IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 26, 2019, accepted August 13, 2019, date of publication August 15, 2019, date of current version August 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2935686

On the Distribution of an Effective Channel Estimator for Multi-Cell Massive MIMO

FELIPE A. P. DE FIGUEIREDO[1], CLAUDIO FERREIRA DIAS[2], FABBRYCCIO A. C. M. CARDOSO[3], AND GUSTAVO FRAIDENRAICH[2]

[1]IDLab, Department of Information Technology, Ghent University – imec, 9052 Ghent, Belgium
[2]DECOM/FEEC–State University of Campinas (UNICAMP), Campinas, Brazil
[3]CPqD–Research and Development Center on Telecommunication, Campinas 13086-902, Brazil

Corresponding author: Felipe A. P. de Figueiredo (felipe.pereira@ugent.be)

ABSTRACT Accurate channel estimation is of utmost importance for massive MIMO systems to provide significant improvements in spectral and energy efficiency. In this work, we present a study on the distribution of a simple but yet effective and practical channel estimator for multi-cell massive MIMO systems suffering from pilot-contamination. The proposed channel estimator performs well under moderate to aggressive pilot contamination scenarios without previous knowledge of the inter-cell large-scale channel coefficients and noise power, asymptotically approximating the performance of the linear MMSE estimator as the number of antennas increases. We prove that the distribution of the proposed channel estimator can be accurately approximated by the circularly-symmetric complex normal distribution, when the number of antennas, M, deployed at the base station is greater than 10.

INDEX TERMS Massive MIMO, multi-cell, pilot-contamination, channel estimation.

I. INTRODUCTION

The foreseen demand increase in data rate has triggered a research race for discovering new ways to increase the spectral efficiency of the next generation of mobile and wireless networks [1]. The report in [1] predicts that data rates will easily reach peaks of 10 Gbps, reaching a staggering 49 Exabytes of data transfer per month. Additionally, the next generation of networks is also expected to serve a higher number of devices, including human-type communication (HTC) devices and machine-type communication devices (MTC). One of the performance requirements defined by the International Mobile Communications (IMT) requires a connection density of 1×10^6 devices/km^2 for a network to be considered 5G [2]. According to [3], the number of connected devices is forecast to be nearly 30 billion by 2023, where around 20 billion are forecast to be MTC devices.

In that sense, several approaches have been proposed to increase the spectral efficiency of such next-generation networks, such as cell densification through the reduction of the cell size [4], high-order modulations such as 128-QAM or 256-QAM [5], larger transmission bandwidths available on the millimeter-wave bands and aggregation of industrial, scientific and medical (ISM) bands [6], [7] simultaneous access using a single time-frequency resource employing non-orthogonal multiple access (NOMA) through different transmission power allocation or spreading-codes [8], and simultaneous multiple users access in the same time-frequency resources by exploiting a higher number of degrees-of-freedom provided by massive MIMO techniques [9]. Among all these approaches, massive MIMO is one of the most important, prominent and studied ones, and is already established on the initial deployments of the 5-th generation of mobile networks (5G) [10].

In a multi-cell massive MIMO system, the base station (BS) has to estimate the wireless channels of all its connected devices. These channel estimates obtained during the uplink communication are used to calculate the decoding and the pre-coding matrices, which are used to receive and transmit user data, respectively. Therefore, accurate estimation of the channels is an essential task in such systems. The linear minimum mean square error (LMMSE) channel estimator is optimum when it comes to MSE minimization [11]. However, the channel statistics at the BS must be known beforehand as a premise for the estimator work properly, *i.e.*, it needs to have previous knowledge of both the intra/inter-cell large-scale fading coefficients and the noise power [12].

The optimum estimation strategy of the intra/inter-cell large-scale fading coefficients and noise power might be unjustified in practice due to the excessive overhead it

The associate editor coordinating the review of this article and approving it for publication was Chenhao Qi.

imposes on the system [12]. For instance, in the case where there are L cells serving K users, each one of the L-th BSs would have to acquire $K(L-1)$ inter-cell large-scale fading coefficients plus the noise power estimation. In [12], we proposed a channel estimator for multi-cell multi-user systems that does not require the estimation of the inter-cell large-scale fading coefficients plus noise power. In this work, we find expressions for the probability density function (PDF), cumulative distribution function (CDF), and central moments of the proposed channel estimator in [12]. The statistics derived and discussed in this work are used to assess the difference between the proposed estimator and the LMMSE estimator, which is the optimum estimator when both the channels and noise vectors are jointly Gaussian distributed [13]. Additionally, the derived statistics presented here, make it possible to find bounds to the spectral efficiency of a system employing the proposed estimator and compare its performance with that of a system employing optimal LMMSE channel estimation.

This article is organized as follows. In section II we list and discusses some related pieces of work. In section III, we present the uplink channel model and define the method for tackling the problem. In section IV, we describe the uplink training scheme employed in this study. In section V we briefly present two of the most well-known channel estimators in the Massive MIMO literature. Next, in section VI, we present a simple and effective channel estimator for the multi-cell scenario. In section VII, we present a thorough study on the distribution of the proposed channel estimator, which is the main contribution of this work. Then, in section VIII, we discuss the results of simulations proving the findings of section VII. Finally, in section IX, we present our final conclusions.

II. RELATED WORK

Here we present and discuss some related work on the channel estimation subject. Channel estimation for massive MIMO systems can be split into three main approaches, namely, non-coherent detection, data-aided and pilot-based channel estimation.

Coherent detection in massive MIMO systems requires that devices send training signals (*i.e.*, pilot sequences) towards the BS so that it can estimate the channels from the devices to each one of the antennas at the BS. However, the transmission of pilot signals increases the signaling overhead considerably as they occupy part of the coherence-interval that could be used for uplink or downlink data transmission. Additionally, as each one of the devices has to be assigned an orthogonal pilot sequence, the interval used by the pilots has to be increased as the number of devices increases. Moreover, coherent detection is susceptible to performance degradation caused by pilot estimation errors and pilot-contamination [14], [15]. Therefore, the first approach discussed here is the non-coherent detection, where in fact, there is no channel estimation involved. With this approach, the disadvantages mentioned earlier disappear, however, other ones arise. Non-coherent detection is based on the careful design of the modulation constellations and allows for systems approaching the channel capacity of their coherent-detection counterparts at moderate to high signal-to-noise ratios (SNR) regimes [16], [17]. The main disadvantages presented by non-coherent detection are a 3 dB penalty in SNR and less design freedom, once the transmitted signal has to be carefully designed based on some constraints necessary for the non-coherent detection (*e.g.*, isotropically-distributed Grassmannian constellations) [18], [19].

Another approach employed for estimation is the data-aided channel estimation, which is a technique where decoded modulation symbols, carrying data, are used to estimate the channel [20]. The body of literature shows approaches where data-aided techniques are used alone and others where it helps to improve the quality of the pilot-based channel estimates [21], [22]. Some works present results showing that it is possible to achieve high spectral efficiency with data-aided schemes using a very small number of pilot sequences [23], [24]. Other works show that iterative joint channel estimation and data decoding has the potential to effectively mitigate the pilot-contamination effect [20], [25], which is a phenomenon inherent to massive MIMO systems [20]. The adoption of hard-decoded symbols for channel estimation might cause significant performance degradation due to the fact that many of the hard decisions are highly probable to be incorrect, especially in low SNR cases, and therefore, producing channel estimates with low accuracy, directly affecting the achievable spectral-efficiency.

A sub-branch of the data-aided techniques is divided into blind and semi-blind estimation schemes. These schemes employ statistical properties of the transmitted signals to estimate the channels. These techniques generally employ second-order statistics of the received data without any a priori knowledge of pilot sequences or channel estimates [26], [27]. Blind estimation presents no overhead penalty (*i.e.*, there is no need to transmit pilot sequences for training purposes) and therefore has high spectral efficiency but, however, they are only suitable for slowly time-varying channels, as they present long convergence times due to the time it takes for them to produce reliable estimates [26]–[28]. On the other hand, semi-blind techniques are based on a hybrid approach between blind and pilot-based channel estimation, and therefore, are more suitable for channel estimation in slowly time-varying channels [29]. They have the advantage of more accurate estimates but still present longer estimation processing than pilot-based techniques [30] and might be computationally more complex than pure pilot- or blind- based estimation algorithms [31].

Pilot-based channel estimation is the most used estimation approach found in the body of the surveyed literature. A myriad of works employ pilot-based channel estimation and use linear channel estimators such as Least-Squares (LS) and the Minimum Mean Squared Error (MMSE) [32]–[36], [40]. Among the linear estimators, MMSE is known for being the

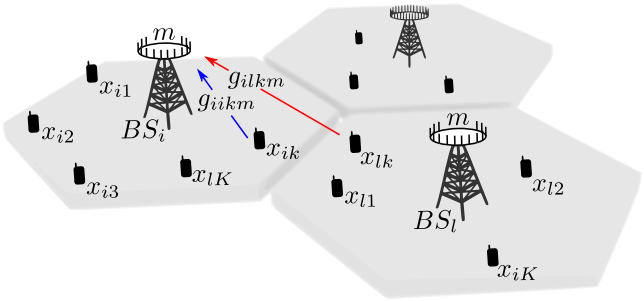

FIGURE 1. A multi-cell system diagram where K is the number of devices per cell, M is the number of co-located antennas, and L is the number of cells. The blue arrow (←) represents the channel from the k-th user within the i-th cell to the i-th BS, and the red arrow (←) represents the channel from the k-th user within the l-th cell to the i-th BS.

