Optimum Performance of Short Block Length Codes Under Multivariate Stationary Rayleigh Fading

Xavier Mestre, Senior Member, IEEE, David Gregoratti, Senior Member, IEEE, and Peng Zhang

Abstract—The performance of short block length codes in the presence of multivariate stationary Rayleigh fading coherent channels is studied. The channel model is inspired by multi-symbol OFDM transmissions for ultra-reliable low latency communication (URLLC) services, according to which information is sent in a small number of OFDM symbols to reduce latency. More specifically, the dispersion of the coherent fading channel is generalized from the scalar case, which is well known in the literature, to the multivariate one. The obtained expressions are then particularized to the Rayleigh fading statistics and expressed in closed form as a series expansion. Finally, a high SNR approximation of the ergodic capacity and channel dispersion is derived for this particular fading choice. Results show that the code performance depends on the time-frequency fading correlation through very simple functions of the channel correlation coefficients. In particular, it is shown that the channel dispersion converges to a constant plus the sum of the dilogarithm of the correlation coefficients of the channel process. These formulas provide a very useful tool to design physical layer system parameters from channel correlation information measured at the receiver.

Index Terms—Short block length, URLLC, channel dispersion, correlated Rayleigh fading.

I. INTRODUCTION

5G COMMUNICATION systems introduce two new use cases, namely massive machine type communications (MMTC) and ultra-reliable low latency communications (URLLC), whose data traffic is served by short packets, giving the little amount of information contained in each sporadic message (MMTC) or the extreme requirements in terms of latency (URLLC). Unfortunately, short-packet based communications are not well represented by classical information theory results, which are known to be significant only for asymptotically large code block lengths. The interest in these new type of communication strategies has motivated the introduction of coding theorems for not-so-large codewords, which can be seen as second-order refinements of conventional channel coding theorems.

In order to introduce this theory, consider an arbitrary noisy communication channel and denote by \( N_{\text{cw}}^* (N, \epsilon) \) the maximum cardinality (number of codewords) of a codebook with block length \( N \) that can be decoded with block error probability no greater than \( \epsilon \). Denote by \( R^* (N, \epsilon) \) the associated coding rate, defined as

\[
R^* (N, \epsilon) = \frac{\log N_{\text{cw}}^* (N, \epsilon)}{N}
\]

in nats per channel access. The evaluation of \( R^* (N, \epsilon) \) in closed form is extremely difficult in channels of interest, so that most of the insights can be gained when the block length converges to infinity, namely \( N \to \infty \). Recent information theory results [1] state that, for a large number of channels, this quantity accepts the following asymptotic expansion:

\[
R^* (N, \epsilon) = C - \sqrt{\frac{V}{N}} Q^{-1} (\epsilon) + o \left( N^{-1/2} \right) \tag{1}
\]

where \( C \) denotes the channel capacity, \( Q(\cdot) \) is the Gaussian Q-function\(^1\) and \( V \) is a certain channel dependent quantity, usually referred to as channel dispersion. The above formula can be seen as a refinement of the conventional Shannon channel coding theorem, which states that the maximum achievable coding rate that guarantees an arbitrarily low \( \epsilon \) for a sufficiently large block length \( N \) is given by the channel capacity \( C \). According to this formula, the channel dispersion \( V \) establishes a penalty term of order \( 1/\sqrt{N} \) that separates the maximum achievable rate from the actual capacity. This type of expansion had been known for discrete memoryless channels since the 60s [2], [3], but it was only recently that equivalent expressions were proven for more interesting channel models [4].

One of the most relevant channels for which an expansion like the one in (1) has been established is the AWGN channel.

\[ Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt \]

\(^1\)
If \( x \in \mathbb{C}^N \) denotes the transmitted codeword, the observation \( y \in \mathbb{C}^N \) can be expressed as
\[
y = x + w
\]
where \( w \sim \mathcal{CN}(0, \sigma^2 I_N) \) is a noise vector (assumed circularly symmetric white Gaussian distributed) and where we impose the average power constraint \( \|x\|^2 \leq NP \). The capacity and dispersion for this channel are respectively given by [4]
\[
C_{\text{AWGN}}(\rho) = c(\rho) \\
V_{\text{AWGN}}(\rho) = 1 - d^2(\rho)
\]
where \( \rho = P/\sigma^2 \) is the signal to noise ratio and where
\[
c(\rho) = \log(1 + \rho) \quad \text{and} \quad d(\rho) = \frac{dc(\rho)}{d\rho} = (1 + \rho)^{-1}
\]
(2)
The availability of this new type of short block length results facilitates the design of communication strategies that have traditionally been optimized for capacity maximization. As illustrated in [5], the approximation in (1) allows one to optimize system parameters (such as the duplexing time in TDD schemes, the size of the slot in random access procedures or the throughput of ARQ schemes) so as to guarantee a certain end-to-end code-agnostic system performance. In [6] the performance of different codes is compared in the URLLC regime, and it is concluded that modern codes such as LDPC and turbo codes show a certain performance gap with respect to the bound in (1). This gap can be considerably reduced by employing more classical code designs such as BCH at the expense of decoding complexity.

More interesting to wireless communication problems are channel models subject to variability, that is, fading. The small block-length achievable rates under different fading channel models have also been the subject of deep research during the past decade. In this general setting the observation \( y \in \mathbb{C}^N \) takes the more general form
\[
y = h \odot x + w
\]
(3)
where \( x \) and \( w \) are as before, \( \odot \) denotes element-wise multiplication and \( h \in \mathbb{C}^N \) is the channel response, assumed here to be random and known to the receiver (coherent detection). Depending on how the randomness of \( h \) is modeled, different channel fading models are obtained.

For example, the quasi-static fading channel model assumes that \( h = h_1 \) where \( h \) is a random coefficient that is independently chosen for each codeword and kept constant for the whole transmission. It is well known that the Shannon capacity of this channel model is zero [7], but it was shown in [8] that one can still establish an asymptotic expansion of the maximum achievable rate as the one in (1) with \( C \) equal to the outage capacity
\[
C_\epsilon(\rho) = \sup_{R} \left\{ R : \mathbb{P} \left[ c \left( |h|^2 \rho \right) < R \right] \leq \epsilon \right\}
\]
and \( V = 0 \). In other words, the channel dispersion of the quasi-static fading channel model is identically zero, so that \( R^*(N, \epsilon) = C_\epsilon(\rho) + O(N^{-1} \log N) \). This shows that the maximum achievable rate converges to the outage capacity at a much higher rate \( O(N^{-1} \log N) \) as compared to the conventional channel dispersion term of the AWGN or other fading channel models, which is of order \( O(1/\sqrt{N}) \). The same fading channel model was analyzed in [9] in a MIMO setting assuming that the number of antennas at both sides of the communication link converged to infinity. On the other hand, [10] considers the problem of power control for this channel model under the assumption of an average power constraint on the whole codebook.

A different channel model was studied in [11], where the channel \( h \) was assumed to be an \( N \)-length realization of a scalar stationary fading process. It was shown in [11] that, under several technical conditions on the fading process, the maximum achievable rate for this channel accepts an expansion as in (1), with \( C \) and \( V \) replaced with
\[
C_{\text{ERG}}(\rho) = \mathbb{E} \left[ c \left( |h|^2 \rho \right) \right] \\
V_{\text{ERG}}(\rho) = \mathbb{L} \left( c \left( |h_n|^2 \right) \right)_{n \in \mathbb{N}} + \mathbb{E}^2 \left[ d \left( |h|^2 \rho \right) \right]
\]
respectively, where \( \mathbb{E} \) denotes expectation and \( h \) denotes a general entry of \( h \). In the above expression, we have introduced the quantity
\[
\mathbb{L} \left( (x_n)_{n \in \mathbb{N}} \right) = \lim_{N \to \infty} \frac{1}{N} \mathbb{V} \left[ \sum_{n=1}^{N} x_n \right]
\]
which is usually referred to as the long term variance of the process \( x = (x_n)_{n \in \mathbb{N}} \), with \( \mathbb{V} \) denoting variance. Thus, according to this fading channel model, the maximum achievable rate converges to the ergodic capacity \( C_{\text{ERG}}(\rho) \) as \( N \to \infty \), which only depends on the individual distribution of the entries of \( h \) and is therefore independent of the fading dynamics. The variability of the fading process plays a fundamental role in the channel dispersion, which clearly depends on the cross-correlation coefficients of the channel fading process through the long term variance \( \mathbb{L} \left( c \left( |h_n|^2 \right) \right)_{n \in \mathbb{N}} \).

A similar fading model, although simpler to analyze, is the coherent block-fading channel model. According to this model, the codeword is divided into \( L \) sub-blocks of length \( M \), that is \( N = ML \). The channel is assumed to be constant during the duration of each sub-block and independent and identically distributed across sub-blocks, so that in practice we can use (3) where \( h = h_1 \otimes M \), \( \otimes \) representing the Kronecker product and \( h_1 \in \mathbb{C}^L \) denoting a certain random vector of independent and identically distributed entries. In fact, assuming \( L \to \infty \), the result in [11] can be reformulated for the block fading model with the same capacity and dispersion, where the long term variance takes the simplified form [12]
\[
\mathbb{L} \left( c \left( |h_n|^2 \right) \right)_{n \in \mathbb{N}} = MV \left[ c \left( |h|^2 \rho \right) \right]
\]
This was later generalized to different MIMO settings in [13]–[15].

This paper is concerned with OFDM transmissions supporting URLLC communications, where the codeword spans a sufficiently high bandwidth and therefore the channel in (3)
may experiment strong variations. In this sense, it appears that the channel model in [11] is the most relevant for this type of transmission. In fact, we consider here a slightly more general version of this channel, the main difference with respect to [11] being that the codeword is allowed to span several OFDM symbols, which may experiment similar or completely uncorrelated channel frequency responses. Some of these results were previously presented in a conference version of this paper [16].

Before going into the technical details, it should be pointed out that there exist other powerful techniques to analyze the coding performance in the short block length regime that are not considered herein. In particular, saddlepoint techniques have recently been shown to provide more accurate asymptotic expansions than those established in (1), see e.g. [17]–[19]. Thus, a future research direction could target the refinement of the results presented here by using more sophisticated techniques such as those based on saddlepoint approximations.

The rest of the paper is organized as follows. Section II introduces the multivariate coherent stationary channel and justifies its relevance in current OFDM-based wireless standards. It also provides a coding theorem for this type of channel and particularizes this result to the case where the channel is Rayleigh distributed. Section II then analyzes the performance under Rayleigh fading at large values of the SNR, showing that in this regime the finite block length performance depends on some specific functions of the channel correlation coefficients. Finally, Section IV provides some numerical studies and Section V concludes the paper.

