposed pilot-less TDD massive MIMO scheme is intended to be used in time-varying channels, we have to analyze its performance in those channels. For this reason, in this section, we first characterize the effect of time correlated channels on the mean square error (MSE) of the channel estimation, which depends on the symbol error rate (SER) of the differentially decoded symbol in the UL [8] and the time variability. The main difference with the cited work in this regard comes from the fact that in this work we are considering channel time variability, so the estimated channel in time instant n will be used in a future time instant n' different from the one where the channel has been estimated. Second, we characterize the symbol-error-rate (SER) of the differentially encoded symbol of the DL of each user. This section utilizes the mathematical properties defined in Appendix A.

A. MEAN SQUARE ERROR OF THE CHANNEL ESTIMATION

An additional error term in the channel estimation, with respect to the classical PSAM [28], is produced by a possible mismatch between transmitted $x_{u,ul}^{n,k}$ and reconstructed differential symbols $\hat{x}_{u,ul}^{n,k}$. This error was characterized in [8], without considering channel time variability and in [23] including channel time variability but without considering the spatial multiplexing of the users in the UL. Therefore, following the same approach, we calculate the MSE of the channel estimation in the UL of a NC stream, now considering channel time variability and the additional error caused by the multiplexing of the users in the UL. The estimated channel in time instant n will be used in another time instant n' different from the one where the channel has actually been estimated. Hence, the channel estimation error is composed of two independent

components [8, Eq. (24)] as shown below

$$e_d^2 = \mathbb{E}\left\{\left|\hat{\mathbf{h}}_u^{n,k} - \mathbf{h}_u^{n+d,k}\right|^2\right\} = \sigma_{x,d}^2 + \sigma_b^2, \tag{18}$$

where $\sigma_{x,d}^2$ is the channel estimation error that comes from compensation and estimation in different time instants with a possible mismatch between transmitted and reconstructed differential symbol. We compute

$$\begin{aligned}
\sigma_{x,d}^2 &= \mathbb{E}\left\{\left|\mathbf{h}_u^{n,k}\frac{x_{u,ul}^{n,k}}{\hat{x}_{u,ul}^{n,k}} - \mathbf{h}_u^{n+d,k}\right|^2\right\} \\
&= \mathbb{E}\left\{\left|\mathbf{h}_u^{n,k}\frac{x_{u,ul}^{n,k}}{\hat{x}_{u,ul}^{n,k}} - \alpha_d\mathbf{h}_u^{n,k} - \mathbf{w}_{u,h}^{n+d,k}\right|^2\right\} \\
&= \mathbb{E}\left\{\left|\mathbf{h}_u^{n,k}\right|^2\left|\frac{x_{u,ul}^{n,k}}{\hat{x}_{u,ul}^{n,k}} - \alpha_d\right|^2 + \left|\mathbf{w}_{u,h}^{n+d,k}\right|^2\right\} \\
&= \mathbb{E}\left\{\left|\mathbf{h}_u^{n,k}\right|^2\right\}\mathbb{E}\left\{\left|\frac{x_{u,ul}^{n,k}}{\hat{x}_{u,ul}^{n,k}} - \alpha_d\right|^2\right\} + \mathbb{E}\left\{\left|\mathbf{w}_{u,h}^{n+d,k}\right|^2\right\} \\
&= \mathbb{E}\left\{\left|\frac{x_{u,ul}^{n,k}}{\hat{x}_{u,ul}^{n,k}} - \alpha_d\right|^2\right\} + (1-\alpha_d^2),
\end{aligned} \tag{19}$$

where $\mathbb{E}\left\{\left|\mathbf{h}_u^{n,k}\right|^2\right\} = 1$ since the channel power is normalized, $\left|\mathbf{h}_u^{n,k}\right|^2$ is uncorrelated and independent of $\left|x_{u,ul}^{n,k}/\hat{x}_{u,ul}^{n,k} - \alpha_d\right|^2$ and $x_{u,ul}^{n,k}/\hat{x}_{u,ul}^{n,k}$ and $\mathbf{w}_{u,h}^{n+d,k}$ are uncorrelated and independent. Developing the first term of the last part of (19), we have

$$\begin{aligned}
&\mathbb{E}\left\{\left|\frac{x_{u,ul}^{n,k}}{\hat{x}_{u,ul}^{n,k}} - \alpha_d\right|^2\right\} \\
&= \mathbb{E}\left\{\left|\frac{x_{u,ul}^{n,k}}{\hat{x}_{u,ul}^{n,k}}\right|^2 - 2\alpha_d\Re\left\{\frac{x_{u,ul}^{n,k}}{\hat{x}_{u,ul}^{n,k}}\right\} + \alpha_d^2\right\} \\
&= 1 + \alpha_d^2 - 2\alpha_d\mathbb{E}\left\{\cos\left(\angle(x_{u,ul}^{n,k}) - \angle(\hat{x}_{u,ul}^{n,k})\right)\right\}.
\end{aligned} \tag{20}$$

We define

$$\beta_u^{n,k} = \mathbb{E}\{\cos\left(\angle(x_{u,ul}^{n,k}) - \angle(\hat{x}_{u,ul}^{n,k})\right)\} \tag{21}$$

and by following the same approach as that of [8, Appendix B] (omitted here for the sake of conciseness), we define

$$\beta_u^{n,k} \approx \begin{cases} 1, & P_{u,ul}^{n,k} = 0 \\ \dfrac{1 - P_{u,ul}^{n,k} - \left(1 - P_{u,ul}^{n,k}\right)^{N_u}}{(N_u - 1)P_{u,ul}^{n,k}}, & 0 < P_{u,ul}^{n,k} \leq 1 \end{cases}, \tag{22}$$

where $P_{u,ul}^{n,k}$ is the error probability for the UL of each user, calculated in (56) with the conrresponding mean and variance values. This error $P_{u,ul}^{n,k}$ is greater in this paper than it was in [8] because that manuscript did not consider several users multiplexed in the UL. Following the same approach, and by substituting from (22) to (18) successively, we define an upper bound (UB) for e_d^2, as stated in [8, Appendix B], and we can write $\sigma_{x,d}^2$ and e_d^2 as

$$\sigma_{x,d}^2 = 1 + \alpha_d^2 - 2\alpha_d\beta_u^{n,k} + 1 - \alpha_d^2 = 2(1 - \alpha_d\beta_u^{n,k}), \tag{23}$$

$$e_d^2 = \mathbb{E}\left\{\left|\hat{\mathbf{h}}_u^{n,k} - \mathbf{h}_u^{n+d,k}\right|^2\right\} = 2(1 - \alpha_d\beta_u^{n,k}) + \sigma_{v,u}^2. \tag{24}$$

By inspecting (24), it can be seen that in case that either α_d or β_n^k is zero, the channel error estimation is the highest, while both need to be 1 to avoid any increment in the channel estimation error with respect to the PSAM.

B. RECEIVED SYMBOL SER IN UL AND DL

In order to analyze the effect of an imperfectly estimated channel in the SER of the UL and DL, we focus on the MRT/MRC and we extend the channel vector products $\mathbf{h}_u^{n'}\mathbf{b}_{u'}^n$ in (6) following the definition in (1) and (13) as

$$\begin{aligned}
\mathbf{h}_u^{n'}\left(\hat{\mathbf{h}}_{u'}^n\right)^H &= \left(\alpha_d\mathbf{h}_u^n + \mathbf{w}_{u,h}^{n'}\right)\left(\mathbf{h}_{u'}^n\frac{x_{u'}^n}{\hat{x}_{u'}^n} + \frac{\boldsymbol{\nu}^n}{\hat{x}_{u'}^n}\right)^H \\
&= \alpha_d\mathbf{h}_u^n(\mathbf{h}_{u'}^n)^H\left(\frac{x_{u'}^n}{\hat{x}_{u'}^n}\right)^* + \alpha_d\mathbf{h}_u^n\frac{(\boldsymbol{\nu}^n)^H}{(\hat{x}_{u'}^n)^*} \\
&\quad + \mathbf{w}_{u,h}^{n'}(\mathbf{h}_{u'}^n)^H\left(\frac{x_{u'}^n}{\hat{x}_{u'}^n}\right)^* + \mathbf{w}_{u,h}^{n'}\frac{(\boldsymbol{\nu}^n)^H}{(\hat{x}_{u'}^n)^*},
\end{aligned} \tag{25}$$

where we drop the subcarrier k and the UL/DL notation for ease of understanding. It can be seen that the product $\mathbf{h}_u^{n'}\mathbf{b}_u^n$ is composed by the sum of 4 terms in which each of them is the product of 2 complex circularly symmetric normally distributed terms. Below we detail the distribution of each term by following the mathematical properties defined in Appendix A. The first term in (25) has a different distribution whether $u = u'$ or $u \neq u'$, shown below for $u = u'$

$$\alpha_d\mathbf{h}_u^n(\mathbf{h}_u^n)^H \sim \Gamma(R, \alpha_d) \xrightarrow{R\to\infty} \mathcal{N}\left(R\alpha_d, R\alpha_d^2\right), \tag{26}$$

and below for $u \neq u'$

$$\alpha_d\mathbf{h}_u^n(\mathbf{h}_{u'}^n)^H \sim CVG\left(2R, 0, \frac{\alpha_d}{\sqrt{2}}, 0\right) \xrightarrow{R\to\infty} \mathcal{CN}\left(0, R\alpha_d^2\right). \tag{27}$$

