Massive MIMO: How many antennas do we need?

Jakob Hoydis∗†, Stephan ten Brink‡, and Mérouane Debbah†

∗Department of Telecommunications, Supélec, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France
†Alcatel-Lucent Chair on Flexible Radio, Supélec, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France
‡Bell Labs, Alcatel-Lucent, Lorenzstr. 10, 70435 Stuttgart, Germany
{jakob.hoydis, merouane.debbah}@supelec.fr, stephan.tenbrink@alcatel-lucent.com

Abstract

We consider a multicell MIMO uplink channel where each base station (BS) is equipped with a large number of antennas $N$. Recent work has shown that, as $N \rightarrow \infty$, (i) the simplest form of user detection, i.e., the matched filter (MF), becomes optimal, (ii) the transmit power per user terminal (UT) can be made arbitrarily small, (iii) the system performance is limited by pilot contamination. The aim of this paper is to assess to which extent the above conclusions hold true for large, but finite $N$. In particular, we derive how many antennas per UT are needed to achieve $\eta \%$ of the ultimate performance. We then study how much can be gained through more sophisticated minimum-mean-square-error (MMSE) detection and how many more antennas are needed with the MF to achieve the same performance. Our analysis relies on novel results from random matrix theory which allow us to derive tight approximations of the signal-to-interference-plus-noise-ratio (SINR) of a class of linear receivers.

I. INTRODUCTION

As wireless networks are inherently limited by their own interference, a lot of research focuses on interference reduction techniques, such as multiuser MIMO [1], interference alignment [2] or multi-cell processing [3]. Although these techniques can lead to considerable performance gains, it is unlikely that they will be able to carry the exponentially growing wireless data traffic. Due to this reason, a significant network densification, i.e., increasing the number of antennas per unit area, seems inevitable. One way of densifying the network consists in cell-size shrinking, such as the deployment of femto or small cells [4], which comes at the cost of additional equipment and increased interference. Another, seemingly simpler, but also less explored option is the use of very large antennas arrays at the base stations (BSs). However, it is well known that accurate channel state information must be acquired to reap the benefits of additional antennas. This poses, in particular in fast fading channels, a challenge as the number of antennas grows. Thus, massive MIMO is only feasible in time-division-duplexing (TDD) systems where channel reciprocity can be exploited. Here, the channels are estimated based on orthogonal pilot tones which are sent by the user terminals (UTs). In such systems, additional antennas do not increase the training overhead and, therefore, “always help” [5]. Nevertheless, for a given coherence time, the number of orthogonal pilot sequences is limited and they must be reused in adjacent cells. This leads to channel estimation impairments known as “pilot contamination” [6].

The use of massively many antennas was advocated for the first time in [7] and has since then received growing research interest [8], [9], [10]. It is particularly intriguing that with an infinite number of antennas, the simplest form of user detection/beamforming, i.e., maximum-ratio-combining or matched filtering (MF), becomes optimal, the transmit power can be made arbitrarily small, and the performance is ultimately limited by pilot contamination. Therefore, several works address the problem of how the negative impact of pilot contamination can be reduced [8], [6].

But what is the difference between massive MIMO and classical MIMO techniques whose benefits are well-known since more than a decade [11]? This paper tries to give an answer to this question. We first provide a definition of massive MIMO as an operating condition of cellular systems where multi-user interference and noise are small compared to pilot contamination. Whether or not this condition is satisfied depends in general on several factors: the number of antennas $N$, the number of UTs
per degree of freedom (DoF) offered by the channel (we denote by DoF the rank of the antenna correlation matrix which might be smaller than $N$), the signal-to-noise-ratio (SNR) and the path loss. We then study if more antennas can compensate for sub-optimal user detection, i.e., how many more antennas does the MF need to achieve minimum-mean-square-error (MMSE) detection performance. The difference between both techniques increases as one approaches the asymptotic performance limit.

The technical contributions of this work are summarized as follows: We consider a very general channel model which allows us to assign a different path loss and antenna correlation matrix to each UT. Based on this model, we derive asymptotically tight, deterministic approximations of the signal-to-interference-plus-noise-ratio (SINR) at the output of a general class of linear receivers, accounting explicitly for pilot contamination. Simulations suggest that our approximations are valid for even small $N$. Although we do not fully exploit the generality of this channel model in the current paper, it enables future studies on massive MIMO under more realistic channel models with the goal of analyzing the impact of antenna correlation, spacing and aperture. Lastly, the techniques developed in this work can be directly applied to the downlink [12].