**Notation:** Matrix and vector quantities are denoted by boldfaced upper case (A) and boldfaced lower case letters (a) respectively. The $M \times M$ identity matrix is denoted by $I_M$, whereas $I_M$ denotes an $M$-dimensional column vector of all ones. Transpose and conjugate transpose operators are denoted as $(\cdot)^T$ and $(\cdot)^H$ respectively, whereas $(\cdot)^*$ denotes complex conjugate (without transposition). The symbols $\otimes$ and $\odot$ denote Kronecker and Hadamard (element-wise) product between matrices. Furthermore, $\text{vec}(A)$ is a column vector formed by piling the columns of A on top of each other, whereas $\|A\|$ denotes its spectral norm. Expectation is denoted as $E$, whereas $\mathbb{V}$ denotes variance when it has a single argument, i.e. $\mathbb{V}(a) = E(|a|^2) - |Ea|^2$, or covariance if it has two, i.e. $\mathbb{V}[a,b] = E(ab^*) - E(a)E(b^*)$.

**II. SYSTEM MODEL AND MAXIMUM ACHIEVABLE RATE**

Let us consider the transmission of a codeword over a set of $L$ OFDM symbols subject to time and frequency channel selectivity. We assume that channel state information is not available at the transmitter but perfectly known at the receiver (coherent reception). Assuming that $M$ denotes the number of occupied subcarriers, the block length will be given by $N = ML$, and therefore we may represent the codeword as an $M \times L$ complex matrix $X$. The encoder output is split into $L$ different parts of equal dimension $M$ (see further Figure 1), which correspond to the columns of $X$, denoted by $x_l$, $l = 1, \ldots, L$. Each column vector $x_l$ is OFDM-modulated onto $M$ subcarriers and transmitted through a frequency selective channel with frequency response given by the $M$-dimensional column vector $h_l$. The received signal in the frequency domain for the $l$th OFDM symbol can be written as

$$y_l = h_l \odot x_l + w_l$$

where $w_l$ denotes a noise vector. As before, we model $w_l$ as independent realizations of a random vector $w_l \sim \mathcal{CN}(0, \sigma^2 I_M)$. Furthermore, we will assume that the power of the transmitted signal is limited for each OFDM symbol, that is

$$\|x_l\|^2 \leq MP, \quad l = 1, \ldots, L.$$  

By gathering the $L$ column vectors into the corresponding $M \times L$ matrices (e.g. $Y = [y_1, \ldots, y_L]$, $H = [h_1, \ldots, h_L]$, etc.) we can express the observed signal as

$$Y = H \odot X + W.$$

Observe that this is fundamentally different from a conventional MIMO signal model, where the Hadamard product is replaced with a conventional matrix product (i.e. $Y =HX + W$).

In this paper, we will model the channel matrix $H$ as bi-dimensional random process that has identically distributed entries but presents correlation in both dimensions. These correlations will model the practical dynamics of the fading across both frequency and time domains. Furthermore, in order to model practical URLLC transmissions, we will generally assume that the matrix $H$ is tall, in the sense that $M \gg L$. More specifically, we assume that $H$ is a realization of an $L$-variate stationary process that evolves along the frequency domain. Indeed, let $\tilde{h}^{(m)} \in \mathbb{C}^{1 \times L}$, $m = 1, \ldots, M$, denote the $m$th row of the channel matrix $H$, so that

$$H = \left[ (\tilde{h}^{(1)})^T, \ldots, (\tilde{h}^{(M)})^T \right]^T.$$  

In this paper, we will model the sequence $\{\tilde{h}^{(m)}\}_{m \in \mathbb{Z}}$ as a (discrete-index) stationary multivariate process with zero mean
and $L \times L$ covariance matrix
\[ E \left[ \left( \hat{H}^{(m_0)} \right)^T \left( \hat{H}^{(m_0+m)} \right)^* \right] = R_m. \] (6)

Note that the covariance does not depend on $m_0$ because of the stationarity of the process. We also assume that the channel distribution is absolutely continuous and that the fading coefficients are normalized so that the diagonal entries of $R_0$ are all equal to one, whereas the off-diagonal ones are strictly lower than one in magnitude. Let $h_{m,t}$ denote the $(m,t)$th entry of the channel matrix $H$. We denote by $\rho^{h}_{m,t}(m)$ the $\rho$-mixing coefficients of the associated bidimensional process $\{h_{m,t}\}_{m,t}$ defined as follows [20]. Letting $F^{m_1,m_2}_{m_3,m_4}$ denote the sigma-algebra generated by the random variables $\{h_{m_1,t}, \ldots , h_{m_2,t}\}$, we define (for $m \in \mathbb{N}$)
\[ \rho^{h}_{m,t}(m) = \sup_{f \in L_2(F^{-\infty,-k},g \in L_2(F^{k,+\infty})} \left| E[f g^*] \right| \]
where the supremum is over all square-integrable functions $f,g$ defined on these sigma-algebras, such that $E[f] = E[g] = 0$ and $E[|f|^2] = E[|g|^2] = 1$. We assume that these $\rho$-mixing coefficients decay sufficiently fast when they are considered as functions of their lag $m$, namely
\[ \max_{l,t'} \rho^{h}_{l,t'}(m) < K_1 e^{-K_2 m} \] (7)
for some positive $K_1$, $K_2$. The above condition is verified in practice by a wide range of channel models. In fact, it can be shown that the $\rho$-mixing coefficient decays exponentially when the channel has a rational spectrum density function [21]. Therefore, any channel obtained as a finite-order auto-regressive moving average (ARMA) process satisfies this condition.

We have now all the ingredients to characterize the maximum achievable rate of this channel. The following result is a generalization of the one in [11] to the case $L \geq 1$.

**Theorem 1:** Under the above assumptions, there exists a code with codeword length $ML$ and maximum probability of error $\epsilon$ with minimum rate
\[ R^* (ML, \epsilon) = C(\rho) - \sqrt{\frac{V_L(\rho)}{ML}} Q^{-1}(\epsilon) + o\left( \frac{1}{\sqrt{M}} \right) \] (8)
where $C(\rho) = E[c(\rho|h|^2)]$ is the channel ergodic capacity, where
\[ V_L(\rho) = L_L(\rho) + 1 - E^2[d(\rho|h|^2)] \] (9)
is the associated channel dispersion and where $L_L(\rho)$ denotes the long term variance of the bidimensional capacity process, that is
\[ L_L(\rho) = \lim_{M \to \infty} \frac{1}{ML} \sum_{l=1}^L \sum_{m=1}^M c(\rho|h_m|^2) \] (10)
Reciprocally, any code of block length $ML$ and maximum error probability of error $\epsilon$ such that $\max_{1 \leq l \leq M} \epsilon_x l \leq \sqrt{PM} \theta^2$ for some $0 \leq \theta < 1/12$ has rate bounded by (8).

**Proof:** See Appendix A. \[ \square \]

1) **Particularization to the Rayleigh Case:** We can particularize the above expressions for the specific case where the fading is Rayleigh distributed, that is when the channel coefficients $h_{m,t}$ follow a normalized circularly symmetric Gaussian random law. Let us denote by $E_1(x)$ the exponential integral, defined as
\[ E_1(x) = \int_x^{\infty} \frac{\exp(-t)}{t} dt. \]
We will also let $\gamma$ denote the Euler-Mascheroni constant, $\Gamma(r,x)$ the upper incomplete gamma function and $\eta(F_k)$ the generalized hypergeometric function [22, 9.14]. For any sequence of $L \times L$ matrices $(A_n)_{n \in \mathbb{Z}}$, we define the scalar-valued operator
\[ T(A_n) = \frac{1}{L} \sum_{l=1}^L \sum_{l'=1}^L \{A_0\}_{l,l'} + \frac{1}{L} \sum_{m=-\infty}^{\infty} \mathbf{1}_L^T A_m \mathbf{1}_L. \]
This operator simply returns the normalized sum of all the entries of the matrix sequence, except for the diagonal of the matrix with index $n = 0$.

**Corollary 1:** In addition to the assumptions in Theorem 1, assume that the channel entries $h_{m,t}$ are zero-mean circularly symmetric Gaussian distributed (Rayleigh fading). Define the absolute correlation coefficient as $\phi_{l,t'}(m) = \langle (R_m)_{l,l'} \rangle$. We can express the channel capacity as
\[ C(\rho) = \exp \left( \frac{1}{\rho} \right) E_1 \left( \frac{1}{\rho} \right) \] (11)
whereas the channel dispersion takes the form
\[ V_L(\rho) = \mathbb{V} \left[ c(\rho|h|^2) \right] + T(V_c(\mu)) \]
\[ = 1 - \frac{1}{\rho^2} \exp \left( \frac{2}{\rho} \right) E_1^2 \left( \frac{1}{\rho} \right) \] (12)
where
\[ \mathbb{V} \left[ c(\rho|h|^2) \right] = \exp \left( \frac{1}{\rho} \right) \left[ \log^2(\rho) - 2 \gamma \log \rho + \gamma^2 + \frac{\pi^2}{6} \right] - \frac{2}{\rho^3} F_3 \left( \left( 1, 1, 1 ; (2, 2, 2) ; -\frac{1}{\rho} \right) \right) - C^2(\rho) \] (13)
and where we have defined the $L \times L$ matrix $V_c(\mu)$ with $(l,l')$th entry equal to (14), shown at the bottom of the next page.

**Proof:** Indeed, the capacity expression in (11) follows directly from partial integration. On the other hand, we can also trivially see that
\[ E \left[ d(\rho|h|^2) \right] = E \left[ \frac{1}{1 + \rho|h|^2} \right] \]
\[ = 1 - \rho \frac{dc(\rho)}{d\rho} = \frac{1}{\rho} \exp \left( \frac{1}{\rho} \right) E_1 \left( \frac{1}{\rho} \right) \] (15)
\[ \square \]
where we have used the fact that \( E_1' (x) = -x^{-1} \exp (-x) \).

