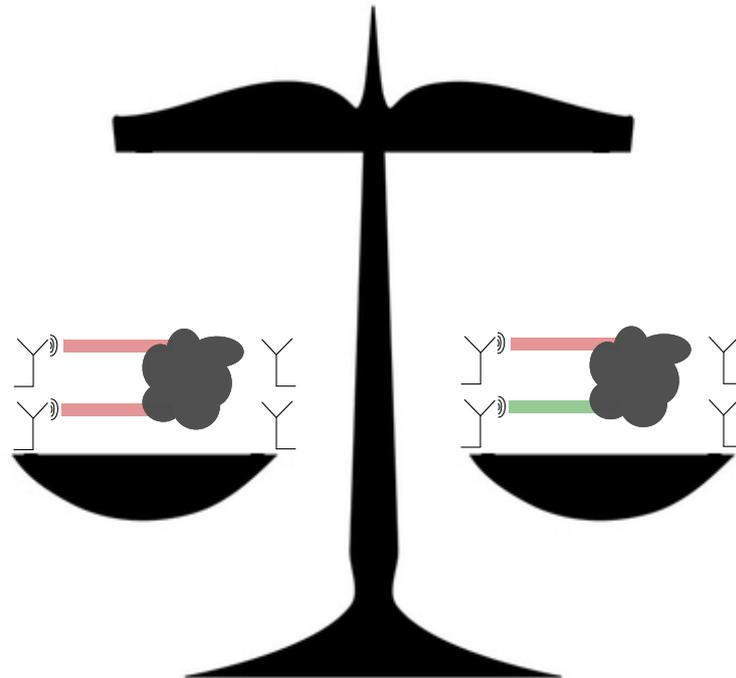


CHALMERS



The tradeoff between transmit diversity and spatial multiplexing at finite blocklength

Master's Thesis in Communication Engineering

JOHAN ÖSTMAN

Department of Signals and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2014
Master's Thesis EX047/2014

REPORT NO. EX047/2014

The tradeoff between transmit diversity and spatial multiplexing at finite blocklength

JOHAN ÖSTMAN



Department of Signals and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2014

Master Thesis for the Master Programme "Communication Engineering"
The tradeoff between transmit diversity and spatial multiplexing at finite
blocklength
JOHAN ÖSTMAN

©JOHAN ÖSTMAN, 2014

Technical report no EX047/2014
Department of Signals and Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 4673 561 97 64

Abstract

In today's society, communication plays a vital role. The mobile data traffic has experienced a tremendous increase in recent years and it is expected to continue on the same path. Therefore, methods that support high data rates with small spectral footprints are highly appealing. A powerful such, is to utilize multiple antennas at the transmitter- and the receiver side. By doing so, one gets to choose between utilizing spatial selectivity, in order to increase reliability, or to use multiplexing and hence increase the rate.

A common performance metric is the channel capacity, the largest rate of information that can be exchanged at an arbitrary low probability of error. However, this metric is asymptotic in the blocklength and it may therefore be questionable whether it is an accurate metric or not in the packet based networks of today. Instead, one may argue that a more suitable metric would be the maximal rate for a given blocklength and probability of error.

This thesis investigates communication over a Rayleigh block-fading channel when multiple antennas are utilized under the assumption that neither the transmitter nor the receiver has any channel state information but knows the statistical properties of the channel perfectly. The channel capacity is characterized in terms of bounds which are shown to be tight for SNR as low as 0 dB. Non-asymptotic bounds are also presented and it is shown that the maximal achievable rate is not monotonically increasing in the coherence block but that there is a rate-maximizing coherence block for a given blocklength and probability of error. These bounds also give insight in how costly it is to learn a fading channel and when the lack of time-frequency selectivity becomes detrimental. Furthermore, the maximal achievable rate is compared to diversity-exploiting schemes that are in use today. This illustrates for what degree of channel selectivity, diversity-exploiting schemes are close to optimal, and when instead the available spatial degrees of freedom should be used to provide spatial multiplexing.

Acknowledgements

First, I would like to extend my deepest gratitude's towards my supervisor Wei Yang, whom I have been working side by side to, for his efforts in supporting me whenever I have been in need. His friendly attitude and many good ideas have been a vital part of writing this thesis and my joy in doing so.

I am also very grateful towards Ass. Prof. Giuseppe Durisi, my examiner and supervisor, who have taught me so much in what it is like to do research and have kept pushing me throughout the thesis. Your advice and good will have been very encouraging.

Last, I would like to thank all my friends who have contributed with belief and support throughout my studies. It has been a delight and without you, this would not have been possible.

Johan Östman, Gothenburg 13/08/14

Contents

1	Introduction	1
1.1	History	1
1.2	Contributions	2
1.3	Organization	2
2	Theory	3
2.1	The channel	3
2.2	Information theoretical tools	4
3	Assumptions	7
4	Capacity bounds	8
4.1	Capacity upper bound	8
4.1.1	Distribution of singular values	10
4.1.2	The square case	11
4.2	Capacity lower bound	19
5	Bounds on the maximal achievable rate	22
5.1	Rate upper bound	22
5.2	Rate lower bound	24
6	Space-Time Block Codes	25
6.1	Alamouti lower bound	25
6.2	SFBC+FSTD lower bound	28
7	Results	30
7.1	Tightness of capacity bounds	30
7.2	Tightness of Rate Bounds	31
7.3	STBC performance	32
8	Discussion	34
9	Conclusion	36
A	Appendix A	37
B	Appendix B	40
C	Appendix C	43

Nomenclature

Variables

x	Deterministic scalar
X	Random scalar
\mathbf{x}	Deterministic vector
\mathbf{X}	Random vector
\mathbf{X}	Deterministic matrix
\mathbb{X}	Random matrix
\mathbf{I}_a	Identity matrix of size $a \times a$
\mathbf{C}	Covariance matrix
$\Lambda(\mathbf{X})$	Diagonal matrix consisting of the eigenvalues of \mathbf{X}
$\Sigma(\mathbf{X})$	Diagonal matrix consisting of the singular values of \mathbf{X}

Constants

n_t	Number of transmit antennas
n_r	Number of receive antennas
T	Coherence block
n	Blocklength
L	Selectivity branches
ρ	Average SNR at a receive antenna
ϵ	Block error rate

Functions

$C_{\text{erg}}(\rho)$	Ergodic capacity
$\Gamma(\beta)$	The Gamma function
$\Gamma_\alpha(\beta)$	Multivariate complex gamma function
$\psi(\beta)$	The Digamma function
$\det(V(\mathbf{X}))$	The determinant of a Vandermonde matrix \mathbf{X}
$M(t)$	Moment generating function
$C_{\text{out},\epsilon}(\rho)$	ϵ -outage capacity
$R^*(n, \epsilon)$	Maximal achievable rate for fixed n and ϵ
$f_X(x)$	Probability density function of the scalar random variable, X
$f_{\mathbb{X}}(\mathbf{X})$	Probability density function of the random matrix, \mathbb{X}
$P_{\mathbb{X}}$	Probability measure of \mathbb{X}
$\Pr\{X\}$	Probability of the event X
$\mathbb{E}_X[\cdot]$	Expected values with respect to X
$[a]^+$	The maximum of a and 0

1 Introduction

1.1 History

Wireless communication plays an important role in today's society. Among other things it has connected the world by enabling the use of cellphones, laid the foundation for GPS and made it possible to gather information from other planets. The original idea of wireless communication dates back to the pre-industrial age where the signalling was made by, for example, smoke, light or flags and even till this day, we are using, for example, smoke, when a new pope is elected. Wireless communication, as we know it today, was not invented until 1896 when Guglielmo Marconi, an Italian physicist, made the first analogue wireless radio transmission, for which he, in 1909, was rewarded the Nobel Prize. After this discovery, the development of wireless communication systems took off and a rapid evolution have been ongoing ever since.

In 1948, Claude E Shannon published his famous work, *A Mathematical Theory of Communication*, laying the foundation of information theory [1]. One of the contributions was the derivation of the *Channel Capacity* which is the largest rate at which information can be transmitted over an unreliable channel with arbitrarily low error rate. Since channel capacity is the ultimate limit on the rate of reliable communication, the interest of finding it for all possible scenarios is clear and for several channels of practical interest, capacity is indeed known. However, capacity is an asymptotic quantity, requiring the use of codes with very long blocklength to be approached. In real life though, communication is done with finite, sometimes short, blocklengths. In this regime, arbitrarily low error rates can not be guaranteed and the relevant quantity is instead the maximal achievable rate for a given error rate and blocklength.

The finite blocklength regime is, until this day, not completely understood although the subject has raised new interest in the last years due to the work by Polyanskiy et. al. where tools for finding converse- and achievability bounds on the maximal achievable rate are provided [2]. These tools have been successfully utilized for several channels and setups of interest such as the single-input-single-output (SISO) Gaussian channel [2], Multiple-input-multiple-output (MIMO) Rayleigh quasi-static channel [3] [4] and for the SISO Rayleigh block-fading channel [5].

Another approach to unveil the performance of communication systems in the finite blocklength regime is to derive the random coding error exponent (RCEE), which characterizes the exponential decay of the error probability as a function of the rate and the blocklength [6]. Unfortunately, the RCEE does not tell the whole story in the finite blocklength regime since it is useful only for codewords greater than about 200 symbols [7, pp 18.].

The aim of this thesis is to characterize the maximal achievable rate in the finite blocklength regime for a MIMO setup in a Rayleigh block-fading environment when neither the receiver nor the transmitter have any information about the fading coefficients but knows the statistics of the channel perfectly, the so called noncoherent setting. Under these assumptions, to the extent of the authors knowledge, there are no closed form expressions for neither the capacity nor the maximal achievable rate.

Lower bounds on the non-coherent capacity in the Rayleigh block-fading environment are reported in [8] where independent and identically distributed (iid) Gaussian inputs was used and in [9] where unitary space time modulation (USTM) was used as the signal input. Furthermore, the output pdf induced by some input distribution was expressed in terms of hypergeometrical functions in [10]. The output pdf can then be used to evaluate the information density and hence the mutual information. An upper and a lower bound on the maximal achievable rate for the Rayleigh block-fading channel is reported for SISO in [5].

1.2 Contributions

In this thesis, bounds on the maximal achievable rate and the channel capacity in the Rayleigh block-fading channel are presented for a MIMO setup with equal number of transmit- and receive antennas. The extension to the configuration of different number of transmit- and receive antennas is straightforward. A new upper bound on the noncoherent capacity, based on a high SNR assumption, in the MIMO Rayleigh block-fading channel is presented. The lower bound presented is equivalent to the one reported in [9] but has proven to be more numerically stable. The bounds on the non-coherent capacity are proven to be tight for signal to noise ratios (SNR) as low as 0 dB. Furthermore, two upper bounds are presented on the maximal achievable rate. One is tight for fast-fading channels while the other is tight for slow-fading channels. A lower bound on the maximal achievable rate is also derived. Diversity-exploiting space-time block codes (STBC) that are employed today are compared to the optimal performance. This illustrates for what degree of channel selectivity, diversity-exploiting schemes are close to optimal, and when instead the available spatial degrees of freedom should be used to provide spatial multiplexing.

1.3 Organization

The thesis is organized as follows. Chapter 2 provides an introduction to the tools and quantities that are utilized throughout the thesis. The aim of this chapter is to review some fundamental properties of the wireless channel and the information theoretic tools needed in later chapters. Chapter 3 explains the assumptions that are made throughout the thesis. In the Chapter 4, an upper and a lower bound on the channel capacity are derived. The fifth chapter focuses on the maximal achievable rate and derives two upper bounds and one lower bound on this quantity. Lower bounds on the rates achievable with STBC's are derived in Chapter 6. In Chapter 7, numerical results are presented. Chapter 8 consists of a discussion based on the results and the last part, Chapter 9, concludes the thesis. There are several appendices included, which contain the detailed mathematical derivations of some of the results.

2 Theory

The purpose of this Chapter is to do a review of important channel parameters, introduce information theoretic tools and to present performance metrics relevant to the communication problem for different scenarios. If the reader is familiar to the concepts in wireless communication and blockfading as well as the fundamentals of information theory, this Chapter may be skipped.

2.1 The channel

When communicating over a wireless channel, one of the main challenges is to mitigate the fading. The fading is usually associated to two different phenomena; large- and small-scale fading. Large scale fading occurs on the order of several wavelengths and is caused by for example buildings, while the small-scale fading occurs on the order of a wavelength. In this thesis, only small-scale fading will be considered, which occurs mainly due to the receiver being in motion and to multipath propagation.