Please note the term $\left(x_{u'}^n/\hat{x}_{u'}\right)^*$ affects the behaviour of (26) due to the fact that it generates a phase rotation in case $x_{u'}^n \neq \hat{x}_{u'}$. In this case, we have $\left(x_{u'}^n/\hat{x}_{u'}\right)^* = e^{j\phi}$ where, for analysis purposes, we assume the following distribution for the angle $\phi \sim \mathcal{N}(0, \sigma_\phi^2)$. The specific value of σ_ϕ will be derived later.

We assume that (26) rotates in the imaginary axis according to σ_ϕ. This is done to characterize the distribution of the

received symbol $x_u^{n,k}$ by considering that the imaginary part of (25) together with the effect of $\left(x_{u'}^n/\hat{x}_{u'}\right)^*$ is

$$\Im\left\{\alpha_d \mathbf{h}_u^n (\mathbf{h}_u^n)^H \left(\frac{x_{u'}^n}{\hat{x}_{u'}^n}\right)^*\right\} \sim \mathcal{N}\left(0, 4\,R^2\alpha_d^2 \tan^2\left(\sigma_\phi\right)\right), \quad (28)$$

which has been obtained straightforwardly using trigonometry.

For both $u = u'$ and $u \neq u'$, the second term is the product of 2 complex circularly symmetric normally distributed terms, so following the properties defined in Appendix B

$$\alpha_d \mathbf{h}_u^n \frac{(\boldsymbol{\nu}^n)^H}{(\hat{x}_{u'}^n)^*} \sim \mathcal{CN}\left(0, R\alpha_d^2\sigma_b^2\right), \quad (29)$$

similarly, the third term is distributed as

$$\mathbf{w}_{u,h}^{n'} (\mathbf{h}_{u'}^n)^H \left(\frac{x_{u'}^n}{\hat{x}_{u'}^n}\right)^* \sim CVG\left(2R, 0, \sqrt{\frac{1-\alpha_d^2}{2}}, 0\right) \quad (30)$$

that is approximated for $R \to \infty$ as

$$\mathbf{w}_{u,h}^{n'} (\mathbf{h}_{u'}^n)^H \left(\frac{x_{u'}^n}{\hat{x}_{u'}^n}\right)^* \xrightarrow{R\to\infty} \mathcal{CN}\left(0, R(1-\alpha_d^2)\right), \quad (31)$$

and the fourth term

$$\mathbf{w}_{u,h}^{n'} \frac{(\boldsymbol{\nu}^n)^H}{(\hat{x}_{u'}^n)^*} \sim CVG\left(2R, 0, \sqrt{\frac{(1-\alpha_d^2)\sigma_b^2}{2}}, 0\right), \quad (32)$$

which is also approximated for $R \to \infty$ as

$$\mathbf{w}_{u,h}^{n'} \frac{(\boldsymbol{\nu}^n)^H}{(\hat{x}_{u'}^n)^*} \xrightarrow{R\to\infty} \mathcal{CN}\left(0, R(1-\alpha_d^2)\sigma_b^2\right). \quad (33)$$

The second, third and fourth terms are multiplied by either $x_{u'}^n$ or $\hat{x}_{u'}^n$, but their distribution is not changed since it is circularly symmetric with zero mean. Thus, the distribution only rotates and does not vary.

The summation (second term) in (6) has $U-1$ summands, each of them multiplied by the symbol for each user $u' \neq u$, which only rotates the distribution over its center and does not change it. At the same time, each summand is composed of the sum of 4 terms, them being (27), (29), (31) and (33). The first term in (6) is composed of three terms distributed as (29), (31) and (33) and one term whose real part is distributed as (26) and whose imaginary part is distributed as (28). Therefore, the distribution of (6) has to take into account all these terms and the noise term as well. Thus, we have

$$\Re\left\{[\mathbf{y}_{dl}^n]_u\right\} \sim \mathcal{N}\left(\mu_{\Re\{[\mathbf{y}_{dl}^n]_u\}}, \sigma^2_{\Re\{[\mathbf{y}_{dl}^n]_u\}}\right), \quad (34)$$

with

$$\begin{aligned}\mu_{\Re\{[\mathbf{y}_{dl}^n]_u\}} &= R\alpha_d \\ \sigma^2_{\Re\{[\mathbf{y}_{dl}^n]_u\}} &= \frac{R(U(1+\sigma_b^2)+\alpha_d^2)+\sigma_u^2}{2},\end{aligned} \quad (35)$$

and we also have

$$\Im\left\{[\mathbf{y}_{dl}^n]_u\right\} \sim \mathcal{N}\left(0, \sigma^2_{\Im\{[\mathbf{y}_{dl}^n]_u\}}\right), \quad (36)$$

with

$$\sigma^2_{\Im\{[\mathbf{y}_{dl}^n]_u\}} = \frac{R(U(1+\sigma_b^2)-\alpha_d^2)+\sigma_u^2}{2} + 4R^2\alpha_d^2\tan^2\left(\sigma_\phi\right). \quad (37)$$

The previous process can be repeated similarly for the UL in (7) and the distribution would be the same with the only difference of changing σ_u^2 by $R\sigma_b^2$ for the previous equations, thus

$$\sigma^2_{\Re\{[\mathbf{y}_{dl}^n]_u\}} = \frac{R(U(1+\sigma_b^2)+\alpha_d^2+\sigma_b^2)}{2}, \quad (38)$$

and

$$\sigma^2_{\Im\{[\mathbf{y}_{dl}^n]_u\}} = \frac{R(U(1+\sigma_b^2)-\alpha_d^2+\sigma_b^2)}{2} + 4R^2\alpha_d^2\tan^2\left(\sigma_\phi\right). \quad (39)$$

Assuming $x_{u,dl}^n = 1$ without loss of generality,[1] we can extend (34) as

$$\Re\left\{[\mathbf{y}_{dl}^n]_u\right\} \sim \mu_{\Re\{[\mathbf{y}_{dl}^n]_u\}} + \mathcal{N}\left(0, \sigma^2_{\Re\{[\mathbf{y}_{dl}^n]_u\}}\right), \quad (40)$$

and we have $\Im\left\{[\mathbf{y}_{dl}^n]_u\right\} \sim \mathcal{N}\left(\mu_\Im, \sigma_\Im^2\right)$, so

$$\Re\left\{[\mathbf{y}_{dl}^n]_u\right\} \sim R\alpha_d + \mathcal{N}\left(0, \sigma^2_{\Re\{[\mathbf{y}_{dl}^n]_u\}}\right) \quad (41)$$

$$\Im\left\{[\mathbf{y}_{dl}^n]_u\right\} \sim \mathcal{N}\left(0, \sigma^2_{\Im\{[\mathbf{y}_{dl}^n]_u\}}\right). \quad (42)$$

