

On the Performance Optimization of Linear Receivers in Uplink Massive MIMO FBMC Systems

Elena Vasquez and Ryan Thompson

Elena Vasquez, Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA; Ryan Thompson, Department of Communication and Signal Processing, University of California, San Diego, CA, USA

Abstract—Offset-QAM-based filterbank multicarrier (FBMCOQAM) has been shown to be a promising alternative to cyclic prefix-orthogonal frequency division multiplexing (CP-OFDM). More recently, the use of FBMC-OQAM has been proposed in combination with massive MIMO communications. In this paper, the performance of an FBMC-OQAM uplink massive MIMO system is theoretically characterized in terms of the output mean squared error (MSE) of the estimated transmitted symbols and for three types of linear receivers, namely, zero forcer (ZF), linear minimum mean squared error (LMMSE) and matched filter (MF). Using random matrix theory, the output MSE of these receivers is asymptotically characterized as the number of base station (BS) antennas N and the number of users K grow large, while keeping a finite ratio N/K . The obtained expressions allow to draw many conclusions, some of which were already noticed in the literature but not yet theoretically proven. First, the MSE becomes uniform across the frequency band as a result of the channel hardening effect. Secondly, it is shown that a good synchronization of the users is crucial in a massive MIMO scenario. Finally, if the users are well synchronized, the different terms that compose the MSE, such as noise, inter-user interference (IUI) and the distortion caused by the channel frequency selectivity, become negligible for large values of the ratio N/K . This effect was previously referred to as “self-equalization” in the literature.

I. INTRODUCTION

In recent years, massive MIMO systems have received a lot of attention in the wireless communication literature. These architectures are able to achieve the high data rate requirements targeted by 5G communication networks in terms of capacity and energy efficiency [1]–[3]. By using a very large number of antennas at the base station, these systems are able to simultaneously serve multiple users occupying the same time-frequency

resources and separated using space division multiple access (SDMA).

The idea of using FBMC-OQAM modulations in combination with massive MIMO communications was first introduced in [4]. FBMC-OQAM technologies inherently have significant advantages compared to CP-OFDM modulations [5]. First, the absence of cyclic prefix increases the spectral efficiency. Furthermore, the very good time-frequency localization of the prototype filter provides a lot of flexibility for spectrum allocation, as well as a high robustness to synchronization errors [6]. One should note that, to achieve a higher spectral selectivity, the prototype pulse needs to be more spread out in the time domain, which induces more latency and might be detrimental for short burst transmissions. Pre-loading techniques and tail shortening algorithms were proposed to compensate for this effect [7]–[10].

In general, the FBMC-OQAM orthogonality is well known to be progressively destroyed as the channel frequency selectivity becomes large [11], i.e., when the channel cannot be approximated as flat at the subcarrier level. This is due to the generation of inter-symbol interference (ISI) and inter-carrier interference (ICI). However, it appears that this distortion decreases as the number of BS antennas increases, even in the case of strong channel selectivity and for simple linear receivers [4]. By simple linear receivers, we mean that equalization is performed on a subcarrier basis and with single-tap decoding matrices, leading to a low implementation complexity. It is stated in [4] that, thanks to this so called “self-equalization effect”, the subcarrier spacing can be increased in massive MIMO FBMC systems, reducing the number of subcarriers and leading to a lower sensitivity to carrier frequency offset and peak-to-average-power-ratio. In [12], the same authors propose to use the frequency spreading implementation of FBMC [13], [14] in the massive MIMO setting. This would allow for a more accurate channel equalization, which would allow for a further reduction of the number of subcarriers. In [15], the prototype filter is modified in order to remove the distortion floor as the number of BS antennas grows to infinity.

As explained above, the combination of FBMC and massive MIMO has been studied in several previous works. However, to the best of our knowledge, there does not exist a theoretical analysis of the FBMC performance in a massive MIMO scenario. In [15], bounds for the



signal-to-interference-plus-noise ratio (SINR) are derived. However, these bounds only hold when the number of BS antennas N grows to infinity, for a fixed number of users K . In practice however, N is finite and may not be sufficiently large with respect to K , such that the derived bounds are very far from the actual SINR value for practical K and N . Furthermore, the work of [15] assumes a common power delay profile (PDP) per user, equal channel gains for each user and does not consider spatial channel correlation at the BS.

In this paper, the performance of an FBMC-OQAM uplink

K users, N antennas per base station

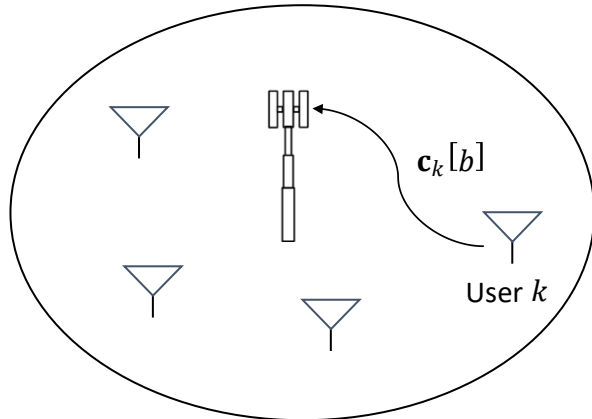


Fig. 1. Uplink massive MIMO system.

massive MIMO system is theoretically characterized in terms of the mean squared error (MSE) of the estimated transmitted symbols and at the output of the ZF, the LMMSE and the MF linear receivers. Using random matrix theory, the MSE approximation derived in [11] is asymptotically analyzed as N and K grow to infinity, while keeping a finite ratio N/K , for each type of receiver. Even if the results are only asymptotically true, it is shown that for practical values of N and K , the MSE is very close to its asymptotic equivalent, so that the approximations are very accurate and useful in practice. The obtained asymptotic MSE expressions are very compact and give a lot of insight on how the different scenario parameters affect the MSE, i.e., the number of BS antennas N , the number of users K , the spatial correlation between the BS antennas, the small-scale and large-scale fading and the different PDP of each user. More specifically, we draw the following conclusions.

First, for each type of receiver under study, the asymptotic MSE almost surely converges to a deterministic quantity that does not depend on the subcarrier index, meaning that the MSE becomes flat across the frequency band.

Secondly, a good synchronization of the users and the BS is crucial in the considered SDMA uplink massive MIMO scenario. Indeed, it is shown here that a sizable part of the distortion is related to the average delay of the channel or in other words, the timing synchronization. This part of the output distortion does not go to zero as N/K grows large. This performance floor was earlier noticed in [4], [15], though not related to timing synchronization.

Finally, it is shown that, if users are well synchronized, the different terms that compose the MSE, such as noise, interuser interference (IUI) and the distortion caused by the channel frequency selectivity, become negligible for large values of the ratio N/K . This effect was previously referred to as “self-equalization” in [4], [15].

The rest of this paper is structured as follows. Section II provides details on the system model of the considered uplink massive MIMO scenario. A per-subcarrier MSE approximation previously derived in the literature is recalled, which holds for general types of linear receivers. Furthermore, the channel model under study in the current massive MIMO scenario is described. In Section III, the asymptotic performance of FBMC is studied for the three types of linear receivers. Section IV validates the theoretical study of the self-equalization effect in massive MIMO FBMC systems in practical situations. Finally, Section V concludes the paper.

A. Notations

Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively (resp.). Superscripts $*$, T and H stand for conjugate, transpose and Hermitian transpose operators. The symbols tr , \mathbb{E} , Im and Re denote the trace, expectation, imaginary and real parts, respectively. \mathbb{I} is the imaginary unit. \mathbf{I}_N denotes the identity matrix of order N . \otimes stands for the Kronecker product and $\delta[n]$ is the Kronecker delta. The $\text{diag}(\cdot)$ operator applied to a vector returns a diagonal matrix whose k -th diagonal entry is equal to the k -th entry of this vector. The $\text{diag}(\cdot)$ operator applied to a squared matrix returns the same matrix with off-diagonal elements set to zero.

II. SYSTEM MODEL

A. Massive MIMO FBMC-OQAM Transmission Model

Let us consider a massive MIMO uplink transmission with one BS, equipped with N antennas, and K single-antenna users. The system topology and the transmission chain considered in the following are depicted in Fig. 1 and Fig. 2 respectively. Each user simultaneously

transmits one data stream to the the BS at the same frequency resources, using SDMA. It is assumed that the BS has knowledge of the channel response of all the users in the system. The number of subcarriers is denoted by $2M$. The vector $\mathbf{d}_{m,l} \in \mathbb{R}^{K \times 1}$ is made of the real-valued symbols transmitted at subcarrier m and multicarrier symbol l by each user. The symbols $\mathbf{d}_{m,l} \in \mathbb{R}^{K \times 1}$ are assumed bounded, independent and identically distributed random variables with zero mean and variance¹ $P_s/2$. The symbols $\mathbf{d}_{m,l}$ are FBMC-OQAM modulated using a prototype pulse $p[n]$ with unit energy and of length $2M\kappa$, where κ is the overlapping factor. The aggregated transmitted signal $\mathbf{s}[n] \in \mathbb{C}^{K \times 1}$ can be written as

$$\mathbf{s}[n] = \sum_{m=0}^{2M-1} \sum_{l=0}^{N_s-1} \mathbf{d}_{m,l} p_{m,l}[n]$$

where $p_{m,l}[n] = j^{l+m} p[n - lM] e^{j \frac{2\pi}{2M} m(n - \frac{2M\kappa-1}{2})}$. We denote by $\mathbf{C}[b] \in \mathbb{C}^{N \times K}$ the channel impulse response, which is assumed to be finite of length² L . The aggregated received

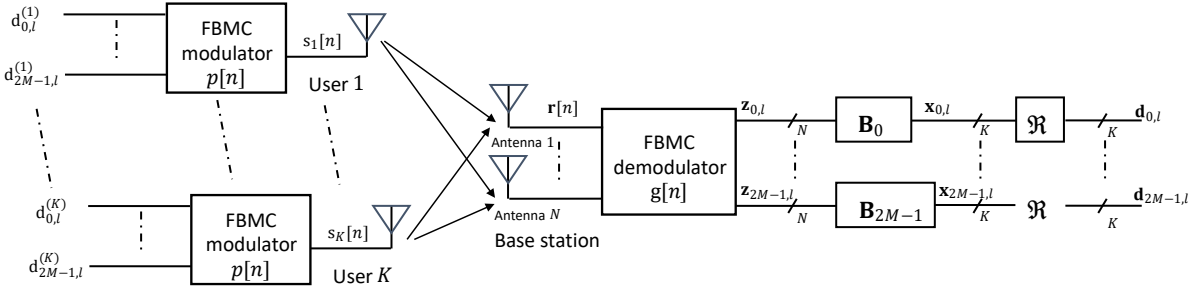


Fig. 2. Uplink massive MIMO offset-QAM-based filterbank multicarrier (FBMC-OQAM) transmission chain.