If the receiver is moving, the received signal will experience frequency dispersion due to Doppler shift. The Doppler spread is a measure of the spectral broadening and is defined as the band of frequencies, around the carrier frequency, where the Doppler shifts are significant [11]. The Doppler spread is inversely proportional to the channels coherence time which is a statistical measure of how long the channel remains invariant. If the symbol time is short in comparison to the coherence time, the channel is slow fading otherwise it is called fast fading.

When a signal travels different paths, the receiver will be receiving signals over some time. This time is called the channels delay spread and is defined as the time between the first and last significant received signal component. Its frequency dual is called the coherence bandwidth, the smallest frequency band for which the channel remains invariant. If the symbol time is larger than the delay spread, the multipath components will not interfere with the next sample. This is equivalent as to say that if the signal bandwidth is smaller than the coherence bandwidth there will be no dispersion, this is called flat fading. If the signal bandwidth is larger than the coherence bandwidth, the signal experiences frequency selective fading [11].

A model that is commonly adopted for wireless fading without line-of-sight (LOS) is the Rayleigh fading model. It is based on the assumption that there are a large number of independent signals taking different paths with random amplitude and phase. By this assumption, the channel fading may be modelled as a zero-mean circular complex Gaussian random variable (rv) [11, pp 36.]. Also, a common method of modelling the noise at the receiver such as thermal noise etc., is by adding white Gaussian noise to the received signal.

Important concept throughout the thesis is non-coherent- and coherent communication. In the former, neither the transmitter nor the receiver has channel state information (CSI) but knows the statistics of the channel perfectly while in the latter, the receiver have full knowledge of the channel realization while the transmitter have the same knowledge as in the non-coherent case. In non-coherent communication, there is a need of estimating the channel in order to mitigate the effect of fading. This can be done in several ways, for example by sending deterministic sequences called pilots [12]. If the channel's coherence time is not much larger than the symbol time, intuitively, there is going to be a significant decrease in exchanged information between transmitter and receiver due to the need of frequently having to estimate the channel.

As previously discussed, it is the Doppler spread that dictates the coherence time. However, to simplify the model, a Rayleigh block-fading channel will be assumed, originally proposed

in [13]. This model assumes a fixed coherence time for which the channel remains invariant where after it changes to a new independent realization. Although this is a very simplified model of reality, it captures the big picture of the fading, not least under the commonly adopted Clarke's spectrum [14], and yields results close to the continuous-time fading model [15].

The concept of selectivity is now easily understood by considering a codeword spanning several coherence times or coherence bandwidths. Since each realization of the channel is independent of the others, time-selectivity may be utilized by proper interleaving. Analogously, frequency selectivity may be exploited by transmitting over several frequencies separated in frequency by at least the coherence bandwidth. We define a coherence block, T as the number of symbols that can be transmitted before the channel changes. The coherence block may be defined in time, frequency or both. A codeword of length n will experience $L = n/T$ coherence blocks during a transmission. The number of coherence blocks, L , will be referred to as the number of selectivity branches. Furthermore, when multiple antennas are utilized, a third type of selectivity is introduced; spatial selectivity. Spatial selectivity offers the user to choose between using diversity exploiting schemes to increase the reliability or to use multiplexing in order to increase the information rate. There is a fundamental trade-off between diversity gain and multiplexing gain. This trade-off is presented for the coherent Rayleigh block-fading channel in [16] where it is shown that the optimal trade-off depend only on the number of transmit- and receive antennas. For the non-coherent Rayleigh block-fading channel, it has been shown that the penalty to pay is precisely the number of unknown channel coefficient [17]. However, in today's standards, such as LTE, a point of operation is chosen with a fixed error rate rather than aiming for as low error target as possible. It has been shown that from a performance perspective, there is basically no decision to be made but one should always go for multiplexing [18].

2.2 Information theoretical tools

The fundamental challenge in communication is to extract information from an observation, which may be viewed as a random variable, Y . The differential entropy of a random variable, Y , is a measure of the uncertainty in a random variable and is given as [19]

$$h(Y) \triangleq - \int_S f_Y(y) \log(f_Y(y)) dy \quad (1)$$

where S is the support set of the random variable and $f_Y(y)$ is the probability density function (pdf). A very useful measure, based on entropy, is the mutual information of two random variables, X and Y , given as [19]

$$\mathcal{I}(X; Y) \triangleq h(Y) - h(Y | X). \quad (2)$$

This is interpreted as the shared information between X and Y . The mutual information may also be seen from another point of view, by considering the information density, a random variable, given as [20]

$$i(x; y) \triangleq \log \left(\frac{f_{X,Y}(x, y)}{f_X(x) f_Y(y)} \right). \quad (3)$$

This quantity will be useful when deriving bounds on the maximal achievable rate. The mutual information is given as the expected value of the information density. Also, a measure of the similarity between two probability measures, P_X and P_Y will be useful. This is called the relative entropy and given as [19]

$$D(P_X || P_Y) = \int f_X(x) \log \left(\frac{f_X(x)}{f_Y(y)} \right) dx. \quad (4)$$

As previously stated, the aim of communication is to extract information out of an observed symbol and determine what was originally transmitted. The largest information rate, at vanishing error rates and large blocklengths, one could hope for is called the channel capacity [1]. For a memoryless channel, the capacity, C , is given as [21]

$$C = \max_{P_X} \{\mathcal{I}(X; Y)\} \quad (5)$$

where P_X is an arbitrary probability measure on X . For the coherent setting, there are mainly two different measures of capacity; ergodic- and ϵ -outage capacity. The ergodic capacity applies in fast fading scenarios, i.e. when the codeword experiences a large number of fades. For the Rayleigh fading channel the MIMO ergodic capacity is given as [21]

$$C_{\text{erg}} = \mathbb{E}_{\mathbb{H}} \left[\log \left(\det \left(\mathbf{I} + \frac{\rho}{n_t} \mathbb{H} \mathbb{H}^H \right) \right) \right] \quad (6)$$

where ρ is the average SNR at each receive antenna and \mathbb{H} is the channel fading matrix. The outage capacity is the counterpart to the ergodic capacity; it is the relevant capacity measure in a slow fading scenario, i.e. when the codeword experiences a small number of fades. The ϵ -outage capacity is based on the probability that the target rate is larger than the instantaneous capacity [11, pp. 187]

$$P_{\text{out}}(R) \triangleq \Pr \left\{ \log \left(\det \left(\mathbf{I} + \frac{\rho}{n_t} \mathbb{H} \mathbb{H}^H \right) \right) < R \right\}. \quad (7)$$

The ϵ -outage capacity, $C_{\text{out},\epsilon}$, is given as the largest rate, R , for which the outage probability is less than ϵ [11, pp.188]

$$C_{\text{out},\epsilon} = \sup_R \{R : P_{\text{out}}(R) \leq \epsilon\}. \quad (8)$$

In the case of a Rayleigh block-fading channel, the codeword experiences L independent fadings during a transmission which can be seen as parallel channels. Hence, the outage probability has to be modified as [11, pp. 198]

$$P_{\text{out}}(R) \triangleq \Pr \left\{ \sum_{l=1}^L \log \left(\det \left(\mathbf{I} + \frac{\rho}{n_t} \mathbb{H}_l \mathbb{H}_l^H \right) \right) < R \right\}. \quad (9)$$

As aforementioned, the capacities in (6), (8) and (9) assumes perfect CSI at the receiver. In the noncoherent setting, the capacity for the Rayleigh block-fading channel is, to the extent of the authors knowledge, not known. A lower bound have been presented by Rusek *et al.* under the assumption of independent and identically distributed (iid) Gaussian inputs [8]. Marzetta and Hochwald have presented a lower bound based on the unitary space time modulation (USTM) input distribution [9], which has been proven to be the input distribution that achieves capacity in the regime of large SNR [13].

Although the channel capacity is the ultimate limit of reliable communication, it is in today's packet oriented systems questionable whether the capacity is a relevant performance metric or not when the packet size is small. The reason is that capacity is an asymptotic measure which may be written as

$$C = \lim_{n \rightarrow \infty} \lim_{\epsilon \rightarrow 0} R^*(n, \epsilon) \quad (10)$$

where $R^*(n, \epsilon)$ is the maximal achievable rate for a given blocklength, n , and error target, ϵ . For large packets, however, the capacity is a reasonable performance metric. Since wireless

communication is packet oriented rather than stream oriented, one would instead consider $R^*(n, \epsilon)$ to be the performance metric of interest. To the extent of the authors knowledge, there are no closed form expressions for $R^*(n, \epsilon)$ in the non-coherent setting.

To characterize communication in the finite blocklength regime, there are mainly two different approaches; by using the Random coding error exponent (RCEE) or by finding bounds on $R^*(n, \epsilon)$. The RCEE gives a bound on the error as a function of a given blocklength and rate [6]. Studies of the RCEE in the coherent Rayleigh block-fading channel have shown that, unlike the capacity in (6), the coherence block plays an important role in the RCEE; the longer the coherence block, the larger the exponent and the smaller the block error probability [22]. Abou-Faycal and Hochwald has shown that the same result applies for the non-coherent setting [23]. Interestingly, in the same paper, it is shown that the input distribution that maximizes the RCEE has the same product distribution as the non-coherent capacity achieving distribution [23].

For the method of bounding $R^*(n, \epsilon)$, Polyanskiy et al. recently presented several tools for finding achievability- and converse bounds in the finite blocklength regime [2]. In this thesis, an achievability bound called the Dependence Testing (DT) bound will be utilized to derive a lower bound on $R^*(n, \epsilon)$. The bound sequentially tests messages and return the first message whose likelihood exceeds a pre-determined threshold. The DT-bound holds for any input distribution and is given as [2, Thm. 17]

$$\epsilon \leq \mathbb{E}_{X,Y} \left[e^{-\left[i(X;Y) - \log\left(\frac{M-1}{2}\right)^+ \right]} \right] \quad (11)$$

where $i(X;Y)$ is the information density, introduced earlier, and M is the cardinality of the codebook. In order to derive an upper bound on $R^*(n, \epsilon)$, a converse bound called the Meta-Converse (MC) bound will be utilized. It is based on the Neyman-Pearson lemma which states that the likelihood ratio test is the optimal test between two hypothesis [24, pp. 89]. The shorthand for the bound is given as [2, Thm. 30]

$$M \leq \inf_{Q_Y} \sup_{x \in F} \frac{1}{\beta_{1-\epsilon}(x, Q_Y)} \quad (12)$$

where F is the space of input signals fulfilling some power constraint, Q_Y is an arbitrary output distribution independent of x , M is the cardinality of the codebook and $\beta_{1-\epsilon}(x, Q_Y)$ is the minimum probability of error under the hypothesis Q_Y if the probability of error under P_X is not greater than ϵ .

One may also obtain an upper bound on $R^*(n, \epsilon)$ by making use of Fano's inequality as [19, Thm 2.10.1].

$$R^*(n, \epsilon) \leq \bar{R}(n, \epsilon) \triangleq \frac{C(\rho) + \frac{H_b(\epsilon)}{n}}{1 - \epsilon} \quad (13)$$

where $C(\rho)$ is the non-coherent capacity and $H_b(x)$ is the binary entropy function defined as

$$H_b(x) = -x \log(x) - (1-x) \log(1-x). \quad (14)$$

However, as previously mentioned, there are no closed form expressions for the non-coherent capacity under the Rayleigh-blockfading assumption. Therefore, to upper bound $R^*(n, \epsilon)$ using (13), we also need to have an upper bound on the capacity.

3 Assumptions

Throughout the thesis, it will be assumed that the channel follows a Rayleigh block-fading channel without memory. This model assumes that the channel is constant over a coherence block of T consecutive symbols after which it changes to a new independent realization. The coherence block, T , can be thought of as a block of symbols transmitted in time, frequency or in a time-frequency block, for which the channel remains invariant.

The channel, \mathbb{H} , will be modelled as a matrix with entries iid $\mathcal{CN}(0, 1)$. At the receiver side, the entries of the thermal noise matrix, \mathbb{W} , will also be modelled with entries iid $\mathcal{CN}(0, 1)$ and independent of \mathbb{H} . This is a realistic channel model for systems using some form of time-division multiplexing or frequency hopping [9].