The differential decoding performed in reception for the received signal at each user (10) results in the product of complex normally distributed variables, where in order to find the distribution of the received symbol, we have to consider the product of two complex variables $x = a + jb$ and $y = c + jd$, so the product $(x)^*y = (ac + bd) + j(ad - bc)$. Using again the properties of the product of normal variables and VG distributions, found in Appendix VI, we have that

$$\Re\{z_u^n\} = \Re\{[\mathbf{y}_{dl}^{n-1}]_u\}\Re\{[\mathbf{y}_{dl}^n]_u\} + \Im\{[\mathbf{y}_{dl}^{n-1}]_u\}\Im\{[\mathbf{y}_{dl}^n]_u\}, \quad (43)$$

$$\Im\{z_u^n\} = \Re\{[\mathbf{y}_{dl}^{n-1}]_u\}\Im\{[\mathbf{y}_{dl}^n]_u\} - \Im\{[\mathbf{y}_{dl}^{n-1}]_u\}\Re\{[\mathbf{y}_{dl}^{n-1}]_u\}. \quad (44)$$

Thus, the first term of (43) is composed of three terms as

$$\Re_1\{z_u^n\} = R^2\alpha_d^2, \quad (45)$$

$$\Re_2\{z_u^n\} \sim R\alpha_d\mathcal{N}\left(0, 2\sigma^2_{\Re\{[\mathbf{y}_{dl}^n]_u\}}\right), \quad (46)$$

$$\Re_3\{z_u^n\} \sim VG\left(1, 0, \sigma^2_{\Re\{[\mathbf{y}_{dl}^n]_u\}}, 0\right), \quad (47)$$

while the second term of (43) is distributed as

$$\Im\{[\mathbf{y}_{n-1}^k]_u\}\Im\{[\mathbf{y}_n^k]_u\} \sim VG\left(1, 0, \sigma^2_{\Im\{[\mathbf{y}_{dl}^n]_u\}}, 0\right). \quad (48)$$

Each term of (44) is composed of two terms, them being

$$\Im_1\{z_u^n\} = R\alpha_d\mathcal{N}\left(0, \sigma^2_{\Im\{[\mathbf{y}_{dl}^n]_u\}}\right), \quad (49)$$

$$\Im_2\{z_u^n\} \sim VG\left(1, 0, \sigma_{\Re\{[\mathbf{y}_{dl}^n]_u\}}\sigma_{\Im\{[\mathbf{y}_{dl}^n]_u\}}, 0\right). \quad (50)$$

[1]The error is computed for $[\mathbf{x}_n]_u = 1$ for simplicity but is the same for the rest of the symbols due to the symmetry of the DMPSK constellation.

The VG distributions can be approximated to normal distributions as

$$VG\left(1,0,\sigma^2_{\Re\{[\mathbf{y}^n_{dl}]_u\}},0\right)\approx\mathcal{N}\left(0,\sigma^4_{\Re\{[\mathbf{y}^n_{dl}]_u\}}\right), \quad (51)$$

$$VG\left(1,0,\sigma^2_{\Im\{[\mathbf{y}^n_{dl}]_u\}},0\right)\approx\mathcal{N}\left(0,\sigma^4_{\Im\{[\mathbf{y}^n_{dl}]_u\}}\right), \quad (52)$$

and

$$VG\left(1,0,\sigma_{\Re\{[\mathbf{y}^n_{dl}]_u\}}\sigma_{\Im\{[\mathbf{y}^n_{dl}]_u\}},0\right)\approx\mathcal{N}\left(0,\sigma^2_{\Re\{[\mathbf{y}^n_{dl}]_u\}}\sigma^2_{\Im\{[\mathbf{y}^n_{dl}]_u\}}\right). \quad (53)$$

We rewrite $\Re\left\{[\mathbf{y}^n_{dl}]_u\right\}$ as $\Re\{y\}$, $\Im\{y\}$ as $\Im\{y\}$, $\Re\left\{z^n_u\right\}$ as $\Re\{z\}$ and$\Im\left\{z^n_u\right\}$ as $\Im\{z\}$. for space economy. Summarizing, the distribution of z^n_u for $\mathbf{s}^{n,k}_u=1$ is defined as

$$\Re\{z^n_u\}\sim\mathcal{N}\left(R^2\alpha_d^2,2R^2\alpha_d^2\sigma^2_{\Re\{y\}}+\sigma^4_{\Re\{y\}}+\sigma^4_{\Im\{y\}}\right), \quad (54)$$

$$\Im\{z^n_u\}\sim\mathcal{N}\left(0,2\sigma^2_{\Im\{y\}}\left(R^2\alpha_d^2+\sigma^2_{\Re\{y\}}\right)\right), \quad (55)$$

so the SER for the DL of user u can be computed following the approach in [8, Appendix A] as shown in (56), as shown at the bottom of the page.

Referring to Appendix A, we can simplify the double integral of (56) into a one dimensional integral as

$$P^{n,k}_u \approx 1-\int_{-\pi/M}^{\pi/M}\frac{e^{-c}\sqrt{\pi}be^{\frac{b^2}{4a}}\left(\operatorname{erfc}\left(\frac{b}{2\sqrt{a}}\right)+1\right)+2\sqrt{a}}{8\pi\sigma^2_{\Re}\sigma^2_{\Im}a^{3/2}}d\gamma. \quad (57)$$

where a, b and c are defined in Appendix A.

C. DERIVATION OF σ_ϕ

First of all, we have to take into account that $\left|\angle(x^{n,k}_{u,ul})-\angle(\hat{x}^{n,k}_{u,ul})\right|$ is very small, since an erroneous $\hat{x}^{n,k}_{u,ul}$ is caused by immediate neighboring symbols with a very high probability. Recalling (21) and using simple trigonometry

$$\cos\left(\angle(x^{n,k}_{u,ul})-\angle(\hat{x}^{n,k}_{u,ul})\right)=1-\frac{\left|\angle(x^{n,k}_{u,ul})-\angle(\hat{x}^{n,k}_{u,ul})\right|^2}{2}, \quad (58)$$

which applying $\mathbb{E}\{\cdot\}$ on both sides of the equation, can be rewritten as

$$\beta^{n,k}_u=1-\frac{\mathbb{E}\left\{\left|\angle(x^{n,k}_{u,ul})-\angle(\hat{x}^{n,k}_{u,ul})\right|^2\right\}}{2}. \quad (59)$$

Since $\mathbb{E}\left\{\left|\angle(x^{n,k}_{u,ul})-\angle(\hat{x}^{n,k}_{u,ul})\right|\right\}=0$, due to the fact that the mean of the angle of the error is 0 (due to the symmetry of the distribution of the received symbol and the symmetry of the DMSPK constellation), we have

$$\sigma^2_\phi=\mathbb{E}\left\{\left|\angle(x^{n,k}_{u,ul})-\angle(\hat{x}^{n,k}_{u,ul})\right|^2\right\}, \quad (60)$$

so joining (59) and (60) we have

$$\beta^{n,k}_u=1-\frac{\sigma^2_\phi}{2}\rightarrow\sigma_\phi=\sqrt{2(1-\beta^{n,k}_u)}, \quad (61)$$

which combined with (22) and via straightforward manipulations, results in

$$\sigma_\phi=\sqrt{\frac{2\left(N_uP^{n,k}_{u,ul}+(1-P^{n,k}_{u,ul})^{N_u}\right)}{(N_u-1)P^{n,k}_{u,ul}}}, \quad (62)$$

which may be substituted back in (37).

D. CHANNEL PREDICTION QUALITATIVE ANALYSIS

For a detailed analysis on how the channel prediction affects the performance under noisy channel estimations with CA effects the reader is referred to [19, Sec. IV]. We perform a qualitative analysis below to give some insights on the effect that it will have on the proposed pilot-less TDD massive MIMO system.

Combining (1) and (13) for $d=1$ into a single equation as done in [19, Eq. (20)], we have

$$\mathbf{h}^{n+1}_u=\alpha_1\hat{\mathbf{h}}^n_u+\mathbf{w}^{n+1}_{u,h}=\alpha_1\mathbf{h}^n_u\frac{x^n_{u,ul}}{\hat{x}^n_{u,ul}}+\underbrace{\left(\alpha_1\frac{\boldsymbol{\nu}^{n,k}_b}{\hat{x}^n_{u,ul}}+\mathbf{w}^{n+1}_{u,h}\right)}_{\mathbf{W}^n_u}, \quad (63)$$

which is simplified (assuming $x^n_{u,ul}=\hat{x}^n_{u,ul}$) as

$$\mathbf{h}^{n+1}_u=\alpha_1\hat{\mathbf{h}}^n_u+\mathbf{w}^{n+1}_{u,h}=\alpha_1\mathbf{h}^n_u+\mathbf{W}^n_u, \quad (64)$$

where $\mathbf{W}^n_u$ is an equivalent noise source.

