Performance Analysis of Large-Scale MU-MIMO With a Simple and Effective Chanel Estimator

Felipe A. P. de Figueiredo‡*, Claudio F. Dias‡, Eduardo R. de Lima† and Gustavo Fraidenraich‡

*Ghent University - imec, IDLab, Department of Information Technology, Ghent, Belgium.
†Eldorado Research Institute, Campinas, Brazil.
‡DECOM/FEEC–State University of Campinas (UNICAMP), Campinas Brazil.
Email: *felipe.pereira@ugent.be, †eduardo.lima@eldorado.org.br, ‡[aplnx, gf]@decom.fee.unicamp.br.

Abstract—Accurate channel estimation is of utmost importance for massive MIMO systems that allow providing significant improvements in spectral and energy efficiency. In this work, we investigate the spectral efficiency performance and present a channel estimator for multi-cell massive MIMO systems subjected to pilot-contamination. The proposed channel estimator performs well under moderate to aggressive pilot contamination scenarios without prior knowledge of the inter-cell large-scale channel coefficients and noise power. The estimator approximates the performance of a linear Minimum Mean Square Error (MMSE) as the number of antennas increases. Following, we derive a lower bound closed-form spectral efficiency of the Maximum Ratio Combining (MRC) detector in the proposed channel estimator. The simulation results highlight that the proposed estimator performance approaches the linear minimum mean square error (LMMSE) channel estimator asymptotically.

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 precoding 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 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 to 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 estimation of the intra/inter-cell large-scale fading coefficients and noise power might be unjustified in practice due to the excessive overhead it imposes on the system [12]. For instance, in the case where there are L cells serving K users (or MTC devices for example), 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 the foreseen case where cells have to serve a huge number of MTC devices along with human-type communication (HTC) devices, the estimation of the inter-cell large-scale fading coefficients plus the noise power estimation becomes unfeasible.

In this work, we present 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. We derive closed-form lower-bound expressions for the spectral efficiency of a system employing MRC detection with MMSE and proposed channel estimation. The simulation results presented here show that the performance of the proposed estimator asymptotically approaches that of the LMMSE channel estimator.

II. 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 devices. Let g_{ilkm} represents the complex channel gain from the k-th device within the l-th cell to the m-th antenna of the BS located at the ith 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 device located at the l-th cell to the ith cell, β_{ilk}, are assumed constant for all the M antennas at the ith BS, once path-loss and shadow fading change slowly along the space [14, 15]. 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 [14]. 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 device in the l-th cell to the ith 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 [15]. The overall channel matrix $\mathbf{G}_{il}$ can be defined directly by the channel coefficients as $\mathbf{G}_{il} = \mathbf{H}\mathbf{D}^{1/2}$, where $\mathbf{H}$ is a $M \times K$ matrix containing the small-scale fading coefficient vectors of the K channels and $\mathbf{D}$ is the diagonal matrix with the large-scale coefficients for all K devices, $\beta_{ilk}, k = 1, \ldots, K$.

III. UPLINK TRAINING

We assume that the devices of all cells use the same set of pilot sequences at the same time (*i.e.*, all devices' transmissions aligns to the BS uplink) and that the pilot reuse factor is equal to one, which is the most aggressive case for pilot reuse. 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 devices 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 [16, 17], and (ii) the pilot overhead cost is only proportional to the number of terminals, K, and not to the number of antennas, M [18, 19]. Finally, we assume Orthogonal frequency-division duplexing multiplexing (OFDM), which is a well-known modulation with a straightforward interpretation of the coherence interval concept [15].

The pilot sequences of K devices 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 ith BS are represented by a $M \times \tau$ matrix, $\mathbf{Y}_i$, defined as

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

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

As the length of the pilot sequences has to be much smaller than the coherence time, 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* [15]. Due to the pilot contamination phenomenon, channel estimates obtained within a given cell get contaminated by pilots transmitted by the devices 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 1, 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 [15].

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 ith BS is given by

$$\mathbf{z}_{ik} = \frac{1}{\sqrt{\rho}}\mathbf{Y}_i\phi_k = \sum_{l=1}^{L}\mathbf{G}_{il}\Phi^H\phi_k + \frac{1}{\sqrt{\rho}}\mathbf{N}_i\phi_k$$

$$= \underbrace{\mathbf{g}_{iik}}_{\text{Desired channel}} + \underbrace{\sum_{l=1,l\neq i}^{L}\mathbf{g}_{ilk}}_{\text{Inter-cell interference}} + \underbrace{\mathbf{w}_{ik}}_{\text{Noise}}, \quad (2)$$

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 $\zeta_{ik} = \sum_{l=1}^{L}\beta_{ilk} + \frac{1}{\rho}$. During the detection phase, the ith BS has to estimate the channels of its devices, *i.e.*, $\mathbf{g}_{ilk}, \forall k$, based on the transmitted pilot sequences.