Regarding the long term variance, we can use stationarity to write

\[
\mathcal{L}_L (\rho) = \mathbb{V} \left[ \left( \rho | h_1 |^2 \right) \right] + \frac{1}{L} \sum_{l,l' = 1}^{L} \{ \mathbf{V}_c (0) \}_{l,l'}
\]

\[
+ \lim_{M \to \infty} \frac{1}{L} \sum_{l,l' = 1}^{L} \sum_{m = -M}^{M - 1} \left( 1 - \frac{|m|}{M} \right) 1_l^T \mathbf{V}_c (m) 1_l
\]

where

\[
\{ \mathbf{V}_c (m) \}_{l,l'} = \mathbb{V} \left[ \left( \rho | h_{m,l} |^2 \right) , \left( \rho | h_{m',l'} |^2 \right) \right].
\]

Observe that, under condition (7) and given the definition of the matrix \( \mathbf{V}_c (m) \) above, we have

\[
\frac{1}{M} \sum_{m = -M}^{M - 1} |m| \{ \mathbf{V}_c (m) \}_{l,l'} \leq \frac{1}{M} \sum_{m = -M}^{M - 1} |m| \rho_{l,l'}^h (|m|) \to 0
\]

and this shows (12).

The final result follows from the evaluation of the second moments of the capacity process \( c^2 (\rho | h_{m,l} |^2) \).

namely the evaluation of \( \mathbb{E} \left[ c^2 (\rho | h_1 |^2) \right] \) when \( |h_1|^2 \) is exponentially distributed with mean \( \lambda \) and the evaluation of \( \mathbb{E} \left[ c^2 (\rho | h_1 |^2) c (\rho | h_2 |^2) \right] \) when \( (h_1, h_2) \) is a circularly symmetric complex Gaussian vector with zero mean and unit variances \( \mathbb{E} |h_1|^2 = \mathbb{E} |h_2|^2 = 1 \). Regarding the first, we can trivially use [23, eq. (15)] and directly obtain (13).

For the cross-covariance computation, we can follow the approach [24] and use a series representation of the modified Bessel function of the first kind \( I_0 \) in the formula

\[
\mathbb{E} \left[ c (\rho | h_1 |^2) c (\rho | h_2 |^2) \right] = \int_0^\infty \int_0^\infty c (\rho x) c (\rho y) \exp \left( -\frac{\phi}{1 - \phi^2} \sqrt{xy} \right) dxdy
\]

where \( \phi = |\mathbb{E} [h_1 h_2]| \). Indeed, we have

\[
I_0 \left( 2 \frac{\phi}{1 - \phi^2} \sqrt{xy} \right) = \sum_{k = 0}^{\infty} \frac{1}{(k!)^2} \left( \frac{\phi}{1 - \phi^2} \right)^{2k} (xy)^k
\]

Inserting the Taylor expansion into the above integral, we obtain

\[
\mathbb{E} \left[ c (\rho | h_1 |^2) c (\rho | h_2 |^2) \right] = \frac{1}{1 - \phi^2} \sum_{k = 0}^{\infty} \left( \frac{\phi}{1 - \phi^2} \right)^{2k} \mathcal{I}_k^2
\]

(16)

where

\[
\mathcal{I}_k = \frac{1}{k!} \int_0^\infty x^k \log (1 + \rho x) \exp \left( -\frac{x}{1 - \phi^2} \right) dx
\]

Using integration by parts and [22, 3.381.3] in order to solve \( \mathcal{I}_k \) and inserting the result into (16), we obtain (14).

The above expression can be evaluated numerically by taking a finite number of terms in the sum of (14). In practice, one can increase the number of terms of the series until the contribution of additional terms falls below a certain error. Obviously, the number of terms that need to be computed increases as the correlation coefficient \( \phi \) becomes close to one.

In any case, it appears difficult to draw any insight from the complicated form of the resulting variance/covariance expressions. Furthermore, in order to evaluate the long term variance, one needs the collection of all the correlation coefficients of the time-frequency grid, namely \( \phi_{l,l'} (m) \) for \( 1 \leq l, l' \leq L \) and \( m = 1, \ldots, M \), which should be measured at the receiver and fed back to the transmitter in order to make system design decisions. We will see in the next subsection that a much simpler expression (which depend on some functions of the coefficients \( \phi_{l,l'} (m) \), rather than the whole collection) can be obtained by investigating the asymptotic behavior of these quantities for large signal to noise ratios (SNR).

### III. High SNR Analysis

In this section, we analyze the behavior of the dispersion in (9) as the signal to noise ratio (SNR) converges to infinity, namely \( \rho \to \infty \). The asymptotic behavior of the capacity \( C (\rho) \) can be obtained by noticing that the exponential integral can be expanded as

\[
E_1 \left( \frac{1}{\rho} \right) = \log (\rho) - \gamma - \sum_{k = 1}^{\infty} \frac{(-1)^k}{k k!} \left( \frac{1}{\rho} \right)^k
\]

and therefore

\[
C (\rho) = \log (\rho) - \gamma + \frac{\log (\rho) - \gamma + 1}{\rho} + O \left( \frac{\log (\rho)}{\rho^2} \right).
\]

Regarding the channel dispersion \( \mathcal{V}_L (\rho) \), we can expand it as established in the following result.

\[
\{ \mathbf{V}_c (m) \}_{l,l'} = \frac{\exp \left( \frac{2}{\rho (1 - \phi_{l,l'}^2 (m))} \right)}{1 - \phi_{l,l'}^2 (m)} \times \prod_{i = 0}^{\infty} \left( \frac{\phi_{l,l'} (m)}{1 - \phi_{l,l'}^2 (m)} \right)^{2i}
\]

\[
\times \sum_{i = 0}^{\infty} \left( \frac{(-\rho)^{-i}}{(i - r)!} \left( 1 - \phi_{l,l'}^2 (m) \right)^{r+1} \right) \sum_{k = 0}^{r} \frac{1}{(r - k)!} \left( \frac{1}{\rho (1 - \phi_{l,l'}^2 (m))} \right)^{r+k} - C^2 (\rho).
\]

(14)
Proposition 1: As \( \rho \to \infty \), the channel dispersion \( \mathcal{V}_L(\rho) \) accepts the expansion
\[
\mathcal{V}_L(\rho) = \mu_1 - \frac{(\log(\rho) - \gamma + 1)^2}{\rho} + 2\frac{\log(\rho) - \gamma + 1}{\rho} \mu_2 + \frac{1}{\rho} \mu_3 + o\left(\frac{1}{\rho}\right)
\]
where we have defined the three coefficients
\[
\mu_1 = 1 + \frac{\pi^2}{6} + T \log(1 + |\mathbf{R}_m|^2)
\]
\[
\mu_2 = T + \log(1 + |\mathbf{R}_m|^2)
\]
\[
\mu_3 = -1 + \frac{\pi^2}{6} + T \log^2(1 + |\mathbf{R}_m|^2) + 2L_i|\mathbf{R}_m|^2
\]
where \( |\mathbf{R}_m|^2 \) is a matrix with the modulus squared entries of \( \mathbf{R}_m \) defined in (6), \( \mathbf{L}_1 \) is an \( L \)-dimensional column vector with ones, \( L_i(x) \) is the dilogarithm function
\[
L_i(x) = -\int_0^x \frac{\log(1-t)}{t} dt = \sum_{k=1}^{\infty} \frac{x^k}{k^2}
\]
and where matrix functions are applied entry-wise.

Proof: In order to analyze the asymptotic behavior of the channel dispersion, we recall the expression in (12). The last term can be directly analyzed by using the expansion in (17), so we only need to evaluate the asymptotic behavior of \( \mathcal{V}[c(\rho | h|^2)] \) and the matrix \( \mathbf{V}_c(m) \) when \( \rho \to \infty \). Regarding the variance \( \mathcal{V}[c(\rho | h|^2)] \) term, using the fact that \( \gamma_3((1, 1, 1), (2, 2, 2); -\rho^{-1}) = 1 + o(1) \) together with the expansion in (17), we have
\[
\mathbb{V}[c(\rho | h|^2)] = \frac{\pi^2}{6} - \frac{(\log(\rho) - \gamma + 1)^2}{\rho} + \frac{1}{\rho} \left( \frac{\pi^2}{6} - 1 \right) + o\left(\frac{1}{\rho}\right). \tag{19}
\]
For the covariance matrix \( \mathbf{V}_c(m) \) the expansion follows directly from the following result, which is proven in Appendix B.

Lemma 1: Assume that \( h_1, h_2 \) are jointly complex circularly symmetric Gaussian distributed with standardized marginals and correlation coefficient \( \phi = |\mathbb{E}[h_1^* h_2]| \). Then, as \( \rho \to \infty \),
\[
\mathbb{V}[c(\rho | h_1|^2), c(\rho | h_2|^2)] = L_i(\phi^2) + 2\frac{\log(\rho) - \gamma + 1}{\rho} \log(1 - \phi^2) + \frac{1}{\rho} \left( \log^2(1 - \phi^2) + 2L_i(\phi^2) \right) + o\left(\frac{1}{\rho}\right)
\]
where \( L_i(x) \) is the dilogarithm function defined above.

Inserting all these expansions into the dispersion expression of (12), we obtain the result.

Interestingly enough, the above asymptotic approximation depends on the fading dynamics only through three specific parameters \( \mu_1, \mu_2 \) and \( \mu_3 \) that can be trivially computed from the average dilogarithm and complementary logarithm of the fading correlation coefficients in the time-frequency grid. These quantities can be easily measured at the receiver in the long run and fed back to the transmitter for system optimization. Since all these parameters depend on the channel correlation, they do not need to be updated very often. Now, if we denote by \( R(\rho) \) the actual code rate in a practical system, we may asymptotically approximate the error probability of the optimum decoder by disregarding the error terms in (9) as
\[
P_e = Q \left( \sqrt{ML} \frac{C(\rho) - R(\rho)}{\mathcal{V}_L(\rho)} \right). \tag{20}
\]
We can use the large SNR analysis above to study the asymptotic behavior of the argument of the \( Q \) function. Indeed, by using the expansion in Proposition 1 for \( \mathcal{V}_L(\rho) \) and the corresponding approximation of \( C(\rho) \) in (18) we obtain the high SNR approximation in (21), shown at the bottom of this page, which contains all the error terms up to the order \( O(\rho^{-1}) \). It will be shown in the next section that the approximation in (21) provides very accurate evaluations of the original quantity for a relevant range of block error rates in typical URLLC requirements. This is because URLLC services typically require very high reliability (in the order of \( 10^{-5} \)), which in the absence of additional diversity (e.g. produced by multi-antenna transceivers) can only be achieved at significantly high SNR.