signal at the BS antenna, denoted by $\mathbf{r}[n] \in \mathbb{C}^{N \times 1}$, is given by

$$\mathbf{r}[n] = \sum_{b=0}^{L-1} \mathbf{C}[b] \mathbf{s}[n-b] + \mathbf{w}[n], \quad (1)$$

where $\mathbf{w}[n]$ represents zero mean additive circularly white Gaussian noise samples with variance σ_w^2 . The received signal $\mathbf{r}[n]$ is FBMC-OQAM demodulated using a prototype pulse $g[n]$ of length $2M\kappa$ and of unit energy. The signal after demodulation, at subcarrier m_0 and multicarrier symbol l_0 , is denoted by $\mathbf{z}_{m_0,l_0} \in \mathbb{C}^{N \times 1}$, and can be written as

$$\mathbf{z}_{m_0,l_0}$$

$$\mathbf{z}^{m_0,l_0} = \sum_{n=0} \mathbf{r}[n] g_{m_0,l_0}^*[n], \quad (2)$$

where $g_{m,l}[n] = j^{l+m} g[n - lM] e^{j \frac{2\pi}{2M} m(n - \frac{2M\kappa-1}{2})}$. In classical

FBMC approaches, the channel is assumed to remain constant over the frame duration and frequency flat at the subcarrier level. Under these conditions and if the pulses $p[n]$ and $g[n]$ are well localized in time and frequency, a good approximation of \mathbf{z}_{m_0,l_0} is given by

$$\mathbf{z}_{m_0,l_0} \approx \sum_{m=0}^{2M-1} \sum_{l=0}^{N_s-1} \mathbf{H}_{m_0} \mathbf{C}[b] \mathbf{d}_{m,l} p_{m,l}[n] g_{m_0,l_0}^*[n] + \mathbf{w}_{m_0,l_0}$$

where $\mathbf{H}_{m_0} = \sum_{b=0}^{L-1} \mathbf{C}[b] e^{-j \frac{2\pi}{2M} b m_0}$ is the channel frequency

response evaluated at the subcarrier of interest and \mathbf{w}_{m_0,l_0} is the filtered noise. Commonly, the channel is then

equalized by applying a one-tap matrix multiplication $\mathbf{B}_{m_0} \in \mathbb{C}^{K \times N}$, i.e.,

$$\mathbf{x}_{m_0,l_0} = \mathbf{B}_{m_0} \mathbf{z}_{m_0,l_0}$$

$$\mathbf{x}_{m_0,l_0} \approx \sum_{m,l} \mathbf{B}_{m_0} \mathbf{H}_{m_0} \mathbf{C}[b] \mathbf{d}_{m,l} p_{m,l}[n] g_{m_0,l_0}^*[n] + \mathbf{w}_{m_0,l_0}$$

The pulses $p[n]$ and $g[n]$ are assumed to be perfect reconstruction (bi-orthogonal) pulses. In this situation, one has $\Re\{\sum_{n=0}^{2M\kappa-1} p_{m,l}[n] g_{m_0,l_0}^*[n]\} = \delta_{m-m_0, l-l_0}$ and the trans-

mitted symbols \mathbf{d}_{m_0,l_0} are finally estimated as $\hat{\mathbf{d}}_{m_0,l_0} = \langle \mathbf{x}_{m_0,l_0} \rangle$. If the noise level is low, if the channel is not too

¹ The factor $1/2$ comes from the fact that P_s is the variance of the complex QAM symbol while $P_s/2$ is the variance of the related real PAM symbol. Furthermore, in practice, users may have a different transmit power. This can be taken into account by the channel gain of each user, that is specified in the channel model presented in the next section.

² Note that the channel of each user may have a different length. L would then be the supremum of all the channel lengths, which we assume bounded.

frequency selective and if $\mathbf{B}_{m_0}\mathbf{H}_{m_0} \approx \mathbf{I}_K$, one will have

$\hat{\mathbf{d}}_{m_0,l_0}$ close to \mathbf{d}_{m_0,l_0} .

B. MSE Approximation under High Channel Frequency Selectivity

In practice, the presence of channel frequency selectivity will deteriorate the symbol estimate, which will be affected by IUI, ICI and ISI. Moreover if the channel is not perfectly inverted at the receiver, i.e., $\mathbf{B}_m\mathbf{H}_m \neq \mathbf{I}_K$, IUI coming from the other users will impact the symbol estimate. We define the aggregated MSE of all streams at subcarrier m as

$$\text{MSE}_m = E\|\mathbf{d}_{m,l} - \hat{\mathbf{d}}_{m,l}\|_2^2,$$

where the expectation is taken over transmitted symbols and noise samples. The exact expression of the MSE is an intricate function of the channel impulse response. An MSE approximation was first derived in [16] and later extended to any form of linear receivers in [11]. For the approximation to hold, it should be assumed that the transmit and receive prototype pulses $p[n]$ and $g[n]$ are either symmetric or anti-symmetric and real-valued. Moreover, $g[n]$ is assumed to be obtained by the discretization of a smooth real-valued analog waveform $g(t)$, which is only non zero for $t \in [-\kappa/2, \kappa/2]$ and with bounded derivatives, so that,

$$g[n] = g\left(\left(n - \frac{2M\kappa - 1}{2}\right) \frac{1}{2M}\right),$$

for $n = 0, \dots, 2M\kappa - 1$ and zero elsewhere. Under these assumptions and when no pre-equalizer is used at the transmitter side, an approximation of the MSE is given by [11]

$$\begin{aligned} \text{MSE}^m &= \frac{P_T}{2K} \text{tr} \left[(\mathbf{B}_m\mathbf{H}_m - \mathbf{I}_K) (\mathbf{B}_m\mathbf{H}_m - \mathbf{I}_K)^H \right] \\ &+ \frac{P_T}{K} \frac{\eta}{(2M)^2} \text{tr} \left[\left(\mathbf{B}_m\mathbf{H}_m^{(1)} \right) \left(\mathbf{B}_m\mathbf{H}_m^{(1)} \right)^H \right] \\ &+ \frac{P_T}{K} \frac{\eta}{(2M)^2} \Re \text{tr} \left[\left(\mathbf{B}_m\mathbf{H}_m - \mathbf{I}_K \right) \left(\mathbf{B}_m\mathbf{H}_m^{(2)} \right)^H \right] \\ &+ \frac{\sigma_w^2}{2} \text{tr} \left[\mathbf{B}_m\mathbf{B}_m^H \right] + \epsilon, \end{aligned} \tag{3}$$

where $P_T = KP_s$ is the total transmit power, η is a pulserelated quantity properly defined in Appendix VI-A. For an integer r , $\mathbf{H}_m^{(r)}$ denotes the r -th order derivative of the channel in the frequency domain and evaluated at subcarrier m , i.e.,

$$\mathbf{H}_m^{(r)} = \sum_{b=0}^{L-1} (-jb)^r \mathbf{C}[b] e^{-j\frac{2\pi}{2M}mb} \tag{4}$$

Note that $\mathbf{H}_m^{(0)} = \mathbf{H}_m$. The MSE approximation is obtained by a Taylor approximation of the demodulated signal, truncated to its first order terms [11]. The term is an error contribution that depends on the degree of frequency selectivity of the channel, which can be characterized by the ratio $\frac{L}{2M}$. As $\frac{L}{2M} \rightarrow 0$, the approximation can be shown to be asymptotically exact [11]. As will be shown in the simulations, for practical channel lengths and number of subcarriers, the approximation is very accurate. Therefore, we will neglect the error term. Note that the approximation takes into account the effect of IUI, ICI, ISI and additive noise. The first term is for instance due to intrinsic interference and IUI, leaking on the symbol of interest, if the channel is not perfectly inverted at the receiver. The fourth term is related to the additive noise while the second and third ones are caused by the channel frequency selectivity. As the channel becomes more frequency selective, its derivatives $\mathbf{H}_m^{(r)}$ are amplified, progressively destroying the FBMC-OQAM orthogonality.

C. Channel Model

To study the MSE behavior when the number of antennas grows large, we first need to define a channel model that is relevant in the considered uplink massive MIMO scenario in Fig. 1. We denote by $\mathbf{c}_k[b] \in \mathbb{C}^{N \times 1}$, $k = 1, \dots, K$, the b -th tap of the channel impulse response between the k -th user and the BS antennas. In the following, we will assume that vector $\mathbf{c}_k[b]$ is modeled as

$$\mathbf{c}_k[b] = \underbrace{\mathbf{C}_{\text{BS}}^{1/2}}_{\text{BS correlation}} \underbrace{\mathbf{g}_k[b]}_{\text{Small-scale fading}} \underbrace{\left\{ \sqrt{\beta_k} \sqrt{p_k[b]} \right\}}_{\text{Channel gain}} \underbrace{\text{PDP}}_{\text{PDP}} \tag{5}$$

The positive matrix $\mathbf{C}_{\text{BS}} \in \mathbb{C}^{N \times N}$ models the spatial correlation between the BS antennas, normalized such that $\text{tr}[\mathbf{C}_{\text{BS}}] = N$. The vector $\mathbf{g}_k[b] \in \mathbb{C}^{N \times 1}$ is made of i.i.d. zero mean and unit variance Gaussian entries that accounts for

the small-scale fading. The coefficient β_k accounts for the channel gain related to user k , normalized such that $\sum_{k=1}^K \beta_k = K$. The coefficient $p_k[b]$ models the power of the b -th tap, normalized such that $\sum_{b=0}^{L-1} p_k[b] = 1$. Note that since $p_k[b]$ depends both on k and b , the model can take into account different PDP's for each user, depending on its own environment. Collecting the vectors $\mathbf{c}_k[b]$ for all users $k = 1, \dots, K$, the matrix $\mathbf{C}[b] \in \mathbb{C}^{N \times K}$ can be formulated as

$$\mathbf{C}[b] = \mathbf{C}_{\text{BS}}^{1/2} \mathbf{G}[b] \mathbf{D}_{\beta,b}^{1/2}, \quad b = 0, \dots, L-1, \quad (6)$$

diagonal matrix $\mathbf{D}_{\beta,b} \in \mathbb{C}^{K \times K}$ is given by where the

$\mathbf{D}_{\beta,b}^{1/2} = \text{diag}(\beta_1 p_1[b], \dots, \beta_K p_K[b])$ and $\mathbf{G}[b] = (\mathbf{g}_1[b], \dots, \mathbf{g}_K[b])$. Inserting (6) in (4), the r -th derivative of the channel frequency response at subcarrier m can be expressed as

$$\begin{aligned} \mathbf{H}_m^{(r)} &= \mathbf{C}_{\text{BS}}^{1/2} \sum_{b=0}^{L-1} (-jb)^r \mathbf{G}[b] \mathbf{D}_{\beta,b}^{1/2} e^{-j\frac{2\pi}{2M}mb} \\ &= \mathbf{C}_{\text{BS}}^{1/2} \mathbf{G} \Phi^r \mathbf{D}_{1/2,m}, \end{aligned} \quad (7)$$

³The coefficient β_k includes the effect of large-scale fading and power control.

where we defined

$$\mathbf{G} = [\mathbf{G}[0], \dots, \mathbf{G}[L-1]], \quad (8)$$

$$\Phi = (-j\mathbf{1} \Psi \otimes \mathbf{I}_K),$$

$$\Psi = \text{diag}(0, \dots, L-1),$$

$$\mathbf{D}_{1/2,m} = \left(e^{-j\frac{2\pi}{2M}m0} \mathbf{D}_{\beta,0}^{1/2}, \dots, e^{-j\frac{2\pi}{2M}m(L-1)} \mathbf{D}_{\beta,L-1}^{1/2} \right)^T$$

and matrix Φ^0 is defined as the identity \mathbf{I}_{LK} . Furthermore, we define key channel statistics that will help characterize the distortion induced by the channel frequency selectivity. The average delay $\tau_{k,\text{av}}$ and the delay spread $\tau_{k,\text{rms}}$ associated to user k are defined as

$$\begin{aligned} \tau_{k,\text{av}} &= \sum_{b=0}^{L-1} b p_k[b], \\ \tau_{k,\text{rms}} &= \sqrt{\sum_{b=0}^{L-1} b^2 p_k[b] - \tau_{k,\text{av}}^2}. \end{aligned}$$

Further, we define the three following diagonal matrices of size $K \times K$

$$\mathbf{D}_\beta = \text{diag}(\beta_1, \dots, \beta_K),$$

$$\mathbf{D}_{\text{av}} = \text{diag}(\tau_{1,\text{av}}, \dots, \tau_{K,\text{av}}),$$

$$\mathbf{D}_{\text{rms}} = \text{diag}(\tau_{1,\text{rms}}, \dots, \tau_{K,\text{rms}}),$$

which respectively contain on their diagonal the channel gain, the average delay and the delay spread of each user.