IV. 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 ith BS is defined as [11]

$$\hat{\mathbf{g}}_{iik}^{\mathrm{LS}} = \mathbf{z}_{ik}, \quad (3)$$

with distribution given by $\mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$ [12]. Additionally, by using (3), we can show that $\hat{\mathbf{g}}_{ilk}^{\mathrm{LS}} = \hat{\mathbf{g}}_{iik}^{\mathrm{LS}}, \forall l$, meaning that the channel estimates acquired by the ith 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 ith BS, the optimum LMMSE channel estimator is defined as

$$\hat{\mathbf{g}}_{iik}^{\text{MMSE}} = \frac{\beta_{iik}}{\zeta_{ik}} \mathbf{z}_{ik}, \qquad (4)$$

with distribution given by $\mathcal{CN}(\mathbf{0}_M, \frac{\beta_{iik}^2}{\zeta_{ik}} \mathbf{I}_M)$ [12]. Additionally, by using (4), 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, $\frac{\beta_{ilk}}{\beta_{iik}}$.

V. A SIMPLE AND EFFECTIVE CHANNEL ESTIMATOR

After analyzing the LMMSE channel estimator equation, given by (4), we realize we do not need to know the individual large-scale fading coefficients plus the noise power, but, in fact, what is necessary is their summation, denoted by ζ_{ik}. The term ζ_{ik} can be estimated by using the maximum-likelihood (ML) method [11] and noticing that $\mathbf{z}_{ik} \sim \mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$, therefore a estimator for ζ_{ik} is defined as

$$\hat{\zeta}_{ik} = \frac{\|\mathbf{z}_{ik}\|^2}{M}, \qquad (5)$$

which has $\mathbb{E}\left[\hat{\zeta}_{ik}\right] = \zeta_{ik}$ and $\text{var}\left[\hat{\zeta}_{ik}\right] = \zeta_{ik}^2/M$. The Cramer-Rao bound of the estimator is given by

$$\text{var}\left[\hat{\zeta}_{ik}\right] \geq \zeta_{ik}^2/M. \qquad (6)$$

Therefore, we conclude that the estimator of ζ_{ik} is the minimum variance unbiased estimator (MVUE) [11], *i.e.*, it is an unbiased estimator once its mean is equal to the value being estimated and it presents the lowest variance among all other possible estimators of ζ_{ik}.

After replacing ζ_{ik} by $\hat{\zeta}_{ik}$ in (4), a very simple but yet effective channel estimator for the multi-cell case is defined by

$$\hat{\mathbf{g}}_{iik}^{\text{prop.}} = M\beta_{iik} \frac{\mathbf{z}_{ik}}{\|\mathbf{z}_{ik}\|^2}, \qquad (7)$$

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

The channel estimator presented in Equation (7) has $\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop}}] = \mathbf{0}_M$ and covariance matrix given by

$$\mathbb{E}[\hat{\mathbf{g}}_{iik}^{\text{prop}}(\hat{\mathbf{g}}_{iik}^{\text{prop}})^H] = \left(\frac{M}{M-1} \frac{\beta_{iik}^2}{\zeta_{ik}}\right) \mathbf{I}_M. \qquad (8)$$

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 (8), 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 1. *The average normalized squared Euclidean distance between $\hat{\mathbf{g}}_{iik}^{prop}$ and $\hat{\mathbf{g}}_{iik}^{MMSE}$ is given by*

$$\frac{1}{M}\mathbb{E}\left[\|\hat{\mathbf{g}}_{iik}^{\text{prop}} - \hat{\mathbf{g}}_{iik}^{\text{MMSE}}\|^2\right] = \frac{1}{M-1} \frac{\beta_{iik}^2}{\zeta_{ik}} = \epsilon_{ik}. \qquad (9)$$

The proof of (9) is given in Appendix B of [12]. From ζ_{ik} and (9), it is easily noticeable that the average distance decreases with increasing M, decreasing ρ, increasing $\beta_{ilk}, i \neq l$, and decreasing β_{iik}.