We will consider a MIMO setup, consisting of n_t transmit antennas and n_r receive antennas that communicates in a non-coherent setting. It will be assumed that $T \geq n_r + n_t$ i.e., the number of symbols in a coherence block is larger than the total number of antennas. Furthermore, it has been shown that using $n_t > n_r$ or $n_t > \frac{T}{2}$ does not provide any capacity gain in the high SNR regime [25]. Hence, it will be assumed that $n_t \leq \min\{n_r, \frac{T}{2}\}$. The case of large-MIMO systems, i.e., $T \leq n_r + n_t$, is treated in [26] where it is shown that the capacity-achieving distributions are not the same for large SNR in the two cases, i.e., USTM is not capacity achieving for large-MIMO systems.

A codeword will be assumed to consist of n symbols, transmitted over L consecutive coherence blocks, T , making the length of the codeword, $n = LT$ symbols. For each coherence block, the channel input-output relation is modelled as

$$\mathbb{Y}_l = \sqrt{\frac{\rho}{n_t}} \mathbb{X}_l \mathbb{H}_l + \mathbb{W}_l, \quad l = 1, \dots, L \quad (15)$$

where each of the matrices above is given as

$$\begin{aligned} \mathbb{Y} &= \begin{bmatrix} Y_{11} & \dots & Y_{1n_r} \\ \vdots & \ddots & \vdots \\ Y_{T1} & \dots & Y_{Tn_r} \end{bmatrix}, \quad \mathbb{X} = \begin{bmatrix} x_{11} & \dots & x_{1n_t} \\ \vdots & \ddots & \vdots \\ x_{T1} & \dots & x_{Tn_t} \end{bmatrix} \\ \mathbb{H} &= \begin{bmatrix} H_{11} & \dots & H_{1n_r} \\ \vdots & \ddots & \vdots \\ H_{n_t 1} & \dots & H_{n_t n_r} \end{bmatrix}, \quad \mathbb{W} = \begin{bmatrix} W_{11} & \dots & W_{1n_r} \\ \vdots & \ddots & \vdots \\ W_{T1} & \dots & W_{Tn_r} \end{bmatrix} \end{aligned} \quad (16)$$

and ρ can be thought of as the SNR at each receive antenna. Finally, the communication over each codeword is assumed to be power limited as

$$\frac{1}{L} \sum_{i=1}^L \text{tr}\{\mathbb{X}_i \mathbb{X}_i^H\} \leq T n_t. \quad (17)$$

4 Capacity bounds

The focus of this chapter is to derive upper- and lower bounds on the channel-capacity of a non-coherent Rayleigh block-fading channel. The bounds will thereafter be used for deriving an upper bound on the maximal achievable rate given in (13).

4.1 Capacity upper bound

This section aims to generalize the capacity upper bound that was attained in [5] for the SISO-case in the high SNR-regime by using duality. This method of upper bounding the capacity has previously been used in e.g., the non-coherent additive white Gaussian noise (AWGN) channel and in the optical direct-detection channel [27][28].

Since the channel is assumed to be block-memoryless the ergodic channel capacity is given by (5) as

$$C(\rho) = \frac{1}{T} \sup_{P_{\mathbb{X}}} \mathcal{I}(\mathbb{X}; \mathbb{Y}) \quad (18)$$

where the normalization by T is carried out since the channel needs to be utilized T times in order to transmit \mathbb{X} . The supremum in (18) is taken over all probability measures on \mathbb{X} that satisfies the cost constraint

$$\mathbb{E}_{P_{\mathbb{X}}} [\text{tr} \{ \mathbb{X} \mathbb{X}^H \}] \leq T n_t \quad (19)$$

We start by noting that the mutual information in (18) can be upper bounded by

$$\begin{aligned} \mathcal{I}(\mathbb{X}; \mathbb{Y}) &= D(P_{\mathbb{X}, \mathbb{Y}} \| P_{\mathbb{X}} P_{\mathbb{Y}}) \\ &= \int_{\mathbb{X}, \mathbb{Y}} f_{\mathbb{X}, \mathbb{Y}}(\mathbb{X}, \mathbb{Y}) \log \left(\frac{f_{\mathbb{X}, \mathbb{Y}}(\mathbb{X}, \mathbb{Y})}{f_{\mathbb{X}}(\mathbb{X}) f_{\mathbb{Y}}(\mathbb{Y})} \right) d\mathbb{X} d\mathbb{Y} \\ &= \int_{\mathbb{X}, \mathbb{Y}} f_{\mathbb{Y}|\mathbb{X}}(\mathbb{Y}|\mathbb{X}) f_{\mathbb{X}}(\mathbb{X}) \log \left(\frac{f_{\mathbb{Y}|\mathbb{X}}(\mathbb{Y}|\mathbb{X}) f_{\mathbb{X}}(\mathbb{X}) q_{\mathbb{Y}}(\mathbb{Y})}{f_{\mathbb{X}}(\mathbb{X}) f_{\mathbb{Y}}(\mathbb{Y}) q_{\mathbb{Y}}(\mathbb{Y})} \right) d\mathbb{X} d\mathbb{Y} \\ &= \int_{\mathbb{X}} f_{\mathbb{X}}(\mathbb{X}) \left[\int_{\mathbb{Y}} f_{\mathbb{Y}|\mathbb{X}}(\mathbb{Y}|\mathbb{X}) \log \left(\frac{f_{\mathbb{Y}|\mathbb{X}}(\mathbb{Y}|\mathbb{X}) q_{\mathbb{Y}}(\mathbb{Y})}{q_{\mathbb{Y}}(\mathbb{Y}) f_{\mathbb{Y}}(\mathbb{Y})} \right) d\mathbb{Y} \right] d\mathbb{X} \\ &= \mathbb{E}_{P_{\mathbb{X}}} [D(P_{\mathbb{Y}|\mathbb{X}} \| Q_{\mathbb{Y}})] - \underbrace{D(P_{\mathbb{Y}} \| Q_{\mathbb{Y}})}_{\geq 0} \\ &\leq -\mathbb{E}_{P_{\mathbb{Y}}} [\log(q_{\mathbb{Y}}(\mathbb{Y}))] - h(\mathbb{Y} | \mathbb{X}). \end{aligned} \quad (20)$$

Here, $q_{\mathbb{Y}}(\mathbb{Y})$ is an arbitrary probability density function of \mathbb{Y} . Using (20) in (18), the capacity can be upper-bounded as

$$C(\rho) \leq \frac{1}{T} \sup_{P_{\mathbb{X}}: \mathbb{E}_{\mathbb{X}}[\text{tr}\{\mathbb{X}\mathbb{X}^H\}] \leq T n_t} \left\{ \inf_{Q_{\mathbb{Y}}} \{ -\mathbb{E}_{\mathbb{Y}}[\log(q_{\mathbb{Y}}(\mathbb{Y}))] \} - h(\mathbb{Y} | \mathbb{X}) \right\}. \quad (21)$$

To limit the search space, we will choose $q_{\mathbb{Y}}(\mathbb{Y})$ heuristically rather than finding the infimum. By relaxing the constraints on $P_{\mathbb{X}}$, we end up with the Lagrangian

$$L(P_{\mathbb{X}}, \lambda) = -\mathbb{E}_{\mathbb{Y}}[\log(q_{\mathbb{Y}}(\mathbb{Y}))] - h(\mathbb{Y} | \mathbb{X}) + \lambda (T n_t - \mathbb{E}_{\mathbb{X}}[\text{tr} \{ \mathbb{X} \mathbb{X}^H \}]) \quad (22)$$

and the capacity upper-bound is now given by the dual problem

$$C(\rho) \leq \frac{1}{T} \inf_{\lambda \geq 0} \sup_{P_{\mathbb{X}}} \{ L(P_{\mathbb{X}}, \lambda) \}. \quad (23)$$

To evaluate (23), we need to choose a suitable output probability distribution, $Q_{\mathbb{Y}}$. We emphasize the fact that for $T \geq n_r + n_t$, the input distribution that achieves capacity in the high SNR regime is the scaled unitary isotropic distribution (id) [13]. Such an input may be constructed by letting Φ be a unitary id matrix i.e., a matrix for which

$$\begin{aligned} f_{\Phi}(\Phi) &= f_{\Phi}(\mathbf{Q}^H \Phi) \quad \forall \mathbf{Q} : \mathbf{Q}^H \mathbf{Q} = \mathbf{I} \\ \Phi \Phi^H &= \Phi^H \Phi = \mathbf{I} \end{aligned} \quad (24)$$

where $f_{\Phi}(\Phi)$ denotes the probability density function of the random matrix Φ , which is defined over the set of unitary matrices. Now, we take \mathbb{X} distributed as follows

$$\mathbb{X} = \Phi \mathbf{D}, \quad \Phi \in \mathbb{C}^{T \times n_t}, \mathbf{D} \in \mathbb{R}^{n_t \times n_t}. \quad (25)$$

where \mathbf{D} is diagonal and denotes the scaling. By using (25) in (15), we can rewrite the input-output relation as

$$\mathbb{Y} = \sqrt{\frac{\rho}{n_t}} \Phi \mathbf{D} \mathbf{H} + \mathbb{W}, \quad \mathbf{D} = \sqrt{T} \mathbf{I}_{n_t}. \quad (26)$$

Now, we need to choose the output distribution $q_{\mathbb{Y}}(\mathbb{Y})$. We express the output matrix by its singular value decomposition (SVD) according to

$$\mathbb{Y} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H \quad (27)$$

where $\mathbf{U} \in \mathbb{C}^{T \times T}$ and $\mathbf{V} \in \mathbb{C}^{n_r \times n_r}$ are unitary matrices and $\mathbf{\Sigma} \in \mathbb{R}^{T \times n_r}$ is a diagonal matrix containing the singular values of \mathbb{Y} . Heuristically, we choose \mathbf{U} and \mathbf{V} to be independent of each other and isotropically distributed on their corresponding manifolds. They are also assumed to be independent of $\mathbf{\Sigma}$. The rationale behind choosing $Q_{\mathbb{Y}}$ this way is that it resembles the USTM distribution at high SNR.

In order to find the pdf of the output, we want to do a mapping of the distribution of \mathbb{Y} onto the distribution in the coordinates of the SVD as

$$Q_{\mathbb{Y}} \rightarrow P_{\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}} \left| \frac{d\mathbf{U} d\mathbf{\Sigma} d\mathbf{V}}{d\mathbb{Y}} \right| = P_{\mathbf{U}} P_{\mathbf{\Sigma}} P_{\mathbf{V}} \left| \frac{d\mathbf{U} d\mathbf{\Sigma} d\mathbf{V}}{d\mathbb{Y}} \right| \quad (28)$$

where the last equality follows because \mathbf{U} , $\mathbf{\Sigma}$ and \mathbf{V} are independent. The last term is the Jacobian of the transformation. By assumption, \mathbf{U} and \mathbf{V} are unitary id. Furthermore, to make the SVD unique, we constrain the diagonal elements of \mathbf{U} to be real and non negative. Also, since $\mathbf{\Sigma}$ has rank n_r , we are only interested in the the first n_r columns of \mathbf{U} which we denote by $\tilde{\mathbf{U}} \in \mathbb{C}^{T \times n_r}$. Since $\tilde{\mathbf{U}}$ is unitary, it is an element of a constrained Stiefel manifold [25]

$$S^C(T, n_r) = \{ \mathbf{A} \in \mathbb{C}^{T \times n_r} | \mathbf{A}^H \mathbf{A} = \mathbf{I}_{n_r}, \mathbf{A}_{ii} \in \mathbb{R}^+ \}, i = 1, \dots, n_r \quad (29)$$

and is isotropically distributed over this. Hence, in the probability space of all matrices belonging to $S^C(T, n_r)$, the pdf of the matrix $\tilde{\mathbf{U}}$ is given by

$$f_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}) = \frac{1}{\text{Vol}(S^C(T, n_r))}. \quad (30)$$