III. FBMC PERFORMANCE IN MASSIVE MIMO SYSTEMS

In this section, we aim at characterizing the performance of FBMC in an uplink massive MIMO scenario for three types of classical linear receivers, namely, the ZF, the LMMSE and the MF receivers. These

are designed based on the assumption of a frequency flat channel at the subcarrier level, i.e., $\mathbf{H}_m^{(1)} \approx \mathbf{0}$

and $\mathbf{H}_m^{(2)} \approx \mathbf{0}$. In practice, it will not be the case and this simple design will lead to ICI, ISI and IUI, which is taken into account by the second and third terms of (3). In the following, we study the asymptotic behavior of the MSE in (3) for each type of linear receiver. We will show that (3) converges in each case almost surely to an asymptotic equivalent as N and K grow infinitely large, for a finite ratio N/K . Of course, in a massive MIMO scenario, the number of BS antennas N and the number of users K are finite but they are sufficiently large so that the actual MSE becomes very close to its asymptotic equivalent, as will be shown via simulations.

In the sequel, as N and K grow large, we will assume that

$$\text{(As1): } 0 < \liminf \frac{N}{K} \leq \limsup \frac{N}{K} < \infty,$$

$$\text{(As2): } \limsup_N K \mathbf{C}_{\text{BS}k} < \infty,$$

$$\text{(As3): } \limsup_K K \mathbf{D}_{\beta,bk} < \infty \text{ for all } b.$$

A. ZF and LMMSE Receivers

We will here study the performance of two particular receivers. The first one, the ZF receiver, is obtained by minimizing

(3) under a channel inversion constraint and assuming a frequency flat channel at the subcarrier level ($\mathbf{H}_m^{(1)} = \mathbf{H}_m^{(2)} \approx \mathbf{0}$),

i.e.,

$$\begin{aligned} \mathbf{B}_m^{\text{ZF}} &= \text{argmin}_{\mathbf{B}_m} \text{MSE}_m \text{ s.t. } \mathbf{B}_m \mathbf{H}_m = \mathbf{I}_K \\ &= \left(\mathbf{H}_m^H \mathbf{H}_m \right)^{-1} \mathbf{H}_m^H. \end{aligned}$$

It is well known that inverting the channel may lead to noise amplification, which is particularly detrimental for low signal-to-noise ratios (SNR). An alternative is to use an LMMSE receiver, obtained by minimizing (3) assuming a frequency flat channel at the subcarrier level ($\mathbf{H}_m^{(1)} = \mathbf{H}_m^{(2)} \approx \mathbf{0}$), i.e., $\mathbf{B}_{\text{LMMSE}m} = \text{argmin}_{\mathbf{B}_m} \text{MSE}_m$

$$= \left(\mathbf{H}_m^H \mathbf{H}_m + \frac{\sigma_w^2 K}{P_T} \mathbf{I}_K \right)^{-1} \mathbf{H}_m^H.$$

As soon as the channel becomes significantly frequency selective, we have $\mathbf{H}_m^{(1)} \neq \mathbf{0}, \mathbf{H}_m^{(2)} \neq \mathbf{0}$. Hence, \mathbf{B}_m^{ZF} and $\mathbf{B}_{\text{LMMSE}m}$ are not optimal anymore, as was shown in [11], [17]. Still, in the massive MIMO setting, we will show further that classical linear receivers, designed for the frequency flat case, such as \mathbf{B}_m^{ZF} and $\mathbf{B}_{\text{LMMSE}m}$, can lead to very small MSE values for large values of the ratio N/K , even for highly frequency selective channels. In the

following, we will more generally study a linear receiver of the form

$$\mathbf{B}^m = \left(\mathbf{H}_m^H \mathbf{H}_m + K \lambda \mathbf{I}_K \right)^{-1} \mathbf{H}_m^H \tag{9}$$

Note that if $\lambda = 0$, we retrieve the expression of the ZF receiver whereas if we choose $\lambda = \frac{\sigma_w^2}{P_T}$, we find the LMMSE receiver. Inserting the expression of this receiver in (3), the MSE becomes

$$\begin{aligned} \text{MSE}_m(\lambda) &= \frac{P_T}{2K} \lambda^2 \text{tr} \left[\left(\frac{\mathbf{H}_m^H \mathbf{H}_m}{K} + \lambda \mathbf{I}_K \right)^{-2} \right] \\ &+ \frac{P_T \tilde{\eta}}{K} \text{tr} \left[\left(\frac{\mathbf{H}_m^H \mathbf{H}_m}{K} + \lambda \mathbf{I}_K \right)^{-2} \frac{\mathbf{H}_m^H \mathbf{H}_m^{(1)} (\mathbf{H}_m^H \mathbf{H}_m^{(1)})^H}{K} \right] \\ &- \frac{P_T \tilde{\eta}}{K} \lambda \Re \text{tr} \left[\left(\frac{\mathbf{H}_m^H \mathbf{H}_m}{K} + \lambda \mathbf{I}_K \right)^{-2} \frac{(\mathbf{H}_m^H \mathbf{H}_m^{(2)})^H}{K} \right] \\ &+ \frac{\sigma_w^2}{2K} \text{tr} \left[\left(\frac{\mathbf{H}_m^H \mathbf{H}_m}{K} + \lambda \mathbf{I}_K \right)^{-2} \left(\frac{\mathbf{H}_m^H \mathbf{H}_m}{K} \right) \right], \end{aligned} \tag{10}$$

where $\tilde{\eta} = \frac{\eta}{(2M)^2}$. The MSE expression can be rewritten as a function of the matrix of Gaussian entries \mathbf{G} defined in (8).

The following theorem shows that the expression of $\text{MSE}_m(\lambda)$ given in (10), almost surely (a.s.) converges to a deterministic equivalent in a massive MIMO scenario.

Theorem 1. Under (As1)-(As3), as N and K grow infinitely large and if $\lambda \in \mathbb{C} \setminus (\mathbb{R}^- \cup \{0\})$, we have

$$\text{MSE}_m(\lambda) - \text{MSE}(\lambda) \xrightarrow{\text{a.s.}} 0,$$

and the expression of $\text{MSE}(\lambda)$ is given by

$$\begin{aligned} \overline{\text{MSE}}(\lambda) &= \frac{P_T}{2} \left(1 - \tilde{\delta}(\lambda) \delta(\lambda) \right) + \frac{\sigma_w^2 - \lambda P_T}{2} \frac{\alpha(\lambda) \tilde{\alpha}(\lambda)}{1 - \gamma(\lambda) \tilde{\gamma}(\lambda)} \\ &+ \frac{P_T \tilde{\eta}}{K} \text{tr} \left[\tilde{\delta}(\lambda) \mathbf{D}_\beta \left(\tilde{\delta}(\lambda) \mathbf{D}_\beta + \lambda \mathbf{I}_K \right)^{-1} \mathbf{D}_{\text{av}}^2 \right] \\ &+ \frac{P_T \tilde{\eta}}{K} \frac{\alpha(\lambda) \tilde{\gamma}(\lambda)}{1 - \gamma(\lambda) \tilde{\gamma}(\lambda)} \text{tr} \left[\mathbf{D}_\beta \mathbf{D}_{\text{rms}}^2 \right] \\ &+ \frac{P_T \tilde{\eta}}{K} \frac{\lambda \tilde{\alpha}(\lambda)}{1 - \gamma(\lambda) \tilde{\gamma}(\lambda)} \text{tr} \left[\mathbf{D}_\beta \left(\tilde{\delta}(\lambda) \mathbf{D}_\beta + \lambda \mathbf{I}_K \right)^{-2} \mathbf{D}_{\text{rms}}^2 \right], \end{aligned} \tag{11}$$

where $\tilde{\delta}(\lambda)$ and $\delta(\lambda)$ are the unique positive solutions of the following system of equations in $(\tilde{\delta}(\lambda), \delta(\lambda))$,

$$\delta(\lambda) = \frac{1}{K} \text{tr} \left[\mathbf{D}_\beta \left(\tilde{\delta}(\lambda) \mathbf{D}_\beta + \lambda \mathbf{I}_K \right)^{-1} \right] \tag{12}$$

$$\tilde{\delta}(\lambda) = \frac{1}{K} \text{tr} \left[\mathbf{C}_{\text{BS}}(\delta(\lambda)) \mathbf{C}_{\text{BS}} + \mathbf{I}_N \right]^{-1} \tag{13}$$

Furthermore, we have defined,

$$\alpha(\lambda) = \frac{1}{K} \text{tr} \left[\mathbf{D}_\beta \left(\tilde{\delta}(\lambda) \mathbf{D}_\beta + \lambda \mathbf{I}_K \right)^{-2} \right] \tag{14}$$

$$\gamma(\lambda) = \frac{1}{K} \text{tr} \left[\mathbf{D}_\beta^2 \left(\tilde{\delta}(\lambda) \mathbf{D}_\beta + \lambda \mathbf{I}_K \right)^{-2} \right] \tag{15}$$

$$\tilde{\alpha}(\lambda) = \frac{1}{K} \text{tr} \left[\mathbf{C}_{\text{BS}}(\delta(\lambda)) \mathbf{C}_{\text{BS}} + \mathbf{I}_N \right]^{-2} \tag{16}$$

$$\tilde{\gamma}(\lambda) = \frac{1}{K} \text{tr} \left[\mathbf{C}_{\text{BS}}^2(\delta(\lambda)) \mathbf{C}_{\text{BS}} + \mathbf{I}_N \right]^{-2} \tag{17}$$

The above properties also hold for $\lambda = 0$ when $\inf \frac{N}{K} > 1$.

Proof. The proof is given in Appendix VI-B. \square

The expression of $\text{MSE}(\lambda)$ in (11) is not trivial to interpret due to the dependence in the variables $\tilde{\delta}(\lambda)$ and $\delta(\lambda)$. A first remark is that it does not depend on the subcarrier index m , meaning that the MSE becomes flat over frequency. Further, the two first terms proportional to σ_w^2 and P_T are terms of additive noise and IUI. The terms proportional to $\tilde{\eta}$, are ICI and ISI caused by the channel frequency selectivity, related to the average delay and the delay spread of the channel. By making further assumptions on the channel statistics, the next corollary will give a better intuition of the different terms of

$\text{MSE}(\lambda)$.