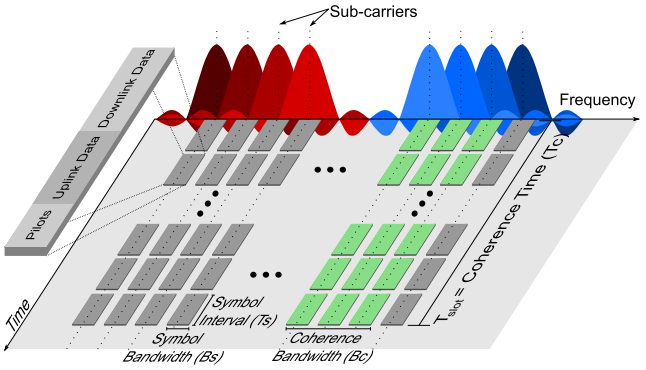

FIGURE 2. OFDM modulation and TDD transmission protocol employed in a massive MIMO system.

optimum estimator when both the channels and the noise are jointly Gaussian [13]. The pilot-based estimation has the advantage of producing good estimates in moderate to low SNR scenarios due to the good correlation properties of the training sequences [13], however, as the pilot sequences have to occupy a portion of the coherence interval, the spectral efficiency decreases with the length of the sequences [37]. Additionally, it is also important to highlight that in the great majority of works, for the sake of clarity, the authors assume perfect knowledge of the channel statistics (*i.e.*, small- and large- scale fading coefficients and noise power), which is not practical [12]. In [12], a simple and practical channel estimator that does not assume a priory knowledge of the channel statistics is presented. The advantage of that estimator is that it does not require the heavy overhead created by the estimation of the channel statistics once they are obtained from the received signal at the BS.

III. UPLINK SIGNAL MODEL

In this work, we consider a multi-cell system with L cells, each cell has a BS, and each BS operates with M antennas and K randomly located single antennas users as depicted in Figure 1. Let g_{ilkm} represents the complex channel gain from the k-th user within the l-th cell to the m-th antenna of the BS located at the i-th cell. The channel gain, g_{ilkm}, can be re-written as $g_{ilkm} = \sqrt{\beta_{ilk}} h_{ilkm}$ where β_{ilk} represents the large-scale coefficients (taking into account both path-loss and shadowing) and h_{ilkm} represents a frequency-flat fading channel coefficients. The large-scale fading coefficients from the k-th user located at the l-th cell to the i-th cell, β_{ilk}, are assumed constant for all the M antennas at the i-th BS, once path-loss and shadow fading change slowly along the space [36], [37]. The channel can be considered flat Rayleigh fading as the small-scale fading is considered to follow a complex Gaussian distribution with zero mean and unitary variance and the transmitted signal's bandwidth is smaller than the coherence bandwidth [36]. The overall channel matrix is denoted by $\mathbf{G}_{il}$ and has dimension $M \times K$, where its k-th column, $\mathbf{g}_{ilk} = [g_{ilk1}, \cdots, g_{ilkM}]^{\mathrm{T}}$, represent the channel gain from the k-th user in the l-th cell to the i-th BS. Additionally, we consider the set of large-scale coefficients, $\{\beta_{ilk}\}$, as being deterministic during the estimation phase. This assumption is reasonable due to the slow change of large-scale fading in comparison with the small-scale fading coefficients [37]. The overall channel matrix $\mathbf{G}_{il}$ can be defined directly by the channel coefficients as showed next,

$$\mathbf{G}_{il} = \begin{pmatrix} g_{il11} & g_{il21} & \cdots & g_{ilK1} \\ g_{il12} & g_{il22} & \cdots & g_{ilK2} \\ \vdots & \vdots & \ddots & \vdots \\ g_{il1M} & g_{il2M} & \cdots & g_{ilKM} \end{pmatrix} = \mathbf{H}\mathbf{D}^{1/2}$$
$$= \begin{pmatrix} h_{il11} & h_{il21} & \cdots & h_{ilK1} \\ h_{il12} & h_{il22} & \cdots & h_{ilK2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{il1M} & h_{il2M} & \cdots & h_{ilKM} \end{pmatrix} \begin{pmatrix} \sqrt{\beta_{il1}} & 0 & \cdots & 0 \\ 0 & \sqrt{\beta_{il2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\beta_{ilK}} \end{pmatrix}, \quad (1)$$

where $\mathbf{D}$ is the diagonal matrix with the large-scale coefficients for all K devices, $\beta_{ilk}, k = 1, \ldots, K$.

IV. UPLINK TRAINING

We assume that the users of all cells use the same set of pilot sequences at the same time (*i.e.*, all users' transmissions aligns to the BS uplink) and that the pilot reuse factor is equal to one which is the most aggressive. Figure 2 shows the signaling structure in more detail. If we assume a static channel during coherence time T_c then the proper signal multiplexing scheme is Time-Division Duplexing (TDD) protocol. Furthermore, we can additionally say that uplink and downlink channels are reciprocal, *i.e.*, they are identical within the channel coherence time, which minimizes the pilot transmission overhead since only users need to transmit pilots in the uplink direction. Therefore, two direct consequences of the TDD protocol adoption are: (i) the channel coherence time limits the TDD frames [38], [39], and (ii) the pilot overhead cost is only proportional to the number of terminals, K, and not to the number of antennas, M [40], [41]. Finally, we assume Orthogonal frequency-division duplexing multiplexing (OFDM), which is a well-known modulation with a straightforward interpretation of the coherence interval concept [37].

The pilot sequences of K users are represented by a $\tau \times K$ matrix Φ with the orthogonality property, $\Phi^H \Phi = \mathbf{I}_K$, where $K \leq \tau$. The received pilot sequences at the i-th BS are represented by a $M \times \tau$ matrix, $\mathbf{Y}_i$, defined as

$$\mathbf{Y}_i = \sqrt{\rho} \sum_{l=1}^{L} \mathbf{G}_{il} \Phi^H + \mathbf{N}_i, \tag{2}$$

where ρ is the average pilot transmit power of each user and $\mathbf{N}_i$ is a $M \times \tau$ matrix with i.i.d. elements following the distribution $\mathcal{CN}(0, 1)$. Equation (2) can be re-written as showed next in Equation (3). Equation (3) clearly shows the coherent inter-cell interference, which is represented by the second term in the equation and is caused by users employing the same pilot sequences within other BSs.

$$\mathbf{Y}_i = \underbrace{\sqrt{\rho}\mathbf{G}_{ii}\Phi^H}_{\text{Desired pilot signals}} + \underbrace{\sqrt{\rho} \sum_{l=1, l\neq i}^{L} \mathbf{G}_{il}\Phi^H}_{\text{Undesired pilot signals}} + \underbrace{\mathbf{N}_i}_{\text{Noise}}. \tag{3}$$

As the length of the pilot sequences has to be much smaller than the coherence time (see Figure 2), the maximum number of orthogonal sequences within a cell is limited, and due to that, other cells end up reusing the same pilot sequences. This reuse of pilot sequences brings about what is known in the literature as *pilot contamination* [37]. Due to the pilot contamination phenomenon, channel estimates obtained within a given cell get contaminated by pilots transmitted by the users located within neighbour cells. The worst-case scenario for pilot contamination is when the frequency reuse factor is equal to 1 (*i.e.*, all cells reuse the same frequency) and all transmissions coming from the different cells are perfectly synchronized. Pilot contamination leads to the coherent interference term in Equation 3, which is hard to mitigate in the case of spatially uncorrelated channels [13]. Under this specific channel condition, pilot contamination is an impairment that does not disappear, even with an infinite number of antennas [13]. On the other hand, the coherent interference caused by pilot contamination becomes negligible when large frequency reuse factors are adopted [37].

Let ϕ_k denote the k-th column of Φ^H. Hence, a sufficient statistic for the estimation of the channel vectors, $\mathbf{g}_{iik}$, at the i-th BS is given by

$$\begin{aligned} \mathbf{z}_{ik} &= \frac{1}{\sqrt{\rho}}\mathbf{Y}_i\phi_k = \sum_{l=1}^{L} \mathbf{G}_{il}\Phi^H\phi_k + \frac{1}{\sqrt{\rho}}\mathbf{N}_i\phi_k \\ &= \sum_{l=1}^{L} \mathbf{g}_{ilk} + \mathbf{w}_{ik} \\ &= \underbrace{\mathbf{g}_{iik}}_{\text{Desired channel}} + \underbrace{\sum_{l=1, l\neq i}^{L} \mathbf{g}_{ilk}}_{\text{Inter-cell interference}} + \underbrace{\frac{\mathbf{N}_i\phi_k}{\sqrt{\rho}}}_{\text{Noise}}, \end{aligned} \tag{4}$$

where $\mathbf{w}_{ik} = \frac{1}{\sqrt{\rho}}\mathbf{N}_i\phi_k$ has distribution $\mathcal{CN}(\mathbf{0}_M, \frac{1}{\rho}\mathbf{I}_M)$ and $\mathbf{z}_{ik}$ follows the distribution $\mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$ where

$$\zeta_{ik} = \sum_{l=1}^{L} \beta_{ilk} + \frac{1}{\rho}. \tag{5}$$

During the detection phase, the i-th BS has to estimate the channels of its users, *i.e.*, $\mathbf{g}_{ilk}, \forall k$, based on the transmitted pilot sequences.