Applying channel prediction results in a smaller α_1 in (64) and a $\mathbf{W}^n_u$ with a smaller variance than $1-\alpha_1^2(1+\sigma_b^2)$. Therefore, the damaging effects of noise in the estimation and CA in the precoding and filtering processes are decreased.

V. NUMERICAL RESULTS

In this section, we first validate the theoretical analysis previously shown in Sec. IV, we then show the performance of the UL and DL with different techniques applied, and last we compare the classical coherent scheme with the proposed pilot-less system in a realistic deployment scenario. Unless otherwise stated, $R=100$, $M_u=M_d=4$. The SNR in the simulations is defined as the inverse of the noise

$$P^{n,k}_u\approx 1-\frac{\int_{-\pi/M}^{\pi/M}\int_0^\infty e^{-\left(\frac{r\cos(\gamma)-\mu_{\Re\{z\}}}{\sqrt{2}\sigma_{\Re\{z\}}}\right)^2}e^{-\left(\frac{r\sin(\gamma)-\mu_{\Im\{z\}}}{\sqrt{2}\sigma_{\Im\{z\}}}\right)^2}rdrd\gamma}{2\pi\sigma_{\Re\{z\}}\sigma_{\Im\{z\}}} \quad (56)$$

TABLE 1. Ratio of T_c and OFDM T_s (n_c), for $v = 500$ km/h for different carrier frequencies f_c in GHz and carrier spacing Δ_f in KHz.

n_c	$\Delta_f = 15$	$\Delta_f = 30$	$\Delta_f = 60$	$\Delta_f = 120$	$\Delta_f = 240$
$f_c = 0.7$	7	15	29	-	-
$f_c = 3.6$	1.4	2.8	5.6	-	-
$f_c = 27$	-	-	-	1.5	3
$f_c = 54$	-	-	-	-	1.5

TABLE 2. Parameters, description and values utilized in the numerical results of Sec. V, unless otherwise stated.

Parameters	Description	Values
R	Number of antennas in the BS	100
M_u	Constellation size in the UL	4
M_d	Constellation size in the DL	4
SNR in the UL	Signal-to-noise ratio in the UL	$-10\log_{10}(\sigma_b^2)$
SNR in the DL	Signal-to-noise ratio in the DL	$-10\log_{10}(\sigma_u^2)$
σ_u^2	Noise power in DL	0.25
T_c	Coherence time	$0.15 f_D^{-1}$
α_d	Time correlation factor	0.13, 0.53, 0.87, 1
U	Number of users	2
τ_u	Uplink slot duration	1, 5
τ_d	Downlink slot duration	1, 5

power (SNR $= -10\log_{10}(\sigma_b^2)$ for the UL and SNR $= -10\log_{10}(\sigma_u^2)$ for the DL). In sections V-B and V-C, the classical and the proposed schemes are compared using Monte Carlo simulations. The coherence time n_c is defined in number of OFDM symbols of the TDD scheme, according to the frame shown in Figs. 2-3. It is calculated as $T_c = 0.15\, f_D^{-1}$ [26], where f_D is the maximum Doppler frequency. To implement time correlation effects we use the autocorrelation model of (2) in [19]. We also use the fact that the duration of an OFDM symbol is the inverse of the separation between subcarriers $T_s = 1/\Delta_f$. We can obtain the ratio of coherence time to the OFDM symbol duration as $n_c = T_c/T_s$, which is given in Table 1 for 5G scenarios with very high mobility at 500 km/h (maximum speed for 5G). Only the values that are compatible with the allowed combinations of carrier frequency (f_c) and subcarrier spacing (Δ_f) in the 5G standard are shown; otherwise they are marked with "-" in the table. The coherent scheme uses LS channel estimation with pilot symbol assisted modulation (PSAM), as done in [8].

A summary of the parameters for the numerical results utilized (unless otherwise stated) in the following sections are shown in Table 2.

A. CORROBORATION OF MSE AND UL/DL PERFORMANCE ANALYSIS

It can be seen in Fig. 4, that the theoretical analysis for the channel estimation error behaves as an upper bound with respect to the Monte Carlo simulations, which confirms the validity of the analysis. The theoretical results are an upper bound of the simulations and the lower bound is given by the performance of the PSAM with a pilot in every coherence time. That is, without any degradation due to time variability (at the expense of a large pilot overhead). It can be observed

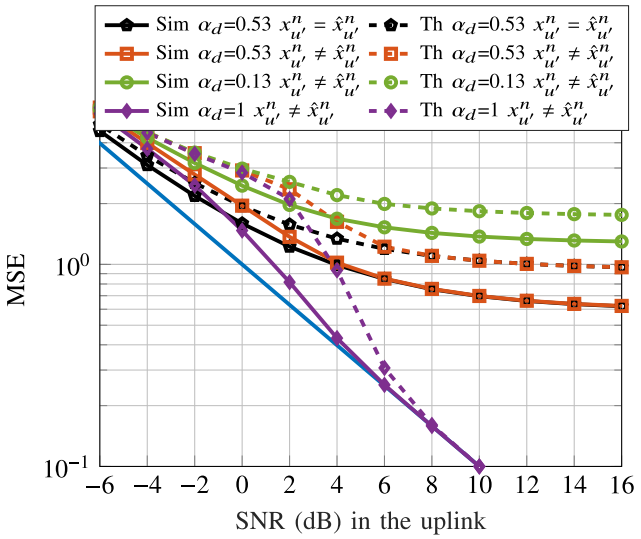

FIGURE 4. MSE of channel estimation for $M_{UL} = 16$ and $R = 100$. Continuous line is the Monte Carlo simulation while dashed line is the theoretical upper bound (obtained in Sec. IV) and pentagon curves are those with perfect detection ($x^n_{u'} = \hat{x}^n_{u'}$) with CA. Blue line is the MSE with perfect channel estimation ($\hat{\mathbf{h}}^n_u = \mathbf{h}^{n+d}_u$), and serves as a lowest bound.

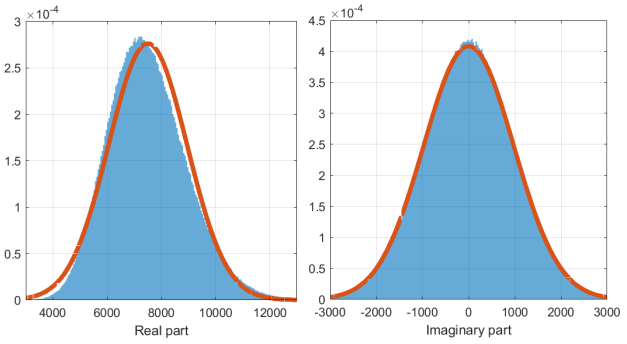

FIGURE 5. Monte Carlo histogram (blue) versus theoretical PDF (red) of the real (left) and imaginary (right) part of $z^{n,k}_{u,dl}$ for 2 users, $s^{n,k}_{u,dl} = 1$, $R = 100$, $M_{DL} = 4$, $\alpha_d = 0.87$, $\sigma_b^2 = 0.5$ and $\sigma_u^2 = 0.25$.

that a correlation α_d caused by time difference between the estimated channel and the real one results in a MSE floor, caused by the time variability of the channel. When there is no time variability, the MSE is affected only by the error probability in the detection of (10). It is worth noting that the error probability is also affected by the imperfect spatial multiplexing caused by $\hat{\mathbf{h}}^n_u \neq \mathbf{h}^{n+d}_u$. Last but not least, the case of PSAM with time variability is shown, were it can be seen that it is below the proposed channel estimation with channel time variability, but is equal to it for high SNR. This is caused by the fact that a high SNR makes no error probability of the symbol detection and thus it is equivalent to using PSAM. Additionally, and to show that the mathematical analysis is correct, in Fig. 5 we have included the histograms of the real and imaginary parts of $\mathbf{z}^{n,k}_{u,dl}$ for 2 users, $\mathbf{s}^{n,k}_{u,dl} = 1$, $R = 100$, $M_{UL} = 4$, $\alpha_d = 0.87$, $\sigma_b^2 = 0.5$ and $\sigma_u^2 = 0.25$.