II. System model

Consider a multi-cellular system consisting of $L$ cells with one BS and $K$ UTs in each cell. Each BSs is equipped with $N$ antennas, the UTs have a single antenna. The focus of this paper is on the uplink without any form of BS-cooperation. The received base-band signal vector $y_j \in \mathbb{C}^N$ at BS $j$ at a given time reads

$$y_j = \sqrt{\rho} \sum_{l=1}^{L} H_{jl} x_l + n$$  \hspace{1cm} (1)

where $
H_{jl} \in \mathbb{C}^{N \times K}$ is the channel matrix from the UTs in cell $l$ to BS $j$, $x_l \sim \mathcal{CN} (0, I_K)$ is the message-bearing transmit vector from the UTs in cell $l$, $n_j \sim \mathcal{CN} (0, I_N)$ is a noise vector and $\rho > 0$ denotes the transmit SNR. We model the $k$th column vector $h_{jlk} \in \mathbb{C}^N$ of the matrix $H_{jl}$ as

$$h_{jlk} = R_{jlk} w_{jlk}$$  \hspace{1cm} (2)

where $R_{jlk} \triangleq R_{jlk} R_{jlk}^H \in \mathbb{C}^{N \times N}$ are deterministic correlation matrices and $w_{jlk} \sim \mathcal{CN} (0, I_N)$ are fast fading channel vectors. The above channel model is very versatile as it allows to assign a different antenna correlation (including path loss) to each channel vector. This is especially important for large antenna arrays with a significant amount of antenna correlation due to either insufficient antenna spacing or a lack of scattering.

Remark 2.1: In particular, (2) can be used to represent a physical channel model with a fixed number of dimensions or angular bins $P$ as in [9], by letting $R_{jlk} = \sqrt{\ell_{jlk}} [A 0_{N \times N-P}]$, where $A \in \mathbb{C}^{N \times P}$ and $\ell_{jlk}$ is the inverse path loss from UT $k$ in cell $l$ to BS $j$. We will use a particular form of this model in Section IV.

A. Channel estimation

During a dedicated training phase (whose length we ignore in this work), the UTs in each cell transmit orthogonal pilot sequences which allow the BSs to compute estimates $\hat{H}_{jj}$ of their local channels $H_{jj}$. The same set of orthogonal pilot
sequences is reused in every cell so that the channel estimate is corrupted by pilot contamination from adjacent cells [13], [7].

Under these assumptions, BS \( j \) estimates the channel vector \( h_{jjk} \) based on the observation

\[
y_{jk}^\tau = h_{jjk} + \sum_{l \neq j} h_{jlk} + \frac{1}{\sqrt{\rho_\tau}} n_{jk}
\]

where \( n_{jk} \sim \mathcal{CN}(0, I_N) \) and \( \rho_\tau > 0 \) is the effective training SNR. Assuming MMSE estimation [14], we can decompose the channel as \( h_{jjk} = \hat{h}_{jjk} + \tilde{h}_{jjk} \), where \( \hat{h}_{jjk} \sim \mathcal{CN}(0, \Phi_{jjk}) \) is the channel estimate and \( \tilde{h}_{jjk} \sim \mathcal{CN}(0, R_{jjk} - \Phi_{jjk}) \) is the independent estimation error, where the matrices \( \Phi_{jlk}, 1 \leq j, l \leq L, 1 \leq k \leq K \), are defined as

\[
\Phi_{jlk} = R_{jjk} Q_{jk} R_{jlk}, \quad Q_{jk} = \left( \frac{1}{\rho_\tau} I_N + \sum_l R_{jlk} \right)^{-1}.
\]