1) *ZF receiver:* Theorem 1 can be particularized to the ZF receiver case by setting $\lambda = 0$ and ensuring that $\inf \frac{N}{K} > 1$.

Let us define $\text{MSE}^{\text{ZF}} = \text{MSE}(0)$.

Corollary 1. If we assume that the channel is spatially uncorrelated at the receiver side, i.e., $\mathbf{C}_{\text{BS}} = \mathbf{I}_N$, the expression

of MSE^{ZF} simplifies to

$$\begin{aligned} \text{MSE}^{\text{ZF}} &= \frac{P}{2} \frac{\frac{1}{\frac{N}{K} - 1}}{K} + P_T \tilde{\eta} \frac{\text{tr} \left[\mathbf{D}_\beta^{-1} \right] \text{tr} \left[\mathbf{D}_\beta \mathbf{D}_{\text{rms}}^2 \right]}{K^{\text{av}}} \\ &+ P_T \tilde{\eta} \frac{1}{\frac{N}{K} - 1} \frac{1}{K} \end{aligned}$$

If further, we assume a common PDP for all users but different per-user large-scale fading coefficients, we

have $p_k[b] = p[b]$, which does not depend on k . We also have $\mathbf{D}_{\beta,b} = p[b]\mathbf{D}_\beta$.

The expression of MSE then becomes

$$\begin{aligned} & \frac{P_T \sigma_w^2}{2} \frac{1}{\frac{N}{K} - 1} \frac{\text{tr}[\mathbf{D}_\beta^{-1}]_2}{K} \text{MSE} \\ & + P_T \eta \tau_{\text{av}} \\ & + \frac{P_T \tilde{\eta}}{K} \frac{\text{tr}[\mathbf{D}_\beta^{-1}]_2}{\frac{N}{K} - 1} \tau_{\text{rms}} P \end{aligned}$$

where we defined $\tau_{\text{av}} = \tau_{\text{av},k}$, which does not depend on k .

Proof. The proof is simply given by particularizing the result of Theorem 1 to the special cases mentioned above. First, if $\mathbf{C}_{\text{BS}} = \mathbf{I}_N$, the solution of the system of equations (12) and (13) admits a closed-form expression ($\lambda = 0$), given by $\tilde{\delta} = \frac{N}{K} - 1$ and $\delta = \tilde{\delta}^{-1}$. Replacing this expression of δ and $\tilde{\delta}$ in (14)-(17) for $\lambda = 0$, we find the first result of the Corollary. Moreover, if $p_k[b] = p[b]$, we have $\mathbf{D}_{\text{av}} = \tau_{\text{av}}\mathbf{I}_K$ and $\mathbf{D}_{\text{rms}} = \tau_{\text{rms}}\mathbf{I}_K$. \square

By adding some assumptions, Corollary 1 provides additional insights into the physical meaning of the different terms of the MSE. The first term is simply related to the additive noise power after equalization and goes to zero as N/K becomes large. This makes sense since with a larger N/K , the channel inversion will be better conditioned and leads to lower noise amplification. As explained before, the two other terms are related to the channel frequency selectivity. Note that the term related to the delay spread of the channel τ_{rms} also goes to zero as N/K grows large. This is part of the so-called ‘‘selfequalization’’ effect, previously noticed in [4].

Surprisingly, the term related to the delay average of the channel τ_{av} is independent of N and K . Hence, this corresponds to the remaining error floor as N/K grows to infinity, previously noticed in [15]. The presence of this term motivates the need for a strict synchronization of the users at the transmit side in the considered SDMA uplink scenario. In practice, if the users and the BS are well synchronized, this term is negligible relatively to the other terms. Further, it can even be made very close to zero. This can be done by redefining the channel $\mathbf{C}[b]$ as being not causal and having a $\tau_{\text{av}} \approx 0$. In other words, the BS can re-center the demodulation window to minimize the average channel delay.

2) *LMMSE Receiver:* Theorem 1 can also be particularized to the LMMSE receiver case by setting $\lambda = \frac{\sigma_w^2}{P_T}$. Let us define

$$\text{MSE}^{\text{LMMSE}} = \text{MSE}(\sigma_w^2/P_T)$$

Unfortunately, the asymptotic expression MSE is not trivial to interpret due to the dependence in the variables $\delta(\lambda)$ and $\tilde{\delta}(\lambda)$. Even by making more assumptions as in Corollary 1, the expression does not simplify a lot. However, we will see via simulations that the performance of the LMMSE receiver gets very close to the one of the ZF for relatively high values of N/K and/or high SNR. Hence, in these regimes, the conclusions that were written for the ZF receiver, approximately hold for the LMMSE receiver.

B. MF Receiver

A disadvantage of the ZF and LMMSE receivers is that they require one matrix inversion, which significantly increases their implementation complexity. To overcome this, an alternative receiver known as MF or maximum-ratio combining, has received a lot of attention in the massive MIMO literature. If

we denote by $\mathbf{h}_{j,m} \in \mathbb{C}^{N \times 1}$ the channel frequency response at subcarrier m between user j and the BS antennas, the BS simply processes this stream by a multiplication by its Hermitian transpose $\mathbf{h}_{j,m}^H / \|\mathbf{h}_{j,m}\|^2$. The normalization factor $\|\mathbf{h}_{j,m}\|^2$ ensures a unit gain for the channel of interest. We can write the matrix form of this receiver as,

$$\mathbf{B}^{\text{MF}_m} = \text{diag}^{-1} \left(\mathbf{H}_m^H \mathbf{H}_m \right) \mathbf{H}_m^H,$$

where the operator $\text{diag}(\cdot)$ applied to a square matrix \mathbf{A} returns a diagonal matrix with diagonal entries equal to the diagonal entries of \mathbf{A} and $\text{diag}^r(\mathbf{A})$ is the r -th power of matrix $\text{diag}(\mathbf{A})$.

1) *MSE expression:* Inserting the expression of \mathbf{B}^{MF_m} into

(3), the MSE expression for the MF receiver becomes

$$\begin{aligned} \text{MSE}_m^{\text{MF}} &= \frac{P_T}{2K} \text{tr} \left[\text{diag}^{-2} \left(\mathbf{H}_m^H \mathbf{H}_m \right) \left(\mathbf{H}_m^H \mathbf{H}_m \right)^2 \right] - \frac{P_T}{2} \\ &+ \frac{P_T \tilde{\eta}}{K} \text{tr} \left[\text{diag}^{-2} \left(\mathbf{H}_m^H \mathbf{H}_m \right) \left(\mathbf{H}_m^H \mathbf{H}_m^{(1)} \right) \left(\mathbf{H}_m^H \mathbf{H}_m^{(1)} \right)^H \right] \\ &+ \frac{P_T \tilde{\eta}}{K} \langle \text{tr}^h \text{diag}^{-2} \left(\mathbf{H}_m^H \mathbf{H}_m \right) \left(\mathbf{H}_m^H \mathbf{H}_m \right) \left(\mathbf{H}_m^H \mathbf{H}_m^{(2)} \right)^H \rangle \\ &- \frac{P_T \tilde{\eta}}{K} \langle \text{tr}^h \text{diag}^{-1} \left(\mathbf{H}_m^H \mathbf{H}_m \right) \left(\mathbf{H}_m^H \mathbf{H}_m^{(2)} \right)^H \rangle \\ &+ \frac{\sigma_w^2}{2} \text{tr} \left[\text{diag}^{-1} \left(\mathbf{H}_m^H \mathbf{H}_m \right) \right], \end{aligned} \quad (18)$$

which can be written as a function of \mathbf{G} using (7).

2) *MSE deterministic equivalent for the MF receiver:* The following theorem gives the expression of

MSE^{MF}_m as the number of BS antennas N and the number of users K grow large for a finite ratio N/K .

Theorem 2. Under (As1)-(As3), as N and K grow infinitely large, we have,

$$\lim_{m \rightarrow \infty} \text{MSE}_m^{\text{MF}} \xrightarrow{\text{a.s.}} 0$$

such that the deterministic equivalent of MSE^{MF}_m is given by

$$\begin{aligned} \text{MSE}_m^{\text{MF}} &= \frac{P_T}{2} \text{tr}[\mathbf{D}_\beta^{-1}] \frac{\text{tr}[\mathbf{C}_{\text{BS}}^2]}{(\text{tr}[\mathbf{C}_{\text{BS}}])^2} + P_T \tilde{\eta} \frac{\text{tr}[\mathbf{D}_{\text{av}}^2]}{K} \\ &+ \frac{\sigma_w^2 \text{tr}[\mathbf{D}_\beta^{-1}]}{2 \text{tr}[\mathbf{C}_{\text{BS}}]} \end{aligned}$$

Proof. The proof is given in Appendix VI-C. □

Note again that MSE does not depend on the subcarrier index m . MSE is composed of three terms. The first one is related to IUI due to the fact that the channel is not perfectly inverted, $\mathbf{B}^{\text{MF}} \mathbf{H}_m \mathbf{6} = \mathbf{I}_K$. This term is generally dominant at mid to high SNR values. The second term is related to the average delay of the PDP of each user, whereas the last term is due to the additive noise. Surprisingly, in contrast to the ZF and LMMSE receivers, the deterministic equivalent

MSE does not depend on the delay spread of the channel, i.e., MSE does not depend on \mathbf{D}_{rms} .

Corollary 2. If we assume that the channel is spatially uncorrelated at the receiver side, i.e., $\mathbf{C}_{\text{BS}} = \mathbf{I}_N$, the expression

of MSE simplifies to

$$\text{MSE} = \frac{P_T}{2} \frac{\text{tr}[\mathbf{D}_\beta^{-1}]}{N} + P_T \tilde{\eta} \frac{\text{tr}[\mathbf{D}_{\text{av}}^2]}{K} + \frac{\sigma_w^2}{2} \frac{\text{tr}[\mathbf{D}_\beta^{-1}]}{N}$$

If further, we assume a common PDP for all users, as in

Corollary 1, MSE becomes,

$$\text{MSE} = \frac{P_T \text{tr} \mathbf{D}_\beta}{2N} + P_T \tilde{\eta} \tau_{\text{av}}^2 + \frac{\sigma_w^2 \text{tr} \mathbf{D}_\beta}{2N},$$

where τ_{av} and τ_{rms} are defined as in Corollary 1.

Proof. The proof is trivially given by the particularization of Theorem 2 to the special cases mentioned above. □

By making some additional assumptions, Corollary 2 allows for a very intuitive understanding of the behavior of the different terms. As mentioned earlier, the first term is related to IUI. As in the ZF and LMMSE cases, we notice the selfequalization impact of massive MIMO on FBMC: for large values of the ratio N/K , the first and third terms will go to zero. The only term that does not go to zero, is the second one, proportional to $\tilde{\eta}$ and τ_{av}^2 , similarly as in the ZF and LMMSE cases. This again emphasizes the fact that in an SDMA situation, the BS and the different users should be well synchronized. If this is the case, this term is negligible compared to the other two.