By the same reasoning, \mathbf{V} belongs to a Stiefel manifold defined as

$$S(n_r, n_r) = \{ \mathbf{A} \in \mathbb{C}^{n_r \times n_r} | \mathbf{A}^H \mathbf{A} = \mathbf{I}_{n_r} \} \quad (31)$$

and exactly as in (30) we get

$$f_{\mathbf{V}}(\mathbf{V}) = \frac{1}{\text{Vol}(S(n_r, n_r))}. \quad (32)$$

Plugging these two simplifications into (34) results in the eigenvalue-distribution

$$f_{\Lambda}(\lambda_1, \dots, \lambda_{n_t}) = \frac{\pi^{m(m-1)} c^{-n_t n_r}}{\Gamma_{n_t}(n_t) \Gamma_{n_t}(n_r)} e^{-\frac{1}{c} \sum_{i=1}^{n_t} \lambda_i} \prod_{i=1}^{n_t} \lambda_i^{n_r - n_t} \prod_{i < j}^{n_t} (\lambda_i - \lambda_j)^2. \quad (37)$$

Since we are interested in the singular values, $\{\sigma_i\}_{i=1}^{n_t}$, of the Gaussian matrix, i.e. the square root of the eigenvalues, $\{\lambda_i\}_{i=1}^{n_t}$, of the Wishart matrix, the transformation that is to be made is

$$\lambda_i = \sigma_i^2, \quad i = 1, \dots, n_t \quad (38)$$

and the joint pdf of $\{\sigma_i\}_{i=1}^{n_t}$ is given as [31, Pg. 216-220]

$$f_{\Sigma}(\sigma_1, \dots, \sigma_{n_t}) = \frac{f_{\Lambda}(\lambda_1, \dots, \lambda_{n_t})}{\det(\mathbf{J})} \Big|_{\lambda_i = \sigma_i^2}, \quad i = 1, \dots, n_t \quad (39)$$

where \mathbf{J} is the Jacobian of the transformation whose determinant is given by

$$\det(\mathbf{J}) = \det(\text{diag}(2\sigma_1, \dots, 2\sigma_{n_t})) = 2^{n_t} \prod_{i=1}^{n_t} \sigma_i. \quad (40)$$

Finally, the joint pdf of the singular values, $\{\sigma_i\}_{i=1}^{n_t}$, can be written as

$$f_{\Sigma}(\sigma_1, \dots, \sigma_{n_t}) = \frac{2^{n_t} \pi^{n_t(n_t-1)}}{\Gamma_{n_t}(n_r) \Gamma_{n_t}(n_t) c^{n_t n_r}} e^{-\frac{1}{c} \sum_{i=1}^{n_t} \sigma_i^2} \prod_{i=1}^{n_t} \sigma_i^{2(n_r - n_t) + 1} \prod_{i < j} (\sigma_i^2 - \sigma_j^2)^2 \quad (41)$$

Using (41), we conclude that the singular values in Σ_1 follows the distribution in (41) with $c = \frac{\rho T}{n_t}$. The same goes for Σ_2 if we do the substitutions $c = 1$, $n_t = n_r$ and $n_r = T$ and instead of i starting at 1 in the summation and product, it would start at $n_t + 1$.

4.1.2 The square case

Now, for simplicity, we investigate the square case, assuming the same number of transmit- and receive antennas i.e., $n_t = n_r$. For this scenario, Σ_1 contains all the singular values of \mathbf{Y} and is full rank. We have the following distributions for the matrices in the singular value decomposition coordinates

$$\begin{aligned} f_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}) &= \frac{1}{\text{Vol}(\mathcal{S}^C(T, n_r))} = \frac{\Gamma_{n_r}(T)}{\pi^{n_r(T-1)}} \\ f_{\Sigma}(\sigma_1, \dots, \sigma_{n_t}) &= \frac{2^{n_r} \pi^{n_r(n_r-1)}}{\Gamma_{n_r}(n_r)^2 \beta^{n_r^2}} e^{-\frac{1}{\beta} \sum_{i=1}^{n_r} \sigma_i^2} \prod_{i=1}^{n_r} \sigma_i \prod_{i < j} (\sigma_i^2 - \sigma_j^2)^2, \quad \beta = \frac{\rho T}{n_r} \\ f_{\mathbf{V}}(\mathbf{V}) &= \frac{1}{\text{Vol}(\mathcal{S}(n_r, n_r))} = \frac{\Gamma_{n_r}(n_r)}{(2\pi^{n_r})^{n_r}} \\ J_{T, n_r}(\sigma_1, \dots, \sigma_{n_t}) &= \left(\prod_{i=1}^{n_r} \sigma_i^{2(T-n_r)+1} \prod_{i < j} (\sigma_i^2 - \sigma_j^2)^2 \right)^{-1} \end{aligned} \quad (42)$$

where the Jacobian follows from [25]. Plugging this into (20), we get

$$\begin{aligned}
\mathcal{I}(\mathbb{X}; \mathbb{Y}) &\leq -\mathbb{E}_{\mathbb{Y}}[\log(q_{\mathbb{Y}}(\mathbb{Y}))] - h(\mathbb{Y}|\mathbb{X}) \\
&= -\mathbb{E}_{\tilde{\mathbb{U}}, \Sigma, \mathbb{V}} \left[\log(f_{\tilde{\mathbb{U}}}(\tilde{\mathbb{U}}) f_{\Sigma}(\sigma_1, \dots, \sigma_{n_t}) f_{\mathbb{V}}(\mathbb{V}) \mathbf{J}_{T, n_r}(\sigma_1, \dots, \sigma_{n_t})) \right] - h(\mathbb{Y}|\mathbb{X}) \\
&= -\mathbb{E}_{\tilde{\mathbb{U}}} \left[\log(f_{\tilde{\mathbb{U}}}(\tilde{\mathbb{U}})) \right] - \mathbb{E}_{\Sigma} [\log(f_{\Sigma}(\sigma_1, \dots, \sigma_{n_t})) - \log(\mathbf{J}_{T, n_r}(\sigma_1, \dots, \sigma_{n_t}))] - \mathbb{E}_{\mathbb{V}} [\log(f_{\mathbb{V}}(\mathbb{V}))] - h(\mathbb{Y}|\mathbb{X}) \\
&= \log(\text{Vol}(S^C(T, n_r)) \text{Vol}(S(n_r, n_r))) - \log\left(\frac{2^{n_r} \pi^{n_r(n_r-1)}}{\Gamma_{n_r}(n_r)^2 \beta^{n_r^2}}\right) + \mathbb{E}_{\Sigma} \left[\frac{1}{\beta} \sum_{i=1}^{n_r} \sigma_i^2 \right] \\
&\quad + \mathbb{E}_{\Sigma} \left[\log\left(\frac{\prod_{i=1}^{n_r} \sigma_i^{2(T-n_r)+1} \prod_{i < j} (\sigma_i^2 - \sigma_j^2)^2}{\prod_{i=1}^{n_r} \sigma_i \prod_{i < j} (\sigma_i^2 - \sigma_j^2)^2}\right) \right] - h(\mathbb{Y}|\mathbb{X}) \\
&= \log\left(\frac{\pi^{n_r(T-1)} (2\pi^{n_r})^{n_r} \Gamma_{n_r}(n_r)^2 \beta^{n_r^2}}{\Gamma_{n_r}(T) \Gamma_{n_r}(n_r) 2^{n_r} \pi^{n_r(n_r-1)}}\right) + \mathbb{E}_{\Sigma} \left[\frac{1}{\beta} \sum_{i=1}^{n_r} \sigma_i^2 \right] + \mathbb{E}_{\Sigma} \left[\log\left(\prod_{i=1}^{n_r} \sigma_i^{2(T-n_r)}\right) \right] - h(\mathbb{Y}|\mathbb{X}) \\
&= \log\left(\frac{\pi^{n_r T} \beta^{n_r^2} \Gamma_{n_r}(n_r)}{\Gamma_{n_r}(T)}\right) + \frac{1}{\beta} \sum_{i=1}^{n_r} \mathbb{E}_{\Sigma}[\sigma_i^2] + \mathbb{E}_{\Sigma} \left[\sum_{i=1}^{n_r} \log(\sigma_i^{2(T-n_r)}) \right] - h(\mathbb{Y}|\mathbb{X}) \\
&= \log\left(\frac{\pi^{n_r T} \beta^{n_r^2} \Gamma_{n_r}(n_r)}{\Gamma_{n_r}(T)}\right) + \frac{1}{\beta} \sum_{i=1}^{n_r} \mathbb{E}_{\Sigma}[\sigma_i^2] + (T - n_r) \sum_{i=1}^{n_r} \mathbb{E}_{\Sigma}[\log(\sigma_i^2)] - h(\mathbb{Y}|\mathbb{X})
\end{aligned} \tag{43}$$

Now we need to evaluate $h(\mathbb{Y}|\mathbb{X})$ and in order to do so, we use the fact that the columns in \mathbb{Y} given \mathbb{X} are Gaussian and independent. The pdf of $\mathbb{Y}|\mathbb{X}$ is given as

$$f_{\mathbb{Y}|\mathbb{X}}(\mathbb{Y}|\mathbb{X}) = \frac{1}{\pi^{n_r T} \det\left(\frac{\rho}{n_t} \mathbb{X} \mathbb{X}^H + \mathbf{I}_T\right)^{n_r}} e^{-\text{tr}\left\{\mathbb{Y}^H \left(\frac{\rho}{n_t} \mathbb{X} \mathbb{X}^H + \mathbf{I}_T\right)^{-1} \mathbb{Y}\right\}}. \tag{44}$$

From this, the conditional entropy is given as

$$\begin{aligned}
h(\mathbb{Y}|\mathbb{X}) &= - \int_{\mathbb{X}, \mathbb{Y}} f_{\mathbb{X}, \mathbb{Y}}(\mathbb{X}, \mathbb{Y}) \log(f_{\mathbb{Y}|\mathbb{X}}(\mathbb{Y}|\mathbb{X})) d\mathbb{Y} d\mathbb{X} \\
&= n_r \sum_{i=1}^{n_t} \mathbb{E}_{\mathbb{X}} \left[\log\left(1 + \frac{\rho}{n_t} \|\mathbf{X}_i\|^2\right) \right] + n_r T \log(\pi e)
\end{aligned} \tag{45}$$

where \mathbf{X}_i is the i th column of \mathbb{X} . Substituting (45) in (43), we obtain

$$\begin{aligned}
\mathcal{I}(\mathbb{X}; \mathbb{Y}) &\leq \log\left(\frac{\beta^{n_r^2} \Gamma_{n_r}(n_r)}{e^{n_r T} \Gamma_{n_r}(T)}\right) + \underbrace{\frac{1}{\beta} \mathbb{E}_{\Sigma} \left[\sum_{i=1}^{n_r} \sigma_i^2 \right]}_A + \underbrace{(T - n_r) \sum_{i=1}^{n_r} \mathbb{E}_{\Sigma} [\log(\sigma_i^2)]}_B \\
&\quad - n_r \sum_{i=1}^{n_r} \mathbb{E}_{\mathbb{X}} \left[\log\left(1 + \frac{\rho \|\mathbf{X}_i\|^2}{n_r}\right) \right].
\end{aligned} \tag{46}$$

Note that, since the $\{\sigma_i\}_{i=1}^{n_r}$ are singular values of \mathbb{Y} , we have that

$$\begin{aligned}
A &= \mathbb{E}_{\Sigma} \left[\sum_{i=1}^{n_r} \sigma_i^2 \right] = \mathbb{E}_{\mathbb{Y}} \left[\sum_{i=1}^T \sum_{j=1}^{n_r} |\mathbb{Y}_{i,j}|^2 \right] \\
&= \mathbb{E}_{\mathbb{X}} \left[\mathbb{E}_{\mathbb{Y}|\mathbb{X}} [\text{tr}\{\mathbb{Y}^H \mathbb{Y}\}] \mid \mathbb{X} \right].
\end{aligned} \tag{47}$$