V. CHANNEL ESTIMATION

In this section we briefly present two of the most well-known and heavily used channel estimators in the Massive MIMO literature [12], namely, Least-Squares (LS) and Linear Minimum Mean-Square Error (LMMSE) channel estimators.

A. LEAST-SQUARES

A sufficient statistic for estimating the channel vectors $\mathbf{g}_{iik}, \forall k$ at the i-th BS is defined as [11]

$$\hat{\mathbf{g}}_{iik}^{\text{LS}} = \mathbf{z}_{ik}, \tag{6}$$

with distribution given by $\mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$ [12]. Additionally, by using (6), we can show that $\hat{\mathbf{g}}_{ilk}^{\text{LS}} = \hat{\mathbf{g}}_{iik}^{\text{LS}}, \forall l$, meaning that the channel estimates acquired by the i-th BS are parallel vectors.

This estimator is known as the Least Squares (LS) estimator and as can be seen, it does not rely on any prior information on the channel statistics, such as the large-scale fading coefficients. Moreover, the LS estimator is known to have poorer performance than the MMSE estimator [12].

B. LINEAR MINIMUM MEAN-SQUARE ERROR

In case the channel statistics (*i.e.*, large-scale fading coefficients and noise power) are assumed perfectly known at the i-th BS, the optimum LMMSE channel estimator is defined as

$$\hat{\mathbf{g}}_{iik}^{\text{MMSE}} = \frac{\beta_{iik}}{\zeta_{ik}}\mathbf{z}_{ik}, \tag{7}$$

with distribution given by $\mathcal{CN}(\mathbf{0}_M, \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M)$ [12]. Additionally, by using (7), we can show that $\hat{\mathbf{g}}_{ilk}^{\text{MMSE}} = \frac{\beta_{ilk}}{\beta_{iik}}\hat{\mathbf{g}}_{iik}^{\text{MMSE}}, \forall l$, meaning that the channel estimates are parallel vectors that only differ by the scaling factor, $\frac{\beta_{ilk}}{\beta_{iik}}$.

VI. A SIMPLE AND EFFECTIVE CHANNEL ESTIMATOR

As proposed in [12], a very simple but yet effective channel estimator for the multi-cell case is defined by

$$\hat{\mathbf{g}}_{iik}^{\text{prop.}} = M\beta_{iik}\frac{\mathbf{z}_{ik}}{\|\mathbf{z}_{ik}\|^2}, \tag{8}$$

which asymptotically approaches the performance of the linear LMMSE estimator as $M \to \infty$ [12].

Remark 1: If $\mathbf{g}_{iik}$ is an i.i.d. complex Gaussian vector, then, we see that $\hat{\mathbf{g}}_{ilk}^{\text{prop.}} = \frac{\beta_{ilk}}{\beta_{iik}}\hat{\mathbf{g}}_{iik}^{\text{prop.}}, \forall l$, meaning that the channel estimates are parallel vectors that only differ by the scaling factor, $\frac{\beta_{ilk}}{\beta_{iik}}$, and therefore, showing the BS's inability

to separate users transmitting the same pilot sequence within other cells.

The channel estimator presented in Equation (8) has $\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop}}] = \mathbf{0}_M$ and covariance matrix given by

$$\begin{aligned}\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop}}(\hat{\mathbf{g}}_{iik}^{\text{prop}})^H] &= \left(\frac{M}{M-1}\frac{\beta_{iik}^2}{\zeta_{ik}}\right)\mathbf{I}_M \\ &= \left(\frac{\beta_{iik}^2}{\zeta_{ik}} + \frac{1}{(M-1)}\frac{\beta_{iik}^2}{\zeta_{ik}}\right)\mathbf{I}_M \\ &= \left(\frac{\beta_{iik}^2}{\zeta_{ik}} + \epsilon_{ik}\right)\mathbf{I}_M, \end{aligned} \tag{9}$$

where $\epsilon_{ik} = \frac{1}{(M-1)}\frac{\beta_{iik}^2}{\zeta_{ik}}$. The mean of $\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop}}]$ is found by using the symmetry property of the distribution of $z_{ikm} \sim \mathcal{CN}(0, \zeta_{ik})$, we conclude that $\mathbb{E}\left[\hat{g}_{iikm}^{\text{prop}}\right] = 0$, once $\mathbb{E}\left[\frac{z_{ikm}}{\|\mathbf{z}_{ik}\|^2}\right] = \mathbb{E}\left[\frac{-z_{ikm}}{\|\mathbf{z}_{ik}\|^2}\right], \forall m$. As can be seen by analyzing equation (9), as $M \to \infty$, $\text{Cov}[\hat{\mathbf{g}}_{iik}^{\text{prop}}] \to \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M$, which is the covariance matrix of the LMMSE estimator [12].

Remark 2: The average normalized squared Euclidean distance between $\hat{\mathbf{g}}_{iik}^{\text{prop}}$ and $\hat{\mathbf{g}}_{iik}^{\text{MMSE}}$ is given by

$$\frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_{iik}^{\text{prop}} - \hat{\mathbf{g}}_{iik}^{\text{MMSE}}\|^2\right] = \frac{1}{M-1}\frac{\beta_{iik}^2}{\zeta_{ik}} = \epsilon_{ik}. \tag{10}$$

The proof of (10) is given in Appendix B of [12]. From (5) and (10), it is easily noticeable that the average distance decreases with increasing M, decreasing ρ, increasing $\beta_{ilk}, i \neq l$, and decreasing β_{iik}.

VII. DISTRIBUTION OF THE PROPOSED ESTIMATOR

Since $\mathbf{z}_{ik} \sim \mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$, then $\|\mathbf{z}_{ik}\|^2 \sim \Gamma(M, \zeta_{ik})$, therefore, the distribution of the elements, $\hat{g}_{iikm}^{\text{prop.}} = \frac{M\beta_{iik}z_{ikm}}{\|\mathbf{z}_{ik}\|^2}$, of the channel estimator, $\hat{\mathbf{g}}_{iik}^{\text{prop.}}$, is the ratio between a Circularly-symmetric normal distribution and a Gamma distribution.

As $z_{ikm} \forall m$ are i.i.d. complex normal random variables, then it is clear that the distribution of $\hat{g}_{iikm}^{\text{prop.}}$ in $\mathbb{C} = \mathbb{R}^2$ is rotation invariant. Therefore, it suffices to find the distribution of $|\hat{g}_{iikm}^{\text{prop.}}|^2$, which is the same as that of

$$R = \frac{(M\beta_{iik})^2|z_{ikm}|^2}{(|z_{ikm}|^2 + \sum_{m'=1,m'\neq m}^{M} z_{ikm'}|^2)^2} = \frac{bX}{(X+Y)^2} \geq 0, \tag{11}$$

where $b = (M\beta_{iik})^2$, $X = |z_{ikm}|^2$ and $Y = |\sum_{m'=1,m'\neq m}^{M} z_{ikm'}|^2$ are independent random variables exhibiting the distributions $\Gamma(1, \zeta_{ik})$ and $\Gamma(M-1, \zeta_{ik})$, respectively. The proof of the distributions of X and Y are given in Lemmas 1 and 2 of Appendix A, respectively. The random variables X and Y are independent due to the fact that the M elements of $\mathbf{z}_{ik}$ and uncorrelated (diagonal covariance matrix), and therefore, as they are Complex Gaussian random variables, they are also independent.

Next, in order to find the PDF of R, we consider the transformation of random variables from (X, Y) to (R', S), where $S = X$ and $R' = \frac{X}{(X+Y)^2}$. After finding the joint distribution of (R', S), $f_{R',S}(r, s)$, then the PDF of R' is defined by $\int_0^\infty f_{R',S}(r', s)ds$. The joint PDFs of $f_{X,Y}(x, y)$ and $f_{R',S}(r', s)$ are derived in Appendix B. After applying the substitution $v = \sqrt{r's}$ to the previous integral, we find (see Appendix C for the proof.)

$$f_{R'}(r') = \frac{1}{\zeta_{ik}^M(M-2)!r'^{M+1}}\int_0^1 v^M(1-v)^{M-2}e^{-\frac{v}{\zeta_{ik}r'}}dv, \quad 0 \leq r' < \infty. \tag{12}$$