VI. SPECTRAL EFFICIENCY

The received signal at the ith BS is denoted as

$$\mathbf{y}_i = \sqrt{\rho} \sum_{l=1}^{L} \sum_{k=1}^{K} \mathbf{g}_{ilk} x_{lk} + \mathbf{n}_i, \qquad (12)$$

where x_{lk} are the data symbols, which are mutually uncorrelated and have zero mean and unit power and $\mathbf{n}_i$ represents noise whose elements are i.i.d. $\mathcal{CN}(0,1)$. During the data transmission phase, the received signal $\mathbf{y}_i \in \mathbb{C}^M$ at the ith BS is linearly processed by $\mathbf{v}_{ik} \in \mathbb{C}^M$, *i.e.*, the receiving combining vector assigned by BS i to its kth device, resulting in

$$r_k = \mathbf{v}_{ik}^H \mathbf{y}_i = \sqrt{\rho} \sum_{l=1}^{L} \sum_{k=1}^{K} \mathbf{v}_{ik}^H \mathbf{g}_{ilk} x_{lk} + \mathbf{v}_{ik}^H \mathbf{n}_i \qquad (13)$$

In order to quantify the spectral efficiency (SE) under imperfect channel statistics information, we need to use a capacity lower-bound that does not require LMMSE estimates. The use-and-then-forget (UatF) bound can be applied to the case at hand [13], where the received combined signal is re-written as

$$r_k = \sqrt{\rho}\mathbb{E}\left[\mathbf{v}_{ik}^H \mathbf{g}_{iik}\right] x_{ik} + \sqrt{\rho}\left(\mathbf{v}_{ik}^H \mathbf{g}_{iik} - \mathbb{E}\left[\mathbf{v}_{ik}^H \mathbf{g}_{iik}\right]\right) x_{ik}$$

$$+\sqrt{\rho} \sum_{k'=1, k'\neq k}^{K} \mathbf{v}_{ik}^H \mathbf{g}_{iik'} x_{ik'} + \sqrt{\rho} \sum_{l=1, l\neq i}^{L} \sum_{k=1}^{K} \mathbf{v}_{ik}^H \mathbf{g}_{ilk} x_{lk} + \mathbf{v}_{ik}^H \mathbf{n}_i. \qquad (14)$$

By treating $\mathbb{E}\left[\mathbf{v}_{ik}^H \mathbf{g}_{iik}\right]$ as a known deterministic channel and noticing that the other terms are uncorrelated with $\mathbb{E}\left[\mathbf{v}_{ik}^H \mathbf{g}_{iik}\right] x_{ik}$, then, the following result is obtained.

Lemma 1. *The channel capacity of the kth device at the ith BS is lower bounded by*

$$SE_{ik} = \log_2\left(1 + \gamma_{ik}\right) \quad [bits/s/Hz], \qquad (15)$$

with γ_{ik} given by

$$\gamma_{ik} = \frac{\rho|\mathbb{E}\left[\mathbf{v}_{ik}^H \mathbf{g}_{iik}\right]|^2}{\rho \sum_{l=1}^{L} \sum_{k=1}^{K} \mathbb{E}\left[|\mathbf{v}_{ik}^H \mathbf{g}_{ilk}|^2\right] - \rho|\mathbb{E}\left[\mathbf{v}_{ik}^H \mathbf{g}_{iik}\right]|^2 + \mathbb{E}\left[\|\mathbf{v}_{ik}\|^2\right]}, \qquad (16)$$

where the expectations are with respect to the channel realizations.

Closed-form expressions for the lower-bound can be found with the MRC type of detection scheme, where the receiving combining vector is expressed as $\mathbf{v}_{ik} = w_{ik}\mathbf{z}_{ik}$, where w_{ik} is defined as

$$w_{ik} = \begin{cases} \frac{\beta_{iik}}{\zeta_{ik}}, & \text{LMMSE estimator}, \\ \frac{M\beta_{iik}}{\|\mathbf{z}_{ik}\|^2}, & \text{Proposed estimator}. \end{cases} \qquad (17)$$

$$\gamma_{ik}^{\text{LMMSE}} = \frac{\frac{\rho M \beta_{iik}^2}{\zeta_{ik}}}{\rho \sum_{l=1}^{L} \sum_{k'=1}^{K} \beta_{ilk'} + \frac{\rho M}{\zeta_{ik}} \sum_{l=1,l\neq i}^{L} \beta_{ilk}^2 + 1} \tag{10}$$