Since the squared singular values of \mathbb{Y} are the eigenvalues of the matrix $\mathbb{Y}^H \mathbb{Y}$, we start from this matrix and try to simplify it, following the same principles as in [25]. We let $\mathbb{Z} \in \mathbb{C}^{T \times n_r}$ where each entry is iid $\mathcal{CN}(0, 1)$. We can now express the distribution of the matrix, $(\mathbb{Y}^H \mathbb{Y} | \mathbb{X}) \in \mathbb{C}^{n_r \times n_r}$, as

$$(\mathbb{Y}^H \mathbb{Y} | \mathbb{X}) \stackrel{(d)}{=} \mathbb{Z}^H \left(\frac{\rho}{n_t} \mathbb{X} \mathbb{X}^H + \mathbf{I}_T \right) \mathbb{Z} \quad (48)$$

where (d) means that the equality holds in distribution. Now we express \mathbb{X} in (25) as

$$\begin{aligned} \mathbb{X} &= \Phi \mathbb{D} \\ &= [\Phi_1 \quad \Phi_2] \begin{bmatrix} \mathbb{D} \\ \mathbf{0} \end{bmatrix} \end{aligned} \quad (49)$$

where $\Phi_1 \in \mathbb{C}^{T \times n_t}$ is the same as in (25) and $\Phi_2 \in \mathbb{C}^{T \times (T-n_t)}$ is a matrix that makes $\Phi \in \mathbb{C}^{T \times T}$ unitary. We can now write (48) as

$$\begin{aligned} (\mathbb{Y}^H \mathbb{Y} | \mathbb{X}) &\stackrel{d}{=} \mathbb{Z}^H \left([\Phi_1 \quad \Phi_2] \begin{bmatrix} \sqrt{\frac{\rho}{n_t}} \mathbb{D} \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{\rho}{n_t}} \mathbb{D}^H & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Phi_1^H \\ \Phi_2^H \end{bmatrix} + \mathbf{I}_T \right) \mathbb{Z} \\ &= \mathbb{Z}^H \left(\Phi \text{diag} \left(\frac{\rho}{n_t} \|\mathbf{X}_1\|^2, \dots, \frac{\rho}{n_t} \|\mathbf{X}_{n_t}\|^2, 0, \dots, 0 \right) \Phi^H + \underbrace{\Phi \Phi^H}_{\mathbf{I}_T} \right) \mathbb{Z} \\ &= \underbrace{\mathbb{Z}^H \Phi}_{\stackrel{d}{=} \mathbb{Z}^H} \left(\text{diag} \left(\frac{\rho}{n_t} \|\mathbf{X}_1\|^2, \dots, \frac{\rho}{n_t} \|\mathbf{X}_{n_t}\|^2, 0, \dots, 0 \right) + \mathbf{I}_T \right) \underbrace{\Phi^H \mathbb{Z}}_{\stackrel{(d)}{=} \mathbb{Z}}. \end{aligned} \quad (50)$$

Further, decompose \mathbb{Z} as

$$\mathbb{Z} = \begin{bmatrix} \mathbb{Z}_1 \\ \mathbb{Z}_2 \end{bmatrix}, \quad \mathbb{Z}_1 \in \mathbb{C}^{n_t \times n_r} \quad \mathbb{Z}_2 \in \mathbb{C}^{(T-n_t) \times n_r} \quad (51)$$

and express (50) as

$$\begin{aligned} (\mathbb{Y}^H \mathbb{Y} | \mathbb{X}) &\stackrel{d}{=} [\mathbb{Z}_1^H \quad \mathbb{Z}_2^H] \begin{bmatrix} \tilde{\mathbb{D}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{T-n_t} \end{bmatrix} \begin{bmatrix} \mathbb{Z}_1 \\ \mathbb{Z}_2 \end{bmatrix} \\ &= \mathbb{Z}_1^H \tilde{\mathbb{D}} \mathbb{Z}_1 + \mathbb{Z}_2^H \mathbb{Z}_2 \end{aligned} \quad (52)$$

where $\tilde{\mathbb{D}} = \text{diag} \left(\frac{\rho}{n_t} \|\mathbf{X}_1\|^2 + 1 \right), i = 1, \dots, n_t$. Hence, we have decomposed the matrix into two Wishart matrices, the first containing the transmitted information plus noise and the second containing only information about the noise.

We are interested in the n_r non-zero singular values of \mathbb{Y} and will make use of the notation $(\mathbb{Z}_1)_{i,*}$ and $(\mathbb{Z}_1)_{*,i}$ to denote the i :th row and the i :th column of \mathbb{Z}_1 respectively. Starting from (46), we write

$$\begin{aligned} \sum_{i=1}^{n_r} \mathbb{E}_{\Sigma} [\sigma_i^2] &= \mathbb{E}_{\mathbb{X}} [\mathbb{E}_{\mathbb{Y} | \mathbb{X}} [\text{tr} \{ \mathbb{Y}^H \mathbb{Y} \} | \mathbb{X}]] \\ &= \mathbb{E}_{\mathbb{X}} \left[\mathbb{E}_{\mathbb{Z}} \left[\text{tr} \left\{ \mathbb{Z}_1^H \tilde{\mathbb{D}} \mathbb{Z}_1 \right\} + \text{tr} \left\{ \mathbb{Z}_2^H \mathbb{Z}_2 \right\} \right] \right] \\ &= \mathbb{E}_{\mathbb{X}} \left[\mathbb{E}_{\mathbb{Z}} \left[\sum_{i=1}^{n_t} \left(\frac{\rho}{n_t} \|\mathbf{X}_i\|^2 + 1 \right) (\mathbb{Z}_1)_{i,*} (\mathbb{Z}_1)_{i,*}^H + \sum_{i=1}^{n_r} (\mathbb{Z}_2)_{*,i}^H (\mathbb{Z}_2)_{*,i} \right] \right] \\ &= \mathbb{E}_{\mathbb{X}} \left[\sum_{i=1}^{n_t} \left(\frac{\rho}{n_t} \|\mathbf{X}_i\|^2 + 1 \right) \mathbb{E}_{\mathbb{Z}_1} [\|(\mathbb{Z}_1)_{i,*}\|^2 | \mathbb{X}] + \sum_{i=1}^{n_r} \mathbb{E}_{\mathbb{Z}_2} [\|(\mathbb{Z}_2)_{*,i}\|^2] \right]. \end{aligned} \quad (53)$$

Now, since $(\mathbf{Z}_1)_{i,*} \in \mathbb{C}^{1 \times n_r}$ and $(\mathbf{Z}_2)_{*,i} \in \mathbb{C}^{(T-n_t) \times 1}$ are vectors with i.i.d $\mathcal{CN}(0, 1)$ elements, their norm follow the gamma distribution as

$$\|(\mathbf{Z}_1)_{i,*}\|^2 \sim \text{Gamma}(n_r, 1) \quad \|(\mathbf{Z}_2)_{*,i}\|^2 \sim \text{Gamma}(T - n_t, 1) \quad (54)$$

which gives that

$$\mathbb{E}_{\mathbf{Z}_1} [\|(\mathbf{Z}_1)_{i,*}\|^2] = n_r, \quad \mathbb{E}_{\mathbf{Z}_2} [\|(\mathbf{Z}_2)_{*,i}\|^2] = T - n_t. \quad (55)$$

Substituting (55) into (53), we get

$$\mathbb{E}_{\mathbf{X}} \left[\sum_{i=1}^{n_t} \left(\frac{\rho}{n_t} \|\mathbf{X}_i\|^2 + 1 \right) \mathbb{E}_{\mathbf{Z}_1} [\|(\mathbf{Z}_1)_{i,*}\|^2] + \sum_{i=1}^{n_r} \mathbb{E}_{\mathbf{Z}_2} [\|(\mathbf{Z}_2)_{*,i}\|^2] \right] = n_r \mathbb{E}_{\mathbf{X}} \left[\sum_{i=1}^{n_t} \left(\frac{\rho}{n_t} \|\mathbf{X}_i\|^2 + 1 \right) \right] + (T - n_t). \quad (56)$$

Substituting this expression into (46) we get

$$\begin{aligned} \mathcal{I}(\mathbb{X}; \mathbb{Y}) &\leq \log \left(\frac{\beta^{n_r} \Gamma_{n_r}(n_r)}{e^{n_r T} \Gamma_{n_r}(T)} \right) + \frac{1}{\beta} \left(n_r \mathbb{E}_{\mathbf{X}} \left[\sum_{i=1}^{n_t} \left(\frac{\rho}{n_t} \|\mathbf{X}_i\|^2 + 1 \right) \right] + (T - n_t) \right) \\ &\quad + (T - n_r) \underbrace{\sum_{i=1}^{n_r} \mathbb{E}_{\mathbb{Z}} [\log(\sigma_i^2)]}_B - n_r \sum_{i=1}^{n_r} \mathbb{E}_{\mathbf{X}} \left[\log \left(1 + \frac{\rho \|\mathbf{X}_i\|^2}{n_r} \right) \right]. \end{aligned} \quad (57)$$

Next, we will simplify B in (57) which may be expressed as

$$\begin{aligned} \mathbb{E}_{\mathbb{Z}} \left[\sum_{i=1}^{n_r} \log(\sigma_i^2) \right] &= \mathbb{E}_{\mathbb{Z}} \left[\log \left(\prod_{i=1}^{n_r} \sigma_i^2 \right) \right] \\ &= \mathbb{E}_{\mathbf{X}} [\mathbb{E}_{\mathbf{Y}|\mathbf{X}} [\log(\det(\mathbf{Y}^H \mathbf{Y})) | \mathbf{X}]]] \end{aligned} \quad (58)$$

and by using the decomposition in (52), we obtain

$$\mathbb{E}_{\mathbf{X}} \left[\mathbb{E}_{\mathbf{Z}_1, \mathbf{Z}_2 | \mathbf{X}} \left[\log \left(\det \left(\mathbf{Z}_1^H \tilde{\mathbf{D}} \mathbf{Z}_1 + \mathbf{Z}_2^H \mathbf{Z}_2 \right) \right) \right] \right]. \quad (59)$$

Noting that both of the matrices in the determinant is Hermitian and that the sum of two Hermitian matrices is also Hermitian, we can use the property

$$\log(\det(\mathbf{A})) = \text{tr} \{ \text{logm}(\mathbf{A}) \} \quad (60)$$

where \mathbf{A} is Hermitian and $\text{logm}(\cdot)$ denotes the matrix logarithm defined as [32, pp. 525]

$$\text{logm}(\mathbf{A}) = \mathbf{U} \log(\Lambda(\mathbf{A})) \mathbf{U}^H \quad (61)$$

where $\mathbf{U}\Lambda(\mathbf{A})\mathbf{U}^H$ is the eigen-decomposition of \mathbf{A} and the log operates on each of the diagonal elements in $\Lambda(\mathbf{A})$. By utilizing (60) in (59) we can write

$$\mathbb{E}_{\mathbf{X}} \left[\mathbb{E}_{\mathbf{Z}_1, \mathbf{Z}_2 | \mathbf{X}} \left[\log \left(\det \left(\mathbf{Z}_1^H \tilde{\mathbf{D}} \mathbf{Z}_1 + \mathbf{Z}_2^H \mathbf{Z}_2 \right) \right) \right] \right] = \mathbb{E}_{\mathbf{X}} \left[\text{tr} \left\{ \mathbb{E}_{\mathbf{Z}_1 | \mathbf{X}} \left[\mathbb{E}_{\mathbf{Z}_2 | \mathbf{X}} \left[\text{logm} \left(\mathbf{Z}_1^H \tilde{\mathbf{D}} \mathbf{Z}_1 + \mathbf{Z}_2^H \mathbf{Z}_2 \right) \right] \right] \right\} \right] \quad (62)$$

Now, looking in detail at the diagonal of the matrix summation inside the logarithm, the diagonal of the first multiplication is given as

$$\mathbf{Z}_1^H \tilde{\mathbf{D}} \mathbf{Z}_1 = \begin{bmatrix} \tilde{D}_1 |(Z_1)_{1,1}|^2 + \dots + \tilde{D}_{n_t} |(Z_1)_{n_t,1}|^2 & & \\ & \ddots & \\ & & \tilde{D}_1 |(Z_1)_{1,n_r}|^2 + \dots + \tilde{D}_{n_t} |(Z_1)_{n_t,n_r}|^2 \end{bmatrix}.$$