Finally, by applying the change of variable $R = bR'$ to (12) results in

$$\begin{aligned} f_R(r) &= \frac{1}{b}f_{R'}\left(\frac{r}{b}\right) \\ &= \frac{(M\beta_{iik})^{2M}}{\zeta_{ik}^M(M-2)!r^{M+1}}\int_0^1 v^M(1-v)^{M-2}e^{-\frac{(M\beta_{iik})^2 v}{\zeta_{ik}r}}dv, \quad 0 \leq r < \infty, \end{aligned} \tag{13}$$

which by using an integral solver [42] results in

$$\begin{aligned} f_R(r) = \frac{1}{\zeta_{ik}r^2}\sqrt{\frac{(M\beta_{iik})^2\pi}{\zeta_{ik}r}}\frac{e^{-\frac{(M\beta_{iik})^2}{2\zeta_{ik}r}}}{2} \\ \times\left\{\left[2\zeta_{ik}Mr - (M\beta_{iik})^2\right]I_{M-\frac{1}{2}}\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik}r}\right)\right. \\ \left. +(M\beta_{iik})^2 I_{M+\frac{1}{2}}\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik}r}\right)\right\}, \end{aligned} \tag{14}$$

where $I_n(z)$ is the modified Bessel function of the first kind. From (13) and using the Fubini theorem [43] (*i.e.*, switch the order of integration), it is easy to find all the central moments of $R = |\hat{g}_{iikm}^{\text{prop.}}|^2$, which are derived in Appendix D and given as

$$\mathbb{E}\left[R^k\right] = \mathbb{E}\left[|\hat{g}_{iikm}^{\text{prop.}}|^{2k}\right] = \frac{(M\beta_{iik})^{2k}k!(M-k-1)!}{\zeta_{ik}^k(M+k-1)!}, \quad k \in \mathbb{Z}. \tag{15}$$

The variance of $|\hat{g}_{iikm}^{\text{prop.}}|^2$ is defined as

$$\text{var}\left(|\hat{g}_{iikm}^{\text{prop.}}|^2\right) = \frac{(M\beta_{iik})^4[2+M(M-1)]}{\zeta_{ik}^2M^2(M+1)(M-1)^2(M-2)}, \tag{16}$$

The covariance between $|\hat{g}_{iikm}^{\text{prop.}}|^2$ and $|\hat{g}_{iikn}^{\text{prop.}}|^2$ is given by

$$\text{cov}\left(|\hat{g}_{iikm}^{\text{prop.}}|^2, |\hat{g}_{iikn}^{\text{prop.}}|^2\right) = \frac{2(M\beta_{iik})^4}{\zeta_{ik}^2M^2(M+1)(M-1)^2(M-2)}, \tag{17}$$

which tends to 0 when $M \to \infty$. By using the symmetry property of the distribution of $z_{ikm} \sim \mathcal{CN}(0, \zeta_{ik})$, we conclude that $\mathbb{E}\left[\hat{g}_{iikm}^{\text{prop.}}\right] = 0$, once $\mathbb{E}\left[\frac{z_{ikm}}{\|\mathbf{z}_{ik}\|^2}\right] = \mathbb{E}\left[\frac{-z_{ikm}}{\|\mathbf{z}_{ik}\|^2}\right]$. Next, by using the symmetry property of z_{ikm} again, we find that $\hat{g}_{iikm}^{\text{prop.}}$ and $\hat{g}_{iikn}^{\text{prop.}}$, when $m \neq n$, are uncorrelated, *i.e.*, $\text{cov}(\hat{g}_{iikm}^{\text{prop.}}, \hat{g}_{iikn}^{\text{prop.}}) = \mathbb{E}\left[(\hat{g}_{iikm}^{\text{prop.}})^*\hat{g}_{iikn}^{\text{prop.}}\right]$, once

$\mathbb{E}\left[\hat{g}_{iikm}^{\text{prop.}}\right] = \mathbb{E}\left[\hat{g}_{iikn}^{\text{prop.}}\right] = 0$. First, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}\left[|(\hat{g}_{iikm}^{\text{prop.}})^* \hat{g}_{iikn}^{\text{prop.}}|\right] &\leq \sqrt{\mathbb{E}\left[\left|\hat{g}_{iikm}^{\text{prop.}}\right|^2\right]\mathbb{E}\left[\left|\hat{g}_{iikn}^{\text{prop.}}\right|^2\right]} \\ &= \mathbb{E}\left[\left|\hat{g}_{iikm}^{\text{prop.}}\right|^2\right] = \mathbb{E}\left[\left|\hat{g}_{iikn}^{\text{prop.}}\right|^2\right] \\ &= \frac{M\beta_{iik}^2}{\zeta_{ik}(M-1)} < \infty, m \neq n, \quad (18)\end{aligned}$$

hence, $\mathbb{E}\left[(\hat{g}_{iikm}^{\text{prop.}})^* \hat{g}_{iikn}^{\text{prop.}}\right]$ exists and is finite. Therefore and because the joint distribution of the pair $\left((-\hat{g}_{iikm}^{\text{prop.}})^*, \hat{g}_{iikn}^{\text{prop.}}\right)$ is the same as that of $\left((\hat{g}_{iikm}^{\text{prop.}})^*, \hat{g}_{iikn}^{\text{prop.}}\right)$, we conclude that $\mathbb{E}\left[(\hat{g}_{iikm}^{\text{prop.}})^* \hat{g}_{iikn}^{\text{prop.}}\right] = \mathbb{E}\left[(-\hat{g}_{iikm}^{\text{prop.}})^* \hat{g}_{iikn}^{\text{prop.}}\right] = 0$. Therefore, the elements of $\hat{\mathbf{g}}_{iik}^{\text{prop.}}$ are uncorrelated.

The CDF of $R = |\hat{g}_{iikm}^{\text{prop.}}|^2$ is derived in Appendix E and defined as

$$\begin{aligned}F_R(r) &= \int_{-\infty}^{r} f_R(t)dt \\ &= 1 - \sqrt{\frac{(M\beta_{iik})^2\pi}{\zeta_{ik} r}} e^{-\frac{(M\beta_{iik})^2}{2\zeta_{ik} r}} I_{M-\frac{1}{2}}\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik} r}\right), \quad M > 1. \end{aligned} \quad (19)$$

Next, we find the PDF and CDF of $U = \Re\{\hat{g}_{iikm}^{\text{prop.}}\} = \Im\{\hat{g}_{iikm}^{\text{prop.}}\}$. First we recall that $\hat{g}_{iikm}^{\text{prop.}}$ can be expressed as $\sqrt{RC} + j\sqrt{RS}$, where $C = \cos(\Theta)$, $S = \sin(\Theta)$, and Θ is a random variable independent of R and uniformly distributed in the interval $[0, 2\pi]$ or, by symmetry, in $[0, \pi]$. The PDF of C, is given by $f_C(c) = \frac{1}{\pi\sqrt{1-c^2}}$, $-1 < c < 1$. Considering the transformation of random variables from (R, C) to (U, W), where $U = \sqrt{RC}$ and $W = R$, we have the following for the PDF of U

$$f_U(u) = \frac{1}{\pi}\int_{u^2}^{\infty} \frac{1}{\sqrt{w-u^2}} f_R(w)dw, \quad (20)$$

for all real u, where $f_R(r)$ is given by (13). Using the Fubini theorem and an integral solver [42] for the calculation of the iterated double integral, we get

$$\begin{aligned}f_U(u) = \frac{(M\beta_{iik})^2 e^{-\frac{(M\beta_{iik})^2}{2\zeta_{ik}u^2}}}{2\zeta_{ik}|u|^3} &\left[I_{M-1}\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik}u^2}\right)\right. \\ &\left. - I_M\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik}u^2}\right)\right]. \quad (21)\end{aligned}$$

By recalling that $U = \sqrt{R}\cos(\Theta)$ and that R and Θ are independent, we know that all odd moments of U vanish due to the fact that $\mathbb{E}\left[(\cos(\Theta))^{2k+1}\right] = 0, k \in \mathbb{Z}$ and that all even moments of U are given by

$$\begin{aligned}\mathbb{E}\left[U^{2k}\right] &= \mathbb{E}\left[R^k\right]\mathbb{E}\left[(\cos(\Theta))^{2k}\right] \\ &= \frac{(M\beta_{iik})^{2k}}{\zeta_{ik}^k}\frac{(M-k-1)!}{(M+k-1)!}\frac{\Gamma(k+\frac{1}{2})}{\sqrt{\pi}}. \quad (22)\end{aligned}$$

The covariance between $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ and $\Im\{\hat{g}_{iikm}^{\text{prop.}}\}$ is define as

$$\begin{aligned}\text{cov}\left(\Re\{\hat{g}_{iikm}^{\text{prop.}}\}, \Im\{\hat{g}_{iikm}^{\text{prop.}}\}\right) &= \mathbb{E}\left[\Re\{\hat{g}_{iikm}^{\text{prop.}}\}\Im\{\hat{g}_{iikm}^{\text{prop.}}\}\right] \\ &= \mathbb{E}\left[\sqrt{RC}\sqrt{RS}\right] = \mathbb{E}\left[RCS\right] \\ &= \mathbb{E}\left[R\right]\mathbb{E}\left[CS\right] = 0, \quad (23)\end{aligned}$$

once $\mathbb{E}[CS] = \frac{1}{\pi}\int_0^{\pi}\cos(\Theta)\sin(\Theta)d\Theta = 0$.

By using (22), the Kurtosis of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\} = \Im\{\hat{g}_{iikm}^{\text{prop.}}\}$ is defined as

$$\begin{aligned}\text{Kurt}\left[\Re\{\hat{g}_{iikm}^{\text{prop.}}\}\right] &= \frac{\mathbb{E}[U^4]}{(\mathbb{E}[U^2])^2} \\ &= \frac{3M(M-1)}{(M+1)(M-2)} \overset{M\to\infty}{\approx} 3, \quad M > 2, \end{aligned} \quad (24)$$

which approaches 3 as the number of antennas, M, increases. Therefore, as $M \to \infty$, the distribution of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ and $\Im\{\hat{g}_{iikm}^{\text{prop.}}\}$ becomes a mesokurtic distribution, *i.e.*, this distribution has tails shaped the same way as the normal distribution as M increases. Hence, as $M \to \infty$ the distributions of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ and $\Im\{\hat{g}_{iikm}^{\text{prop.}}\}$ tend to that of the normal distribution with mean equal to 0 and variance equal to $\frac{M\beta_{iik}^2}{2(M-1)\zeta_{ik}}$.