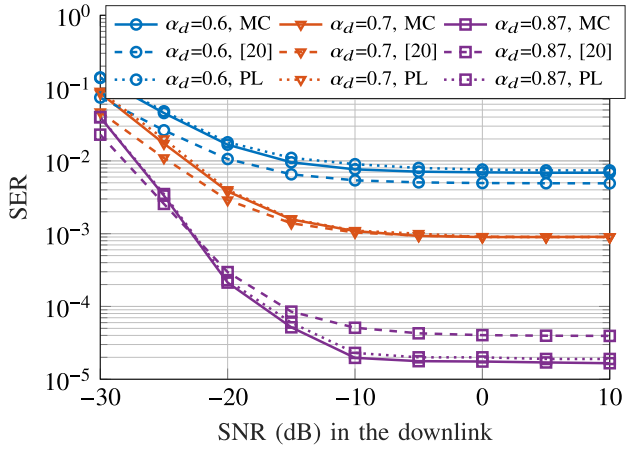

FIGURE 6. Comparison of analysis of DL performance in [23] (dashed line) and this paper (PL, for pilot-less, dotted line) with respect to Monte Carlo (MC, continuous line). $U = 2$, $M_{DL} = 4$ and $R = 100$.

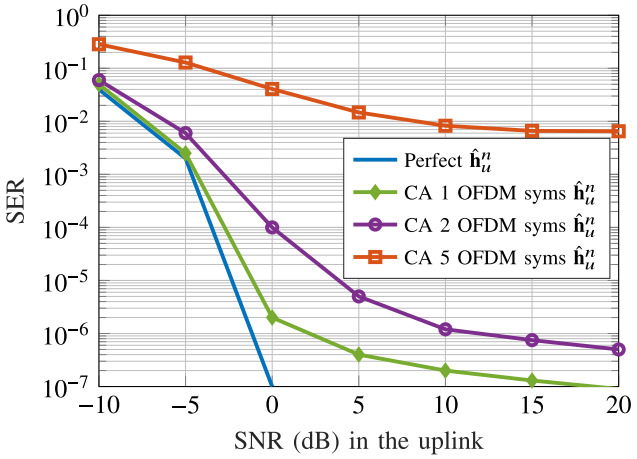

FIGURE 7. UL performance for $M_{UL} = 16$, $R = 100$ and 2 users for the proposed scheme for a coherence time $n_c = 4$, for different CA effects (time difference in OFDM symbols between the estimated and the compensated channel).

A very good agreement is observed between the theoretical and the real PDFs, with a lower agreement in the real part. The approximations from variance gamma to normal distributions are the main reason behind this discrepancy.

Last, Fig. 6 includes additional results to compare the theoretical analysis with the Monte Carlo simulations of the analysis done in the preliminary work [23] and the analysis in this manuscript. To perform a fair comparison, since the assumptions made for the analysis are different, we have adapted the results in [23] by utilizing (64). In this case $\alpha_d = \sqrt{1 - e_d^2}$ and $1 - \alpha_d^2(1 + \sigma_b^2) = e_d^2$ for the variances of the channel components of [23, (5)]. From Fig. 6, we can see that the dotted line is very close to the Monte Carlo simulations, and thus, we can clearly conclude that the analysis in this work is improved with respect to that in [23].

B. SER OF THE UL AND THE DL OF THE PROPOSED SCHEME

By looking at Fig. 7, we can see the performance of the UL for a coherence time $n_c = 4$ for a perfect channel estimation, and for an outdated channel estimation that suffers from channel aging of 1, 2 and 5 OFDM symbols. The curve "Perfect $\hat{\mathbf{h}}_u^n$" assumes perfect CSI, so it serves as a benchmark of the performance of the UL. The curve CA 1 OFDM symbol represents $\hat{\mathbf{h}}_u^n \neq \mathbf{h}_u^{n+1}$ and shows the performance when the channel estimated in a certain time instant is utilized in the immediate next time instant, which happens in the UL slot for all symbols, assuming the data is correctly estimated in the slot used for channel estimation. The curve CA 2 OFDM symbol represents $\hat{\mathbf{h}}_u^n \neq \mathbf{h}_u^{n+2}$ and shows the performance when the channel is estimated with correct data and used for filtering 2 time instants later. This happens for all the UL symbols in the configuration shown in Fig. 1. The curve CA 5 OFDM symbol represents $\hat{\mathbf{h}}_u^n \neq \mathbf{h}_u^{n+5}$, which is obtained for a configuration with a DL slot of 4 symbols, and shows the performance of the immediate next UL slot after a DL slot. In this case, the performance is largely degraded due to CA. In summary, it is shown how the larger the time between the estimated and the compensated channel, results in a decrement in performance, with it being much worse in case the compensated UL symbol is further than the coherence time. This information is useful to define a proper symbol placing strategy and the required TDD slots' duration for UL and DL so that the CA effects are controlled and a good performance is obtained.

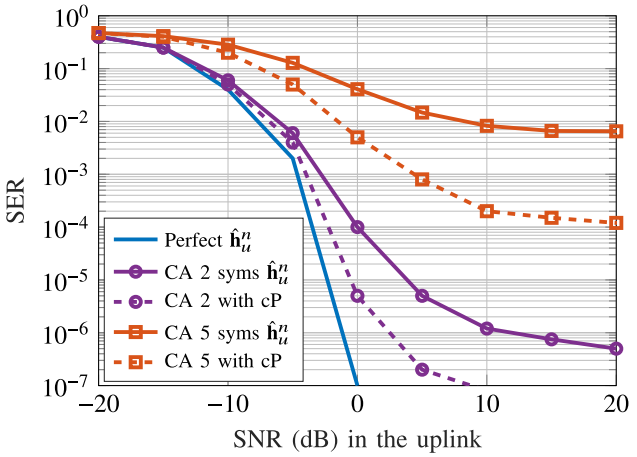

FIGURE 8. UL performance for $M_{UL} = 4$, $R = 100$ and 2 users for the proposed scheme for a coherence time $n_c = 4$, for different d values. "CA" refers to channel aging of OFDM symbols and "cP" refers to channel prediction.

To show the power of the channel prediction proposed in Sec. III-D, we include Fig. 8. It shows 2 curves for each configuration, where "cP" means channel prediction applied. It can be seen that the performance can be largely improved when both techniques are applied, with the channel prediction affecting the most.

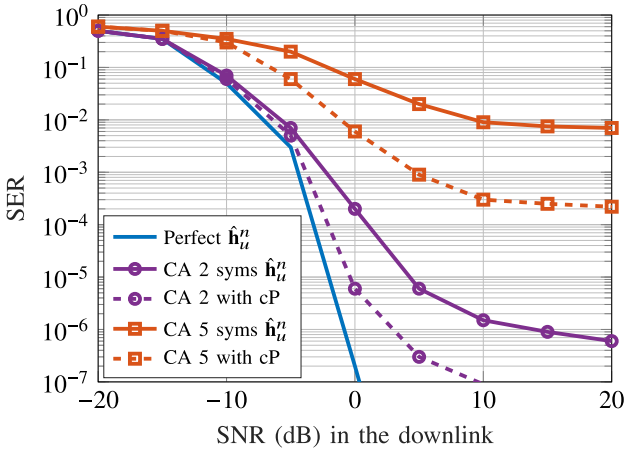

FIGURE 9. DL performance for $M = 4$, $R = 100$ and 2 users for the proposed scheme for a coherence time $n_c = 4$, for different d values and assuming perfect $\hat{x}_{u,ul}^n$ reconstruction. "CA" refers to channel aging of OFDM symbols and "cP" refers to channel prediction.