Since $(\mathbb{Z}_1)_{i,j} \sim \mathcal{CN}(0, 1)$, each diagonal entry is distributed as the sum of n_t scaled Gamma(1, 1) random variables. Analogously, we consider the diagonal of the second product as

$$\mathbb{Z}_2^H \mathbb{Z}_2 = \begin{bmatrix} \|(\mathbb{Z}_2)_{*,1}\|^2 & & \\ & \ddots & \\ & & \|(\mathbb{Z}_2)_{*,n_r}\|^2 \end{bmatrix}$$

where each diagonal element is distributed as a Gamma($T - n_t, 1$) random variable. Hence, the diagonal entries of (48), denoted $\{\kappa_{ii}\}_{i=1}^{n_r}$, are given according to $\sum_{j=1}^{n_t} \tilde{D}_j |(\mathbb{Z}_1)_{j,i}|^2 + \|(\mathbb{Z}_2)_{*,i}\|^2$ for $i = 1, \dots, n_r$. Hence, each diagonal entries are iid as the sum of $n_t + 1$ Gamma distributed random variables. The pdf of each of the diagonal element is given as [33]

$$f_{\kappa_{ii}}(a) = \prod_{i=1}^{n_t} \left(\frac{1}{\xi_i} \right) \sum_{k=0}^{\infty} \frac{\delta_k a^{T+k-1} e^{-a}}{\Gamma(T+k)} \quad (63)$$

where

$$\begin{aligned} \xi_i &= \begin{cases} 1 + \frac{\rho}{n_t} \|\mathbf{X}_i\|^2 & \text{if } i = 1, \dots, n_t \\ 1 & \text{if } i = n_t + 1 \end{cases} \\ \delta_{k+1} &= \frac{1}{k+1} \sum_{i=1}^{k+1} i \gamma_i \delta_{k+1-i}, \quad k = 0, 1, \dots \\ \gamma_i &= \sum_{j=1}^{n_t} \frac{\left(1 - \frac{1}{\xi_j}\right)^i}{i}. \end{aligned} \quad (64)$$

For the SISO case ($n_t = n_r = 1$), the matrix logarithm is just an ordinary logarithm and we have

$$\begin{aligned} \mathbb{E}_{\mathbb{X}} \left[\sum_{i=1}^{n_r} \mathbb{E}_{\kappa_{ii}|\mathbb{X}} \left[\log \left(\underbrace{\mathbb{Z}_1^H \tilde{\mathbf{D}} \mathbb{Z}_1 + \mathbb{Z}_2^H \mathbb{Z}_2}_a \right)_{ii} \mid \mathbb{X} \right] \right] &= \mathbb{E}_{\mathbb{X}} \left[\int_0^{\infty} \log(a) f_{\kappa_{ii}}(a) da \right] \\ &= \mathbb{E}_{\mathbb{X}} \left[\frac{1}{\xi} \sum_{k=0}^{\infty} \frac{\delta_k}{\Gamma(T+k)} \int_0^{\infty} \log(a) a^{T+k-1} e^{-a} da \right] \\ &= \mathbb{E}_{\mathbb{X}} \left[\frac{1}{1 + \|\mathbf{X}\|^2} \sum_{k=0}^{\infty} \delta_k \psi(T+k) \right] \end{aligned} \quad (65)$$

where $\psi(\cdot)$ denotes the digamma function. Now the dual problem in (23) can be completely stated. For the SISO case, one can use (65) to obtain an expression for the upper bound of the mutual information stated in (20) as

$$\begin{aligned} \mathcal{I}(\mathbf{X}; \mathbf{Y}) &\leq \log \left(\frac{\beta}{e^T \Gamma(T)} \right) + \frac{\mathbb{E}_{\mathbf{X}} [\|\mathbf{X}\|^2] + T - 1}{\beta} + (T - 1) \mathbb{E}_{\mathbf{X}} \left[\frac{1}{1 + \|\mathbf{X}\|^2} \sum_{k=0}^{\infty} \delta_k \psi(T+k) \right] \\ &\quad - \mathbb{E}_{\mathbf{X}} [\log(1 + \|\mathbf{X}\|^2)] \\ &= \underbrace{\log \left(\frac{\beta}{e^T \Gamma(T)} \right) + \frac{T - 1}{\beta}}_{e^{\text{SISO}(\rho)}} + \mathbb{E}_{\mathbf{X}} \left[\frac{1}{\beta} (\|\mathbf{X}\|^2) + \frac{T - 1}{1 + \|\mathbf{X}\|^2} \sum_{k=0}^{\infty} \delta_k \psi(T+k) - \log(1 + \|\mathbf{X}\|^2) \right]. \end{aligned} \quad (66)$$

For simplicity, we define $f^{\text{SISO}}(\mathbf{X}) : \mathbb{R}^{T \times 1} \rightarrow \mathbb{R}$ as

$$f^{\text{SISO}}(\mathbf{X}) \triangleq \left(\frac{1}{\beta} (\|\mathbf{X}\|^2) + \frac{T-1}{1+\|\mathbf{X}\|^2} \sum_{k=0}^{\infty} \delta_k \psi(T+k) - \log(1+\|\mathbf{X}\|^2) \right). \quad (67)$$

By substituting (67) and $c^{\text{SISO}}(\rho)$ in (23), we end up with the SISO dual problem

$$\begin{aligned} C(\rho) &\leq \frac{1}{T} \inf_{\lambda \geq 0} \sup_{P_{\mathbf{X}}} \{c^{\text{SISO}}(\rho) + \mathbb{E}_{\mathbf{X}}[f^{\text{SISO}}(\mathbf{X})] + \lambda(T - \mathbb{E}_{\mathbf{X}}[\mathbf{X}\mathbf{X}^H])\} \\ &= \frac{c^{\text{SISO}}(\rho)}{T} - \frac{1}{T} \inf_{\lambda \geq 0} \sup_{P_{\mathbf{X}}} \{\mathbb{E}_{\mathbf{X}}[f^{\text{SISO}}(\mathbf{X})] + \lambda(T - \|\mathbf{X}\|^2)\} \\ &\leq \frac{c^{\text{SISO}}(\rho)}{T} - \frac{1}{T} \inf_{\lambda \geq 0} \sup_{\|\mathbf{X}\|} \{f^{\text{SISO}}(\mathbf{X}) + \lambda(T - \|\mathbf{X}\|^2)\} \\ &\triangleq U^{\text{SISO}}(\rho). \end{aligned} \quad (68)$$

The last inequality can be understood from the fact that the expectation of a function of a rv is never larger than the maximum value taken by the function over the support of the rv.

For the MIMO-case, the procedure used in the SISO case gets complicated due to the matrix logarithm. It is possible to bound the expression in (62) rather than computing it exactly. Unfortunately, though, this bound is not very tight. In fact, with increasing T , the bound exceeds the coherent capacity. For the complete derivation and illustrations, refer to appendix A. We will now try to evaluate (47) in another way.

Again, utilizing that the columns of $\mathbb{Y} | \mathbb{X} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$ where $\mathbf{C} = \mathbf{I}_T + \tilde{\rho} \mathbf{X} \mathbf{X}^H$ and $\tilde{\rho} = \rho/n_t$, we write

$$\mathbb{E}_{\mathbb{X}}[\mathbb{E}_{\mathbb{Y} | \mathbb{X}}[\log(\det(\mathbb{Y}\mathbb{Y}^H)) | \mathbb{X}]] = \mathbb{E}_{\mathbb{X}}[\mathbb{E}_{\mathbb{Z} | \mathbb{X}}[\log(\det(\mathbb{Z}\mathbb{C}\mathbb{Z}^H)) | \mathbb{X}]] \quad (69)$$

where the entries of \mathbb{Z} are iid $\mathcal{CN}(0, 1)$. We notice that (69) is derived in closed form in [34, Lemma 2] where the moment generating function (mgf) is utilized. From [34, Eq. 98] we have the mgf as

$$\begin{aligned} M(t) &= \mathbb{E}_{\mathbb{Z} | \mathbb{X}} \left[e^{t \log(\det(\mathbb{Z}\mathbb{C}\mathbb{Z}^H))} | \mathbb{X} \right] \\ &= \mathbb{E}_{\mathbb{Z} | \mathbb{X}} \left[\det(\mathbb{Z}\mathbb{C}\mathbb{Z}^H)^t | \mathbb{X} \right] \\ &= \frac{1}{\det(\Omega) \prod_{l=1}^{n_r} l!} \int_{\xi_1, \dots, \xi_{n_r}} \left(\prod_{i=1}^{n_r} \xi_i^t \right) \det(\Gamma) \det(\Xi) d\xi_1, \dots, d\xi_{n_r}. \end{aligned} \quad (70)$$

where Ω is a vandermonde matrix containing the extended eigenvalues of the covariance matrix, \mathbf{C} , given as

$$\Omega = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{T-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{T-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_{n_t} & \lambda_{n_t}^2 & \cdots & \lambda_{n_t}^{T-1} \\ 1 & \lambda'_{n_t+1} & \lambda_{n_t+1}^2 & \cdots & \lambda_{n_t+1}^{T-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda'_T & \lambda_T^2 & \cdots & \lambda_T^{T-1} \end{bmatrix} \in \mathbb{R}^{T \times T}. \quad (71)$$

The matrix, Ξ , is also a Vandermonde matrix containing the singular values of $\mathbb{Z}\mathbb{C}\mathbb{Z}^H$ and is

given as

$$\Xi = \begin{bmatrix} 1 & \xi_1 & \xi_1^2 & \cdots & \xi_1^{n_r-1} \\ 1 & \xi_2 & \xi_2^2 & \cdots & \xi_2^{n_r-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_{n_r} & \xi_{n_r}^2 & \cdots & \xi_{n_r}^{n_r-1} \end{bmatrix} \in \mathbb{R}^{n_r \times n_r} \quad (72)$$

and the matrix, Γ is given as

$$\Gamma = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{T-n_r-1} & \lambda_1^{T-n_r-1} e^{-\xi_1/\lambda_1} & \cdots & \lambda_1^{T-n_r-1} e^{-\xi_{n_r}/\lambda_1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{T-n_r-1} & \lambda_2^{T-n_r-1} e^{-\xi_1/\lambda_2} & \cdots & \lambda_2^{T-n_r-1} e^{-\xi_{n_r}/\lambda_2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & \lambda_{n_t} & \lambda_{n_t}^2 & \cdots & \lambda_{n_t}^{T-n_r-1} & \lambda_{n_t}^{T-n_r-1} e^{-\xi_1/\lambda_{n_t}} & \cdots & \lambda_{n_t}^{T-n_r-1} e^{-\xi_{n_r}/\lambda_{n_t}} \\ 1 & \lambda'_{n_t+1} & \lambda_{n_t+1}^2 & \cdots & \lambda_{n_t+1}^{T-n_r-1} & \lambda_{n_t+1}^{T-n_r-1} e^{-\xi_1/\lambda'_{n_t+1}} & \cdots & \lambda_{n_t+1}^{T-n_r-1} e^{-\xi_{n_r}/\lambda'_{n_t+1}} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & \lambda'_T & \lambda_T^2 & \cdots & \lambda_T^{T-n_r-1} & \lambda_T^{T-n_r-1} e^{-\xi_1/\lambda'_T} & \cdots & \lambda_T^{T-n_r-1} e^{-\xi_{n_r}/\lambda'_T} \end{bmatrix} \in \mathbb{R}^{T \times T}. \quad (73)$$

We will consider Γ as a block matrix given as

$$\Gamma = \begin{bmatrix} \Upsilon & \mathbb{G} \\ \mathbb{C} & \mathbb{B} \end{bmatrix}, \quad \Upsilon \in \mathbb{R}^{n_r \times (T-n_r)}, \mathbb{G} \in \mathbb{R}^{n_r \times n_r}, \mathbb{C} \in \mathbb{R}^{(T-n_r) \times (T-n_r)}, \mathbb{B} \in \mathbb{R}^{(T-n_r) \times n_r}. \quad (74)$$

For the expression in (70) to be valid, the eigenvalues of \mathbb{C} must be distinct. This is not the true in our case since the eigenvalues, $\Lambda(\mathbb{C})$, are obtained as

$$\begin{aligned} \mathbb{C} &= \mathbb{I}_T + \tilde{\rho} \mathbf{X} \mathbf{X}^H \\ &= \mathbb{I}_T + \tilde{\rho} \Phi \mathbb{D}^2 \Phi^H \\ &= \Phi \underbrace{(\mathbb{I}_T + \tilde{\rho} \mathbb{D}^2)}_{\Lambda(\mathbb{C})} \Phi^H \end{aligned} \quad (75)$$

where $\mathbb{D} = \text{diag}(\|\mathbf{x}_1\|^2, \dots, \|\mathbf{x}_{n_t}\|^2, 0, \dots, 0)$. Hence the eigenvalues are given as

$$\Lambda(\mathbb{C}) = \text{diag}(1 + \|\mathbf{x}_1\|^2, \dots, 1 + \|\mathbf{x}_{n_t}\|^2, 1, \dots, 1). \quad (76)$$

Obviously, the $T - n_t$ last eigenvalues are not distinct. Therefore, the determinants involving the eigenvalues of \mathbb{C} in [34, Eq. 98] will evaluate to zero. This is circumvented by letting $\Lambda' = \text{diag}(1 + \|\mathbf{x}_1\|^2, \dots, 1 + \|\mathbf{x}_{n_t}\|^2, 1 + \epsilon_{n_t+1}, \dots, 1 + \epsilon_T)$ and instead use this in the computation of the moment generating function (mgf) in [34, Eq. 98] and take the limit. Note that the limit only affects \mathbb{C} and \mathbb{B} in Γ . Now, we write the mgf as

$$\begin{aligned} M(t) &= \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \frac{1}{\det(\Omega) \prod_{l=1}^{n_r} l!} \int_{\xi_1, \dots, \xi_{n_r}} \left(\prod_{i=1}^{n_r} \xi_i^t \right) \det(\Gamma) \det(\Xi) d\xi_1, \dots, d\xi_{n_r} \\ &= \frac{1}{\prod_{l=1}^{n_r} l!} \int_{\xi_1, \dots, \xi_{n_r}} \left(\prod_{i=1}^{n_r} \xi_i^t \right) \left(\lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \frac{\det(\Gamma)}{\det(\Omega)} \right) \det(\Xi) d\xi_1, \dots, d\xi_{n_r} \end{aligned} \quad (77)$$

and evaluate the limit by iteratively applying L'hopitals rule, see Appendix B.