### B. Achievable rate with linear detection

We consider linear single-user detection, where the \( j \)th BS estimates the symbols \( x_{jm}, m = 1, \ldots, K \), of the UTs in its cell by calculating the inner products between the received vector \( y_j \) and the linear filters \( r_{jm} \in \mathbb{C}^N \). Two particular detectors are of practical interest, the matched filter \( r_{jm}^{MF} \) and the MMSE detector \( r_{jm}^{MMSE} \) which we define as

\[
\begin{align*}
    r_{jm}^{MF} &= \hat{h}_{jjm} \\
    r_{jm}^{MMSE} &= \left( \hat{H}_{jj}^H \hat{H}_{jj} + Z_j + N \lambda I_N \right)^{-1} \hat{h}_{jjm}
\end{align*}
\]

where \( \lambda > 0 \) is a design parameter and

\[
Z_j = E \left[ \hat{H}_{jj} \hat{H}_{jj}^H + \sum_{l \neq j} H_{jl} H_{jl}^H \right] = \sum_k (R_{jjk} - \Phi_{jjk}) + \sum_{l \neq j} \sum_k R_{jlk}.
\]

**Remark 2.2:** Note that \( r_{jm}^{MMSE} \) does not correspond to the real MMSE detector. This is because the interference \( \hat{H}_{jj} x_j \) caused by imperfect CSI is not independent of the useful signal \( \hat{H}_{jj} x_j \). However, both terms become asymptotically \( (N \to \infty) \) independent and the performance loss negligible. Moreover, a BS could theoretically estimate all channel matrices \( H_{jl} \) from the observations (3) to further improve the performance. Nevertheless, high path loss to neighboring cells will render these channel estimates unreliable and the potential performance gains are expected to be marginal. Our formulation of \( r_{jm}^{MMSE} \) further allows us to treat \( \lambda \) (and also \( Z_j \)) as a design parameter which could be optimized. A natural choice is \( \lambda = \frac{1}{\rho N} \).

The worst effect the channel estimation error \( \tilde{h}_{jjm} x_{jm} \) can have is to behave as independent noise with covariance matrix \( (R_{jjm} - \Phi_{jjm}) \) [15]. Under this assumption, the received SINR \( \gamma_{jm} \) related to \( x_{jm} \) can be written as

\[
\gamma_{jm} = \frac{\left| r_{jm}^H \hat{h}_{jjm} \right|^2}{r_{jm}^H \left( \frac{1}{\rho} I_N + R_{jjm} - \Phi_{jjm} - h_{jjm} \hat{h}_{jjm} + \sum_l H_{jl} H_{jl}^H \right) r_{jm}}.
\]

Thus, an ergodic achievable rate \( R_{jm} \) of UT \( m \) in cell \( j \) is given by

\[
R_{jm} = E \left[ \log_2 (1 + \gamma_{jm}) \right].
\]
III. ASYMPTOTIC ANALYSIS

In this section, we present our main technical results. All proofs are provided in [12]. Under the assumption that $N$ and $K$ grow infinitely large while keeping a finite ratio $K/N$, we will derive deterministic approximations $\bar{\gamma}_{jm}$ of $\gamma_{jm}$, such that

$$\gamma_{jm} - \bar{\gamma}_{jm} \xrightarrow{a.s.} 0 \quad (9)$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence. We will refer to the quantities $\bar{\gamma}_{jm}$ as deterministic equivalents of $\gamma_{jm}$. In the sequel, we assume that the following technical conditions hold:

A 1: The matrices $R_{jlk}$ have uniformly bounded spectral norm with respect to $N$ for all $j, l, k$.

A 2: \( \liminf_{N} \frac{1}{N} \text{tr} R_{jlk} > 0 \) for all $j, l, k$.

Our first result is a deterministic equivalent of the SINR at the output of the matched filter:

**Theorem 1 (Matched filter):** A deterministic equivalent of $\gamma_{jm}$ for the matched filter is given as

$$\bar{\gamma}_{jm}^{\text{MF}} = \frac{\left( \frac{1}{N} \text{tr} \Phi_{jjm} \right)^2}{\rho N^2 \text{tr} \Phi_{jjm} + \frac{1}{N} \sum_{l,k} \frac{1}{N} \text{tr} R_{jlk} \Phi_{jjm} + \sum_{l \neq j} \left| \frac{1}{N} \text{tr} \Phi_{jlm} \right|^2}.$$  

Our second result is a deterministic equivalent of the SINR at the output of the MMSE detector:

**Theorem 2 (MMSE detector):** A deterministic equivalent of $\gamma_{jm}$ for the MMSE detector is given as

$$\bar{\gamma}_{jm}^{\text{MMSE}} = \frac{\delta_{jm}^2}{\rho N^2 \text{tr} \Phi_{jjm} T'_j + \frac{1}{N} \sum_{l,k} \mu_{jlk} + \sum_{l \neq j} |\vartheta_{jlm}|^2},$$

where

\[
\begin{align*}
\mu_{jlk} &= \frac{1}{N} \text{tr} R_{jlk} T'_j, \\
\vartheta_{jlk} &= \frac{1}{N} \text{tr} \Phi_{jlk} T_j, \\
\vartheta'_{jlk} &= \frac{1}{N} \text{tr} \Phi_{jlk} T'_j
\end{align*}
\]

and where

- $T_j = T(\lambda)$ and $\delta_j = [\delta_{j1}, \ldots, \delta_{jK}]^T = \delta(\lambda)$ are given by Theorem 3 for $S = Z_j/N$, $R_k = \Phi_{jjk} \forall k$,
- $T'_j = T'(\lambda)$ is given by Theorem 4 for $S = Z_j/N$, $\Theta = I_N$, $R_k = \Phi_{jjk} \forall k$,
- $T'_{jm} = T'(\lambda)$, $\delta'_{jm} = [\delta'_{j1m}, \ldots, \delta'_{jKm}]^T = \delta'(\lambda)$ are given by Theorem 4 for $S = Z_j/N, \Theta = \Phi_{jjm}$, $R_k = \Phi_{jjk} \forall k$.

Theorems 3 and 4 can be found in the Appendix.

Interestingly, the performances of matched filter and MMSE detector coincide for an infinite number of antennas:

**Corollary 1 (Performance with an unlimited number of antennas):**

$$\lim_{N \to \infty} \bar{\gamma}_{jm}^{\text{MF}, \text{MMSE}} = \gamma_{\infty} \equiv \frac{\beta_{jm}^2}{\sum_{l \neq j} |\beta_{jlm}|^2},$$

where $\beta_{jlk} = \lim_{N} \frac{1}{N} \text{tr} \Phi_{jlk}$, whenever the limit exists.
IV. ON THE MASSIVE MIMO EFFECT

Let us for now ignore the effects of estimation noise, i.e., $\rho \to \infty$, and consider the simple channel model

$$\begin{align*}
H_{jj} &= \sqrt{\frac{N}{P}} AW_{jj}, \\
H_{jl} &= \sqrt{\frac{N}{P}} \alpha AW_{jl}, \quad l \neq j
\end{align*}$$

(10)

where $A \in \mathbb{C}^{N \times P}$ is composed of $P \leq N$ columns of an arbitrary unitary $N \times N$ matrix, $W_{jl} \in \mathbb{C}^{P \times K}$ are standard complex Gaussian matrices and $\alpha \in [0, 1]$ is an inter-cell interference factor. Note that this is a special case of (2). Under this model, the total energy of the channel grows linearly with the number of antennas $N$ and UTs $K$, since $\mathbb{E} [\text{tr} H_{jj} H_{jj}^H] = \frac{KN}{P} \text{tr} AA^H = KN$. The motivation behind this channel model is two-fold. First, we assume that the antenna aperture increases with each additional antenna element. Thus, the captured energy increases linearly with $N$. This is in contrast to existing works which assume that more and more antenna elements are packed into a fixed volume [16], [17]. An insufficiency of this channel model is that the captured energy grows without bounds as $N \to \infty$. However, we believe that linear energy gains can be achieved up to very large numbers of antennas if the size of the antenna array is scaled accordingly. Second, the number of DoF $P$ offered by the channel does not need to be equal to $N$ [9]. One could either assume $P$ to be large but constant or to scale with $N$, e.g., $P = cN$, where $c \in (0, 1]$. In general, $P$ depends on the amount of scattering in the channel and, therefore, on the radio environment. Under the assumptions made above, we obtain the following corollary from Theorem 1:

**Corollary 2:** For the channel model in (10) and $\rho \to \infty$, $\bar{\gamma}_{km}^{\text{MF}}$, for all $k, m$, can be given in closed form as

$$\bar{\gamma}_{km}^{\text{MF}} \triangleq \frac{1}{\frac{L}{\rho N} \text{noise} + \frac{K}{P} \tilde{L}^2 \text{multi-user interference} + \alpha(L-1) \text{pilot contamination}}$$

(11)

where $\tilde{L} = 1 + \alpha(L - 1)$.