IV. SIMULATION RESULTS

This section aims at assessing the accuracy of the approximations through simulations. Furthermore, the performance of an uplink massive MIMO FBMC-OQAM system is characterized as a function of different parameters.

In the simulations, the subcarrier spacing is fixed to 15 kHz. The number of subcarriers is $2M = 128$. The prototype filter used in all simulations is the PHYDYAS pulse [18] with overlapping factor $\kappa = 4$. To characterize the PDP's $p_k[b]$ and the channel gains β_k of each user, we will assume that the users are clustered in four groups of equal number $K/4$. The PDP of the users of each cluster follow the same channel model, which is given by an ITU - Veh. A, an ITU - Veh. B, a 3GPP - Typical Urban or a 3GPP - Hilly Terrain channel model. Note that these models correspond to both mildly and highly frequency selective channels. Moreover, the users inside a same cluster have an equal channel gain β_k which takes one of the following four values [0.4,0.8,1.2,1.6]. The spatial correlation matrix at the BS \mathbf{C}_{BS} was fixed using the exponential model [19], i.e., its m,n entry is given by $[\mathbf{C}_{\text{BS}}]_{m,n} = \rho^{|m-n|}$. The parameter ρ will be referred to as the correlation coefficient and is fixed to 0.7, if not stated otherwise. The SNR is defined as P_T/σ_w^2 .

$$\text{SNR} = 15 \text{ dB}, N/K = 1.5, N = 6, K = 4$$

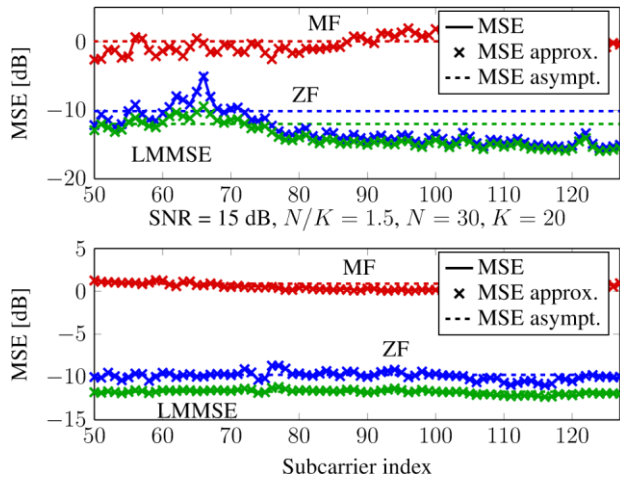


Fig. 3. MSE across the subcarriers. The MSE approximation in crosses perfectly matches the simulated MSE in continuous lines. As the number of users K and the BS antennas N increases (above figure), the MSE gets flat across frequency and converge to the derived asymptotic MSE.

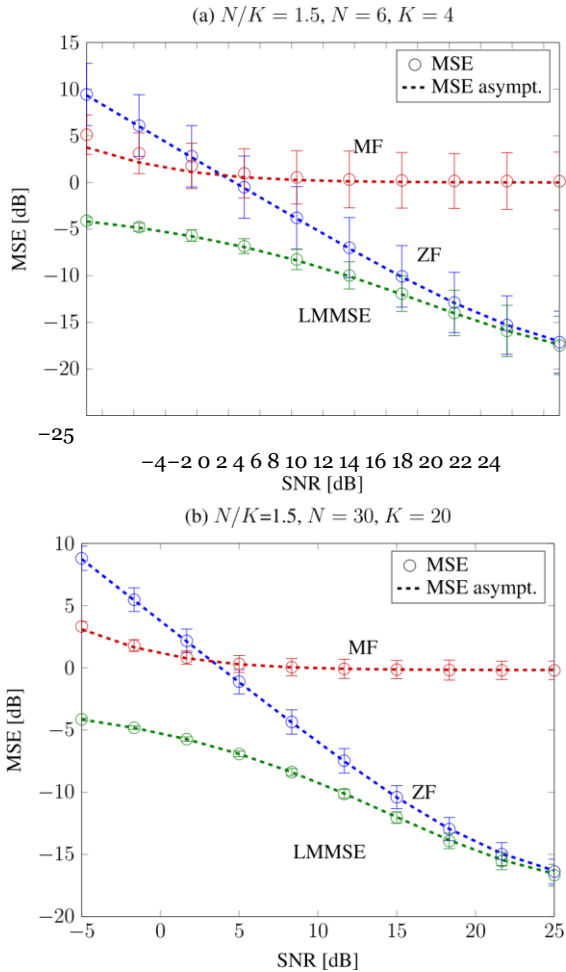


Fig. 4. Simulated MSE vs. asymptotic MSE as a function of the SNR. Vertical bars represent the 90% confidence interval.

A. Accuracy of the MSE approximations and frequency flattening effect

Fig. 3 shows the MSE of the three types of linear receivers as a function of the subcarrier index for an SNR of 15 dB. One

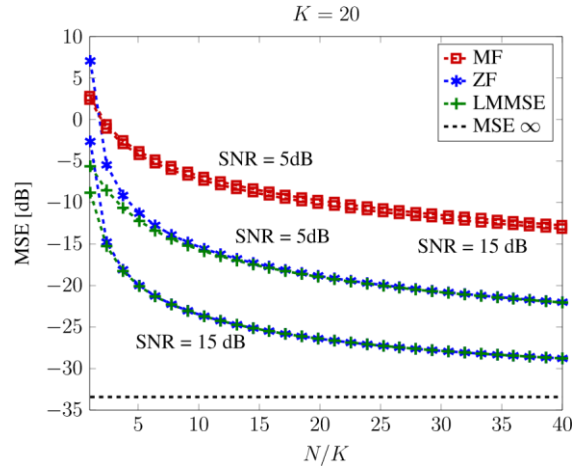


Fig. 5. Asymptotic MSE as a function of the ratio N/K .

can first check that the crosses, which corresponds to the MSE approximation of (3), perfectly match the simulated MSE. As it could be expected, the LMMSE is the best receiver closely followed by the ZF, while the MF has a poorer performance. Note that both Figures 3 (a) and (b) are plotted for the same ratio $N/K = 1.5$, while in (a), $N = 6$ and in (b), $N = 30$. One can clearly see in (b) that, as expected by Theorems 1-2, when N and K grow, the MSE gets flat across frequency and converges to the asymptotic expressions of the MSE derived in Theorems 1-3.

Figures 4 (a) and (b) plot the MSE as a function of the SNR for the same antenna configuration as in the previous figure. The dashed lines correspond to the asymptotic approximation of the MSE of Theorems 1-3 while the circles correspond to the MSE averaged over the subcarriers and the channel realizations. The vertical bars indicate the 90% confidence interval of the measured MSE across the spectrum and channel realizations. In other words, 90% of the subcarriers and channel realizations resulted in a simulated MSE within these vertical intervals. Observe that the asymptotic formulas of the MSE provide a very accurate description of the average MSE, even for relatively low values of N and K (in Fig. 4 (a)). Further, observe in Fig. 4 (b) that, as N and K grow, the variance of the simulated MSE is drastically reduced, confirming the convergence predicted by Theorems 1-3. Finally, note that at high SNR, the MSE begins to saturate due to the distortion induced by the channel selectivity, i.e., the terms related to the delay average and delay spread of the channel in Theorems 1-3. We will see in the following figure that these terms may go to zero as well.

B. On the self-equalization effect

Fig. 5 plots the asymptotic MSE as a function of the ratio N/K for the three receivers and two SNR values, i.e., 5 dB and 15 dB. One can clearly observe the self-equalization effect. As N/K increases, the MSE decreases and the different terms of the MSE progressively go to zero, except for the term related to the average delay of the channel, which corresponds to the performance floor, denoted by “MSE ∞ ” on the graph. Furthermore, one can note that the performance of the

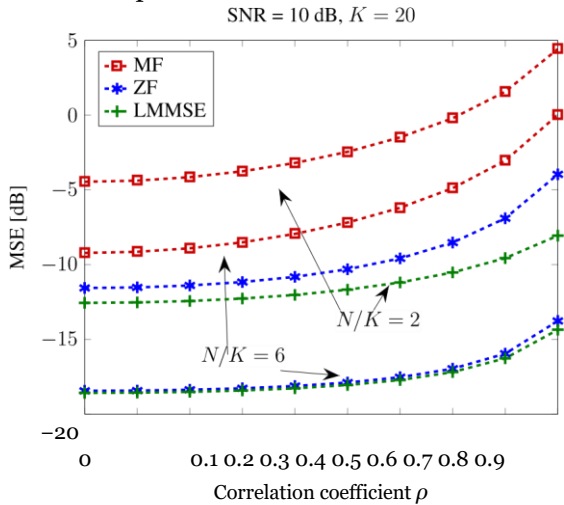


Fig. 6. Asymptotic MSE as a function of the correlation at the BS.

LMMSE and ZF receivers rapidly converge as N/K increases, which makes sense since the inversion of the channel is better conditioned. Moreover, note that increasing the SNR is beneficial for the ZF and LMMSE receivers while it does not help the MF a lot. Indeed, as predicted by Theorem 3, the MF is especially limited by IUI, which does not depend on the SNR.

Fig. 6 plots the asymptotic MSE of the three linear receivers as a function of the correlation coefficient ρ for an SNR of 10 dB and two ratios N/K . Note that ρ models the degree of correlation between the BS antennas. One can see that the ZF and LMMSE receivers are relatively robust to the correlation between BS antennas except for high values of ρ , especially for the ZF receiver with a low ratio N/K . On the other hand, the MF appears to be more sensitive to the BS inter-antenna correlation. Again, this makes sense since, following Theorem 3, the MF is especially limited by IUI, which becomes stronger as the channels of the different users get more correlated.

V. CONCLUSIONS

We have investigated the performance of FBMC-OQAM systems in an uplink massive MIMO scenario. The performance of three types of linear receivers was analyzed as the number of BS antennas and the number

of users grow large. For each receiver, the MSE was shown to converge to a deterministic quantity that does not depend on the subcarrier index. Further, it was shown that a good synchronization of the user is of crucial importance. Finally, the “self-equalization” effect of FBMC-OQAM was verified by theory and simulations. This shows that, under a good synchronization, having a large ratio N/K allows for a low value of the MSE, even in the case of highly selective channels and classical low complexity receivers.

VI. APPENDIX

A. Definition of pulse-related quantity η

Let us define $q[n]$ as

$$q[n] = g^{(1)} \left(\left(n - \frac{2M\kappa - 1}{2} \right) \frac{1}{2M} \right),$$

for $n = 0, \dots, 2M\kappa - 1$ and where $g^{(1)}(t) = \frac{dg(t)}{dt}$. Given $\kappa \inf \frac{N}{K} > 1$ the two pulses $p[n]$ and $q[n]$, let \mathbf{P} and \mathbf{Q} denote two $2M \times$

matrices obtained by arranging the samples of the respective pulses in columns from left to right. We will define,

$$\mathbf{R}(p, q) = \mathbf{P} \sim \mathbf{J}_{2M} \mathbf{Q}$$

$$\mathbf{S}(p, q) = (\mathbf{J}_2 \otimes \mathbf{I}_M) \mathbf{P} \sim \mathbf{J}_{2M} \mathbf{Q},$$

where \sim denotes row-wise convolution, \otimes denotes Kronecker product, \mathbf{I}_M (resp. \mathbf{J}_M) are the identity (resp. exchange) matrices of order M . We define η as

$$\eta = \frac{P_s M}{2} \text{tr} \left[\mathbf{U}^+ \mathcal{R}(p, q) \mathcal{R}^T(p, q) + \mathbf{U}^- \mathcal{S}(p, q) \mathcal{S}^T(p, q) \right],$$

where $\mathbf{U}^\pm = \mathbf{I}_2 \otimes (\mathbf{I}_M \pm \mathbf{J}_M)$.