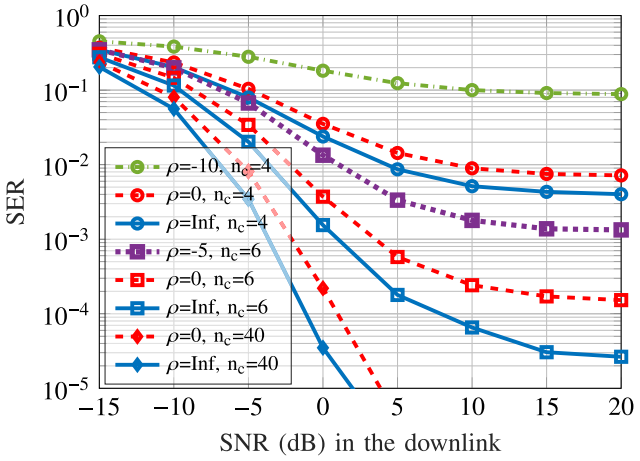

FIGURE 10. DL performance for $M = 4$, $R = 100$ and 2 users for the proposed scheme for different coherence times n_c, and different SNR values ρ in the UL.

Please note that the performance for the DL is very similar to that of the UL since the channels are reciprocal and the equations are very similar, except for the noise term as we stated in the analysis in Sec. IV. Nevertheless, we include some results in Fig. 9 for the DL for illustrations purposes where one can extract the same conclusions as for the UL in Figs. 7 and 8. In all Figs. 8, 7 and 9, it is assumed perfect reconstruction of $\hat{x}_{u,ul}^n = x_{u,ul}^n$.

Additionally, and to show how it is important to control the UL reconstruction error of $\hat{x}_{u,ul}^n$, we show in Fig. 10 the DL performance for TDD slot periods of 4 OFDM symbols for each UL and DL for different coherence times, without utilizing channel prediction and for different SNRs in the UL (referred as ρ in the figure). It can be seen how the lower the SNR, the worse the reconstruction of $\hat{x}_{u,ul}^n$, which results in a worse channel estimation (as also shown in Fig. 4), and thus in a worse DL performance for the same coherence time and SNR in the downlink. Besides, a lower coherence time results in a decrement in performance since the CA effect is more stringent, which can be observed by looking at the performance of a certain ρ and different n_c values. It is remarkable that the performance is close to that of static channels for n_c around 2 times the DL slot duration. It is worth noting that applying channel prediction improves the performance as it happens with the UL, but we are not including the results for space economy and avoidance of repetitive results.

C. COMPARISON WITH CLASSICAL COHERENT PSAM

In order to compare the classical coherent scheme and the proposed pilot-less scheme, we will use the SER as a comparison metric, and both schemes will use channel prediction and data detection improvement. Since the overhead is different in them, for a fair comparison we will ensure the same spectral efficiency in both systems. For such purpose, different constellation sizes are used. For instance, in a TDD slot of 4 OFDM symbols, where 2 are used for the uplink and the other 2 for the downlink, half of the OFDM symbols should be used for pilots in the uplink and only one in the whole OFDM DL slot should be used for downlink. In this case, for a fair comparison, the size of the constellation used in the coherent scheme should be the square of that used in the non-coherent scheme for the UL, and the same size for the DL. Alternatively, a complementary data efficiency compensation consists on modifying the effective SNR as

$$SNR_{eff} = \frac{data_{eff}^{coh}}{data_{eff}^{nc}} SNR, \tag{65}$$

where $data_{eff}^{coh}$ and $data_{eff}^{nc}$ are the data efficiency of the classical and the proposed schemes respectively. The data efficiency is calculated as the number of data resources over the total resources.

The classical coherent scheme is more limited than the proposed pilot-less, as show in Fig. 11. The number of symbols in a TDD slot period for the UL τ_u and the DL τ_d are set to 1 and 2 to see the dependence with this parameter, and the constellation sizes to 4 and 16. The classical (coherent) and the proposed (non-coherent) schemes are referred in Figs. 11 and 12 as C and N, respectively.

It is worth noting that in Fig. 11, we are performing the channel compensation in the immediate next symbol to that where the channel has been really predicted, this is, there is an actual channel measurement in the previous symbol. In this sense, the performance is valid for both the UL symbols in the same UL TDD slot in Fig. 1 and for the DL of Fig. 2. By looking at such figure, it can be observed how, for large n_c (10 in Fig. 11), the scheme C is approximately the same as the scheme N, which reinforces the validity of the proposed scheme even for large n_c. For very fast varying channels ($n_c = 2$), the latter outperforms the former for the same spectral efficiency. More concretely, for $\tau_u = \tau_d = 2$, the scheme C utilizes one symbol pilot while the scheme N does not. Thus, the scheme C transmits a 16-QAM while the

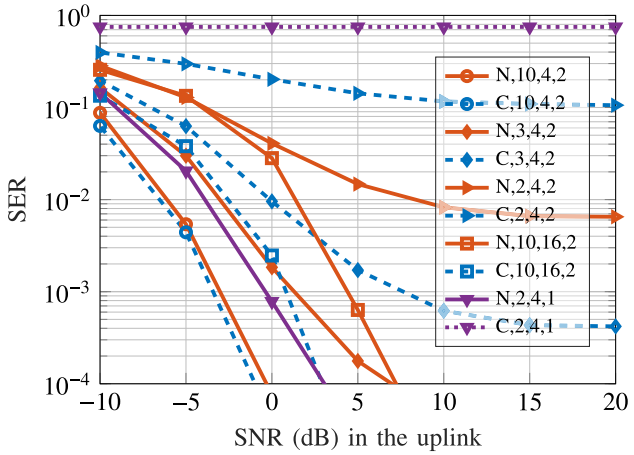

FIGURE 11. Comparison between classical (C, dashed) and proposed (N, continuous) schemes, labelled from left to right with the legend written as "technique (N,C), n_c (2,3,10), M_{UL} (4,16), $\tau_u = \tau_d$ (1,2)" for $R = 100$ for $d = 1$ ($\hat{\mathbf{h}}_u^n \neq \mathbf{h}_u^{n+1}$) for 2 users.

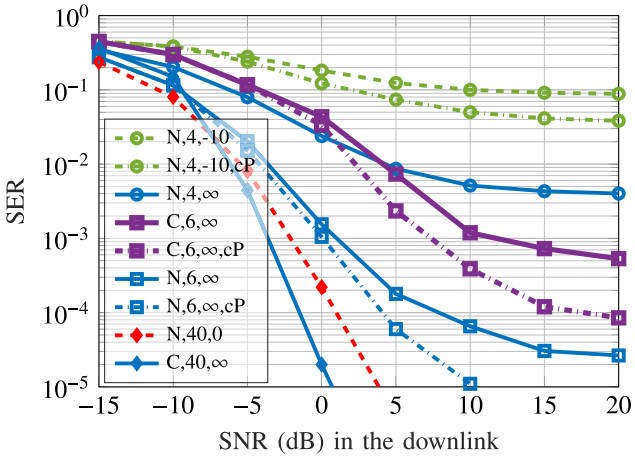

FIGURE 12. Comparison between classical (C, dashed) and proposed (N, continuous) schemes in the DL, labelled from left to right with the legend written as "technique (N,C), n_c (4,6,40), SNR (dB) uplink for channel estimation" for $R = 100$, $\tau_d = 4$, $M_{DL} = 4$ and 2 users.

scheme N transmits a QPSK. For both schemes transmitting a QPSK (scheme C with a lower spectral efficiency than scheme N) and for $n_c = 3$, scheme N outperforms scheme C, which reinforces our proposal. In fact, scheme N can go down to $\tau_u = \tau_d = 1$, while C cannot since then, only pilots would be sent both in UL and DL (with a related null efficiency) and the consequent $SNR_{eff} = 0$, giving the results named "C,2,4,1" in Fig. 11. This is an advantage for the proposed scheme N which can work in extremely fast varying channels with a very good data efficiency.

For a larger delay between the estimated and compensated channel, which is present in the DL of larger TDD slots such as those shown in Fig. 1, the performance worsens. We have different coherence times and we check the performance of the DL with two SNR values of the UL, one sufficiently large to avoid any errors in the UL detection and one that results in some errors, to check if the detection improvement and the channel prediction can also help the scheme in the DL even when there is an error in the UL detection. By looking at Fig. 12, we can see that the proposed scheme (N) works better than the coherent scheme (C) in case the coherence time n_c is smaller than 2 times the TDD slot duration. It is worth noting that the SNR of the uplink of the channel estimation for the coherent scheme affects the MSE as the PSAM in Fig. 4. Even with channel prediction, the coherent performs worse than the proposed scheme, when the latter does not use channel prediction in case the coherence time is 1.5 times the DL slot duration (please look at curves C,6,∞,cP and N,6,∞). This comes from the fact that the proposed scheme is much more robust than the coherent scheme in these scenarios. Besides this, for the proposed scheme, when the SNR does not ensure perfect UL detection, the performance of the DL is degraded. This comes from the fact that the estimated and compensated channels differ more than in case only channel aging is considered. Last but not least, when applying channel prediction (cP), the performance improves in all cases.