We define the associated rate $\bar{R}_{\text{MF}}$ as

$$\bar{R}_{\text{MF}} = \log_2 \left(1 + \bar{\gamma}_{km}^{\text{MF}}\right).$$

(12)

One can make several observations from (11). Obviously, the effective SNR $\rho N$ increases linearly with $N$. Thus, if the number of antennas is doubled, the transmit power can be reduced by a factor of two to achieve the same SNR. Less obvious is that the multi-user interference depends mainly on the ratio $K/P$ (number of UTs per DoF) and not directly on the number of antennas. Moreover, noise and multi-user interference vanish for $N, P \to \infty$ while pilot contamination is the only performance-limiting factor [7], i.e., $\bar{\gamma}_{km}^{\text{MF}} \to \gamma_\infty = \frac{1}{\alpha(L-1)}$. We denote by $R_\infty$ the ultimately achievable rate, defined as

$$R_\infty = \log_2 (1 + \gamma_\infty) = \log_2 \left(1 + \frac{1}{\alpha(L-1)}\right).$$

(13)

It is interesting that even with more sophisticated linear single-user-detection, such as MMSE detection, the ultimate performance limit $\gamma_\infty$ cannot be exceeded (see Corollary 1 and [9]). Note that without pilot contamination, i.e., for $L = 1$ or $\alpha = 0$, the SINR grows without bounds as $P, N \to \infty$. If $P$ is fixed but large, the SINR saturates at a smaller value than $\gamma_\infty$. In this
case, adding additional antennas only improves the SNR. Moreover, MMSE detection is asymptotically superior to the MF.

Based on the above observations, we believe that it is justified to speak about a massive MIMO effect whenever \( \gamma_{jm} \) is close to \( \gamma_\infty \), or in other words, whenever noise and multi-user interference are small compared to the pilot contamination. It becomes evident from (11) that the number of antennas needed for massive MIMO depends strongly on the system parameters \( P, K, L, \alpha \) and \( \rho \). In particular, there is no massive MIMO effect without pilot contamination since \( \gamma_\infty \to \infty \). Thus, massive MIMO can be seen as a particular operating condition in multi-cellular systems where the performance is ultimately limited by pilot contamination and the matched filter achieves a performance close to this ultimate limit. To make this definition more precise, we say that we operate under massive MIMO conditions if, for some desired \( \eta \in (0,1) \),

\[
\bar{R}^{\text{MF}} \geq \eta R_\infty. \tag{14}
\]

This condition implies that we achieve at least the fraction \( \eta \) of the ultimate performance with the simplest form of single user detection. Solving (11) for \( P/K \) leads to

\[
P \geq \frac{1}{L^2 [(1 + \gamma_\infty)^\eta - 1]} - \frac{1}{\rho N L} - \frac{\alpha (L - 1)}{L^2}. \tag{15}
\]

For the particular channel model considered in this section, also Theorem 2 can be significantly simplified:

**Corollary 3:** For the channel model in (10) and \( \rho_r \to \infty \), \( \bar{\gamma}_{km}^{\text{MMSE}} \), for all \( k, m \), can be given in closed-form as

\[
\bar{\gamma}_{km}^{\text{MMSE}} = \frac{1}{\rho N X + \frac{K}{\rho N} L^2 Y + \alpha (L - 1)}
\]

where \( \bar{L} = 1 + \alpha (L - 1) \), \( Z = \lambda \bar{L} (1 + \delta) + \frac{K}{\rho} [1 + (1 + \delta) (\bar{L}^2 - 1)] \),

\[
X = \frac{Z^2}{Z^2 - \frac{K}{\rho}}, \quad Y = X + \frac{[1 + \alpha^2 (L - 1)] (1 - 2Z)}{L^2 (Z^2 - \frac{K}{\rho})}, \quad \delta = \frac{1 - \lambda \bar{L} - \frac{K}{\rho} L^2 + \sqrt{(1 + \lambda \bar{L} + \frac{K}{\rho} L^2)^2 - \frac{4}{\bar{L}^2}}}{2 \left[ \lambda \bar{L} + \frac{K}{\rho} (L^2 - 1) \right]}. \]

The last result reveals that also the SINR of the MMSE detector depends on \( P, K, N \) and \( \rho \) only through the ratio \( P/K \) and the effective SNR \( \rho N \). Hence, we can use Corollary 3 to determine the ratio \( P/K \) necessary to satisfy the condition

\[
\bar{R}^{\text{MMSE}} = \log \left( 1 + \bar{\gamma}^{\text{MMSE}} \right) \geq \eta R_\infty \tag{16}
\]

which can be very efficiently solved by a simple line search, e.g. via bi-section. We assume in the sequel \( \lambda = 1/(\rho N) \).