IV. NUMERICAL VALIDATION

The objective is to analyze the latency-reliability tradeoff of the transmission scheme that has been depicted in Figure 1. In particular, we study the possibility of reducing the BLER by transmitting parts of the codeword in OFDM symbols that are located far apart in the time domain, which therefore go through more uncorrelated channels. This leads to increased diversity, which is of course achieved at the expense of end-to-end latency. Indeed, separating the codeword transmission into widely spaced OFDM symbols improves performance because of the higher channel diversity, but increases latency because of the longer transmission time. More specifically, we consider a transmission of an \( ML \)-dimensional codeword transmitted in \( L \) OFDM symbols, equispaced in the time domain but transmitted every \( T \) OFDM symbols (see further Figure 1). Therefore, the \( ML \)-dimensional codeword is transmitted in a total of \((L-1)T+1\) OFDM symbols. The channel fading
correlation was fixed as \( \{ R_m \}_{l,l'} = \rho_f^l \rho_t^{l-l'} \) where \( \rho_f \) and \( \rho_t \) represent the correlation coefficients in the time and frequency domains. In this numerical study, we fixed the correlation coefficient in the frequency domain to \( \rho_f = 0.9 \), whereas the correlation coefficient in the time domain was fixed to \( \rho_t = (0.9)^T \) where here again \( T \) is the number of OFDM symbols between two consecutive transmissions as in Figure 1.

In order to investigate the performance prediction capabilities of the derived formulas, we considered first a simple scenario where the transmitted codeword occupied \( M = 128 \) subcarriers and \( L = 2 \) consecutive OFDM symbols (so that \( T = 1 \) in this particular example). Figure 2 represents the block error rate (BLER) performance of the LDPC code proposed for the 5G New Radio standard (graph 1) [25]. Both simulated BLER (solid lines) and performance bounds (semi-dotted lines) are shown in the figure, and results are given for AWGN and Rayleigh fading channels. Three different modulation and coding rates are considered, namely \( R = 1 \) bps/Hz with 4-QAM (blue lines), \( R = 2 \) bps/Hz with 16-QAM (red lines) and \( R = 3 \) bps/Hz with 64-QAM (green lines). These modulations were selected from the set of square QAM constellations as those that provided the lowest BLER in the SNR region of interest. It can be seen that in all cases the performance of the real codes is around \( 1 - 2 \) dB worse than the best performance as predicted by the Gaussian approximation, both for AGWN and fading channels. It can be concluded from these figures that the performance gap for the Rayleigh fading case is similar to the performance gap observed in other well studied channels, such as AWGN.

In order to investigate the tradeoff between transmission latency and coding performance, we considered next a situation where the total number of used subcarriers was fixed to \( M = 512 \) and the number of OFDM symbols was fixed to \( L = 3 \) (corresponding to a 1536-dimensional codeword).

The transmission rate (in number of information bits per channel access) was fixed according to the signal to noise ratio as 90% of the total achievable capacity, that is \( R(\rho) = 0.9C(\rho) \).

Performance was measured in terms of the approximate BLER bound as given by (20), together with its corresponding high SNR approximation in (21).

Figure 3 represents the BLER as a function of the signal to noise ratio (SNR) for different values of the number \( T \) of OFDM symbols between consecutive transmissions. Apart from the performance for the exact \( C(\rho) \) and \( V_L(\rho) \), we also represent the asymptotic approximation as given in (21) in dotted line. Observe that increasing the separation between consecutive OFDM symbols always increases performance, basically because the \( L \) different parts of the codeword go through a more uncorrelated channel and are therefore received with a higher degree of diversity. It should also be pointed out that, for conventional block error rates around \( 10^{-5} \), the high SNR approximations are very good approximations of the actual predicted BLER. This is important, since the high SNR approximation can be very easily evaluated, whereas the original one is much more complex and usually requires evaluating a high number of terms in the series of (14), especially when the correlation factor is high.

In Figure 4 we represent the BLER as a function of the number \( T \) of OFDM separation symbols\(^3\) for different values of the SNR. Given a value of the SNR, this figure provides us with the minimum latency (in number \( T \) of OFDM separation symbols) that is required in order to achieve a certain reliability (BLER). It therefore illustrates the fundamental tradeoff between latency and reliability in a practical stationary coherent channel. For example, if the SNR is fixed to 25dB, the minimum number of OFDM separation symbols to achieve a reliability of \( 10^{-5} \) is \( T = 2 \) (which correspond to a total latency of 5 OFDM symbols), whereas if the SNR drops to 20dB, we can only guarantee this performance by

\(^3\)Note that since both \( M \) and \( L \) are kept constant, the total codeword length is constant in this example.
increasing the symbol separation to \( T = 8 \) (corresponding to 17 OFDM symbols). The high SNR approximation provides a very accurate prediction of the original bound, especially when the SNR is above 20dB.

\section{Conclusion}

The performance of short block-length codes in the presence of stationary multi-variate Rayleigh fading has been studied in terms of its capacity and asymptotic channel dispersion. The channel model has been inspired by multi-symbol OFDM transmissions for URLLC, where transmissions are typically concentrated in the frequency domain, occupying a large number of subcarriers in order to minimize latency, and span different OFDM symbols that can be separated in the time domain in order to provide diversity. A novel high SNR approximation of these quantities has been derived, and results provide a very simple design rule that depends on different functions of the fading correlation in the time-frequency axis. The usefulness of these results has been illustrated in the context of time-frequency transmit strategies for this type of channel, where it has been shown that the optimum transmission configuration can be chosen through a very simple feedback based on long-term fading statistics. As a final remark, we should point out that, even if the paper has focused on single antenna channels, the results presented here could be generalized to include other relevant transceiver architectures for URLLC services, such as multi-antenna (MIMO) techniques. Furthermore, the assumption that channel state information is available at the receiver is quite restrictive in practical terms, since acquiring this information typically requires the insertion of pilot symbols that penalize the transmission rate. There are several recent contributions in the literature trying to characterize this for block fading channels [12], [26] and it would be quite interesting to generalize the analysis in this paper to the case of unknown CSI. All these aspects clearly deserve specific attention and are left for future research.

\section{Appendix A}

\textbf{Proof of Theorem 1}

The proof is an extension of the corresponding ones in [11] and [15]; therefore, only the main differences are highlighted here. We divide the exposition into two subsections, corresponding to the achievability and converse parts of the theorem respectively.

The development strongly relies on concepts from binary hypothesis tests that are summarized here. Given two probability distributions \( P, Q \), consider the set of binary hypothesis tests designed in order to distinguish these two distributions. We will denote by \( \beta_\alpha (P, Q) \) the minimum achievable probability of false alarm under the null hypothesis \( (Q) \) when the probability of correct detection under the alternative one \( (P) \) is higher than or equal to \( \alpha \). Here, \( P \) will be always fixed to be the joint distribution of output and channel, given the input \( X \), which will be denoted as \( P = P_{YH|X=x} \). Therefore, with some abuse of notation we can write \( \beta_\alpha (X, Q) = \beta_\alpha (P_{YH|X=x}, Q) \). In some occasions, we may want to consider hypothesis tests between \( Q \) and a collection of alternatives \( \mathcal{P}_X \) indexed by \( X \in F \), with \( F \) a given channel input subset. In this case, we define \( \kappa_{\tau} (F, Q) \) as the minimum achievable probability of false alarm under the null hypothesis \( (Q) \) when the infimum (over all \( X \in F \)) of the probability of correct detection under the alternative \( P_X \) is higher than or equal to \( \tau \).

\subsection{Achievability}

The achievability is established according to the \( \kappa/\beta \) bound [1, Theorem 25], which showed that, given any output probability distribution \( Q \) and any subset of the input space \( F \), for each \( 0 < \tau < \epsilon < 1 \), there exists an \((N_{cw}, \epsilon)\) code with codewords chosen from \( F \) that satisfies

\[
N_{cw} \geq \sup_{0 < \tau < \epsilon} \sup_{X \in F} \frac{\kappa_\tau (F, Q)}{\beta_{1-\epsilon+\tau} (X, Q)}.
\]

As mentioned above, the alternative distribution is always fixed to \( P = P_{YH|X=x} \), whereas \( Q \) is fixed in our case to the joint distribution of channel and observation when the inputs are standardized complex Gaussian distributed, denoted here by \( Q = Q_{YH} \). In order to prove the result, we will first find an upper bound on \( \beta_{1-\epsilon+\tau} (X, Q_{YH}) \) and then a lower bound on \( \kappa_\tau (F, Q_{YH}) \). In this process, we will have to work with the Radon–Nikodym derivative of the distribution \( P_{YH|X=x} \) with respect to \( Q_{YH} \), which is usually referred to as the information density and can be expressed as

\[
i_{N} (X) = \log \frac{dP_{YH|X=x}}{dQ_{YH}} = \sum_{m=1}^{M} i_m (x^{(m)}). \tag{22}
\]

where

\[
i_m (x^{(m)}) = \sum_{l=1}^{L} \log \left( 1 + \frac{1}{\rho |h_{m,l}|^2} \right)
+ \frac{|h_{m,l}|^2 |x_{m,l}|^2 + 2 \Re (h_{m,l}x_{m,l}w_{m,l}^*) - \rho |w_{m,l}|^2 |h_{m,l}|^2}{P|h_{m,l}|^2 + \sigma^2} \tag{23}
\]
and where we have denoted $y_{m,l}$ (respectively $h_{m,l}$, $x_{m,l}$ and $w_{m,l}$) as the $(m,l)$th entry of the matrix $Y$ (respectively $H$, $X$, $W$). In particular, we begin by investigating the first three absolute moments of $i_N(X)$ conditioned on the input $X$, which will be necessary in order to bound $\beta_{1-\epsilon^\tau}(X, Q_{YH})$.

1) Moments of the Information Density: Observe that the stationarity of the channel implies that, for any measurable frequency index $i$, we are not able to eliminate the dependence among channel entries.