D. DISCUSSION OF THE RESULTS

In this subsection we provide a discussion of the obtained results, to serve as a summary that can be useful for the reader.

The analysis of the MSE and the UL/DL performance is corroborated in Figs. 4, 5 and 6. By looking at those figures, it can be clearly seen that the numerical results obtained with Monte Carlo simulation match the theoretical analysis very well. More concretely, the MSE of the channel error shown in Fig. 4 for different time correlation factors (α_d) and SNRs, shows that the theoretical analysis serves as an upper bound of the Monte Carlo results, as was expected from the explanation given in Sec. IV. Figs. 5 and 6 show the histogram of the DL received symbol and performance respectively. It can be seen that there is a good agreement between the analysis and the Monte Carlo simulations, and the discrepancies mainly come from the distribution of the real part, which is not exactly the same between the simulation and the analysis. This discrepancy comes from the approximation of the variance gamma to the normal distribution made in (51)-(53).

In Figs. 7, 8, 9 and 10 we can see the UL and DL performance of the proposed scheme. It can be seen that the larger the CA value, the worse the performance, which makes sense due to the fact that the estimation time instant and the compensation time instant are further apart, so the channel differs more. Nevertheless, this effect is partially compensated by the channel prediction (cP), which improves the overall performance. Also, it is shown how a worse performance in the UL degrades the channel estimation and thus affects the DL performance negatively. Therefore, it is important to properly design the UL given the channel characteristics. However, this is out of the scope of this work.

Last, Figs. 11 and 12 show a comparison in performance for the UL and the DL of the proposed scheme and the classical PSAM based scheme (coherent, C in the figures). It can be clearly seen that the faster the channel changes

(smaller n_c) and the smaller the SNR (ρ), or both, the worse the performance of the classical PSAM based scheme. For slowly varying channels or high SNR, the performance of the proposed scheme is still close to that of the coherent scheme. This demonstrates the feasibility and the validity of the proposed scheme based on estimating the channel utilizing differentially encoded data in the UL and combining it with channel prediction for estimation improvement.

VI. CONCLUSION

In this work, we propose a pilot-less massive MIMO TDD scheme where the data is detected via non-coherent processing utilizing a differential MPSK and spatial multiplexing is included in both the UL and the DL, with the channel being estimated using the data received in the UL. We have proposed several symbol placing strategies over an OFDM grid and, to improve the performance a channel prediction has been proposed by means of a linear Wiener predictor. The proposed scheme has the main advantage that it avoids the use of pilot signals for channel estimation and only needs a few reference signals for the MDPSK data streams in both the UL and DL. This is a great benefit with respect to classical PSAM based TDD massive MIMO since it maintains a very good data efficiency even for very fast changing channels.

We have analyzed the MSE of the channel estimation for time-varying channels using the reconstructed differentially encoded data of the UL, and the performance of using spatial multiplexing together with differentially encoded data in both the UL and the DL. Furthermore, we have performed a qualitative analysis of the advantages brought by the channel prediction. The analysis has been validated via simulations. Additionally, we have simulated the performance of the four different strategies utilized for the proposed pilot-less scheme where we show that the system largely benefits from performing channel prediction. Last but not least, the proposed pilot-less scheme has been compared with the PSAM based coherent counterpart to demonstrate that our proposal is better and much more convenient than the classical one for fast varying channels.

APPENDICES–MATHEMATICAL PRELIMINARIES

In section IV, the properties of VG and Gamma distributions are used, which can be found in [30], [31], and [32]:

- The product of two zero mean uncorrelated normal random variables X and Y with standard deviation σ_X and σ_Y respectively, follows a variance-gamma distribution with parameters $VG(1, 0, \sigma_X\sigma_Y, 0)$.
- A $CVG(k, 0, \sigma, 0)$ is a circularly symmetric distribution in which the real and imaginary parts are each distributed as $VG(k, 0, \sigma, 0)$.
- The sum of R variables distributed according to $VG(k, 0, \sigma_i, 0)$ is distributed according to

$$VG\left(kR, 0, \sqrt{\frac{\sum_{i=1}^{R}\sigma_i^2}{R}}, 0\right).$$

- Scaling a $VG(1, 0, \sigma, 0)$ random variable by any parameter k results in a variable distributed as $VG(1, 0, k\sigma, 0)$.
- A distribution $VG(R, 0, \sigma, 0)$ with $R \to \infty$ can be regarded as a $\mathcal{N}(0, R\sigma^2)$.
- If a variable that follows a Gamma distribution $\Gamma(\delta, \beta)$ is scaled by a parameter k, the new variable is distributed as $\Gamma(\delta, k\beta)$.

The distribution of the sum of n independent Gamma variables $\Gamma(\delta_i, \beta_i)$ is approximated as a $\Gamma(\delta_m, \beta_m)$ with

$$\delta_m = \frac{\left(\sum_{i=0}^{n}\delta_i\beta_i\right)^2}{\sum_{i=0}^{n}\delta_i\beta_i^2}, \quad \text{and} \quad \beta_m = \frac{\sum_{i=0}^{n}\delta_i\beta_i^2}{\sum_{i=0}^{n}\delta_i\beta_i}, \tag{66}$$

and when $\delta_m \gg \beta_m$ (true when there are several APs with small correlation), according to [33],

$$\Gamma(\delta_m, \beta_m) \approx \mathcal{N}\left(\delta_m\beta_m, \delta_m\beta_m^2\right) = \mathcal{N}\left(\sum_{i=0}^{n}\delta_i\beta_i, \sum_{i=0}^{n}\delta_i\beta_i^2\right). \tag{67}$$

Besides this, $\Re$ and $\Im$ are the real and imaginary parts of a random variable, and thus we can decompose the product of two complex variables as

$$\begin{aligned}&(\Re_1 + j\Im_1)(\Re_2 + j\Im_2)\\ &\quad = (\Re_1\Re_2 - \Im_1\Im_2) + j(\Re_1\Im_2 + \Im_1\Re_2),\end{aligned} \tag{68}$$

composed of the sums and products of real normal variables, distributed according to

$$\Re_1\Re_2 - \Im_1\Im_2 \sim VG\left(2, 0, \sqrt{\frac{\sigma_{\Re_1}^2\sigma_{\Re_2}^2 + \sigma_{\Im_1}^2\sigma_{\Im_2}^2}{2}}, 0\right), \tag{69}$$

$$\Re_1\Im_2 + \Im_1\Re_2 \sim VG\left(2, 0, \sqrt{\frac{\sigma_{\Re_1}^2\sigma_{\Im_2}^2 + \sigma_{\Im_1}^2\sigma_{\Re_2}^2}{2}}, 0\right). \tag{70}$$

APPENDICES–P_s DOUBLE INTEGRAL TO SINGLE INTEGRAL

The double integral of (56) can be simplified by doing the following change of variables

$$\begin{aligned}a &= \frac{\left(\sigma_{\Re\{z\}}^2 - \sigma_{\Im\{z\}}^2\right)\sin^2\gamma + \sigma_{\Im\{z\}}^2}{2\sigma_{\Re\{z\}}^2\sigma_{\Im\{z\}}^2},\\ b &= \frac{2\sigma_{\Im\{z\}}^2\mu_{\Re\{z\}}\cos\gamma}{2\sigma_{\Re\{z\}}^2\sigma_{\Im\{z\}}^2}, \quad c = \frac{\sigma_{\Im\{z\}}^2\mu_{\Re\{z\}}^2}{2\sigma_{\Re\{z\}}^2\sigma_{\Im\{z\}}^2},\end{aligned} \tag{71}$$

and rewriting (56) as below we can solve it

$$\begin{aligned}&\int_0^\infty re^{-ar^2+bx-c}dr\\ &\quad = e^{-c}\frac{\sqrt{\pi}be^{\frac{b^2}{4a}}\left(\operatorname{erfc}\left(\frac{b}{2\sqrt{a}}\right)+1\right)+2\sqrt{a}}{4a^{3/2}},\end{aligned} \tag{72}$$

so the double integral of (56) turns to the integral in (57).