$$\gamma_{ik}^{\text{Prop.}} = \frac{\frac{\rho(M-1)\beta_{iik}^2}{\zeta_{ik}}}{\rho \left[\sum_{l=1}^{L} \sum_{k'=1,k'\neq k}^{K} \beta_{ilk'} + \frac{(M-1)}{(M+1)} \sum_{l=1}^{L} \beta_{ilk} \right] + \frac{\rho M}{\zeta_{ik}} \left[\frac{(M-1)}{(M+1)} \sum_{l=1}^{L} \beta_{ilk}^2 - \frac{(M-1)}{M} \beta_{iik}^2 \right] + 1} \tag{11}$$
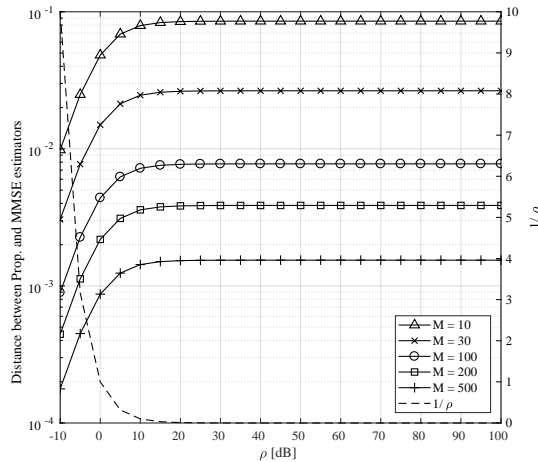

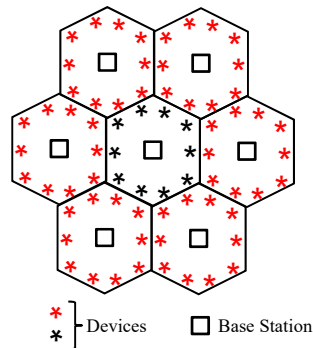Figure 1. Distance between the proposed and MMSE channel estimators (Remark 1).

Figure 2. Multi-cell setup witl $L = 7$ cells and $K = 10$ devices in each cell, where each pilot is reused by the device at the same position in all cells.
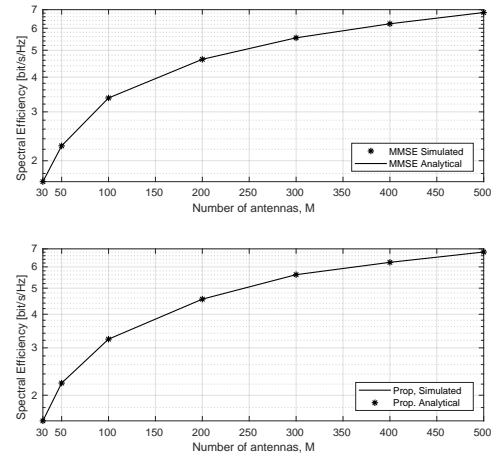

Figure 3. Lower bounds and numerically evaluated values of the spectral efficiency for different numbers of BS antennas for MRC detector.
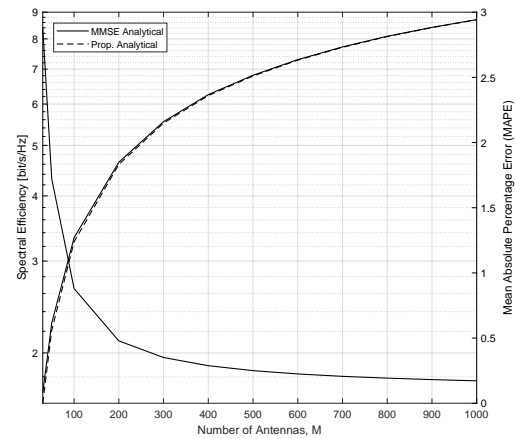

Figure 4. Sum SE achieved with MRC detector with LMMSE and proposed channel estimators over different number of antennas, M.

After computing the expectations in (16) for the LMMSE and proposed channel estimators, then, the channel capacity of the kth device at the ith BS can be lower bounded by (15) with γ_{ik} given by (10) and (11) for the LMMSE and proposed channel estimators respectively. The proof for the LMMSE estimator can be found in [15] and for the proposed channel estimator is available in Appendix A.

Remark 2. *The channel capacity of the MRC combining with the proposed channel estimator asymptotically approaches that of the MRC with the LMMSE channel estimator as the number of antennas increases, i.e.,* $\lim_{M\to\infty} \gamma_{ik}^{Prop.} \to \gamma_{ik}^{MMSE}$.

VII. Simulation Results

In this section, we present simulation results assessing the performance of the proposed estimator and showing that its

performance tends to that of the optimum LMMSE channel estimator.

In Figure 1, 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 1 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 (9) 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.