Observe that, contrary to what happens with the expectation in (2), where $\varphi_{m,l}(X)$ is defined in (23) and where we recall that $X(m)$ denotes the $m$th row of $X$. Let us denote $[\cdot]^\circ = [\cdot] - E[\cdot]$ and observe that we can bound, by Jensen’s inequality and the fact that $|\text{Re}x| \leq |x|$, 

$$
\mathbb{E} \left| i_m(X(m)) \right|^3 \leq 9L^2 \sum_{l=1}^{L} \left( \tau_{m,l}^{(1)} + \tau_{m,l}^{(2)} |x_{m,l}|^3 + \tau_{m,l}^{(3)} |x_{m,l}|^6 \right)
$$

where

$$
\tau_{m,l}^{(1)} = E \left( \log \left( 1 + \rho |h_{m,l}|^2 \right) - \rho \frac{|w_{m,l}|^2 |h_{m,l}|^2}{P |h_{m,l}|^2 + \sigma^2} \right)^3
$$

$$
\tau_{m,l}^{(2)} = 2E \left( \frac{|h_{m,l}|^2}{P |h_{m,l}|^2 + \sigma^2} \right)^3
$$

$$
\tau_{m,l}^{(3)} = E \left( \frac{|h_{m,l}|^2}{P |h_{m,l}|^2 + \sigma^2} \right)^6
$$

These three quantities are trivially upper bounded by quantities independent of $m$. Therefore, we can generally write

$$
\max_{1 \leq m \leq M} \mathbb{E} \left[ i_m(X(m)) - E_i_m(X(m)) \right]^3 \leq K_1 + K_2 \max_{m,l} |x_{m,l}|^3 + K_3 \max_{m,l} |x_{m,l}|^6
$$

for three positive constants $K_1$, $K_2$, $K_3$ independent of $m, l$.

2) Upper Bound on $\beta_{1-\epsilon^\tau}(X, Q_{YH})$: The upper bounds in [1, Theorem 25] can be modified to our setting by simply noting that the conventional Berry-esseen inequality is not applicable here because the different terms in the sum of (22) are not independent. However, we can still apply some variants of the Berry-Esseen theorem that are still valid under the strong mixing conditions imposed here, see e.g. [27], [28]. In particular, we can use [29, Theorem 10], which is valid under exponentially decaying $\rho$-mixing coefficients and provides a relationship with the Berry-Esseen ratio $B_N(X)$ defined in (25). More specifically, a trivial modification of [1, Lemma 58] leads to

$$
\log \beta_{1-\epsilon^\tau}(X, Q_{YH}) \leq -MLC(\rho) - \sqrt{MLV_N(X)}Q^{-1} \left( 1 - \epsilon + \frac{B_N(X) \log^2 N}{\sqrt{N}} \right)
$$

where we assume here that the argument of $Q^{-1}(\cdot)$ is in the interval $[0, 1]$. Now, we choose $F$ as

$$
F = \{X \in \mathbb{C}^{M \times L} : |x_i|^2 = MP, \quad |\mathcal{V}_N(X) - \mathcal{V}_L(\rho)| \leq \delta_1, \quad \max_{m,l} |x_{m,l}| \leq \delta_2 M^\theta \}
$$

\[\text{Observe that, instead of the original term of order } O(\sqrt{N}) \text{ in the Berry-Esseen theorem, we obtain a term in } O(\sqrt{N} \log^2 N) \text{ due to the dependence among channel entries.}\]
where $V_N(X)$ and $V_L(\rho)$ are defined in (24) and (9) respectively, where $\delta_1, \delta_2, \theta, C$ are positive constants with $\theta < 1/12$. We claim that
\[
\sup_{x \in F} \log \beta_{1-\epsilon+\tau} (X, Q_{YH}) \leq -MLC(\rho)
\]
\[+ \sqrt{ML(V_L(\rho) + \delta_1)Q^{-1}(\epsilon - \tau + O\left(\frac{\log^2 M}{M^{1/2-\theta}}\right))}.
\]
Indeed, observe that if $\|x_l\|^2 = MP$ we can lower bound $V_N(X)$ in (24) by
\[
\inf_{X \in F} V_N(X) \geq 1 - \frac{1}{L} \sum_{l=1}^L \mathbb{E}\left[ d^2(\rho|n|^2) \right] > 0.
\]
Therefore, from (26) we can see that
\[
B_N(X) \leq C_1 + C_2 \max_{m,l} |x_{m,l}|^3 + C_3 \max_{m,l} |x_{m,l}|^6
\]
with $C_1, C_2, C_3$ three positive constants independent of $M$. Hence,
\[
\sup_{F \in F} \frac{\log^2 M}{\sqrt{M}} B_N(X) \leq C_1 \frac{\log^2 M}{\sqrt{M}} + C_2 \delta_1 \frac{\log^2 M}{M^{1/2-\theta}}
\]
\[+ C_3 \delta_2^2 \frac{\log^2 M}{M^{1/2-\theta}} = O\left(\frac{\log^2 M}{M^{1/2-\theta}}\right).
\]

3) Lower Bound on $\kappa_\tau(F, Q_{YH})$: We can use [11, Lemma 4] to state that, for any input distribution $P_X$ we can write
\[
\kappa_\tau(F, Q_{YH}) \geq \beta_{P_X[F]}(P_{YH}|X=F \circ P_X, Q_{YH}).
\]
In particular, we can select $P_X$ to be the $L$-fold product of the uniform distribution on the sphere $\mathbb{S}^{M-1}$. Using the procedure in [11, eq. (57)-(58)] we find that
\[
\kappa_\tau(F, Q_{YH}) \geq \exp\left(-\frac{K}{\tau P_X(F)}\right)
\]
for some constant $K > 0$ independent of $M$.

Lemma 2: Let $F$ be defined as above and assume that $P_X$ is the $L$-fold uniform distribution on the sphere $\mathbb{S}^{M-1}$. When $M \to \infty$, we have
\[
P_X(F) \to 1.
\]
Proof: We only need to see that
\[
P_X(\|V_N(X) - V_L(\rho)\| > \delta_1) \to 0
\]
and
\[
P_X\left(\max_{m,l} |x_{m,l}| \leq CM^\theta\right) \to 0
\]
as $M \to \infty$. Regarding the convergence of $V_N(X)$, observe that (7) provides a sufficient condition for the spectral norm of $\Phi_1, \Phi_2$ to be uniformly bounded in $M$. Indeed, observe that $\Phi_1$ is a block-Toeplitz matrix, and the $(l, l')$ block has spectral norm bounded by
\[
\|\Phi_1\|_{l,l'} \leq \sup_{\omega} \sum_{m=-\infty}^{\infty} \mathbb{E}\left[ d\left(\rho|h_{m,l}|^2\right) d\left(\rho|h_{m,l+l'}|^2\right) \right] e^{-\omega m}
\]
\[
\leq \sum_{m=-\infty}^{\infty} \mathbb{E}\left[ d\left(\rho|h_{m,l}|^2\right) d\left(\rho|h_{m,l+l'}|^2\right) \right] + \sum_{m=-\infty}^{\infty} \mathbb{E}\left[ d\left(\rho|h_{m,l}|^2\right) \right] \leq +\infty
\]
where (a) follows from [30, Lemma 4.1], (b) from the triangular inequality and (d) from the $\rho$-mixing assumption in (7). The same argument shows that $\|\Phi_2\| < +\infty$.

Now, using the fact that for $\delta > 0$
\[
\mathbb{P}\left(\sum_{i=1}^K a_i > \delta \right) \leq \sum_{i=1}^K \mathbb{P}(|a_i| > \delta/K)
\]
we see from the definition of $V_N(X)$ in (24) that it is enough to show that, for any $\delta' > 0$, we have
\[
P_X\left(\frac{1}{ML} \left(\frac{|x_l|^2}{P} - 1_{ML}\right)_H \Phi_1 \left(\frac{|x_l|^2}{P} - 1_{ML}\right) \right) - \frac{1}{ML} \right) \to 0
\]
as well as
\[
P_X\left(\frac{1}{ML} \left(\frac{|x_l|^2}{P} - 1_{ML}\right) \Phi_2 \left(\frac{|x_l|^2}{P} - 1_{ML}\right) \right) \to 0
\]
as $M \to \infty$. Now, let $G$ denote an $M \times L$ matrix with i.i.d. standardized complex Gaussian entries, with columns denoted as $g_1, \ldots, g_L$ and entries $g_{m,l}$. Clearly, $x_l$ has the same distribution as $\mathbb{P}(\xi_{l,l'}) = \mathbb{E}(r_{l,l'})\xi_{l,l'}$. To see (30), we define $\xi_l = \|g_l\|^2/M, r_l = g_l \odot g^*_l$ and use the fact that this probability can be upper bounded by (29), shown in (32) at the bottom of the next page, where $\Phi_1^{l,l'}$ is the $M \times M$ matrix corresponding to the $(l, l')$th block of $\Phi_1$.

Let us first deal with the terms in the above sum such that $l \neq l'$. Using the Markov inequality together with the fact that $\mathbb{E}(r_{l,l'}) = 1_M$ and the identity
\[
\mathbb{E}\left[ (r_{l,l'} - 1) (r_{l,l'})^H \right] = 1_M 1_M^H + \frac{M}{M+1} P_{l,M}^d, \delta_l - l'
\]
which is clearly of order $O(M^{-1})$. Regarding the terms for which $l = l'$, we can use the triangular inequality to write
\[
\frac{1}{M} \left| \left( \frac{r_{l}}{\xi_{l}} - 1 \right) M \phi_{l}^{i} \left( \frac{r_{l}}{\xi_{l}} - 1 \right) - \text{tr} \phi_{l}^{i} \right| \\
\leq \frac{1}{M} \left| \left( r_{l} - 1 \right) M \phi_{l}^{i} \left( r_{l} - 1 \right) - \text{tr} \phi_{l}^{i} \right| \\
+ \left| 1 - \frac{1}{\xi_{l}} \right| \frac{r_{l}}{\xi_{l}} M \phi_{l}^{i} r_{l} + 2 \left| 1 - \xi_{l} \right| M \phi_{l}^{i} \xi_{l} \right|.
\tag{33}
\]

Now, observe that
\[
E \left[ \frac{r_{l}}{\xi_{l}} M \phi_{l}^{i} r_{l} \right] = \frac{1}{M+1} \text{tr} \phi_{l}^{i} + \frac{1}{M+1} M \text{tr} \phi_{l}^{i} M
\leq \frac{2M+1}{M+1} \left\| \phi_{l}^{i} \right\|
\leq \left\| \phi_{l}^{i} \right\|
\]
from where it follows that the quantities inside the expectation form tight sequences in $M$. Noting, however, that $\left| 1 - \xi_{l} \right| \rightarrow 0$ and $\left| 1 - \frac{1}{\xi_{l}} \right| \rightarrow 0$ in probability (law of large numbers), we see that the last two terms in (33) converge to zero in probability.