B. Proof of Theorem 1

For the sake of the compactness of the expressions, we define $\mathbf{F}_m = \mathbf{D}_{1/2, m}$ and $\mathbf{F}_{(1)m} = \mathbf{\Phi} \mathbf{D}_{1/2, m}$. Let us first review well known results related to the resolvent $\mathbf{Q}(\lambda)$ defined as

$$\mathbf{Q}(\lambda) = \left(\frac{\mathbf{H}_m^H \mathbf{H}_m}{K} + \lambda \mathbf{I}_K \right)^{-1} \\ = \mathbf{F}_m \frac{\mathbf{H}_m^H \mathbf{C}_{BS} \mathbf{G}}{K} \mathbf{F}_m + \lambda \mathbf{I}_K \quad !^{-1}$$

Theorem 3. Let \mathbf{G} consist of i.i.d. standardized complex Gaussian entries, and assume that both \mathbf{C}_{BS} and $\mathbf{D}_\beta = \mathbf{F}_m^H \mathbf{F}_m$ be matrices of appropriate dimensions with uniformly bounded spectral norm. Assume that $K, N \rightarrow \infty$ with

$$0 < \liminf \frac{N}{K} \leq \limsup \frac{N}{K} < \infty.$$

Then, for each sequence of deterministic matrices \mathbf{A} of size $K \times K$ with uniformly bounded spectral norm and

assuming $\lambda > 0$, we have $\frac{1}{K} \text{tr}[\mathbf{A}(\mathbf{Q}(\lambda) - \mathbf{T}(\lambda))] \xrightarrow{a.s.} 0$, (19)

where

$$\mathbf{T}(\lambda) = \left(\lambda \mathbf{I}_K + \tilde{\delta}(\lambda) \mathbf{D}_\beta \right)^{-1}$$

and where $(\tilde{\delta}(\lambda), \delta(\lambda))$ are the only positive pair of solutions to the system of equations

$$\delta(\lambda) = \frac{1}{K} \text{tr} \left[\mathbf{D}_\beta \left(\lambda \mathbf{I}_K + \tilde{\delta}(\lambda) \mathbf{D}_\beta \right)^{-1} \right]$$

$$\tilde{\delta}(\lambda) = \frac{1}{K} \text{tr} \left[\mathbf{C}_{\text{BS}} (\mathbf{I}_N + \delta(\lambda) \mathbf{C}_{\text{BS}})^{-1} \right].$$

Furthermore, the above convergence can be extended to the case $\lambda = 0$ assuming that $\inf \frac{N}{K} > 1$.

The result is well known in the random matrix theory literature, see e.g. [20] [21] for some early studies of this random matrix model. We here follow the formulation established in [22], [23]. In particular, it follows from [23] and [24] that if the eigenvalues of \mathbf{D}_β and \mathbf{C}_{BS} are uniformly contained in two compact intervals of \mathbb{R}^+ , the eigenvalues of $\mathbf{F}_m^H \mathbf{G}_{\text{BS}, K} \mathbf{G}_m \mathbf{F}_m$ are contained on a compact interval $S \subset \mathbb{R}^+$ for all large K, N . Furthermore, it was shown in [23] that $0 \in S$ if

. This means that, if $\lambda \in \mathbb{C} \setminus (\mathbb{R}^- \cup \{0\})$ we have $\text{dist}(-\lambda, S) > 0$ and that this property also holds for $\lambda = 0$ when $\inf \frac{N}{K} > 1$.

Convergence of (19) has been usually proven for $\lambda \in \mathbb{C} \setminus (\mathbb{R}^- \cup \{0\})$, but it can trivially be extended to all the region of interest according to the following reasoning. Observe that the spectral norm of $\mathbf{Q}(\lambda)$ and $\mathbf{T}(\lambda)$ is upperbounded by $(\text{dist}(-\lambda, S))^{-1}$, because they are Stieljes transforms of positive matrix valued measures (see further [25]). Furthermore these matrices are holomorphic functions on all the region of interest. Therefore, by a standard application of Montel's theorem, one can establish almost convergence in all the region of interest.

Next, we present a result that can trivially be established using the same approach as in [22] (details are therefore omitted).

Proposition 1. *Let $\text{dist}(-\lambda, S) > 0$. Under the above set of assumptions, and assuming that \mathbf{A}, \mathbf{B} and \mathbf{Y} are sequences of rectangular deterministic matrices of appropriate dimensions with uniformly bounded spectral norm, and define*

$$\widehat{\Psi}_0(\lambda) = \frac{1}{K} \text{tr}[\mathbf{A}\mathbf{Q}(\lambda)]$$

$$"$$

$$\mathbf{1} \quad \mathbf{G}\mathbf{Y}\mathbf{G} \quad \Psi_{b_1}(\lambda) = \frac{1}{K} \text{tr}[\mathbf{A}\mathbf{Q}(\lambda)\mathbf{B}]$$

It holds that

$$\mathbb{E} \Psi_0(\lambda) - \Psi_0(\lambda) = \frac{1}{N} \mathcal{O} \left(\frac{1}{N \text{dist}(-\lambda, S)} \right) \quad (20)$$

$$\mathbb{E} \Psi_1(\lambda) - \Psi_1(\lambda) = \frac{1}{N} \mathcal{O} \left(\frac{1}{N \text{dist}(-\lambda, S)} \right) \quad (21)$$

where we have defined the deterministic quantities

$$\Psi_0(\lambda) = \frac{1}{K} \text{tr}[\mathbf{A}\mathbf{T}(\lambda)] \quad \Psi_1(\lambda) = \frac{1}{K} \text{tr}[\mathbf{A}\mathbf{T}(\lambda)\mathbf{B}] \frac{1}{K} \text{tr}[\mathbf{Y}] - \frac{1}{K} \text{tr}[\mathbf{A}\mathbf{T}(\lambda)\mathbf{F}_m^H] \frac{1}{K} \text{tr}[\mathbf{T}(\lambda)\mathbf{B}\mathbf{F}_m] + \frac{1}{K} \text{tr}[(\mathbf{I}_N + \delta(\lambda)\mathbf{C}_{\text{BS}})^{-1} \mathbf{C}_{\text{BS}}\mathbf{Y}].$$

Furthermore, we can establish that

$$\text{var} \Psi_0(\lambda) = \frac{1}{N^2} \mathcal{O} \left(\frac{1}{N \text{dist}(-\lambda, S)} \right)$$

$$\text{var} \Psi_1(\lambda) = \frac{1}{N^2} \mathcal{O} \left(\frac{1}{N \text{dist}(-\lambda, S)} \right).$$

The result in (21) accepts an interesting particularization in the case where $\mathbf{B} = \mathbf{F}_m^H$, for which one trivially obtains

$$\Psi_1(\lambda) = \frac{1}{K} \text{tr}[\mathbf{A}\mathbf{T}(\lambda)\mathbf{F}_m^H] \frac{1}{K} \text{tr}[(\mathbf{I}_N + \delta(\lambda)\mathbf{C}_{\text{BS}})^{-1} \mathbf{Y}].$$

We now have all the ingredients to prove Theorems 1 and 2, which we summarize in the following proposition.

Proposition 2. *Let $\text{dist}(-\lambda, S) > 0$. Consider two sequences of $N \times N$ deterministic matrices $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2$ with uniformly bounded spectral norm, and define*

Under the assumptions of Theorem 1 and 2, we can write

$$\left| \mathbb{E} \widehat{\Psi}_2(\lambda) - \Psi_2(\lambda) \right| = \frac{1}{N} \mathcal{O} \left(\frac{1}{\text{dist}(-\lambda, S)} \right) \quad (22)$$

where

$$\Psi_2(\lambda) = \tilde{\delta}^2(\lambda) \frac{1}{K} \text{tr} \left[\mathbf{T}(\lambda) \mathbf{F}_m^H \mathbf{F}_m^{(1)} \left(\mathbf{F}_m^{(1)} \right)^H \mathbf{F}_m \right] + \delta(\lambda) \frac{1}{K} \text{tr} \left[\mathbf{C}_{\text{BS}}^2 (\mathbf{I}_N + \delta(\lambda) \mathbf{C}_{\text{BS}})^{-1} \right] \frac{1}{K} \text{tr} \left[\mathbf{F}_m^{(1)} (\mathbf{F}_m^{(1)})^H \right].$$

Furthermore,

$$\text{var}(\widehat{\Psi}_2(\lambda)) = \frac{1}{N^2} \mathcal{O} \left(\frac{1}{\text{dist}(-\lambda, \mathcal{S})} \right)$$

Before presenting the proof of this result, let us reason why the above two propositions directly prove almost sure convergence of the MSE quantities in Theorems 1 and 2. To see this, observe that, using (7), all the terms in the expression of MSE_m^λ can be expressed in the form of $\widehat{\Psi}_0(\lambda)$, $\widehat{\Psi}_1(\lambda)$ and

$\widehat{\Psi}_2(\lambda)$ together with their derivatives with respect to λ ,³ for different choices of the deterministic matrices \mathbf{A} , \mathbf{B} , \mathbf{Y} , $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$ (all of which have bounded spectral norm). Hence, almost sure convergence of all these quantities and their derivatives will ultimately establish the result. Now, from Propositions 1 and 2, we obtain that

$$\mathbb{E} \left| \widehat{\Psi}_i(\lambda) - \Psi_i(\lambda) \right|^2 = \left| \mathbb{E} \widehat{\Psi}_i(\lambda) - \Psi_i(\lambda) \right|^2 + \text{var} \left(\widehat{\Psi}_i(\lambda) \right) = \mathcal{O} \left(\frac{1}{N^2} \right)$$

because $\text{dist}(-\lambda, \mathcal{S}) > 0$. A direct application of the Markov inequality and the Borel Cantelli lemma shows that $\widehat{\Psi}_i(\lambda) - \Psi_i(\lambda) \rightarrow 0$ almost surely to zero for every fixed $\lambda \in \mathbb{C} \setminus (\mathbb{R} \cup \{0\})$ (or $\lambda = 0$ when $\inf \frac{N}{K} > 1$). On the other hand, it can readily be seen that all the terms in the MSE expression are analytic and almost surely bounded for all $\lambda \in \mathbb{C} \setminus (\mathbb{R} \cup \{0\})$ (or $\lambda = 0$ when $\inf \frac{N}{K} > 1$). Hence, Montel's theorem shows that convergence holds uniformly in compact sets in the same region, which implies convergence of the corresponding complex derivatives.