As $\text{cov}\left(\hat{g}_{iikm}^{\text{prop.}}, \hat{g}_{iikn}^{\text{prop.}}\right) = 0$, $m \neq n$, $\text{cov}\left(\Re\{\hat{g}_{iikm}^{\text{prop.}}\}, \Im\{\hat{g}_{iikm}^{\text{prop.}}\}\right) = 0, \forall m$, and $\text{Kurt}\left[\Re\{\hat{g}_{iikm}^{\text{prop.}}\}\right] \to 3$ as M increases, then we conclude that the distribution of random vector $\hat{\mathbf{g}}_{iik}^{\text{prop.}}$, approaches that of a circularly-symmetric complex normal vector, $\hat{\mathbf{g}}_{iik}^{\text{prop.}} \sim \mathcal{CN}\left(\mathbf{0}_M, \frac{M\beta_{iik}^2}{(M-1)\zeta_{ik}}\mathbf{I}_M\right)$ as M increases.

Remark 3: As the number of antennas, M grows without bound, the distribution of the proposed channel estimator tends asymptotically to the distribution of LMMSE estimator.

$$\lim_{M\to\infty} \mathcal{CN}\left(\mathbf{0}_M, \frac{M\beta_{iik}^2}{(M-1)\zeta_{ik}}\mathbf{I}_M\right) = \mathcal{CN}\left(\mathbf{0}_M, \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M\right) \quad (25)$$

The CDF of U, $F_U(u) = 2\int_0^u f_U(t)dt$, results in a quite complicated expression involving Hypergeometric functions, *i.e.*, it does not have a neat closed form expression, however, as we just discussed, it can be approximated by the CDF of a normal random variable as M increases.

VIII. SIMULATION RESULTS

In this section, we present simulation results assessing the performance of proposed estimator and showing that its distribution tends to that of a complex Gaussian random variable.

In Figure 3, we compare the MSE versus the number of antennas, M, deployed at the BS under the settings $\beta_{iik} = 1$, $\beta_{ilk} = 0.05, \forall l \neq i$ and $\rho = 10$ dB. As can be seen, with the increase of M, the MSE of the proposed estimator approaches that of the optimum MMSE estimator, while the MSE of both the LS and MMSE estimators are independent of the number of antennas and therefore, stay constant over all values of M. Based on these results and the ones presented next, we will

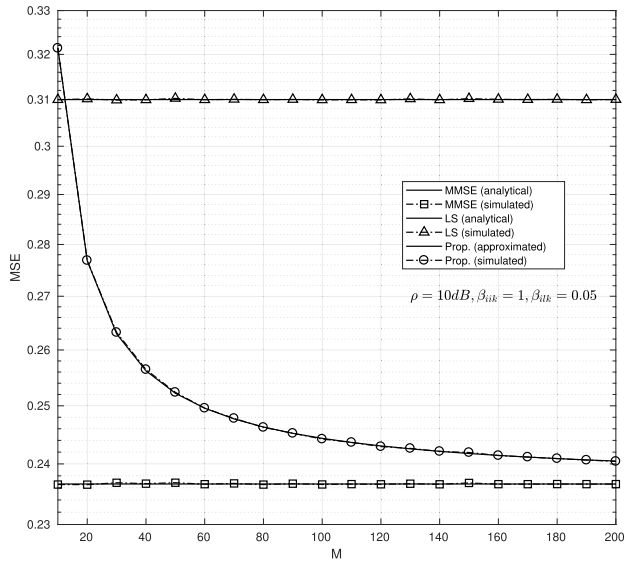

FIGURE 3. MSE performance versus number of antennas deployed at the BS.

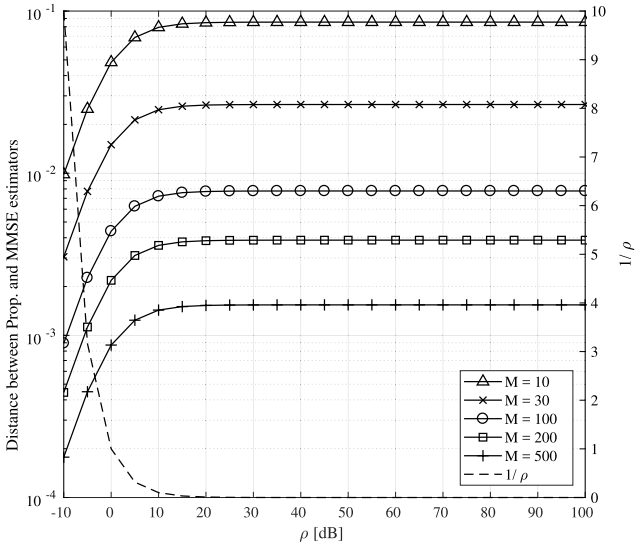

FIGURE 4. Distance between the proposed and MMSE channel estimators (Remark 2).

conclude that the proposed channel estimator and its distribution tends to that of the MMSE channel estimator, which has complex Gaussian distribution given by $\mathcal{CN}\left(\mathbf{0}_M, \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M\right)$.

In Fig. 4, we compare the distance between the proposed and MMSE channel estimators for different number of antennas, M, with $\beta_{iik} = 1$ and $\beta_{ilk} = 0.05, \forall l \neq i$. As the Remark 2 states, the distance is small at low values of ρ, increasing with it until a ceiling value is reached. As can be also noticed, the ceiling value decreases with the number of antennas, M. The figure also depicts the variation of $1/\rho$, showing that it has a direct impact on the distance as it direct affects ζ_{ik}. For low ρ values the term $\frac{\beta_{iik}^2}{\zeta_{ik}}$ in (10) tends to zero quite fast due to the impact of $1/\rho$ and when ρ increases, the effect of $1/rho$ in ζ_{ik} vanishes and the distance tends to a constant value for a fixed value of M.

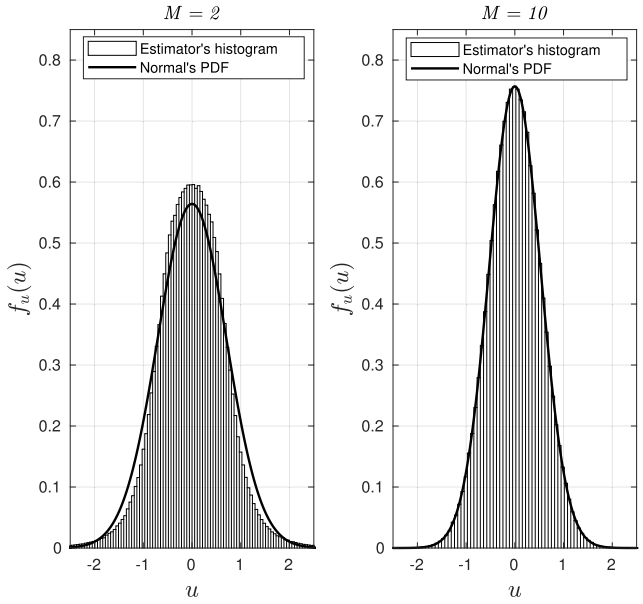

FIGURE 5. Proposed estimator's histogram versus Normal's PDF.

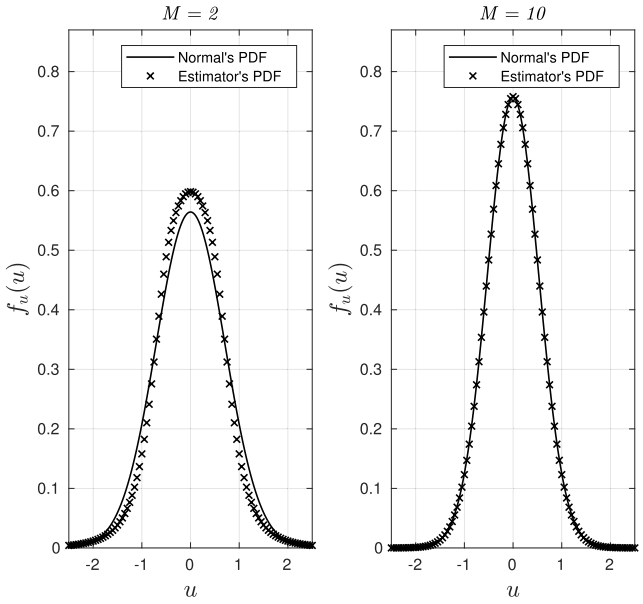

FIGURE 6. Eq. (21) versus the Normal's PDF.