Before we proceed, let us verify the accuracy of the approximations \( \bar{R}^{\text{MF}} \) and \( \bar{R}^{\text{MMSE}} \). In Fig. 1, we depict the ergodic achievable rate \( R_{jm} \) of an arbitrary UT with MF and MMSE detection as a function of the number of antennas \( N \) for \( K = 10 \) UTs, \( L = 4 \) cells, \( \rho = 0 \) dB and intercell-interference factor \( \alpha = 0.1 \). We compare two different cases: \( P = N \) and \( P = N/3 \).

As expected, the performance in the latter scenario is inferior due to stronger multi-user interference. Most importantly, our closed-form approximations are almost indistinguishable from the simulation results over the entire range of \( N \).

Fig. 2 and 3 show the necessary DoF per UT \( P/K \) for a given effective SNR \( \rho N \) to achieve a spectral efficiency of \( \eta R_\infty \) with either the MF (solid lines) or MMSE detector (dashed lines). We consider \( L = 4 \) cells and an intercell-interference factor
$\alpha = 0.3$ and $\alpha = 0.1$, respectively. The plots must be understood in the following way: Each curve corresponds to a particular value of $\eta$. In the region above each curve, the condition (14) or (16) is satisfied.

Let us first focus on Fig. 2 with $\alpha = 0.3$. For an effective SNR $\rho N = 20$ dB (e.g. $\rho = 0$ dB and $N = 100 = 20$ dB), we need about $P/K = 90$ DoF per UT to achieve 90\% of the ultimate performance $R_\infty$, i.e., $0.9 \times 2.2 \approx 2$ b/s/Hz. If $P \approx N$, only a single UT could be served. However, if we had $N = 1000 = 30$ dB antennas, the transmit power $\rho$ could be decreased by 10 dB and 10 UTs could be served with the same performance. At the same operating point, the MMSE detector requires only $\sim 60$ DoF per UT to achieve 90\% of the ultimate performance. Thus, MMSE detection would allow us to increase the number of simultaneously served UTs by a factor $\frac{90}{60} = 1.5$. This example also demonstrates the importance of the relation between $N$ and $P$. In particular, if $P$ saturates for some $N$, adding more antennas increases the effective SNR but does not reduce the multi-user interference. Thus, the number of UTs which can be simultaneously supported depends significantly on the radio environment. We can further see that adding antennas shows diminishing returns. This is because the distances between the curves for different values of $\eta$ grow exponentially fast. Remember that for $\eta = 1$, a ratio of $P/K = \infty$ would be needed. A last observation we can make is that the difference between MF and MMSE detection is marginal for small values of $\eta$ but gets quickly pronounced as $\eta \to 1$.

Turning to Fig. 3 for $\alpha = 0.1$, we can see that for the same effective SNR $\rho N = 20$ dB and the same number of DoF per UT $P/K = 90$ as in the previous example, only 80\% of the ultimate performance are achieved by the MF. However, since the intercell-interference is significantly smaller compared to the previous example, this corresponds to $0.9 \times 5.1 \approx 4.6$ b/s/Hz. Thus, although we operate further away from the ultimate performance limit, the resulting spectral efficiency is still higher. With MMSE detection, only 35 DoF per UT are necessary to achieve the same performance and, consequently, $90/35 \approx 2.5$ times more UTs could be simultaneously served. With decreasing intercell interference (and hence decreasing pilot contamination) the advantages of MMSE detection become more and more significant.

V. Conclusions

We have studied the uplink of a MIMO multi-cell system where channel estimates are prone to pilot contamination. Under a very general channel model, allowing for an individual correlation matrix for each UT, and the assumption that the number of BS-antennas $N$ grows large, we have derived deterministic equivalents of the SINR at the output of the matched filter and MMSE detector. These results have been used to study the performance of both detectors in the large but finite $N$ regime. Interestingly, their performance depends mainly on the DoF per UT the channel offers and the effective SNR. Moreover, we have determined (i) how many antennas are needed to achieve $\eta\%$ of the ultimate performance limit and (ii) how many more antennas are needed by the MF to achieve MMSE performance. Our results can be directly applied for the downlink and could be further used to analyze the effects of more realistic channels models, such as antenna correlation, spacing, and aperture.