We devote the rest of the appendix to the proof of Proposition 2. We first establish the bound of the variance by using the Nash-Poincare inequality [26], which establishes that (if g_{ij} denotes the (i,j) -th entry of \mathbf{G})

$$\begin{aligned} \widehat{\Psi}_2(\lambda) &= \frac{1}{K} \text{tr} \left[\mathbf{Q}(\lambda) \mathbf{F}_m^H \frac{\mathbf{G} \mathbf{\Gamma}_1 \mathbf{G}^H}{K} \mathbf{F}_m^{(1)} (\mathbf{F}_m^{(1)})^H \frac{\mathbf{G} \mathbf{\Gamma}_2 \mathbf{G}^H}{K} \mathbf{F}_m \right] \sum_{i=1}^{LK} \sum_{j=1}^N \mathbb{E} \left[\left| \frac{\partial \widehat{\Psi}_2(\lambda)}{\partial g_{ij}^*} \right|^2 \right] \\ &\leq \sum_{j=1}^{LK} \sum_{i=1}^N \mathbb{E} \left[\left| \frac{\partial \widehat{\Psi}(\mathbf{G}, \mathbf{G}^*)}{\partial g_{ij}^*} \right|^2 \right] + \mathbb{E} \left[\left| \widehat{\Psi}(\mathbf{G}, \mathbf{G}^*) \right|^2 \right] \\ &= \text{h}\Psi(\mathbf{b}, \mathbf{G}, \mathbf{G}^*) \end{aligned}$$

for any continuously differentiable $\widehat{\Psi}(\mathbf{G}, \mathbf{G}^*)$ such that both itself and its derivatives are polynomially bounded functions of \mathbf{G} . One can readily see that

$$\begin{aligned} &= \frac{1}{K^2} \mathbb{E} \text{tr} \left[\mathbf{Q} \frac{\mathbf{G} \mathbf{\Gamma}_1 \mathbf{G}^H}{K} \mathcal{F} \frac{\mathbf{G} \mathbf{\Gamma}_2 \mathbf{G}^H}{K} \mathbf{Q} \frac{\mathbf{G} \mathbf{C}_{\text{BS}}^2 \mathbf{G}^H}{K} \mathbf{Q}^H \right] \\ &+ \frac{1}{K^2} \mathbb{E} \text{tr} \left[\mathcal{F} \frac{\mathbf{G} \mathbf{\Gamma}_2 \mathbf{G}^H}{K} \mathbf{Q} \frac{\mathbf{G} \mathbf{\Gamma}_1 \mathbf{\Gamma}_1^H \mathbf{G}^H}{K} \mathbf{Q}^H \frac{\mathbf{G} \mathbf{\Gamma}_2^H \mathbf{G}^H}{K} \mathcal{F} \right] \\ &+ \frac{1}{K^2} \mathbb{E} \text{tr} \left[\mathbf{Q} \frac{\mathbf{G} \mathbf{\Gamma}_1 \mathbf{G}^H}{K} \mathcal{F} \frac{\mathbf{G} \mathbf{\Gamma}_2 \mathbf{\Gamma}_2^H \mathbf{G}^H}{K} \mathcal{F} \frac{\mathbf{G} \mathbf{\Gamma}_1^H \mathbf{G}^H}{K} \mathbf{Q}^H \right] \end{aligned}$$

where $\mathbf{Q} = \mathbf{F}_m \mathbf{Q}(\lambda) \mathbf{F}_m^H$, $\mathcal{F} = \mathbf{F}_m^{(1)} (\mathbf{F}_m^{(1)})^H$ and where we have used the fact that

$$\frac{\partial \mathbf{Q}}{\partial g_{ij}^*} = -\mathbf{Q}(\lambda) \mathbf{F}_m^H \frac{\mathbf{G} \mathbf{C}_{\text{BS}} \mathbf{e}_j \mathbf{e}_i^T}{K}$$

with \mathbf{e}_j denoting the j -th column of the identity matrix. Now, using the fact that $\|\mathbf{Q}(\lambda)\| \leq \frac{1}{\text{dist}(-\lambda, \mathcal{S})}$, the uniform boundedness of the spectral norm of \mathbf{F} , \mathbf{C}_{BS} , \mathbf{F}_m , $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$ and the result in [27, Lemma 2], we readily obtain that the above term is bounded by a quantity of the type $N^{-2} \mathcal{O} \left(\frac{1}{\text{dist}(-\lambda, \mathcal{S})} \right)$.

The same reasoning can be applied to the second term on the right hand side of (23), leading to the result (details are omitted).

³ Note that the derivative with respect to λ of \mathbf{Q} .

Next, we prove the identity in (22). We will be using the integration by parts formula [26], which establishes that for any continuously differentiable $\widehat{\Psi}(\mathbf{G}, \mathbf{G}^*)$ such that both itself and its derivatives are polynomially bounded functions of \mathbf{G} , we have

$$\mathbb{E} \left[g_{ij} \widehat{\Psi}(\mathbf{G}, \mathbf{G}^*) \right] = \mathbb{E} \left[\frac{\partial \widehat{\Psi}(\mathbf{G}, \mathbf{G}^*)}{\partial g_{ij}^*} \right].$$

More specifically, we consider the decomposition of \mathbf{G} as

$$\mathbf{G} = \sum_{j=1}^{LK} \mathbf{X} \mathbf{X}^H g_{ij} \mathbf{e}_i \mathbf{e}_j^T.$$

Using this decomposition in the definition of $\Psi_{b_2}(\lambda)$, we directly obtain

$$\mathbb{E} \widehat{\Psi}_2(\lambda) = \sum_{j=1}^{LK} \sum_{i=1}^N \frac{1}{K} \mathbb{E} \frac{\partial}{\partial g_{ij}^*} \left[\mathbb{Q} \frac{\mathbf{e}_i \mathbf{e}_j^T \mathbf{\Gamma}_1 \mathbf{G}^H}{K} \mathcal{F} \frac{\mathbf{G} \mathbf{\Gamma}_2 \mathbf{G}^H}{K} \right]$$

and, after some algebra,

$$\begin{aligned} \mathbb{E} \widehat{\Psi}_2(\lambda) &= -\mathbb{E} \left[\frac{1}{K} \text{tr} \left[\mathbb{Q} \frac{\mathbf{G} \mathbf{\Gamma}_1 \mathbf{G}^H}{K} \mathcal{F} \frac{\mathbf{G} \mathbf{\Gamma}_2 \mathbf{G}^H}{K} \right] \frac{1}{K} \text{tr} \mathbb{Q} \right] \\ &+ \mathbb{E} \left[\frac{1}{K} \text{tr} \left[\mathbb{Q} \mathcal{F} \frac{\mathbf{G} \mathbf{\Gamma}_2 \mathbf{G}^H}{K} \right] \frac{1}{K} \text{tr} [\mathbf{\Gamma}_1] \right] \\ &+ \mathbb{E} \left[\frac{1}{K} \text{tr} \left[\mathbf{\Gamma}_1 \frac{\mathbf{G}^H \mathcal{F} \mathbf{G}}{K} \mathbf{\Gamma}_2 \right] \frac{1}{K} \text{tr} \mathbb{Q} \right]. \end{aligned}$$

Observe that the first term on the right hand side of the above equation is very similar to $\widehat{\Psi}_2(\lambda)$. Next, we introduce the following error definition in the first and last terms on the right hand side of the above equation, namely

$$\epsilon_0(\lambda) = \frac{1}{K} \text{tr} \left[\mathbf{F}_m^H \mathbf{F}_m (\mathbf{Q} - \mathbb{E} \mathbf{Q}) \right].$$

After some algebra, and replacing $\mathbf{\Gamma}_1$ in the above expression of $\mathbb{E} \widehat{\Psi}_2(\lambda)$ we obtain

$$\mathbb{E} \frac{1}{K} \text{tr} \left[\mathbb{Q} \frac{\mathbf{G} \mathbf{\Gamma}_1 \mathbf{G}^H}{K} \mathcal{F} \frac{\mathbf{G} \mathbf{\Gamma}_2 \mathbf{G}^H}{K} \right] = \quad (24)$$

$$+ \frac{1}{K} \text{tr} \left[\left(\mathbf{I}_N + \frac{1}{K} \mathbb{E} \text{tr} [\mathbb{Q}] \right.$$

$$\left. \right)^{-1} \mathbf{\Gamma}_1 \mathbf{\Gamma}_2 \left] \frac{\text{tr} [\mathcal{F}]}{K} \frac{\mathbb{E} \text{tr} [\mathbb{Q}]}{K}$$

$$\frac{1}{K} \mathbb{E} \text{tr} \left[\mathbb{Q} \mathcal{F} \frac{\mathbf{G} \mathbf{\Gamma}_2 \mathbf{G}^H}{K} \right] \frac{1}{K} \text{tr} \left[\left(\mathbf{I}_N + \frac{1}{K} \mathbb{E} \text{tr} [\mathbb{Q}] \right)^{-1} \mathbf{\Gamma}_1 \right]$$

+ $\chi_1 + \chi_2$

where

$$\chi_1 = \mathbb{E} [t_1(\lambda) \epsilon_0(\lambda)]$$

$$\chi_2 = -\mathbb{E} [t_2(\lambda) \epsilon_0(\lambda)]$$

$$t_1(\lambda) = \text{tr} \left[\frac{1}{K} \left(\mathbf{I}_N + \frac{1}{K} \mathbb{E} \text{tr} [\mathbb{Q}] \right)^{-1} \mathbf{\Gamma}_1 \frac{\mathbf{G}^H \mathcal{F} \mathbf{G}}{K} \mathbf{\Gamma}_2 \right]$$

$$t_2(\lambda) = \frac{1}{K} \text{tr} \left[\mathbb{Q} \frac{\mathbf{G} \mathbf{C}_{BS} \mathbf{I}_{N+K_1}}{\text{Etr}[\mathbb{Q}] \mathbf{C}_{BS}} \right)^{-1} \mathbf{\Gamma}_1 \mathbf{G}^H \mathcal{F} \frac{\mathbf{G} \mathbf{\Gamma}_2 \mathbf{G}^H}{K} \right].$$

Now, a direct application of the Cauchy-Schwarz inequality shows that

$$|\chi_1| \leq \sqrt{\text{var}(t_1(\lambda)) \text{var}(\epsilon_0(\lambda))}.$$

By Proposition 1, we see that $|\chi_1| = N^{-1} O \left(\frac{1}{N \text{dist}(-\lambda, S)} \right)$

A similar reasoning shows that $|\chi_2| = N^{-1} O \left(\frac{1}{N \text{dist}(-\lambda, S)} \right)$

that. Using this in (24), particularizing the resulting expression for

$\mathbf{\Gamma}_1 = \mathbf{\Gamma}_2 = \mathbf{C}_{BS}$ and observe that, by Proposition 1,

$$\begin{aligned} \frac{1}{K} \mathbb{E} \text{tr} [\mathbb{Q}] - \delta &= N^{-1} O \left(\frac{1}{N \text{dist}(-\lambda, S)} \right), \\ \frac{1}{K} \mathbb{E} \text{tr} \left[\mathbb{Q} \mathcal{F} \frac{\mathbf{G} \mathbf{C}_{BS} \mathbf{G}^H}{K} \right] &- \tilde{\delta} \frac{1}{K} \text{tr} \mathbf{h}^T(\lambda) \mathbf{F}_m^H \mathbf{F}_m \\ &= N^{-1} O \left(\frac{1}{N \text{dist}(-\lambda, S)} \right), \end{aligned}$$

-1

we obtain the result of the proposition.