After taking the limit, we use the mgf in (77) to derive the expected value of (76) as

$$\begin{aligned} \mathbb{E}_{\mathbb{Y}|\mathbb{X}}[\log(\det(\mathbb{Y}\mathbb{Y}^H) | \mathbb{X})] &= \frac{d}{dt} M(t) \Big|_{t=0} \\ &= \frac{(-1)^{\frac{(T-1)(T+4)}{2} - T - n_r - 1}}{\prod_{l=1}^{n_r-1} l! \prod_{i=1}^{n_r} (1 - \lambda_i)^{T - n_r} \det(\mathbb{V}(\Lambda(\mathbb{C})))} \frac{d}{dt} \det(\mathbb{A}) \Big|_{t=0} \end{aligned} \quad (78)$$

where $\det(V(\cdot))$ denotes the determinant of a Vandermonde matrix and \mathbf{A} is given in appendix B. The derivative of the determinant of \mathbf{A} may be expressed by the determinant chain-rule as the sum of n_r determinants as [32, pp 471.]

$$\frac{d}{dt} \det(\mathbf{A}) = \sum_{i=1}^{n_r} \det(\tilde{\mathbf{A}}_i) \quad (79)$$

where $\tilde{\mathbf{A}}_i$ is an $n_r \times n_r$ matrix where the i th column is differentiated at $t = 0$ and the remaining elements are the elements from \mathbf{A} evaluated at $t = 0$. The entries of $\tilde{\mathbf{A}}_i$ for $t = 0$ are

$$(\tilde{\mathbf{A}}_i)_{k,l} = \begin{cases} \lambda_k^{T-n_r-1+l-1} \Gamma(l) - \sum_{d=1}^{T-n_r} \sum_{q=1}^{T-n_r} \left((C')_{d,q}^{-1} \lambda_k^{d-1} \binom{T-q-n_r}{q} \right. \\ \quad \left. \times \sum_{j=0}^{T-q-1} \frac{\Gamma(T-n_r-j)\Gamma(j+l)}{\Gamma(q+2n_r)} - \Gamma(T-q-n_r+l) \right), & l = i \\ \lambda_k^{T-n_r-1+l-1} \Gamma(l) (\log(\lambda_k) + \psi(l)) - \sum_{d=1}^{T-n_r} \sum_{q=1}^{T-n_r} \left((C')_{d,q}^{-1} \lambda_k^{d-1} \binom{T-q-n_r}{q} \right. \\ \quad \left. \times \sum_{j=0}^{T-q-1} \frac{\Gamma(T-n_r-j)\Gamma(j+l)\psi(i+l)}{\Gamma(q+2n_r)} - \Gamma(T-q-n_r+l)\psi(T-q-n_r+l) \right), & l \neq i \end{cases} \quad (80)$$

Hence, we have that

$$\mathbb{E}_{\mathbb{X}} \left[\sum_{i=1}^{n_r} \log(\sigma_i^2) \right] = \underbrace{\frac{(-1)^{\binom{(T-1)(T+4)}{2} - T - n_r - 1}}{\prod_{l=1}^{n_r-1} l! \prod_{i=1}^{n_r} (1 - \lambda_i)^{T-n_r}}}_{c_B} \mathbb{E}_{\mathbb{X}} \left[\frac{\sum_{i=1}^{n_r} \det(\tilde{\mathbf{A}}_i)}{\det(\mathbf{V}(\Lambda(C)))} \right]. \quad (81)$$

Substituting (81) into (57), we get the mutual information upper bound as

$$\begin{aligned} \mathcal{I}(\mathbb{X}; \mathbb{Y}) &\leq \log \left(\frac{\beta^{n_r^2} \Gamma_{n_r}(n_r)}{e^{n_r T} \Gamma_{n_r}(T)} \right) + \frac{1}{\beta} \left(n_r \mathbb{E}_{\mathbb{X}} \left[\sum_{i=1}^{n_t} \left(\frac{\rho}{n_t} \|\mathbf{X}_i\|^2 + 1 \right) \right] + (T - n_t) \right) \\ &\quad + (T - n_r) c_B \mathbb{E}_{\mathbb{X}} \left[\frac{\sum_{i=1}^{n_r} \det(\tilde{\mathbf{A}}_i)}{\det(\mathbf{V}(\Lambda(C)))} \right] - n_r \sum_{i=1}^{n_t} \mathbb{E}_{\mathbb{X}} \left[\log \left(1 + \frac{\rho \|\mathbf{X}_i\|^2}{n_t} \right) \right] \\ &= \underbrace{\log \left(\frac{\beta^{n_r^2} \Gamma_{n_r}(n_r)}{e^{n_r T} \Gamma_{n_r}(T)} \right) \frac{(T - n_t)}{\beta}}_{c^{\text{MIMO}}(\rho)} \\ &\quad + \mathbb{E}_{\mathbb{X}} \left[\frac{n_r}{\beta} \sum_{i=1}^{n_t} \left(\frac{\rho}{n_t} \|\mathbf{X}_i\|^2 + 1 \right) + \frac{(T - n_r) c_B \sum_{i=1}^{n_r} \det(\tilde{\mathbf{A}}_i)}{\det(\mathbf{V}(\Lambda(C)))} - n_r \sum_{i=1}^{n_t} \log \left(1 + \frac{\rho \|\mathbf{X}_i\|^2}{n_t} \right) \right] \end{aligned} \quad (82)$$

For simplicity, we define $f^{\text{MIMO}}(\mathbb{X}) : \mathbb{R}^{T \times n_t} \rightarrow \mathbb{R}$ as

$$f^{\text{MIMO}}(\mathbb{X}) \triangleq \frac{n_r}{\beta} \sum_{i=1}^{n_t} \left(\frac{\rho}{n_t} \|\mathbf{X}_i\|^2 + 1 \right) + \frac{(T - n_r) c_B \sum_{i=1}^{n_r} \det(\tilde{\mathbf{A}}_i)}{\det(\mathbf{V}(\Lambda(C)))} - n_r \sum_{i=1}^{n_t} \log \left(1 + \frac{\rho \|\mathbf{X}_i\|^2}{n_t} \right). \quad (83)$$

By substituting (83) and $c^{\text{MIMO}}(\rho)$ in (23), we end up with the MIMO dual problem

$$\begin{aligned}
C(\rho) &\leq \frac{1}{T} \inf_{\lambda \geq 0} \sup_{P_{\mathbb{X}}} \{c^{\text{MIMO}}(\rho) + \mathbb{E}_{\mathbb{X}}[f^{\text{MIMO}}(\mathbb{X})] + \lambda (n_t T - \mathbb{E}_{\mathbb{X}}[\text{tr}\{\mathbb{X}\mathbb{X}^{\text{H}}\}])\} \\
&= \frac{c^{\text{MIMO}}(\rho)}{T} - \frac{1}{T} \inf_{\lambda \geq 0} \sup_{P_{\mathbb{X}}} \{\mathbb{E}_{\mathbb{X}}[f^{\text{MIMO}}(\mathbb{X})] + \lambda (n_t T - \text{tr}\{\mathbb{X}\mathbb{X}^{\text{H}}\})\} \\
&\leq \frac{c^{\text{MIMO}}(\rho)}{T} - \frac{1}{T} \inf_{\lambda \geq 0} \sup_{\|\mathbf{X}_i\|} \left\{ f^{\text{MIMO}}(\mathbb{X}) + \lambda \left(n_t T - \sum_{i=1}^{n_t} \|\mathbf{X}_i\|^2 \right) \right\}, \quad i = 1, \dots, n_t \\
&\triangleq U^{\text{MIMO}}(\rho).
\end{aligned} \tag{84}$$

The last inequality can be understood from the fact that the expectation of a function of a rv is never larger than the maximum value taken by the function over the support of the rv.

4.2 Capacity lower bound

To lower bound the capacity, a straightforward approach is to simply assume an input distribution, $P_{\mathbb{X}}$, derive the corresponding output distribution and evaluate the mutual information. This would result in a lower bound of the capacity since the supremum in (18) is ignored. As stated in Section 4.1, the input distribution that attains capacity in the high SNR regime is the scaled unitary id called USTM. Furthermore, as mentioned in Section 2, the distribution that maximizes the RCEE is also very similar to the USTM distribution. With these arguments, we assume USTM as the input distribution with the belief that the lower bound will be tight for high SNR. This is in fact the distribution that is used in [9] to obtain the output distribution. Unfortunately, this result is not numerically stable for $T \approx 35$ and larger at an SNR of 6 dB. This is not acceptable for the purpose of the thesis since we are also interested in the slow fading scenario.

We will obtain a lower bound on the capacity by assuming USTM as the input distribution i.e. $\mathbb{X} = \sqrt{T}\Phi$. The approach is the same as in the method used by Hassibi and Marzetta [9] but the methodology of evaluating the expectation in (88) is different from [9, Eq. (12)] and the results are more numerically stable. We start by observing that, under USTM, $(\mathbb{Y} | \mathbb{X}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_T + \tilde{\rho}T)$ where $\tilde{\rho} = \rho/n_t$. By using Woodbury's matrix identity [32, pp. 124], we write the conditional p.d.f. as in [13, Eq. (11)]

$$f_{\mathbb{Y}|\mathbb{X}}(\mathbb{Y} | \mathbb{X}) = \frac{1}{\pi^{n_r T} \det(\mathbf{I}_{n_t} + \tilde{\rho}T)^{n_r}} e^{-\text{tr}\{\mathbb{Y}^{\text{H}}\mathbb{Y}\}} e^{-\text{tr}\{\mathbb{Y}^{\text{H}}\Phi((\tilde{\rho}T)^{-1} + \mathbf{I}_{n_t})^{-1}\Phi^{\text{H}}\mathbb{Y}\}} \tag{85}$$

This implies that

$$f_{\mathbb{Y}}(\mathbb{Y}) = \frac{1}{\pi^{n_r T} \det(\mathbf{I}_{n_t} + \tilde{\rho}T)^{n_r}} \mathbb{E}_{\Phi} \left[e^{\text{tr}\{\mathbb{Y}^{\text{H}}\Phi((\tilde{\rho}T)^{-1} + \mathbf{I}_{n_t})^{-1}\Phi^{\text{H}}\mathbb{Y}\}} \right] \tag{86}$$