Next, we present some results showing that the distribution of the proposed estimator asymptotically approaches that of the circularly-symmetric complex normal distribution. Figures 5 to 8 show several results comparing the distribution of the real (or imaginary) part, $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$, of the proposed estimator and the Normal distribution for $\beta_{iik} = 1$ and $\zeta_{ik} = \beta_{iik} + \sum_{l=1,l\neq i}^{L} \beta_{ilk} + \frac{1}{\rho} = 2$ (*i.e.*, $\sum_{l=1,l\neq i}^{L} \beta_{ilk} + \frac{1}{\rho} = 1$) respectively. The large-scale coefficient of the device within the target cell, β_{iik}, was chosen so that it was greater than the other individual large-scale coefficients from interfering cells. The value of ζ_{ik} was chosen so that the inter-cell large-scale fading coefficients plus noise power term, $\sum_{l=1,l\neq i}^{L} \beta_{ilk} + \frac{1}{\rho}$, is comparable to β_{iik}.

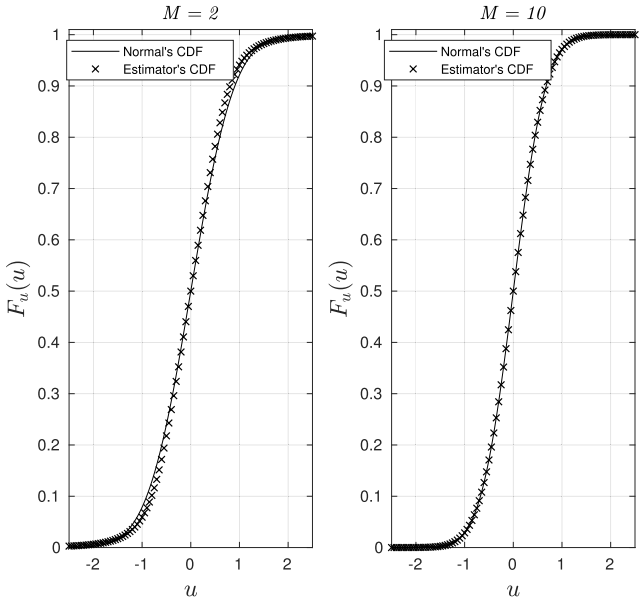

FIGURE 7. Proposed estimator's CDF, $F_U(u) = 2\int_0^u f_U(t)dt$, versus the Normal's CDF.

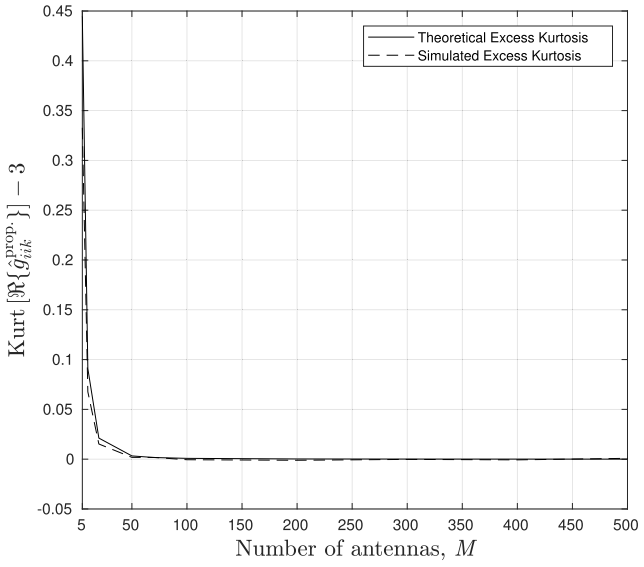

FIGURE 8. Proposed estimator's excess Kurtosis for several number of antennas, M.

Figure 5 compares the normalized histogram of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ with the PDF of the Normal distribution, $\mathcal{N}\left(0, \frac{M\beta_{iik}^2}{2(M-1)\zeta_{ik}}\right)$. The figure shows that for $M = 2$ the distributions differ considerably, however, for $M = 10$ they become very close, showing that the PDF of the random variable U asymptotically approaches that of the normal Gaussian PDF as the number of antennas increases. It is also worth mentioning that the same conclusion can be drawn to the imaginary part of $\hat{g}_{iikm}^{\text{prop.}}$, *i.e.*, $\Im\{\hat{g}_{iikm}^{\text{prop.}}\}$.

Figure 6 compares the PDF of $U = \Re\{\hat{g}_{iikm}^{\text{prop.}}\}$, with the PDF of the Normal distribution. Again the results show that both PDFs become quite close as M increases. The results presented in figure 6 shows that the analytical expression

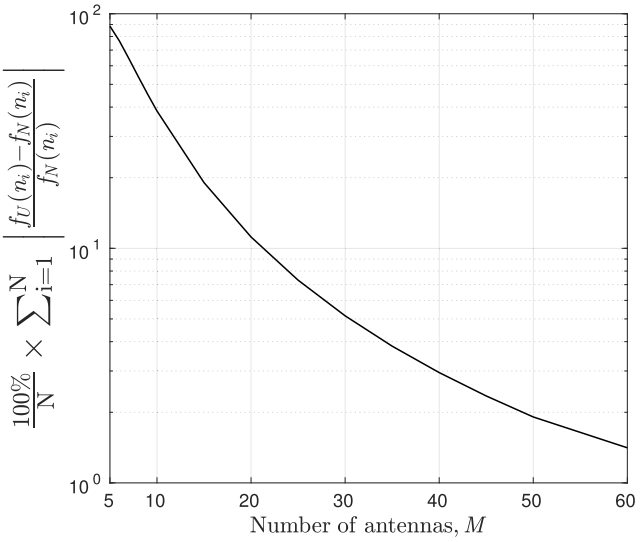

FIGURE 9. Mean absolute percentage error between (21) and the Normal's PDF, $f_N(n)$ for several number of antennas, M.

derived for the the PDF of U in (21) is tight and tends to represent the same distribution of a Normal distribution as M grows.

Figure 7 compares the CDF of U, with the CDF of the Normal distribution. As with the PDF analysis and as expected, the results presented in the figure show that the CDFs as M increases, the difference between the two CDFs disappear.

Figure 8 shows the excess Kurtosis of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$. The excess kurtosis measures the distance of a given distribution to the normal distribution, which has Kurtosis equal to 3. The figure shows that the excess Kurtosis of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ tends to 0 as M increases, showing that, indeed, the distribution of $\Re\{\hat{g}_{iikm}^{\text{prop.}}\}$ asymptotically tends to that of the Normal distribution. It is worth mentioning that the excess Kurtosis of $\Im\{\hat{g}_{iikm}^{\text{prop.}}\}$ also tends to 0 as M increases, as the real and imaginary parts of $\hat{g}_{iikm}^{\text{prop.}}$ have the same distribution.

Figure 9 depicts the mean absolute percentage error (MAPE) between the PDF of the random variable U, *i.e.*, $f_U(u)$, given by (21), and the Normal's PDF, $f_N(n)$. The MAPE is defines as

$$\text{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{f_U(u_i) - f_N(n_i)}{f_N(n_i)}\right|. \tag{26}$$

As it can be seen in the figure, the MAPE decreases as the number of antennas, M, increases, showing once more that $f_U(u)$ asymptotically approaches $f_N(n)$ as M increases.

It is important to emphasize that since the proposed estimator was based on the LMMSE channel estimator, it was expected that the distribution of proposed estimator would approach asymptotically the distribution of LMMSE, $\mathcal{CN}\left(\mathbf{0}_M, \frac{\beta_{iik}^2}{\zeta_{ik}}\mathbf{I}_M\right)$, as the number of antennas, M, grew.

All the results presented in this section prove that the distribution of the proposed channel estimator can be accurately approximated by a circularly-symmetric complex normal

distribution with covariance matrix given by $\frac{M\beta_{ik}^2}{(M-1)\zeta_{ik}}\mathbf{I}_M$ for $M > 10$. Having a BS deployed with an antenna array of more than 10 elements (*i.e.*, antennas) is quite feasible given the recent advances in massive MIMO development. As reported in [10], current deployments have shown BSs equipped with antenna arrays of more than 64 elements.

IX. CONCLUSION

This study derived expressions for the PDF, CDF, and central moments of a channel estimator proposed in [12]. The distribution of the proposed channel estimator asymptotically approaches the circularly-symmetric complex normal distribution as the number of antennas M increases. The results presented here show that the distribution of the proposed channel estimator approximates accurately by a circularly-symmetric complex normal distribution for $M > 10$, which is the case for massive MIMO deployments [10]. Moreover, the findings presented here, clearly show that the distance between the distribution of proposed estimator and that of the optimum LMMSE estimator decreases as the number of antennas increases, getting really close when $M > 10$.

As future work, an interesting direction where the findings of this work can be very useful, is on the derivation of spectral capacity bounds of systems employing the channel estimator presented here. This direction is an important one as it will shed light on the performance gap that there might be between systems employing the optimum MMSE channel estimator and the one presented here.

APPENDIX A

Lemma 1: If $W \sim \mathcal{CN}(0, \sigma^2)$ then $\|W\|^2 \sim \Gamma(1, \sigma^2)$.