For the first term, we have
\[
E \left( \frac{1}{M^2} \right) \left[ \left( r_{l} - 1 \right) M \phi_{l}^{i} \left( r_{l} - 1 \right) - \text{tr} \phi_{l}^{i} \right]^{2}
\leq \frac{8}{M^2} E \left[ \left( r_{l} - 1 \right) M \phi_{l}^{i} \text{diag} \left( r_{l} \right) \phi_{l}^{i} \left( r_{l} - 1 \right) \right]
= \frac{16}{M^2} \text{tr} \left[ \phi_{l}^{i} \otimes \phi_{l}^{i} \right]
+ \frac{8}{M^2} \text{tr} \left[ \left( \phi_{l}^{i} \right)^2 \right] = O \left( M^{-1} \right).
\]

where (a) follows from the Poincaré-Nash inequality on functionals of Gaussian random variables [32] and (b) follows from the identity
\[
E \left[ r_{l}^{2} \phi_{l}^{i} \text{diag} \left( r_{l} \right) \phi_{l}^{i} r_{l} \right]
= \frac{1}{M} \left( \phi_{l}^{i} \right)^2 1_{M}
+ 2 \text{tr} \left[ \phi_{l}^{i} \otimes \phi_{l}^{i} \right] + \text{tr} \left( \phi_{l}^{i} \right)^2
+ 2 \text{tr} \left[ \phi_{l}^{i} \text{diag} \left( \phi_{l}^{i} 1_{M} \right) \right].
\]

Thus, the Markov inequality shows that the first term in (33) converges to zero in probability, concluding the proof of (30).

The proof of (31) follows the same reasoning and is therefore omitted. Together, (30) and (31) imply (27).

In order to prove (28), observe that for any $\delta > 0$ we have [15, Proof of Lemma 18]
\[
P_{X} \left( \text{max}_{m,l} x_{m,l} \leq CM^{\delta} \right)
= P_{G} \left( \text{max}_{m,l} g_{m,l} \leq C \xi_{l} M^{\delta} \right)
\geq P_{G} \left( \text{max}_{m,l} g_{m,l} \leq \frac{C}{2} M^{\delta} \right) - P_{X} \left( \xi_{l} \leq 1/2 \right)
= \left( P_{G} \left( |g_{m,l}| \leq \frac{C}{2} M^{\delta} \right) \right)^{ML} + o(1)
\geq \left( 1 - \frac{2}{CM^{q}} \right)^{2k} E \left[ |g_{m,l}^{2k}| \right] + o(1)
\]
for any $k \in \mathbb{N}$. Noting that $E \left[ |g_{m,l}^{2k}| \right] = k!$ and choosing $k > 1/(2q) > 0$ we see that the right hand side converges to 0 as $M \rightarrow \infty$.

4) Finalizing the Proof of Achievability: Combining the two bounds above, we have just proven that, for each $0 < \tau < \epsilon < 1$ and $\delta_{1} > 0$, there exists an $(N_{cw}, \epsilon)$ code such that
\[
\log N_{cw} \geq -\frac{1}{\tau} + MLC \left( \rho \right) + \sqrt{ML \left( \frac{1}{\tau} + 1 \right) Q^{-1} \left( \epsilon - \tau + o(1) \right)} + O \left( 1 \right).
\]

Therefore, by letting $M \rightarrow \infty$ we see that
\[
\frac{\log N_{cw}}{N} \geq C \left( \rho \right) - \sqrt{\frac{M}{ML} Q^{-1} \left( \epsilon - \tau \right) + o \left( \frac{1}{\sqrt{M}} \right)}.
\]
and since $\tau, \delta_{1}$ can be made arbitrarily low, the proof is complete.

B. Converse

The converse can be established along the lines of [15, Theorem 19]. Consider a specific $(N_{cw}, \epsilon)$ code and let $C$ denote the corresponding codebook. We can assume that the codewords are such that $\|x_{i}\|^{2} = MP$ (otherwise, we can simply add one row to the codeword matrix $X$ so that this constraint is satisfied without affecting the asymptotics of the maximum achievable rate, up to the order that we are considering here [1, Lemma 39]). The maximum number of codewords is thereby bounded by using the meta-converse theorem [1, Theorem 31], that establishes that such an $(N_{cw}, \epsilon)$ code must satisfy
\[
\frac{1}{N_{cw}} \geq \inf_{X \in C} \beta_{1-\epsilon} \left( X, Q \right)
\]

\[
P_{X} \left( \left| \frac{1}{ML} \sum_{l,l'} \left( \frac{r_{l}}{\xi_{l}} - 1 \right) M \phi_{l}^{i} \left( \frac{r_{l}}{\xi_{l}} - 1 \right) - \frac{1}{ML \text{tr} \phi_{l}} \delta_{l,l'} \right| > \delta' \right)
\leq \sum_{l,l'} P_{X} \left( \left| \frac{1}{ML} \left( \frac{r_{l}}{\xi_{l}} - 1 \right) M \phi_{l}^{i} \left( \frac{r_{l}}{\xi_{l}} - 1 \right) - \frac{1}{ML \text{tr} \phi_{l}^{i} \delta_{l,l'}} \right| > \frac{\delta'}{L} \right)
\tag{32}
\]
for any distribution $Q$ on the channel output space. As in the achievability part, we can choose $Q$ as the joint distribution of channel and observation when the inputs are standardized complex Gaussian distributed, namely $Q = Q_{YH}$ according to the notation above. Using the inequality equivalent to [4, (2.87)] after replacing the Berry-Esseen theorem by [29, Theorem 10], we obtain (for any $\Delta > 0$)

$$\log \beta_1 - (X, Q_{YH}) \geq -NC(\rho) + \frac{1}{\sqrt{N}V_N(X)}Q^{-1}(1 - \epsilon - (B_\rho(X) + \Delta) \log^2 N \sqrt{N})$$

where we have chosen $\Delta = (LM)^{1/4}$. Now, the problem with the above bound is the fact that we have no control on $\inf_{X \in C} V_N(X)$ and there might be some codewords in $C$ for which the channel dispersion is abnormally low. This is technically solved in [15, Theorem 19], by dividing the codebook into two disjoint sets, namely $C = C_u \cup C_1$ where $C_u = \{X \in C : \forall N(X) > V_L(\rho) - \delta\}$ for some $\delta > 0$, and where $C_1$ contains the rest of the codewords. The above bound is obviously useful to bound $\log |C_u|$, whereas for $C_1$ one can follow the strategy in [15, Theorem 19].

**APPENDIX B**

**PROOF OF LEMMA 1**

The main idea behind the proof is to deal with the different terms of the series in (14) when $\rho \to \infty$. In order to simplify the expressions below, define $\alpha = (1 - \phi^2)^{-1}$ and observe that we can express (from Corollary 1)

$$\mathbb{E} \left[ c(\rho | h_1|^2) c(\rho | h_2|^2) \right] = \exp \left( \frac{2\alpha}{\rho} \right) \sum_{i=0}^{\infty} (2\rho)^{2i} \times \left( \frac{\Gamma(r - k, \alpha/\rho)}{(k!)(i - r)! \alpha^{r+1} (r - k)!} \right)^2. $$

The key observation lies on the asymptotic behavior of the incomplete gamma function as $\rho \to \infty$. Indeed, the terms with $\Gamma(0, \alpha/\rho) = E_1(\alpha/\rho)$ will scale up as $\log \rho$ -see (17)- whereas for $k > 1$, $\Gamma(k, \alpha/\rho)$ will converge to a constant $\Gamma(k)$. By separating $\Gamma(0, \alpha/\rho)$ from the rest of $\Gamma(k, \alpha/\rho)$ with $k > 0$, we can write the cross-covariance as the sum of three terms, namely $\mathbb{V}[c(\rho | h_1|^2), c(\rho | h_2|^2)] = \vartheta_1(\rho) + \vartheta_2(\rho) + \vartheta_3(\rho)$ where

$$\vartheta_1(\rho) = \frac{1}{\alpha} C^2 \left( \frac{\rho}{\alpha} \right) \sum_{i=0}^{\infty} \phi^{2i} A_i^2 \left( \frac{\alpha}{\rho} \right) - C^2 (\rho)$$

$$\vartheta_2(\rho) = \frac{2}{\alpha} \exp \left( \frac{\alpha}{\rho} \right) C \left( \frac{\rho}{\alpha} \right) \sum_{i=1}^{\infty} \phi^{2i} A_i \left( \frac{\alpha}{\rho} \right) B_i \left( \frac{\alpha}{\rho} \right)$$

$$\vartheta_3(\rho) = \frac{1}{\alpha} \exp \left( \frac{2\alpha}{\rho} \right) \sum_{i=1}^{\infty} \phi^{2i} B_i^2 \left( \frac{\alpha}{\rho} \right)$$

where $C(\rho)$ is defined in (11) and where we have introduced the quantities

$$A_i(x) = \sum_{r=0}^{i} \frac{1}{r!} (-x)^r$$

$$B_i(x) = \sum_{\ell=1}^{i} \left( \frac{-x}{\ell!} \right) \sum_{k=1}^{\ell} \Gamma(k, x).$$

We can analyze the asymptotic behavior of these three terms separately, up to terms $O(\rho^{-1})$.

1) **Analysis of $\vartheta_1$:** Observe that we can write

$$\sum_{i=0}^{\infty} \phi^{2i} A_i^2 \left( \frac{\alpha}{\rho} \right) = 1 + (\alpha - 1) \left( 1 - \frac{\alpha}{\rho} \right)^2 + \xi_{11}(\rho) + \xi_{12}(\rho)$$

where we have defined the two error terms

$$\xi_{11}(\rho) = 2 \left( 1 - \frac{\alpha}{\rho} \right) \sum_{i=2}^{\infty} \phi^{2i} \sum_{r=2}^{i} \frac{1}{r!} \left( \frac{\alpha}{\rho} \right)^r$$

$$\xi_{12}(\rho) = \sum_{i=2}^{\infty} \phi^{2i} \left( \sum_{r=2}^{i} \frac{1}{r!} \left( \frac{\alpha}{\rho} \right)^r \right)^2.$$ 

Next, observe that we can bound these two error terms as

$$|\xi_{11}(\rho)| \leq 2 \left( 1 - \frac{\alpha}{\rho} \right) \frac{\phi^2}{1 - \phi^2} \exp \left( \frac{\alpha}{\rho} \right)$$

$$= O \left( \frac{1}{\rho^2} \right)$$

$$|\xi_{12}(\rho)| \leq \frac{\phi^2}{1 - \phi^2} \left( \frac{\alpha}{\rho} \right)^2 \exp \left( \frac{\alpha}{\rho} \right) = O \left( \frac{1}{\rho^4} \right).$$

Therefore, using the asymptotic approximation of $E_1(1/\rho)$ in (17) we can conclude that

$$\vartheta_1(\rho) = \left( 1 + \frac{2}{\rho} \right) \log^2 \alpha - \frac{2\alpha \log \alpha}{\rho} + 2 \log(\rho - 1) \left( \frac{\alpha - 1 - 2 \log \alpha}{\rho} \right)$$

$$+ o \left( \frac{1}{\rho^2} \right).$$

2) **Analysis of $\vartheta_2$:** We first introduce a result that will be useful in order to analyze this term.