REFERENCES

[1] *5G End to End Key Performance Indicators (KPI)*, Standard 28.554, 3GPP, 2021.

[2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 1–9, Oct. 2019.

[3] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[4] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO-5G system," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 905–923, 2nd Quart., 2016.

[5] F. A. P. De Figueiredo, F. A. C. M. Cardoso, I. Moerman, and G. Fraidenraich, "Channel estimation for massive MIMO TDD systems assuming pilot contamination and frequency selective fading," *IEEE Access*, vol. 5, pp. 17733–17741, 2017.

[6] J. Flordelis, F. Rusek, F. Tufvesson, E. G. Larsson, and O. Edfors, "Massive MIMO Performance—TDD versus FDD: What do measurements say?" *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2247–2261, Apr. 2018.

[7] X. Jiang, A. Decurninge, K. Gopala, F. Kaltenberger, M. Guillaud, D. Slock, and L. Deneire, "A framework for over-the-air reciprocity calibration for TDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 5975–5990, Sep. 2018.

[8] M. J. Lopez-Morales, K. Chen-Hu, and A. Garcia-Armada, "Differential data-aided channel estimation for up-link massive SIMO-OFDM," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 976–989, 2020.

[9] M. Chowdhury, A. Manolakos, and A. J. Goldsmith, "Coherent versus noncoherent massive SIMO systems: Which has better performance?" in *Proc. IEEE Int. Conf. Comms. (ICC)*, Jun. 2015, pp. 1691–1696.

[10] H. Iimori, G. T. F. de Abreu, and K. Ishibashi, "Full-duplex MIMO systems with hardware limitations and imperfect channel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.

[11] S. Malkowsky, J. Vieira, L. Liu, P. Harris, K. Nieman, N. Kundargi, I. C. Wong, F. Tufvesson, V. Öwall, and O. Edfors, "The world's first real-time testbed for massive MIMO: Design, implementation, and validation," *IEEE Access*, vol. 5, pp. 9073–9088, 2017.

[12] R. Chopra, C. R. Murthy, H. A. Suraweera, and E. G. Larsson, "Blind channel estimation for downlink massive MIMO systems with imperfect channel reciprocity," *IEEE Trans. Signal Process.*, vol. 68, pp. 3132–3145, 2020.

[13] K. Mawatwal, D. Sen, and R. Roy, "Performance analysis of a SAGE-based semi-blind channel estimator for pilot contaminated MU massive MIMO systems," *IEEE Access*, vol. 8, pp. 46682–46700, 2020.

[14] C. Hu, H. Wang, and R. Song, "Pilot decontamination in multi-cell massive MIMO systems via combining semi-blind channel estimation with pilot assignment," *IEEE Access*, vol. 8, pp. 152952–152962, 2020.

[15] W.-C. Huang, C.-P. Li, and H.-J. Li, "On the power allocation and system capacity of OFDM systems using superimposed training schemes," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1731–1740, May 2009.

[16] J. C. Estrada-Jiménez, K. Chen-Hu, M. J. F.-G. García, and A. G. Armada, "Power allocation and capacity analysis for FBMC-OQAM with superimposed training," *IEEE Access*, vol. 7, pp. 46968–46976, 2019.

[17] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive MIMO channel prediction: Kalman filtering Vs. machine learning," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 518–528, Jan. 2021.

[18] C. Wu, X. Yi, Y. Zhu, W. Wang, L. You, and X. Gao, "Channel prediction in high-mobility massive MIMO: From spatio–temporal autoregression to deep learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1915–1930, Jul. 2021.

[19] K. T. Truong and R. W. Heath, Jr., "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, Aug. 2013.

[20] A. G. Armada and L. Hanzo, "A non-coherent multi-user large scale SIMO system relaying on M-ary DPSK," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2517–2522.

[21] K. Chen-Hu, Y. Liu, and A. G. Armada, "Non-coherent massive MIMO-OFDM down-link based on differential modulation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11281–11294, Oct. 2020.

[22] H. A. Shehadeh, I. H. Jebril, X. Wang, S.-C. Chu, and M. Y. I. Idris, "Optimal topology planning of electromagnetic waves communication network for underwater sensors using multi-objective optimization algorithms (MOOAs)," *Automatika*, vol. 1, pp. 1–12, Sep. 2022.

[23] M. J. L. Morales, K. Chen-Hu, and A. G. Armada, "Pilot-less massive MIMO TDD system with blind channel estimation using non-coherent DMPSK," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Jun. 2022, pp. 1–6.

[24] T. Hwang, C. Yang, G. Wu, S. Li, and G. Y. Li, "OFDM and its wireless applications: A survey," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1673–1694, May 2009.

[25] M. J. L. Morales, K. Chen-Hu, and A. G. Armada, "Effect of spatial correlation on the performance of non-coherent massive MIMO based on DMPSK," in *Proc. IEEE Global Comms. Conf.*, Dec. 2021, pp. 1–6.

[26] T. S. Rappaport, *Wireless Communications: Principles and Practice*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[27] M. Vu and A. Paulraj, "On the capacity of MIMO wireless channels with dynamic CSIT," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1269–1283, Sep. 2007.

[28] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels (mobile radio)," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[29] K. Chen-Hu, Y. Liu, and A. G. Armada, "Non-coherent massive MIMO-OFDM for communications in high mobility scenarios," *ITU J. Future Evolving Technol.*, vol. 1, no. 1, pp. 13–24, 2020.

[30] R. Gaunt, "Variance-gamma approximation via Stein's method," *Electron. J. Probab.*, vol. 19, p. 33, Jan. 2014, doi: 10.1214/EJP.v19-3020.

[31] R. E. Gaunt, "A note on the distribution of the product of zero-mean correlated normal random variables," *Statisticsa Neerlandica*, vol. 73, no. 2, pp. 176–179, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/stan.12152, doi: 10.1111/stan.12152.

[32] S. Covo and A. Elalouf, "A novel single-gamma approximation to the sum of independent gamma variables, and a generalization to infinitely divisible distributions," *Electron. J. Statist.*, vol. 8, no. 1, pp. 894–926, Jan. 2014.

[33] L. M. Leemis and J. T. McQueston, "Univariate distribution relationships," *Amer. Statistician*, vol. 62, no. 1, pp. 45–53, 2008.

MANUEL J. LOPEZ-MORALES (Member, IEEE) received the B.Eng. degree (Hons.) in electrical engineering from the Polytechnic University of Madrid, in 2015, and the double M.Sc. degree (Hons.) in telecommunications engineering and wireless communications from the Polytechnic University of Madrid and Lund University, respectively, in 2018. He is currently pursuing the Ph.D. (M.S.C.A.) degree with the Universidad Carlos III de Madrid through the TeamUp5G Project. He has been a Visiting Student with the Microwave Vision Group, Satimo, Italy, in 2014, the Tokyo Institute of Technology, Japan, in 2016, Ericsson AB, Sweden, in 2017, ISAE Supaero, France, in 2018, Telefonica Research and Development, Spain, in 2020, UNL, Lisbon, Portugal, in 2021, and Nokia, Spain, in 2022. His research interests include antenna technology, signal processing, massive MIMO, and wireless communications.

ANA GARCIA-ARMADA (Senior Member, IEEE) is currently a Professor with the Universidad Carlos III de Madrid (UC3M), Spain. She has been a Visiting Scholar at Stanford University, the Bell Laboratories, and the University of Southampton. She has published more than 200 papers in conferences and journals and she holds five patents. Her research interest includes signal processing applied to wireless communications. She serves on the Editorial Board of IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, and *ITU Journal on Future and Evolving Technologies*. She has been a member of the Organizing Committee of several conferences, including IEEE GLOBECOM 2021 as the General Chair. She has received several awards from UC3M, the Third Place Bell Laboratories Prize 2014, the Outstanding Service Award from the IEEE ComSoc Signal Processing and Communications Electronics Technical Committee, and the Outstanding Service Award from the IEEE ComSoc Women in Communications Engineering Standing Committee.