Proof: We know that $W = X + jY$ and consequently $\|W\|^2 = X^2 + Y^2$, where X, Y are i.i.d. random variables with distribution $\mathcal{N}(0, \sigma^2/2)$. Now, if we make $Z = X^2+Y^2$, then using the joint pdf of X and Y we have that $\mathbb{P}(Z \leq z)$ is defined by

$$\mathbb{P}(X^2 + Y^2 \leq z) = \frac{1}{\pi\sigma^2}\int_{x^2+y^2\leq z} e^{-\frac{(x^2+y^2)}{\sigma^2}}\, dxdy. \quad (27)$$

Next, switching to polar coordinates we get:

$$\mathbb{P}(Z \leq z) = \frac{1}{\pi\sigma^2}\int_0^{2\pi}\int_0^{\sqrt{z}} re^{-\frac{r^2}{\sigma^2}} drd\theta = \frac{2}{\sigma^2}\int_0^{\sqrt{z}} re^{-\frac{r^2}{\sigma^2}} dr. \quad (28)$$

Now if we set $u = r^2$ then we get

$$\mathbb{P}(Z \leq z) = \frac{1}{\sigma^2}\int_0^{z} e^{-\frac{u}{\sigma^2}} du, \quad (29)$$

so Z is exponentially distributed with rate parameter $\lambda = 1$.

Finally, comparing the pdf given above with the Gamma distribution pdf one can notice that if the shape parameter k is set to 1 and scale parameter θ is set to σ^2 it becomes the exponential pdf, concluding the proof. ∎

Lemma 2: If $X_m \sim \mathcal{CN}(0, \sigma^2)$ $\forall m$ are independent, then $\sum_{m=1}^{M} |X_m|^2 \sim \Gamma(M, \sigma^2)$.

Proof: From Lemma 1 we know that each $Z_m = \|X_m\|^2 \sim \Gamma(1, \sigma^2)$. We also know that Z_m is independent for all m. Therefore, each Z_m has the following characteristic function:

$$\varphi_Z(t) = (1 - j\sigma^2 t)^{-1}. \quad (30)$$

Next, knowing that the characteristic function of the sum of independent random variables is the product of their individual characteristic functions leads to

$$\varphi_{Z_1+Z_2+\ldots+Z_M}(t) = \varphi_{Z_1}(t)\varphi_{Z_2}(t)\ldots\varphi_{Z_M}(t) = (1 - j\sigma^2 t)^{-M}. \quad (31)$$

Eq. (31) defines the characteristic function of a Gamma-distributed random variable with scale parameter $\theta = \sigma^2$ and shape parameter $k = M$, and therefore concluding the proof. ∎

APPENDIX B

In this Appendix we derive the PDF of $f_{R',S}(r', s)$. First, by remembering that $R' = \frac{X}{(X+Y)^2}$ and $S = X$, where $X = |z_{ikm}|^2$ and $Y = |\sum_{m'=1,m'\neq m}^{M} z_{ikm'}|^2$ are independent random variables exhibiting the distributions $\Gamma(1, \zeta_{ik})$ and $\Gamma(M-1, \zeta_{ik})$, respectively, then, the PDF of $f_{X,Y}(x, y)$ is given by

$$f_{X,Y}(x, y) = \frac{1}{\zeta_{ik}\Gamma(1)\Gamma(M-1)} y^{M-2} e^{-(x+y)/\zeta_{ik}}, \quad 0 \leq x < \infty, 0 \leq y < \infty. \quad (32)$$

Next, taking the Jacobian determinant of the transformation $X = S$ and $Y = \sqrt{\frac{S}{R'}} - S$ results in

$$|J| = \left|\frac{\partial(x, y)}{\partial(r', s)}\right| = \begin{vmatrix} \frac{\partial x}{\partial r'} & \frac{\partial x}{\partial s} \\ \frac{\partial y}{\partial r'} & \frac{\partial y}{\partial s} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ -\frac{1}{2r'}\sqrt{\frac{s}{r'}} & \frac{\partial y}{\partial s} \end{vmatrix} = \frac{1}{2r'}\sqrt{\frac{s}{r'}}. \quad (33)$$

Therefore, the joint distribution of R' and S has the pdf given by

$$f_{R',S}(r', s) = f_{X,Y}(x, y)\left|\frac{\partial(x, y)}{\partial(r', s)}\right| = \frac{1}{\zeta_{ik}^M\Gamma(1)\Gamma(M-1)}\frac{1}{2r'}\sqrt{\frac{s}{r'}}\left(\sqrt{\frac{s}{r'}} - s\right)^{M-2} e^{-\sqrt{\frac{s}{r'}}/\zeta_{ik}}, \quad 0 \leq r' < \infty, 0 \leq s < \infty. \quad (34)$$

APPENDIX C

In this Appendix, we derive the PDF of R'. By using (34), the PDF of R' is obtained by integrating $f_{R',S}(r', s)$ over s. We start by writing the integral of $f_{R',S}(r', s)$ over s as

$$f_{R'}(r') = \int_0^{\infty} f_{R',S}(r', s)ds = \int_0^{\infty} \frac{1}{\zeta_{ik}^M\Gamma(1)\Gamma(M-1)}\frac{1}{2r'}\frac{\sqrt{r's}}{r'}\left(\frac{\sqrt{r's} - r's}{r'}\right)^{M-2} \times e^{-\frac{\sqrt{r's}}{\zeta_{ik}r'}} ds, \quad 0 \leq r' < \infty, 0 \leq s < \infty. \quad (35)$$

Next, in order to find $f_{R'}(r')$ we apply the following change of variable to $f_{R',S}(r', s)$, $v = \sqrt{r's}$. Therefore, the PDF of R' is given by

$$f_{R'}(r') = \int_0^1 \frac{\zeta_{ik}^M}{\Gamma(1)\Gamma(M-1)} \frac{1}{r'} \frac{v}{r'} \left(\frac{v-v^2}{r'}\right)^{M-2} e^{-\frac{v}{\zeta_{ik} r'}} \frac{v}{r'} dv$$
$$= \frac{1}{\zeta_{ik}^M (M-2)! r'^{M+1}} \int_0^1 v^M (1-v)^{M-2} e^{-\frac{v}{\zeta_{ik} r'}} dv,$$
$$0 \leq r' < \infty. \quad (36)$$

where the upper limit in the integral is equal to 1 due to the fact that $\sqrt{r's} = \frac{x}{x+y} = \frac{s}{s+y}$ and that $\lim_{s\to\infty} \frac{s}{s+y} = \frac{s}{s} = 1$.

APPENDIX D

In this Appendix, we derive the central moments of R. We start by writing the expectation integral of R, which is defined as (37), as shown at the bottom of this page.

The integral between the square brackets in the last part of (37) can be found by using the following integral identity

$$\int_0^\infty x^a e^{-b/x} dx = b^{1+a} \Gamma(-1-a). \quad (38)$$

By applying the identity (38) into (37), it can be re-written as (39), as shown at the bottom of this page.

The integral in (39) can be found by using the following integral identity

$$\int_0^1 x^a (1-x)^b dx = \frac{\Gamma(1+a)\Gamma(1+b)}{\Gamma(2+a+b)}. \quad (40)$$

Finally, by applying the identity (40) into (39), it can be re-written as

$$\mathbb{E}\left[R^k\right] = \frac{(M\beta_{iik})^{2k}(M-k-1)!}{\zeta_{ik}^k (M-2)!} \frac{\Gamma(1+k)\Gamma(1+M-2)}{\Gamma(k+M)}$$
$$= \frac{(M\beta_{iik})^{2k}(M-k-1)!}{\zeta_{ik}^k (M-2)!} \frac{k!(M-2)!}{(M+k-1)!}$$
$$= \frac{(M\beta_{iik})^{2k} k!(M-k-1)!}{\zeta_{ik}^k (M+k-1)!}, k \in \mathbb{Z}. \quad (41)$$

APPENDIX E

In this Appendix we derive the CDF or R. We start by defining CDF expectation of R, which is defined as

$$F_R(r) = \int_0^r f_R(t) dt$$
$$= \int_0^r \frac{(M\beta_{iik})^{2M}}{\zeta_{ik}^M (M-2)! t^{M+1}} \int_0^1 v^M (1-v)^{M-2} e^{-\frac{(M\beta_{iik})^2 v}{\zeta_{ik} t}} dv dt$$
$$= \frac{(M\beta_{iik})^{2M}}{\zeta_{ik}^M (M-2)!} \int_0^1 v^M (1-v)^{M-2} \left[\int_0^r \frac{1}{t^{M+1}} e^{-\frac{(M\beta_{iik})^2 v}{\zeta_{ik} t}} dt\right] dv, \quad (42)$$

where in the last line of (42) we used the Fubini theorem [43] to switch the order of integration.