Next, we consider the multi-cell setup depicted in Figure 2, with $L = 7$ cells and inter-BS distance of 2 Km. The uplink spectral efficiency is evaluated at the central cell. Each cell serves $K = 10$ devices at 950 m from the serving BS, in a circular pattern as showed in the figure. The devices that are located at the same position in different cells reuse the same pilot sequence. Each device has an average pilot transmit power of $\rho = 20$ dB. The large-scale fading coefficients $\{\beta_{ilk}\}$ are independently generated by $\beta_{ilk} = \psi / \left(\frac{d_{ilk}}{d_0}\right)^v$, where $v = 3.8$, $10 \log_{10}(\psi) \sim \mathcal{N}(0, \sigma_{\text{shadow,dB}}^2)$ with $\sigma_{\text{shadow,dB}} = 8$, and d_{ilk} is the distance of the kth device in the lth cell to the ith BS.

Figure 3 depicts the simulated spectral efficiency and the derived analytical lower bounds for the MRC detector. As it can be clearly seen, all bounds are very tight. Therefore, next, we use these bounds for the numerical work.

Figure 4 shows the sum spectral-efficiency (SE) achieved with the MRC detector with LMMSE and proposed channel estimators as a function of the number of antennas, M. The SE is evaluated at the center cell. As it can be seen, and previously showed by Remark 2, MRC detection with the proposed channel estimator asymptotically approaches the SE performance of the MMSE estimator, being the difference almost imperceptible for $M \geq 500$. The figure also depicts the mean absolute percentage error (MAPE), showing that the error is less than 1% for $M = 100$, and keeps decreasing as the number of antennas increases. For $M = 500$ the error is around 0.25 % and for $M = 1000$, the error is around 0.172 %.

VIII. CONCLUSION

In this work we showed how a BS can estimate the channel statistics based on the received pilot signals and based on that, we proposed a channel estimator that asymptotically approaches the performance of the optimum LMMSE estimator. We derived one SE expression that can be used with any combining and estimation schemes. Moreover, we have also derived closed-form SE expressions for the MRC detector when LMMSE or the proposed channel estimators are used. The numerical results showed that the proposed estimator has performance similar to the LMMSE estimator for a large number of antennas.

APPENDIX A

For the proof of the channel capacity with MRC employing the proposed channel estimator, we first need to present a few Lemmas.

Lemma 2. *If* $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_x^2 \boldsymbol{I}_M)$ *and* $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_y^2 \boldsymbol{I}_M)$ *are independent random variables, then*

$$\mathbb{E}\left\{\Re\left[\frac{(\boldsymbol{x}+\boldsymbol{y})^H \boldsymbol{x}}{\|\boldsymbol{x}+\boldsymbol{y}\|^2}\right]\right\} \approx \frac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2}. \tag{18}$$

Proof. This can be proved by using an trick presented in [21] for normal random variables, which can be applied to complex normal random variables. We first write $\Sigma_{\mathbf{x}} = \mathbb{E}[\mathbf{x}\mathbf{x}^H]$ and

$\Sigma_{\mathbf{y}} = \mathbb{E}[\mathbf{y}\mathbf{y}^H]$ as the covariance matrix of $\mathbf{x}$ and $\mathbf{y}$ respectively. Let $\mathbf{z} = \mathbf{x} + \mathbf{y}$, and $\mathbf{S} = \Sigma_{\mathbf{x}}(\Sigma_{\mathbf{x}} + \Sigma_{\mathbf{y}})^{-1}$. Therefore, we have that $\mathbb{E}[(\mathbf{x} - \mathbf{Sz})\mathbf{z}^H] = 0$, so that $(\mathbf{x} - \mathbf{Sz})$ and $\mathbf{z}$ are uncorrelated and jointly normal, and therefore independent, $\mathbb{E}[\mathbf{x} - \mathbf{Sz}]\mathbb{E}[\mathbf{z}]^H = 0$. Consequently,

$$\mathbb{E}\left[\frac{\mathbf{z}^H \mathbf{x}}{\|\mathbf{z}\|^2}\right] = \mathbb{E}\left[\frac{\mathbf{z}^H(\mathbf{x} - \mathbf{Sz})}{\|\mathbf{z}\|^2}\right] + \mathbb{E}\left[\frac{\mathbf{z}^H \mathbf{Sz}}{\|\mathbf{z}\|^2}\right]$$

$$= \mathbb{E}\left[\frac{\mathbf{z}^H}{\|\mathbf{z}\|^2}\right] \mathbb{E}\left[\mathbf{x} - \mathbf{Sz}\right] + \mathbb{E}\left[\frac{\mathbf{z}^H \mathbf{Sz}}{\|\mathbf{z}\|^2}\right] = \mathbb{E}\left[\frac{\mathbf{z}^H \mathbf{Sz}}{\|\mathbf{z}\|^2}\right] = \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2}. \tag{19}$$

The final result follows from the fact that $\mathbf{S} = \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2}\mathbf{I}_M$. □

Lemma 3. *If* $X \sim \Gamma(k, \theta)$ *and* $\frac{1}{X} \sim \Gamma^{-1}(k, \theta)$, *i.e., the Inverse-Gamma distribution, then* $\mathbb{E}\left\{\frac{1}{X}\right\} = \frac{1}{\theta(k-1)}$.