Appendix

**Theorem 3 ([18, Theorem 1]):** Let $D \in \mathbb{C}^{N \times N}$ and $S \in \mathbb{C}^{N \times N}$ be Hermitian non-negative definite matrices and let $H \in \mathbb{C}^{N \times K}$ be a random matrix with columns $h_k = \frac{1}{\sqrt{N}}R_k^{\frac{1}{2}}u_k$, where $u_k \in \mathbb{C}^N$ are random vectors of i.i.d. elements with
zero mean, unit variance and finite 8th order moment, and $R_k \in \mathbb{C}^{N \times N}$ are deterministic covariance matrices. Assume that $D$, $S$ and the matrices $R_k$, $k = 1, \ldots, K$, have uniformly bounded spectral norms (with respect to $N$). Let $N, K \to \infty$, such that $0 \leq \lim \inf \frac{K}{N} \leq \lim \sup \frac{K}{N} < \infty$. Then, for any $\rho > 0$,
\[
\frac{1}{N} \text{tr} \left( D \left( HH^H + S + \rho I_N \right)^{-1} - \frac{1}{N} \text{tr} \left( T(\rho) \right) \right) \xrightarrow{a.s.} 0
\]
where $T(\rho) \in \mathbb{C}^{N \times N}$ is defined as
\[
T(\rho) = \left( \frac{1}{N} \sum_{k=1}^{K} \frac{R_k}{1 + \delta_k(\rho)} + S + \rho I_N \right)^{-1}
\]
and the following set of $K$ implicit equations
\[
\delta_k(\rho) = \frac{1}{N} \text{tr} R_k T(\rho), \quad k = 1, \ldots, K
\]
has a unique solution $\delta_1(\rho), \ldots, \delta_K(\rho) \geq 0$.

**Theorem 4:** Let $\Theta \in \mathbb{C}^{N \times N}$ be a Hermitian non-negative definite matrix with uniformly bounded spectral norm (with respect to $N$). Under the same conditions as in Theorem 3,
\[
\frac{1}{N} \text{tr} \left( D \left( HH^H + S + \rho I_N \right)^{-1} \Theta \left( HH^H + S + \rho I_N \right)^{-1} - \frac{1}{N} \text{tr} \left( T'(\rho) \right) \right) \xrightarrow{a.s.} 0
\]
where $T'(\rho) \in \mathbb{C}^{N \times N}$ is defined as
\[
T'(\rho) = T(\rho) \Theta T(\rho) + T(\rho) \frac{1}{N} \sum_{k=1}^{K} \frac{R_k \delta'_k(\rho)}{(1 + \delta_k(\rho))^2} T(\rho)
\]
with $T(\rho)$ and $\delta_k(\rho)$ as defined in Theorem 3 and $\delta'(\rho) = [\delta'_1(\rho), \ldots, \delta'_K(\rho)]^T$ given by
\[
\delta'(\rho) = (I_K - J(\rho))^{-1} v(\rho), \quad [J(\rho)]_{kl} = \frac{1}{N} \text{tr} R_k T(\rho) R_l T(\rho), \quad [v(\rho)]_k = \frac{1}{N} \text{tr} R_k T(\rho) \Theta T(\rho)
\]
where $J(\rho) \in \mathbb{C}^{K \times K}$ and $v(\rho) \in \mathbb{C}^K$.

**Proof:** The proof is provided in [12].

**References**


Fig. 1. Ergodic achievable rate with MF and MMSE detection versus number of antennas for \( P = \{ N, N/3 \} \) for \( \rho = 0 \) dB.


Fig. 2. Degrees of freedom per UT $P/K$ necessary to achieve $\eta R_\infty$ versus effective SNR $\rho N$ for $L = 4$ and $\alpha = 0.3$.

Fig. 3. Degrees of freedom per UT $P/K$ necessary to achieve $\eta R_\infty$ versus effective SNR $\rho N$ for $L = 4$ and $\alpha = 0.1$. 