**Lemma 3:** With the above definitions, one can write

$$\sum_{i=1}^{\infty} \phi^{2i} \Gamma(i, \alpha/\rho) = E_1 \left( \frac{1}{\rho} \right) - E_1 \left( \frac{\alpha}{\rho} \right).$$
Proof: We can write $\Gamma \left( i, \frac{\alpha}{\rho} \right) = \Gamma (i) - \gamma \left( i, \frac{\alpha}{\rho} \right)$ where

$$\gamma \left( i, \frac{\alpha}{\rho} \right) = \int_0^{\frac{\alpha}{\rho}} t^{i-1} e^{-t} dt$$

is the lower incomplete gamma function. Clearly,

$$\sum_{i=1}^{\infty} \phi^{2i} \frac{\Gamma (i)}{i!} = \sum_{i=1}^{\infty} \phi^{2i} \frac{\Gamma (i)}{i!} = \sum_{i=1}^{\infty} \phi^{2i} x^{i-1} dx = \log \alpha.$$

On the other hand, using the definition in (34) and the integral in [22, 2.484]

$$\sum_{i=1}^{\infty} \phi^{2i} \frac{\Gamma (i, \frac{\alpha}{\rho})}{i!} = \phi^2 \int_0^{\frac{\alpha}{\rho}} \frac{1}{t} \sum_{i=1}^{\infty} \phi^2 t^{i-1} e^{-t} dt$$

$$= 2 \int_0^{\frac{\alpha}{\rho}} \frac{1}{t} \sinh \left( \frac{\phi^2}{2} t \right) e^{-\left( \frac{\phi^2}{2} t \right)} dt$$

$$= E_1 \left( \frac{\alpha}{\rho} \right) - E_1 \left( \frac{1}{\rho} \right) + \log \alpha$$

and the result follows.

In this case, we are able to write

$$\sum_{i=1}^{\infty} \phi^{2i} A_i \left( \frac{\alpha}{\rho} \right) B_i \left( \frac{\alpha}{\rho} \right)$$

$$= \left( 1 - \frac{\alpha}{\rho} \right) \sum_{i=1}^{\infty} \phi^{2i}$$

$$\times \left[ \left( 1 - \frac{\alpha}{\rho} \right) \sum_{k=1}^{i} \frac{\Gamma \left( k, \frac{\alpha}{\rho} \right)}{k!} + \frac{\alpha}{\rho} \frac{\Gamma (i, \frac{\alpha}{\rho})}{i!} \right]$$

$$+ \xi_{21} \left( \rho \right) + \xi_{22} \left( \rho \right) + \xi_{23} \left( \rho \right)$$

where we have defined the error terms

$$\xi_{21} \left( \rho \right) = \frac{\phi^2}{2} \left( \frac{\alpha}{\rho} \right)^2 \left[ \left( 1 - \frac{\alpha}{\rho} \right) \Gamma \left( 1, \frac{\alpha}{\rho} \right) + \frac{\Gamma \left( 2, \frac{\alpha}{\rho} \right)}{2} \right]$$

$$\xi_{22} \left( \rho \right) = \sum_{i=3}^{\infty} \phi^{2i} B_i \left( \frac{\alpha}{\rho} \right) \sum_{t=2}^{i} \frac{1}{t!} \left( -\frac{\alpha}{\rho} \right)^t$$

$$\xi_{23} \left( \rho \right) = \left( 1 - \frac{\alpha}{\rho} \right) \sum_{i=3}^{\infty} \phi^{2i} \sum_{r=2}^{i-1} \frac{1}{r!} \left( -\frac{\alpha}{\rho} \right)^r \sum_{k=1}^{i-r} \frac{\Gamma \left( k, \frac{\alpha}{\rho} \right)}{k!}.$$

Obviously, $\xi_{21} \left( \rho \right) = O \left( \rho^{-2} \right)$. Regarding $\xi_{22} \left( \rho \right)$, we use the loose bound

$$\left| B_i \left( \frac{\alpha}{\rho} \right) \right| \leq \sum_{r=0}^{i-1} \frac{1}{r!} \left( \frac{\alpha}{\rho} \right)^r \sum_{k=1}^{i-r} \frac{\Gamma \left( k, \frac{\alpha}{\rho} \right)}{k!}$$

$$\leq i \sum_{r=0}^{i-1} \frac{1}{r!} \left( \frac{\alpha}{\rho} \right)^r \leq i^{i-1} e^{\frac{\alpha}{\rho}}.$$

Thus, using Lemma 3 we find that

$$\left| \xi_{22} \left( \rho \right) \right| \leq \left( \frac{\alpha}{\rho} \right)^2 e^{\frac{\alpha}{\rho}} \sum_{j=1}^{\infty} (j + 2) \phi^{2j} \sum_{t=0}^{j} \frac{1}{(t + 2)!} \left( \frac{\alpha}{\rho} \right)^t$$

$$\leq \left( \frac{\alpha}{\rho} \right)^2 e^{\frac{\alpha}{\rho}} \phi^2 \left[ \frac{d}{dx} x^3 \right]_{x=\phi^2}$$

On the other hand, following a similar approach

$$\left| \xi_{23} \left( \rho \right) \right| \leq \left( 1 - \frac{\alpha}{\rho} \right) \left( \frac{\alpha}{\rho} \right)^2 \phi^2 \frac{1}{\phi^2} \left[ \frac{d}{dx} x^3 \right]_{x=\phi^2}$$

and therefore both are of order $O \left( \rho^{-2} \right)$.

A direct use of Lemma 3 and the expansion in (17) show that

$$\sum_{i=1}^{\infty} \phi^{2i} A_i \left( \frac{\alpha}{\rho} \right) B_i \left( \frac{\alpha}{\rho} \right)$$

$$= \alpha \log \alpha + \alpha \frac{1}{\rho} \left( 2 \log \rho - \gamma \right) \left( \log \alpha + \frac{\log \alpha + 1 - \alpha}{\rho} \right) + O \left( \rho^{-2} \right).$$

Finally, inserting this into the expression of $\vartheta_2 \left( \rho \right)$ we can conclude that

$$\vartheta_2 \left( \rho \right) = 2 \left( \log \left( \rho \right) - \gamma \right) \left( \log \alpha + \frac{\log \alpha + 1 - \alpha}{\rho} \right) +$$

$$- 2 \left( \log \alpha \right)^2 + 2 \log \alpha \frac{1}{\rho} \left( 2 \alpha - 1 - \log \alpha \right)$$

$$+ o \left( \rho^{-1} \right).$$

3) Analysis of $\vartheta_3$: We note first that we can express

$$\sum_{i=1}^{\infty} \phi^{2i} \left( \frac{\alpha}{\rho} \right)^2$$

$$= \phi^2 \Gamma^2 \left( 1, \frac{\alpha}{\rho} \right)$$

$$+ \sum_{i=3}^{\infty} \phi^{2i} \left( \frac{1}{i!} \right)^2 \left( \frac{\alpha}{\rho} \right)^2$$

$$+ \xi_{31} \left( \rho \right) + \xi_{32} \left( \rho \right)$$

where we have the two error terms

$$\xi_{31} \left( \rho \right) = \sum_{i=3}^{\infty} \phi^{2i} \left( \sum_{r=2}^{i-1} \frac{1}{r!} \left( -\frac{\alpha}{\rho} \right)^r \sum_{k=1}^{i-r} \frac{\Gamma \left( k, \frac{\alpha}{\rho} \right)}{k!} \right)^2$$

$$\xi_{32} \left( \rho \right) = 2 \sum_{i=3}^{\infty} \phi^{2i} \left( \sum_{r=2}^{i-1} \frac{1}{r!} \left( -\frac{\alpha}{\rho} \right)^r \sum_{k=1}^{i-r} \frac{\Gamma \left( k, \frac{\alpha}{\rho} \right)}{k!} \right)$$

$$+ \xi_{31} \left( \rho \right) + \xi_{32} \left( \rho \right)$$

We can trivially bound

$$\left| \xi_{31} \left( \rho \right) \right| \leq \left( \frac{\alpha}{\rho} \right)^4 e^{\frac{\alpha}{\rho}} \sum_{i=3}^{\infty} \phi^{2i}$$

$$\left| \xi_{32} \left( \rho \right) \right| \leq 2 \left( \frac{\alpha}{\rho} \right)^2 e^{\frac{\alpha}{\rho}} \sum_{i=3}^{\infty} \phi^{2i}$$

so the error is of order $O \left( \rho^{-2} \right)$. 
Lemma 4: If \( H_i \) is the \( i \)th harmonic number and for \( 0 < \phi^2 < 1 \), we have
\[
\sum_{i=1}^{\infty} \phi^{2i} H_i = \frac{-\log (1 - \phi^2)}{(1 - \phi^2)}
\]
\[
\sum_{i=1}^{\infty} \frac{\phi^{2i}}{i} H_i = \operatorname{Li}_2 (\phi^2) + \frac{1}{2} \log^2 (1 - \phi^2)
\]
\[
\sum_{i=1}^{\infty} \phi^{2i} H_i^2 = \frac{\operatorname{Li}_2 (\phi^2) + \log^2 (1 - \phi^2)}{1 - \phi^2}
\]
where \( \operatorname{Li}_2 (x) \) is the dilogarithm function.