The proof for this Lemma is presented in [20].

Lemma 4. *Let* μ_X *and* μ_Y *be the expectations of* X *and* Y, σ_Y^2 *be the variance of* Y, *and* σ_{XY} *be their covariance. Then the expectation,* $\mathbb{E}\{X/Y\}$, *can be approximated by*

$$\mathbb{E}\left\{\frac{X}{Y}\right\} \approx \frac{\mu_X}{\mu_Y}. \tag{20}$$

The proof for this Lemma is presented in [20].

Lemma 5. *Let* $\boldsymbol{z} = [z_1, \ldots, z_M]^T \in \mathbb{C}^{M \times 1}$ *be a complex random vector with distribution* $\boldsymbol{z} \sim \mathcal{CN}(\boldsymbol{0}_M, a\boldsymbol{I}_M)$. *Then* $\mathbb{E}\left[\|\boldsymbol{z}\|^4\right] = a^2 M(M + 1)$.

The proof for this Lemma is presented in [22].

Lemma 6. *If* $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_x^2 \boldsymbol{I}_M)$ *and* $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_y^2 \boldsymbol{I}_M)$ *are independent random variables, then*

$$\mathbb{E}\left[\left|\frac{(\boldsymbol{x}+\boldsymbol{y})^H \boldsymbol{x}}{\|\boldsymbol{x}+\boldsymbol{y}\|^2}\right|^2\right] \approx \frac{(M+1)\sigma_x^4 + \sigma_x^2\sigma_y^2}{(M+1)(\sigma_x^2 + \sigma_y^2)^2}. \tag{21}$$

Proof. By expanding (21) and then applying Lemma 4 to it, we find

$$\mathbb{E}\left[\left|\frac{(\mathbf{x}+\mathbf{y})^H \mathbf{x}}{\|\mathbf{x}+\mathbf{y}\|^2}\right|^2\right] \approx \frac{\mathbb{E}\left[\|\mathbf{x}\|^4\right] + M\sigma_x^2\sigma_y^2}{\mathbb{E}\left[\|\mathbf{x}+\mathbf{y}\|^4\right]}, \tag{22}$$

where $\mathbb{E}\left[\|\mathbf{x}\|^4\right]$ and $\mathbb{E}\left[\|\mathbf{x}+\mathbf{y}\|^4\right]$ are found by using Lemma 5 and remembering that the sum of two i.i.d. complex random variables (r.v.) is also a complex r.v. following $\mathcal{CN}(\boldsymbol{0}_M, (\sigma_x^2 + \sigma_y^2)\mathbf{I}_M)$. □

Lemma 7. *If* $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_x^2 \boldsymbol{I}_M)$ *and* $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}_M, \sigma_y^2 \boldsymbol{I}_M)$ *are independent random variables, then*

$$\mathbb{E}\left[\left|\frac{\boldsymbol{x}^H \boldsymbol{y}}{\|\boldsymbol{x}\|^2}\right|^2\right] = \frac{\sigma_y^2}{(M-1)\sigma_x^2}. \tag{23}$$

Proof. We start by using the trace trick and re-write (23) as

$$\mathbb{E}\left[\frac{\operatorname{tr}\left(\mathbf{x}\mathbf{x}^H \mathbf{y}\mathbf{y}^H\right)}{\|\mathbf{x}\|^4}\right] = \operatorname{tr}\left(\mathbb{E}\left[\frac{\mathbf{x}\mathbf{x}^H}{\|\mathbf{x}\|^4}\right]\sigma_y^2\mathbf{I}_M\right)$$

$$= \mathbb{E}\left[\frac{\|\mathbf{x}\|^2}{\|\mathbf{x}\|^4}\right]\sigma_y^2 = \mathbb{E}\left[\frac{1}{\|\mathbf{x}\|^2}\right]\sigma_y^2. \tag{24}$$

By applying Lemma 3 to the expectation in the last part of (24) we conclude the proof. □