C. Proof of Theorem 2

Re-writing MSE^{MF_m} in (18) as a function of the matrix \mathbf{G} using (7), as we did for the ZF and LMMSE receivers, one will observe terms that have the following general form,

$$\frac{1}{K} \mathbb{E} \text{tr} \left[\hat{\mathbf{Y}}^{-1} \Psi \left(\mathbf{G}, \mathbf{G}^H \right) \right],$$

where $\hat{\mathbf{Y}}$ is a diagonal matrix defined as

$$\hat{\mathbf{Y}} = \text{diag} \left(\mathbf{F}_m^H \mathbf{F}_m \right) = \text{diag} \left(\mathbf{F}_m^H \mathbf{G} \mathbf{C}_{\text{BS}} \mathbf{G} \right)$$

The proof of Theorem 2 consists in showing that $\hat{\mathbf{Y}}$ can be replaced by its asymptotic equivalent \mathbf{Y} . Then, one only needs to compute the expectation of $\mathbb{E} \Psi \left(\mathbf{G}, \mathbf{G}^H \right)$, which can be easily computed. Here below, in Proposition 3, we only give the first part of the proof. The other one is omitted due to space constraints.

Proposition 3. Assume that $\Psi \left(\mathbf{G}, \mathbf{G}^H \right)$ is a function of \mathbf{G} such that

$$\sup_{K, N} \frac{1}{K} \mathbb{E} \text{tr} \left[\Psi \left(\mathbf{G}, \mathbf{G}^H \right) \Psi \left(\mathbf{G}, \mathbf{G}^H \right)^H \right] < \infty$$

and let

$$\mathbf{Y} = \mathbb{E} \left(\hat{\mathbf{Y}} \right) = \text{diag} \left(\mathbf{F}_m^H \mathbf{F}_m \right) \frac{1}{K} \text{tr} \left[\mathbf{C}_{\text{BS}} \right].$$

Assume that the eigenvalues of $\mathbf{F}_m^H \mathbf{F}_m$ are located on a compact set of \mathbb{R}^+ independent of K, N and that $\inf_K \text{tr} \left[\mathbf{C}_{\text{BS}} \right] >$

o. Then, as $K, N \rightarrow \infty$ at the same rate,

$$\frac{1}{K} \text{tr} \left[\left(\mathbf{Y}^{-1} - \hat{\mathbf{Y}}^{-1} \right) \Psi \left(\mathbf{G}, \mathbf{G}^H \right) \right] \rightarrow 0$$

almost surely.

Indeed, observe that we can write, by the Cauchy-Schwarz

inequality,

$$\begin{aligned} & \left| \frac{1}{K} \text{tr} \left[\left(\mathbf{Y}^{-1} - \hat{\mathbf{Y}}^{-1} \right) \Psi \left(\mathbf{G}, \mathbf{G}^H \right) \right] \right| \\ &= \left| \frac{1}{K} \text{tr} \left[\left(\hat{\mathbf{Y}} - \mathbf{Y} \right) \hat{\mathbf{Y}}^{-1} \Psi \left(\mathbf{G}, \mathbf{G}^H \right) \mathbf{Y}^{-1} \right] \right| \\ &\leq \sqrt{\frac{1}{K} \text{tr} \left[\mathbf{Y}^{-2} \left(\hat{\mathbf{Y}} - \mathbf{Y} \right) \hat{\mathbf{Y}}^{-2} \left(\hat{\mathbf{Y}} - \mathbf{Y} \right) \right]} \\ &\quad \sqrt{\frac{1}{K} \text{tr} \left[\Psi \left(\mathbf{G}, \mathbf{G}^H \right) \Psi \left(\mathbf{G}, \mathbf{G}^H \right)^H \right]} \\ &\leq \alpha \left\| \hat{\mathbf{Y}}^{-1} \right\| \left\| \mathbf{Y}^{-1} \right\| \left\| \hat{\mathbf{Y}} - \mathbf{Y} \right\| \end{aligned}$$

for some positive constant α . Observe that $\left\| \mathbf{Y}^{-1} \right\| < \infty$

by assumption, whereas $\left\| \hat{\mathbf{Y}}^{-1} \right\|$ will also be bounded for sufficiently large K, N because of the norm convergence.

Hence, we only we now prove that $\left\| \hat{\mathbf{Y}} - \mathbf{Y} \right\| \rightarrow 0$ almost surely to prove the theorem.

In order to see $\left\| \hat{\mathbf{Y}} - \mathbf{Y} \right\| \rightarrow 0$, let \hat{d}_{ii} and d_{ii} denote the i th diagonal entry of $\hat{\mathbf{Y}}$ and \mathbf{Y} respectively. Then, for an arbitrary $\epsilon > 0$, we have

$$\begin{aligned} P \left[\left\| \hat{\mathbf{Y}} - \mathbf{Y} \right\| > \epsilon \right] &= P \left[\max_i \left| \hat{d}_{ii} - d_{ii} \right| > \epsilon \right] \\ &\leq \sum_{i=1}^K P \left[\left| \hat{d}_{ii} - d_{ii} \right| > \epsilon \right]. \end{aligned}$$

However, the i -th entry of $\hat{\mathbf{Y}}$ can be expressed as

$$\hat{d}_{ii} = \mathbf{x}_i^H \frac{\mathbf{C}_{\text{BS}}}{K} \mathbf{x}_i$$

where \mathbf{x}_i is a circular Gaussian vector with zero mean and variance $\mathbb{E} \left[\mathbf{x}_i \mathbf{x}_i^H \right] = \mathbb{E} \left[\mathbf{G}^H \mathbf{f}_i \mathbf{f}_i^H \mathbf{G} \right] = \left\| \mathbf{f}_i \right\|^2 \mathbf{I}$ with \mathbf{f}_i denoting the i -th column of \mathbf{F}_m . Using, e.g. [28, Lemma 2.7], for $p \geq 2$

$$\begin{aligned} \mathbb{E} \left[\left| \hat{d}_{ii} - d_{ii} \right|^p \right] &\leq \frac{\alpha_1}{K^{p/2}} \left(\frac{1}{K} \text{tr} \left[\left(\left\| \mathbf{f}_i \right\|^2 \right) \right] \right)^{p/2} \\ &\quad + \frac{\alpha_2}{K^p} \text{tr} \left[\left(\left\| \mathbf{f}_i \right\|^2 \right) \right]^p. \end{aligned}$$

Choosing, e.g. $p = 5$ we see that

$$\begin{aligned} P \left[\left\| \hat{\mathbf{Y}} - \mathbf{Y} \right\| > \epsilon \right] &\leq O \left(K^{-3/2} \right) \mathbb{E} \left[\left| \hat{d}_{ii} - d_{ii} \right|^5 \right] \\ &= O \text{ and therefore,} \end{aligned}$$

using Markov inequality,

P. Applying the Borel-Cantelli lemma we obtain the result.

REFERENCES

- [1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.
- [2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.
- [3] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [4] A. Farhang, N. Marchetti, L. E. Doyle, and B. Farhang-Boroujeny, "Filter Bank Multicarrier for Massive MIMO," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Sept 2014, pp. 1–7.

- [5] B. Farhang-Boroujeny, "OFDM versus filter bank multicarrier," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 92–112, May 2011.
- [6] A. Aminjavaheri, A. Farhang, A. RezazadehReyhani, and B. FarhangBoroujeny, "Impact of timing and frequency offsets on multicarrier waveform candidates for 5G," in *2015 IEEE Signal Processing and Signal Processing Education Workshop (SP/SPE)*, Aug 2015, pp. 178–183.
- [7] M. Bellanger, D. Le Ruyet, D. Roviras, M. Terre, J. Nossek, L. Baltar, Q. Bai, D. Waldhauser, M. Renfors, T. Ihalainen *et al.*, "FBMC physical layer: a primer," *PHYDYAS*, January, 2010.
- [8] M. Bellanger, "Efficiency of Filter Bank Multicarrier Techniques in Burst Radio Transmission," in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, Dec 2010, pp. 1–4.
- [9] J. Yli-Kaakinen and M. Renfors, "Optimized burst truncation in fastconvolution filter bank based waveform generation," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015, pp. 71–75.
- [10] M. J. Abdoli, M. Jia, and J. Ma, "Weighted circularly convolved filtering in OFDM/OQAM," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 657–661.
- [11] F. Rottenberg, X. Mestre, F. Horlin, and J. Louveaux, "Single-Tap Precoders and Decoders for Multiuser MIMO FBMC-OQAM Under Strong Channel Frequency Selectivity," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 587–600, Feb 2017.
- [12] A. Aminjavaheri, A. Farhang, N. Marchetti, L. E. Doyle, and B. FarhangBoroujeny, "Frequency spreading equalization in multicarrier massive MIMO," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 1292–1297.
- [13] M. Bellanger, "FS-FBMC: An alternative scheme for filter bank based multicarrier transmission," in *2012 5th International Symposium on Communications, Control and Signal Processing*, May 2012, pp. 1–4.
- [14] D. Mattered, M. Tanda, and M. Bellanger, "Frequency-spreading implementation of OFDM/OQAM systems," in *2012 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2012, pp. 176–180.
- [15] A. Aminjavaheri, A. Farhang, L. E. Doyle, and B. FarhangBoroujeny, "Prototype filter design for FBMC in massive MIMO channels," *CoRR*, vol. abs/1610.06273, 2016. [Online]. Available: <http://arxiv.org/abs/1610.06273>
- [16] X. Mestre and D. Gregoratti, "Parallelized structures for MIMO FBMC under strong channel frequency selectivity," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1200–1215, Mar. 2016.
- [17] F. Rottenberg, X. Mestre, and J. Louveaux, "Optimal zero forcing precoder and decoder design for multi-user MIMO FBMC under strong channel selectivity," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3541–3545.
- [18] M. Bellanger, "Specification and design of a prototype filter for filter bank based multicarrier transmission," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, 2001, pp. 2417–2420.
- [19] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sept. 2001.
- [20] V. Girko, *Theory of Random Determinants*. The Netherlands: Kluwer Academic Publishers, 1990.
- [21] A. Boutet de Monvel, A. Khorunzhiy, and V. Vasilchuk, "Limiting eigenvalue distribution of random matrices with correlated entries," *Markov Processes and Related Fields*, vol. 2, no. 2, pp. 607–636, 1996.
- [22] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.
- [23] R. Couillet and W. Hachem, "Analysis of the limiting spectral measure of large random matrices of the separable covariance type," *Random Matrices: Theory and Applications*, vol. 3, no. 4, 2014.
- [24] D. Paul and J. Silverstein, "No eigenvalues outside the support of limiting empirical spectral distribution of a separable covariance matrix," *Journal of Multivariate Analysis*, vol. 100, no. 1, pp. 37–57, 2009.
- [25] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *The Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007.
- [26] L. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*, ser. Mathematical Surveys and Monographs. American Mathematical Society, 2011, vol. 171.
- [27] F. Rubio, X. Mestre, and W. Hachem, "A CLT on the SNR of diagonally loaded MVDR filters," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4178–4195, Aug. 2012.
- [28] Z. Bai and J. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Annals of probability*, vol. 26, pp. 316–345, 1998.