The integral between the square brackets in the last part of (42) can be found by using the following integral identity

$$\int_0^r \frac{1}{x^a} e^{-b/x} dx = b^{1-a} \Gamma(a-1, b/r). \quad (43)$$

Replacing (43) into (42) results in

$$F_R(r) = \frac{1}{(M-2)!} \int_0^1 (1-v)^{M-2} \Gamma\left(M, \frac{(M\beta_{iik})^2 v}{\zeta_{ik} r}\right) dv. \quad (44)$$

The integral in (44) is found by using an integral solver [42] and is defined as

$$\int_0^1 (1-v)^{M-2} \Gamma\left(M, \frac{(M\beta_{iik})^2 v}{\zeta_{ik} r}\right) dv$$
$$= \left\{1 - e^{-\frac{(M\beta_{iik})^2}{2\zeta_{ik} r}} \sqrt{\frac{\pi (M\beta_{iik})^2}{\zeta_{ik} r}} I_{M-\frac{1}{2}}\left(\frac{(M\beta_{iik})^2}{2\zeta_{ik} r}\right)\right\} \Gamma(M-1). \quad (45)$$

After replacing (45) into (44) and remembering that $\Gamma(M-1) = (M-2)!$ we conclude the proof and find $F_R(r)$, which was defined in (19).

$$\mathbb{E}\left[R^k\right] = \int_0^\infty r^k f_R(r) dr$$
$$= \int_0^\infty R^k \left[\frac{(M\beta_{iik})^{2M}}{\zeta_{ik}^M (M-2)! r^{M+1}} \int_0^1 v^M (1-v)^{M-2} e^{-\frac{(M\beta_{iik})^2 v}{\zeta_{ik} r}} dv\right] dr$$
$$= \frac{(M\beta_{iik})^{2M}}{\zeta_{ik}^M (M-2)!} \int_0^1 v^M (1-v)^{M-2} \left[\int_0^\infty r^{k-M-1} e^{-\frac{(M\beta_{iik})^2 v}{\zeta_{ik} r}} dr\right] dv. \quad (37)$$

$$\mathbb{E}\left[R^k\right] = \frac{(M\beta_{iik})^{2M}}{\zeta_{ik}^M (M-2)!} \int_0^1 v^M (1-v)^{M-2} \left(\frac{(M\beta_{iik})^2 v}{\zeta_{ik}}\right)^{k-M} \Gamma(M-k) dv$$
$$= \frac{(M\beta_{iik})^{2k} \Gamma(M-k)}{\zeta_{ik}^k (M-2)!} \int_0^1 v^k (1-v)^{M-2} dv. \quad (39)$$

ACKNOWLEDGMENT

The authors thank our colleagues from Instituto Eldorado who provided insight and expertise that assisted on this research.

REFERENCES

[1] Cisco Systems, "Cisco visual networking index: Global mobile data traffic forecast update 2015–2020," Cisco, San Jose, CA, USA, White Paper, Feb. 2016.
[2] *IMT Vision—Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*, ITU-R, Geneva, Switzerland, Sep. 2015.
[3] Ericsson. (Jun. 2017). *Ericsson Mobility Report*. [Online]. Available: http://www.ericsson.com/mobility-report
[4] V. M. Nguyen and M. Kountouris, "Performance limits of network densification," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1294–1308, Mar. 2017.
[5] C. Sun, L. Yang, L. Chen, and J. Zhang, "SVR based blind signal recovery for convolutive MIMO systems with high-order QAM signals," *IEEE Access*, vol. 7, pp. 23249–23260, Feb. 2019.
[6] S. Wu, R. Atat, N. Mastronarde, and L. Liu, "Improving the coverage and spectral efficiency of millimeter-wave cellular networks using device-to-device relays," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2251–2265, May 2018.
[7] A. P. F. de Figueiredo, R. Mennes, X. Jiao, W. Liu, and I. Moerman, "A spectrum sharing framework for intelligent next generation wireless networks," *IEEE Access*, vol. 6, pp. 60704–60735, Nov. 2018.
[8] Z. Wu, K. Lu, C. Jiang, and X. Shao, "Comprehensive study and comparison on 5G NOMA schemes," *IEEE Access*, vol. 6, pp. 18511–18519, 2018.
[9] Y. Xin, D. Wang, J. Li, H. Zhu, J. Wang, and X. You, "Area spectral efficiency and area energy efficiency of massive MIMO cellular systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3243–3254, May 2016.
[10] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—What is next?: Five promising research directions for antenna arrays," *Digit. Signal Process.*, to be published.
[11] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[12] 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 flat fading," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, pp. 1–14, Jan. 2018.
[13] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signals Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017.
[14] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential space-time shift keying: A dispersion matrix approach," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3219–3230, Nov. 2010.
[15] S. R. Varea, J. C. Penuelas, D. C. Soler, M. D. Rio, J. Francisco, Y. Fouad, R. H. Gohary, and H. Yanikomeroglu, "Non-coherent MIMO communication for the 5th generation mobile: Overview and practical aspects," *Waves*, vol. 6, pp. 5–15, 2014.
[16] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
[17] A. Schenk and R. F. H. Fischer, "Noncoherent detection in massive MIMO systems," in *Proc. Int. ITG Workshop Smart Antennas*, Mar. 2013, pp. 1–8.
[18] K. M. Attiah, K. G. Seddik, and R. H. Gohary, "Hierarchical coherent and non-coherent communication," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2017, pp. 1242–1247.
[19] M. Chowdhury, A. Manolakos, and A. Goldsmith, "Multiplexing and diversity gains in noncoherent massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 265–277, Jan. 2017.
[20] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3111–3124, Jun. 2014.
[21] M. Liu, M. Crussiere, and J.-F. Hélard, "A novel data-aided channel estimation with reduced complexity for TDS-OFDM systems," *IEEE Trans. Broadcast.*, vol. 58, no. 2, pp. 247–260, Jun. 2012.
[22] W. Liu, K.-K. Wong, S. Jin, and X. You, "A data-aided channel estimation scheme for decoupled systems in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 4987–5000, Aug. 2018.
[23] M. Coldrey and P. Bohlin, "Training-based MIMO systems: Part II—Improvements using detected symbol information," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 296–303, Jan. 2008.
[24] N. Jindal, A. Lozano, and T. L. Marzetta, "What is the value of joint processing of pilots and data in block-fading channels?" in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2009, pp. 2189–2193.
[25] H. Schoeneich and P. A. Hoeher, "Iterative pilot-layer aided channel estimation with emphasis on interleave-division multiple access systems," *EURASIP J. Appl. Signal Process.*, vol. 2006, p. 250, Jan. 2006.
[26] L. Wei, C. Ming, S. Cheng, and H. Wang, "A complexity reduced blind channel equalization scheme for OFDM systems," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2006, pp. 1–4.
[27] R. Boloix-Tortosa, F. J. Simois-Tirado, and J. J. Murillo-Fuentes, "Blind adaptive channel estimation for OFDM systems," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2009, pp. 191–195.
[28] C. Shin, R. W. Heath, Jr., and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 670–685, Mar. 2007.
[29] W. Yang, Y. Cai, and Y. Xun, "Semi-blind channel estimation for OFDM systems," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, May 2006, pp. 226–230.
[30] E. G. Larsson, G. Liu, J. Li, and G. B. Giannakis, "Joint symbol timing and channel estimation for OFDM based WLANs," *IEEE Commun. Lett.*, vol. 5, no. 8, pp. 325–327, Aug. 2001.
[31] K. Mawatwal, D. Sen, and R. Roy, "A semi-blind channel estimation algorithm for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 70–73, Feb. 2017.
[32] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
[33] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, Jun. 2013.
[34] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 815–830, Oct. 2014.
[35] A. Ashikhmi, T. L. Marzetta, and L. Li, "Interference reduction in multi-cell massive MIMO systems I: Large-scale fading precoding and decoding," *IEEE Trans. Inf. Theory*, to be published.
[36] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
[37] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*, 1st ed. Cambridge U.K.: Cambridge Univ. Press, 2016.
[38] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. Asilomar Conf. Signals, Syst. Comput. (ACSSC)*, Oct. 2006, pp. 359–363.
[39] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.
[40] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
[41] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.
[42] *Mathematica Version 11.3*, Wolfram Res., Champaign, IL, USA, 2018.
[43] P. Billingsley, *Probability and Measure*, 3rd ed. New York, NY, USA: Wiley, 1995, pp. 231–240.

FELIPE A. P. DE FIGUEIREDO received the B.S. and M.S. degrees in telecommunications from the Instituto Nacional de Telecomunicações (INATEL), Minas Gerais, Brazil, in 2004 and 2011, respectively. He is currently pursuing the Ph.D. degree with the Internet Technology and Data Science Lab, Ghent University, Ghent, Belgium, where he is currently with the IDLab. He has been working with Research and Development of telecommunications systems for more than ten years. His research interests include digital signal processing, digital communications, mobile communications, MIMO, multicarrier modulations, and FPGA development.

CLAUDIO FERREIRA DIAS was born in Brazil, in 1981. He received the B.S. degree in electrical engineering from the University of Uberlandia, Uberlandia, Brazil, in 2008, and the M.S. degree in communications engineering from the University of Campinas, Campinas, Brazil, in 2010, where he is currently a Ph.D. Researcher. His current research interests include massive MIMO, millimeter-wave propagation, and modeling for 5G communication systems.

FABBRYCCIO A. C. M. CARDOSO received the bachelor's degree in electrical engineering from the University of Brasília, Brazil, in 1995, and the M.Sc. and Ph.D. degrees from the State University of Campinas (UNICAMP), Brazil, in 1998 and 2004, respectively. He has over seven years of experience leading the Research and Development projects focused on physical layer design for 4G (LTE and LTE-A) and defense-related SDR systems. He is currently a Senior Researcher with the Wireless Communications Technologies Division, CPqD Foundation, Brazil. His research interest includes 5G wireless communication technologies for radio access networks.

GUSTAVO FRAIDENRAICH graduated in electrical engineering from the Federal University of Pernambuco, UFPE, Brazil, in 1997. He received the M.Sc. and Ph.D. degrees from the State University of Campinas (UNICAMP), Brazil, in 2002 and 2006, respectively, where he is currently an Assistant Professor. From 2006 to 2008, he was a Postdoctoral Fellow with the STARLAB Group, Stanford University, USA. His research interests include multiple antenna systems, cooperative systems, radar systems, and wireless communications, in general.