A. Proof of the lower-bound with MRC and prop. estimator

Here we show how each one of the terms in (15) are calculated. The term at the numerator is given by

$$
\rho \left| \mathbb{E}\left[\mathbf{v}_{ik}^H \mathbf{g}_{iik} \right] \right|^2
$$

$$
= \rho M^2 \beta_{iik}^2 \left| \mathbb{E} \left[\frac{ \left\{ \mathbf{g}_{iik} + \left(\sum_{l=1,l\neq i}^{L} \mathbf{g}_{ilk} + \mathbf{w}_{ik} \right) \right\}^H \mathbf{g}_{iik} }{ \| \mathbf{g}_{iik} + \left(\sum_{l=1,l\neq i}^{L} \mathbf{g}_{ilk} + \mathbf{w}_{ik} \right) \|^2 } \right] \right|^2,
$$

(25)

which can be found by using Lemma 2 and considering $\mathbf{x} = \mathbf{g}_{iik}$ and $\mathbf{y} = \left(\sum_{l=1,l\neq i}^{L} \mathbf{g}_{ilk} + \mathbf{w}_{ik} \right) \sim \mathcal{CN}(\mathbf{0}_M, (\zeta_{ik} - \beta_{iik})\mathbf{I}_M)$. Therefore, the term in the numerator is found to be equal to

$$
\rho \left| \mathbb{E}\left[\mathbf{v}_{ik}^H \mathbf{g}_{iik} \right] \right|^2 = \frac{\rho M^2 \beta_{iik}^4}{\zeta_{ik}^2}.
$$

(26)

The first term in the denominator can be split into two parts, which are defined as

$$
\rho \sum_{l=1}^{L} \sum_{k=1}^{K} \mathbb{E}\left[|\mathbf{v}_{ik}^H \mathbf{g}_{ilk}|^2 \right]
$$

$$
= \rho \sum_{l=1}^{L} \mathbb{E}\left[|\mathbf{v}_{ik}^H \mathbf{g}_{ilk}|^2 \right] + \rho \sum_{l=1}^{L} \sum_{k'=1,k'\neq k}^{K} \mathbb{E}\left[|\mathbf{v}_{ik}^H \mathbf{g}_{ilk'}|^2 \right].
$$

(27)

The first term in the second line of (27) is found by applying Lemma 6 to it, resulting in

$$
\rho \sum_{l=1}^{L} \mathbb{E}\left[|\mathbf{v}_{ik}^H \mathbf{g}_{ilk}|^2 \right]
$$

$$
= \rho M^2 \beta_{iik}^2 \sum_{l=1}^{L} \mathbb{E}\left[\left| \frac{ \left\{ \mathbf{g}_{ilk} + \left(\sum_{l'=1,l'\neq l}^{L} \mathbf{g}_{il'k} + \mathbf{w}_{ik} \right) \right\}^H \mathbf{g}_{ilk} }{ \| \mathbf{g}_{ilk} + \left(\sum_{l'=1,l'\neq l}^{L} \mathbf{g}_{il'k} + \mathbf{w}_{ik} \right) \|^2 } \right|^2 \right]
$$

$$
= \rho M^2 \beta_{iik}^2 \sum_{l=1}^{L} \frac{(M+1)\beta_{ilk}^2 + \beta_{ilk}(\zeta_{ik} - \beta_{ilk})}{(M+1)\zeta_{ik}^2},
$$

(28)

where $\mathbf{x} = \mathbf{g}_{ilk}$ and $\mathbf{y} = \sum_{l'=1,l'\neq l}^{L} \mathbf{g}_{il'k} + \mathbf{w}_{ik}$. The second term in the second line of (27) is found by applying Lemma 7 to it, where $\mathbf{x} = \mathbf{z}_{ik}$ and $\mathbf{y} = \mathbf{g}_{ilk'}$. Therefore, the second term in the second line of (27) is given by

$$
\rho \sum_{l=1}^{L} \sum_{k'=1,k'\neq k}^{K} \mathbb{E}\left[|\mathbf{v}_{ik}^H \mathbf{g}_{ilk'}|^2 \right] = \frac{\rho M^2 \beta_{iik}^2}{(M-1)\zeta_{ik}} \sum_{l=1}^{L} \sum_{k'=1,k'\neq k}^{K} \beta_{ilk'}.
$$

(29)

The second term in the denominator is equal to the term in the numerator and therefore does not need to be recomputed.