Proof: To prove the first identity, we can simply use the integral representation of the Harmonic number, namely
\[
H_n = \int_0^1 \frac{1 - t^n}{1 - t} \, dt
\]
which leads to
\[
\sum_{i=1}^{\infty} \phi^{2i} H_i = \int_0^1 \frac{\phi^2}{(1 - \phi^2)(1 - t\phi^2)} \, dt = \frac{-\log (1 - \phi^2)}{(1 - \phi^2)}.
\]
Regarding the second identity, observe that we can write
\[
\sum_{i=1}^{\infty} \frac{\phi^{2i}}{i} H_i = \int_0^{\phi^2} \frac{1}{x} \left( \sum_{i=1}^{\infty} x^i H_i \right) \, dx.
\]
Now, the inner series can easily be solved by replacing \( \phi^2 \) with \( x \) in the first identity. Separating the rational function into its components we obtain
\[
\sum_{i=1}^{\infty} \phi^{2i} H_i = -\int \frac{\log (1 - x)}{x} \, dx - \int \frac{\log (1 - x)}{1 - x} \, dx.
\]
The first term of the above equation is equal to \( \operatorname{Li}_2 (\phi^2) \), whereas the second accepts direct integration. The last identity follows from
\[
(1 - \phi^2) \sum_{i=1}^{\infty} \phi^{2i} H_i^2 = \phi^2 + \sum_{i=2}^{\infty} \phi^{2i} (H_i^2 - H_{i-1}^2) = 2 \sum_{i=1}^{\infty} \phi^{2i} \left( \frac{H_i^2}{i} \right) - \sum_{i=1}^{\infty} \phi^{2i} i^2.
\]
The first series can be solved using (36), whereas the second is equal to \( -\operatorname{Li}_2 (\phi^2) \).

Using Lemmas 4 and 3 together with the fact that
\[
\Gamma \left( i, \frac{\alpha}{\rho} \right) = \Gamma (i) - \gamma \left( i, \frac{\alpha}{\rho} \right),
\]
we have
\[
\sum_{i=1}^{\infty} \phi^{2i} B_i^2 \left( \frac{\alpha}{\rho} \right) = \alpha (\operatorname{Li}_2 (\phi^2) + \log^2 (1 - \phi^2)) + \frac{-\alpha \log^2 \alpha - 2\alpha \phi^2 \operatorname{Li}_2 (\phi^2) + \log^2 \alpha}{1 - \phi^2} + 2\alpha \log \alpha + \xi_{33} (\rho) + \xi_{34} (\rho) + O (\rho^{-2})
\]
where we have introduced the two error terms
\[
\xi_{33} (\rho) = -2 \sum_{i=2}^{\infty} \phi^{2i} \left( \left( 1 - \frac{\alpha}{\rho} \right) H_{i-1} + \frac{1}{i} \right)
\]
\[
\times \left( \left( 1 - \frac{\alpha}{\rho} \right) \sum_{k=2}^{i-1} \frac{\gamma (k, \frac{\alpha}{\rho})}{k!} + \gamma (i, \frac{\alpha}{\rho}) \right)^2
\]
and
\[
\xi_{34} (\rho) = \phi^2 \gamma^2 \left( \frac{1}{\rho} \right)
\]
\[
+ \sum_{i=2}^{\infty} \frac{\phi^{2i}}{2} \left( \left( 1 - \frac{\alpha}{\rho} \right) \sum_{k=1}^{i-1} \frac{\gamma (k, \frac{\alpha}{\rho})}{k!} + \gamma (i, \frac{\alpha}{\rho}) \right)^2 \right).
\]
To see that these two terms are of order \( O (\rho^{-2}) \), observe that by definition of \( \gamma (k, \frac{\alpha}{\rho}) \) we have, for \( k \geq 1 \), the inequality
\[
\left| \gamma \left( k, \frac{\alpha}{\rho} \right) \right| \leq \int_0^{\frac{\alpha^2}{\rho}} x^{k-1} \, dx \leq \left( \frac{\alpha}{\rho} \right)^k
\]
and therefore, for sufficiently large \( \rho \) we have
\[
|\xi_{33} (\rho)| \leq 2 \left( \frac{\alpha}{\rho} \right)^2 e^{\frac{\alpha}{\rho}} \sum_{i=2}^{\infty} \phi^{2i}
\]
\[
|\xi_{34} (\rho)| \leq \left( \frac{\alpha}{\rho} \right)^2 e^{\frac{\alpha}{\rho}} \sum_{i=1}^{\infty} \phi^{2i}
\]
so that the two terms are effectively of order \( O (\rho^{-2}) \). We can therefore conclude that
\[
\vartheta_3 (\rho) = \operatorname{Li}_2 (\phi^2) + \log^2 \alpha + \frac{1}{\rho} (2 \operatorname{Li}_2 (\phi^2) + \log^2 \alpha - 2 \alpha \log \alpha) + O (\rho^{-1})
\]
Combining the expression for \( \vartheta_1 (\rho) \), \( \vartheta_2 (\rho) \) and \( \vartheta_3 (\rho) \) we obtain the result of the lemma.

REFERENCES

[1] Y. Polyanisky, H. V. Poor, and S. Verdu, “Channel coding rate in the finite blocklength regime,” IEEE Trans. Inf. Theory, vol. 56, no. 5, pp. 2307–2359, May 2010.
[2] R. L. Dobrushin, “Mathematical problems in the Shannon theory of optimal coding of information,” in Proc. 4th Berkeley Symmp. Math., Statist., Probab., Berkeley, CA, USA, 1961, pp. 211–252.
[3] V. Strassen, “Asymptotische abschatzungen in Shannons informationstheorie,” in Proc. Trans. 3D Prague Conf. Inf. Theory, Prague, Czech Republic, 1962, pp. 689–723.
[4] Y. Polyanisky, “Channel coding: Non-asymptotic fundamental limits,” Ph.D. dissertation, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, USA, 2010.
[5] G. Durisi, T. Koch, and P. Popovski, “Toward massive, ultra-reliable, and low-latency wireless communication with short packets,” Proc. IEEE, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.
[6] M. Shirvani-Heggad et al., “Short-block–length codes for ultra-reliable low latency communications,” IEEE Commun. Mag., vol. 57, no. 2, pp. 130–137, Feb. 2019.
[7] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: Information-theoretic and communications aspects,” IEEE Trans. Inf. Theory, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
[8] W. Yang, G. Durisi, T. Koch, and Y. Polyanisky, “Quasi–static multiple–antenna fading channels at finite blocklength,” IEEE Trans. Inf. Theory, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.
[9] J. Hoydis, R. Couillet, and P. Fliantanda, “The second–order coding rate of the MIMO quasi–static Rayleigh fading channel,” IEEE Trans. Inf. Theory, vol. 61, no. 12, pp. 6591–6622, Dec. 2015.
W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, “Optimum power control at finite blocklength,” IEEE Trans. Inf. Theory, vol. 61, no. 9, pp. 4598–4615, Sep. 2015.

Y. Polyanskiy and S. Verdu, “Scalar coherent fading channel: Dispersion analysis,” in Proc. IEEE Int. Symp. Inf. Theory, Jul. 2011, pp. 2978–2982.

W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, “Diversity versus channel knowledge at finite block-length,” in Proc. IEEE Inf. Theory Workshop, Sep. 2012, pp. 572–576.

A. Collins and Y. Polyanskiy, “Orthogonal designs optimize achievable dispersion for coherent MISO channels,” in Proc. IEEE Int. Symp. Inf. Theory, Jun. 2014, pp. 2524–2528.

A. Collins and Y. Polyanskiy, “Dispersion of the coherent MIMO block-fading channel,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jul. 2016, pp. 1068–1072.

A. Collins and Y. Polyanskiy, “Coherent multiple-antenna block–fading channels at finite blocklength,” IEEE Trans. Inf. Theory, vol. 65, no. 1, pp. 380–405, Jan. 2019.

X. Mestre, D. Gregoratti, and P. P. Zhang, “Short block length transmissions under multi–variate stationary Rayleigh fading,” in Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC), Cannes, France, Jul. 2019, pp. 2–5.

A. Martinez and A. G. I. Fabregas, “Saddlepoint approximation of random-coding bounds,” in Proc. Inf. Theory Appl. Workshop, Feb. 2011.

J. Scarlett, A. Martinez, and A. Guillen I Fabregas, “Mismatched decoding: Error exponents, second–order rates and saddlepoint approximations,” IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2647–2666, May 2014.

T. Erseghe, “Coding in the finite–blocklength regime: Bounds based on Laplace integrals and their asymptotic approximations,” IEEE Trans. Inf. Theory, vol. 62, no. 12, pp. 6854–6883, Dec. 2016.

I. A. Dragamov, “On the spectrum of stationary Gaussian sequences satisfying the strong mixing condition I. Necessary conditions,” Theory Probab. Appl., vol. 10, no. 1, pp. 85–106, Jan. 1965.

A. N. Kolmogorov and Y. A. Rozanov, “On strong mixing conditions for stationary Gaussian processes,” Theory Probab. Appl., vol. 5, no. 2, pp. 204–208, Jan. 1960.

I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series and Products, 6th ed. London, U.K.: Academic, 2000.

C. Müller and M. Mittelbach, “Discrete–time frequency–selective Rayleigh fading channel—Impact of correlated scattering on outage and ergodic capacity,” in Proc. IEEE Int. Conf. Commun., May 2008, pp. 4664–4670.

A. Clark, P. J. Smith, and D. P. Taylor, “Instantaneous capacity of OFDM on Rayleigh–fading channels,” IEEE Trans. Inf. Theory, vol. 53, no. 1, pp. 355–361, Jan. 2007.

Technical Specification Group Radio Access Network, NR; Multiplexing and Channel Coding (Release 15), document TS 38.212, 3GPP, 2018.

J. Ostman, G. Durisi, E. G. Strom, M. C. Coskun, and G. Liva, “Short packets over block–memoryless fading channels: Pilot–assisted or noncoherent transmission?” IEEE Trans. Commun., vol. 67, no. 2, pp. 1521–1536, Feb. 2019.

A. N. Tikhomirov, “On the convergence rate in the central limit theorem for weakly dependent random variables,” Theory Probab. Appl., vol. 25, no. 4, pp. 790–809, Jan. 1981.

T. M. Zuparov, “On the rate of convergence in the central limit theorem for weakly dependent random variables,” Theory Probab. Appl., vol. 36, no. 4, pp. 783–792, Jan. 1992.

J. Sunklodas, “Approximation of distributions of sums of weakly dependent random variables by the Normal distribution,” in Limit Theorems of Probability Theory, Y. V. Prokhorov and V. Stanevicius, Eds. New York, NY: USA: Springer, 2000, ch. 3, pp. 113–165.

R. M. Gray, “Toeplitz and circulant matrices: A review,” FNT Commun Inf. Theory, vol. 2, no. 3, pp. 155–239, 2005.

F. Hiai and D. Petz, “Asymptotic freeness almost everywhere for random matrices,” Acta Scientiarum Mathematicarum (Szeged), vol. 66, nos. 3–4, pp. 809–834, 2000.

L. Pastur and M. Shcherbina, Eigenvalue Distribution of Large Random Matrices of Mathematical Surveys and Monographs, vol. 171. Providence, RI, USA: American Mathematical Society, 2011.