The third term in the denominator is defined as

$$
\mathbb{E}\left[\|\mathbf{v}_{ik}\|^2 \right] = M^2 \beta_{iik}^2 \mathbb{E}\left[\frac{1}{\|\mathbf{z}_{ik}\|^2} \right],
$$

(30)

where the expectation in the second part of (30) is found by using Lemma 3 and remembering that $\mathbf{z}_{ik} \sim \mathcal{CN}(\mathbf{0}_M, \zeta_{ik}\mathbf{I}_M)$. Therefore, the last term is given by

$$
\mathbb{E}\left[\|\mathbf{v}_{ik}\|^2 \right] = \frac{M^2 \beta_{iik}^2}{(M-1)\zeta_{ik}}.
$$

(31)

Finally, replacing the results found here in (15) concludes the proof.

REFERENCES

[1] Cisco, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020*, White paper, February 2016.

[2] ITU-R, *IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond*, September 2015.

[3] Ericsson, *Ericsson mobility report*, [Online]. Available: http://www.ericsson.com/mobility-report, June 2017.

[4] Van Minh Nguyen, and Marios Kountouris, *Performance Limits of Network Densification*, IEEE Journal on Selected Areas in Communications, vol. 35, no. 6, pp. 1294-1308, June 2017.

[5] Chao Sun, Ling Yang, Li Chen, and Jiliang Zhang, *SVR Based Blind Signal Recovery for Convolutive MIMO Systems With High-Order QAM Signals*, IEEE Access, vol. 7, pp. 23249-23260, February 2019.

[6] Shuanshuan Wu, Rachad Atat, Nicholas Mastronarde, and Lingjia Liu, *Improving the Coverage and Spectral Efficiency of Millimeter-Wave Cellular Networks Using Device-to-Device Relays*, IEEE Transactions on Communications, vol. 66, no. 5, pp. 2251-2265, May 2018.

[7] Felipe A. P. 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, November 2018.

[8] Zhanji Wu, Kun Lu, Chengxin Jiang, and Xuanbo Shao, *Comprehensive Study and Comparison on 5G NOMA Schemes*, IEEE Access, vol. 6, pp. 18511-18519, March 2018.

[9] Yuanxue Xin, Dongming Wang, Jiamin Li, Huilin Zhu, Jiangzhou Wang, and Xiaohu You, *Area Spectral Efficiency and Area Energy Efficiency of Massive MIMO Cellular Systems*, IEEE Transactions on Vehicular Technology, vol. 65, no. 5, pp. 3243-3254, May 2016.

[10] Emil Bjornson, Luca Sanguinetti, Henk Wymeersch, Jakob Hoydis, and Thomas L. Marzetta, *Massive MIMO is a Reality - What is Next? Five Promising Research Directions for Antenna Arrays*, submitted to Digital Signal Processing, arXiv:1902.07678, February 2019.

[11] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.

[12] Felipe Augusto P. de Figueiredo et. al., *Channel estimation for massive MIMO TDD systems assuming pilot contamination and flat fading*, EURASIP Journal on Wireless Communications and Networking, January 2018.

[13] Emil Bjornson, Jakob Hoydis, and Luca Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*, Foundations and Trends in Signal Processing: vol. 11, no. 3-4, pp. 154-655, DOI: 10.1561/2000000093, 2017.

[14] 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.

[15] Thomas L. Marzetta, Erik G. Larsson, Hong Yang, Hien Quoc Ngo, *Fundamentals of Massive MIMO*, Cambridge U.K., Cambridge University Press, 1st edition, November 2016.

[16] T. L. Marzetta, *How much training is required for multiuser MIMO?*, in Proc. 40th Asilomar Conf. Signals, Syst., Comput. (ACSSC), Pacific Grove, CA, USA, pp. 359-363, Oct. 2006.

[17] 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.

[18] 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.

[19] Emil Bjornson, Erik G. Larsson, and Thomas L. Marzetta, *Massive MIMO: ten myths and one critical question*, IEEE Commun. Mag., vol. 54, no. 2, Feb. 2016.

[20] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Ingrid Moerman, and Gustavo Fraidenraich, *Channel Estimation for Massive MIMO TDD Systems Assuming Pilot Contamination and Frequency Selective Fading*, IEEE Access, vol. 5, no. 9, pp. 17733-17741, September 2017.

[21] George Marsaglia, *Ratios of Normal Variables*, Journal of Statistical Software, vol. 16, no. 4, May 2006.

[22] Felipe A. P. de Figueiredo, Fabbryccio A. C. M. Cardoso, Ingrid Moerman, Gustavo Fraidenraich, *On the Application of Massive MIMO Systems to Machine Type Communications*, IEEE Access, vol. 7, no. 12, pp. 2589-2611, December 2018.