Now, we consider the eigenvalue decomposition of $\mathbb{Y}\mathbb{Y}^{\text{H}}$

$$\mathbb{Y}\mathbb{Y}^{\text{H}} = \mathbf{U} \underbrace{\begin{pmatrix} \Sigma & \mathbf{0}_{n_r \times (T-n_r)} \\ \mathbf{0}_{(T-n_r) \times n_r} & \mathbf{0}_{(T-n_r)} \end{pmatrix}}_{\Delta} \mathbf{U}^{\text{H}} \tag{87}$$

where $\mathbf{U} \in \mathbb{C}^{T \times T}$ is unitary and Σ is a diagonal matrix containing the singular values of \mathbb{Y} . We

let $\Lambda \triangleq ((\tilde{\rho}T)^{-1}\mathbf{I}_{n_t} + \mathbf{I}_{n_t})^{-1}$ and proceed as follows

$$\begin{aligned} \mathbb{E}_{\Phi} \left[e^{\text{tr}\{\mathbf{Y}^H \Phi \Lambda \Phi^H \mathbf{Y}\}} \right] &= \mathbb{E}_{\Phi} \left[e^{\text{tr}\{\Delta \mathbf{U}^H \Phi \Lambda \Phi^H \mathbf{U}\}} \right] \\ &\stackrel{(a)}{=} \mathbb{E}_{\Phi} \left[e^{\text{tr}\{\Delta \Phi \Lambda \Phi^H\}} \right] \\ &= \frac{1}{|S(T, n_t)|} \int_{S(T, n_t)} e^{\text{tr}\{\Delta \Phi \Lambda \Phi^H\}} d\Phi \end{aligned} \quad (88)$$

where (a) follows because $\Phi \stackrel{d}{=} \mathbf{U}^H \Phi$. The integral on the RHS resembles the Itzykson-Zuber integral [35] with the difference of the integral being over a the Stiefel manifold $S(T, n_t)$ instead of the unitary group $\mathcal{U}(T) \triangleq S(T, T)$. By letting $\tilde{\Phi} = [\Phi \quad \Phi_{\perp}]$ where Φ_{\perp} is chosen such that $\tilde{\Phi}$ is unitary we may write the integral as

$$\int_{S(T, n_t)} e^{\text{tr}\{\Delta \Phi \Lambda \Phi^H\}} d\Phi = \frac{1}{|\mathcal{U}(T - n_t)|} \int_{\mathcal{U}(T)} e^{\text{tr}\{\Delta \tilde{\Phi} \Lambda \tilde{\Phi}^H\}} d\tilde{\Phi}. \quad (89)$$

The Itzykson-Zuber integral is valid only for distinct eigenvalues of the two matrices Δ and Λ . Therefore, we let $\Delta_{\epsilon} \triangleq \text{diag} \left([\sigma_1^2 \cdots \sigma_{n_r}^2 \quad \epsilon_{n_r+1} \cdots \epsilon_T]^T \right)$ and $\Lambda'_{\epsilon} \triangleq \text{diag} \left([\lambda + \epsilon'_1 \cdots \lambda + \epsilon'_{n_t} \quad \epsilon'_{n_t+1} \cdots \epsilon'_T]^T \right)$ and write (89) as

$$\int_{\mathcal{U}(T)} e^{\text{tr}\{\Delta \tilde{\Phi} \Lambda \tilde{\Phi}^H\}} d\tilde{\Phi} = \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \lim_{\epsilon'_1 \rightarrow 0, \dots, \epsilon'_{n_t} \rightarrow 0} \lim_{\epsilon'_{n_t+1} \rightarrow 0, \dots, \epsilon'_T \rightarrow 0} \int_{\mathcal{U}(T)} e^{\text{tr}\{\Delta_{\epsilon} \tilde{\Phi} \Lambda'_{\epsilon} \tilde{\Phi}^H\}} d\tilde{\Phi} \quad (90)$$

The integral on the RHS now resembles the Itzykson-Zuber integral and is evaluated in appendix C. The result, for $n_t = n_r$, is given as

$$\int_{\mathcal{U}(T)} e^{\text{tr}\{\Delta \tilde{\Phi} \Lambda \tilde{\Phi}^H\}} d\tilde{\Phi} = \frac{|\mathcal{U}(T)| \prod_{i=T-n_t+1}^T \Gamma(i)}{\prod_{i=1}^{n_t} (\lambda)^{T-n_t} \prod_{i=1}^{n_t-1} i!} \frac{1}{\prod_{j=1}^{T-n_r-1} j! \prod_{i=1}^{n_r} \sigma_i^{T-n_t}} \frac{\det(\tilde{\mathbf{A}}) e^{\lambda \text{tr}\{\Sigma\}} \prod_{j=1}^{T-n_r-1} j!}{\det(\mathbf{V}(\Sigma))} \quad (91)$$

where σ_i is the i th diagonal element of Σ , $\mathbf{V}(\cdot)$ denotes the Vandermode matrix and the matrix $\tilde{\mathbf{A}}$, is given as

$$\tilde{\mathbf{A}} = \left[\sigma_k^{(n_t-l)} \tilde{\gamma}(T+l-2n_t, \lambda \sigma_k) \right]_{1 \leq l, k \leq n_t} \in \mathbb{R}^{n_t \times n_t}. \quad (92)$$

where $\tilde{\gamma}(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt$ denotes the regularized incomplete gamma function. Finally, substituting (91) into (86), we obtain the output pdf as

$$f_{\mathbf{Y}}(\mathbf{Y}) = \frac{(1 + \tilde{\rho}T)^{n_t(T-n_t-n_r)} \Gamma_{n_t}(T)}{\pi^{n_r T} (\tilde{\rho}T)^{n_t(T-n_t)} \Gamma_{n_t}(n_t)} \frac{e^{-(1+\tilde{\rho}T)^{-1} \text{tr}\{\Sigma\}} \det(\tilde{\mathbf{A}})}{\det(\Sigma^{(T-n_r)}) \prod_{i < j}^{n_r} (\sigma_i - \sigma_j)}. \quad (93)$$

The capacity lower bound may now be obtained from (18), (85) and (93) as

$$L(\rho) \triangleq \frac{1}{T} \mathbb{E}[i(\mathbb{X}; \mathbb{Y})] \quad (94)$$

where $i(\mathbb{X}; \mathbb{Y})$ is the information density given by

$$\begin{aligned}
i(\mathbb{X}; \mathbb{Y}) &\triangleq \log \left(\frac{f_{\mathbb{Y}|\mathbb{X}}(\mathbb{Y}|\mathbb{X})}{f_{\mathbb{Y}}(\mathbb{Y})} \right) \\
&= \log \left(\frac{(\tilde{\rho}T)^{n_t(T-n_t)}}{(1+\tilde{\rho}T)^{(n_t(T-n_t-n_r)+n_r n_t)}} \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} \right) + \log \left(\frac{\det(\Sigma^{(T-n_t)}) \det(V(\Sigma))}{\det(\tilde{\mathbf{A}})} \right) \\
&\quad - \text{tr} \left\{ \mathbf{Y}^H \left(\frac{\rho}{n_t} \mathbf{X}\mathbf{X}^H + \mathbf{I}_T \right)^{-1} \mathbf{Y} \right\} + (1+\tilde{\rho})^{-1} \text{tr} \{ \mathbf{Y}^H \mathbf{Y} \} \\
&\stackrel{(a)}{=} c_{n_t, T} - \text{tr} \{ \mathbf{W}^H \mathbf{W} \} + (1+\tilde{\rho})^{-1} \text{tr} \{ \mathbf{Y}^H \mathbf{Y} \} + \log \left(\frac{\det(\Sigma^{(T-n_t)}) \det(V(\Sigma))}{\det(\tilde{\mathbf{A}})} \right)
\end{aligned} \tag{95}$$

where in (a), we have used that $n_t = n_r$, $c_{n_t, T} = (n_t(T-n_t)) \log \left(\frac{\rho T}{n_t + \rho T} \right) + \log \left(\frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} \right)$ and the matrix \mathbf{W} have iid $\mathcal{CN}(0, 1)$ entries.

5 Bounds on the maximal achievable rate

This section presents the methods used to obtain the upper-and lower bounds on the maximal achievable rate, $R^*(n, \epsilon)$. First, two upper bounds are derived where one is mainly interesting in the fast-fading scenario. The other upper bound is realistic for a typical initial transmission where the sender do not know anything about the channel and transmits from each antenna with equal power. Thereafter, a lower bound on $R^*(n, \epsilon)$ is derived.

5.1 Rate upper bound

To obtain an upper bound on the maximal achievable rate, we will use the bounds presented in (13) and (12). The former is given as

$$R^*(n, \epsilon) \leq \bar{R}(n, \epsilon) \triangleq \frac{C(\rho) + \frac{H_b(\epsilon)}{n}}{1 - \epsilon} \quad (96)$$

Where $H(\epsilon)$ is the binary entropy function defined in Section 2 and $C(\rho)$ is the non-coherent capacity. Since the capacity, $C(\rho)$, is not known in closed form it will be replaced by the upper bound from Section 4.1. Since (96) is based on the non-coherent capacity, it will, for large T converge towards the ergodic capacity. This is intuitive since the non-coherent capacity can be seen as a measure of the penalty in estimating the channel if it is compared to the ergodic capacity. However, as T grow large, the penalty becomes smaller and smaller and the non-coherent capacity converges to the ergodic capacity. However, for large T i.e., in a slow fading scenario, the codeword sees, in the extreme case, only one realization per transmission and the capacity should be zero since the error rate can not be arbitrarily fixed. This upper bound does not catch this behaviour and therefore, the upper bound on $R^*(n, \epsilon)$ is loose for the slow-fading scenario.

One way to get an idea of how $R^*(n, \epsilon)$ performs in the slow-fading regime is to use the channel dispersion, a measure of the variance between the channel capacity and the maximal achievable rate [2]. Using the channel dispersion together with the capacity, Polyanskiy derives a second order approximation of $R^*(n, \epsilon)$ in the limit $n \rightarrow \infty$ for the scalar coherent fading channel using the Meta-converse bound in (12) [36]. In our case, however, we have a codeword power constraint and the power might be allocated differently over coherence blocks. This makes us unable to use the central limit theorem as is done in [36]. Instead, we introduce an additional constraint on the input distribution to follow the USTM distribution. This means that the resulting upper bound is suboptimal and will be valid only for orthogonal signaling where the power is uniformly allocated over the antennas. This is a realistic scenario for initial transmission. Since we are mainly interested in the slow-fading scenario, even though it is valid for all T , we will strengthen the validity of this bound by comparing its agreement to the outage capacity which is the common performance metric in this regime. If they agree, we conclude that the outage capacity is indeed a good performance metric for finite blocklengths and if they do not agree, it is an indicator that it is not.

The Meta-converse bound in (12) is repeated below for convinence

$$\log(M) \leq \inf_{Q_Y} \sup_{\mathbf{X} \in \mathcal{F}} -\log(\beta_{1-\epsilon}(\mathbf{X}, Q_Y)) \quad (97)$$

where $\beta_{1-\epsilon}(\mathbf{X}, Q_Y)$, for $\gamma \geq 0$, can be lower bounded by [2, Eq. 106]

$$\begin{aligned} \beta_{1-\epsilon}(\mathbf{X}, Q_Y) &\geq \sup_{\gamma \geq 0} \frac{1}{\gamma} \left(\Pr \left\{ \frac{f_{Y|\mathbf{X}}(Y|\mathbf{X})}{q_Y(Y)} \leq \gamma \right\} - \epsilon \right) \\ &\stackrel{(a)}{=} e^{-\gamma} (\Pr \{i(\mathbf{X}; Y) \leq \gamma\} - \epsilon) \end{aligned} \quad (98)$$

where in (a) we are using the monotonicity of the logarithm inside the probability to write it wrt the information density. Using (98) in (97) and dividing by the blocklength, n , we get the bound on $R^*(n, \epsilon)$ as

$$\begin{aligned} \bar{R}(n, \epsilon) &\stackrel{(b)}{\leq} \inf_{Q_Y} \sup_{\gamma \geq 0} \left\{ -\log \left(e^{-\gamma} (\Pr \{i(\mathbb{X}; \mathbb{Y}) \leq \gamma\} - \epsilon) \right) \right\} \\ &\stackrel{(c)}{\leq} \sup_{\gamma \geq 0} \left\{ \gamma - \log (\Pr \{i(\mathbb{X}; \mathbb{Y}) \leq \gamma\} - \epsilon) \right\} \end{aligned} \quad (99)$$

where in (b), the supremum over the input is not present due the choice of the inputs following the USTM distribution, which makes the distribution of $i(\mathbb{X}, \mathbb{Y})$ under $P_{\mathbb{Y}|\mathbb{X}=\mathbf{x}}$ independent of \mathbf{X} . In (c), we have chosen the arbitrary output distribution, $Q_{\mathbb{Y}}$, to be the output distribution produced by the USTM distribution on the input given in (93). Since the codeword might span several fadings, the information density need to be adjusted accordingly. The length of the transmitted codeword is given by $n = LT$ where T is the number of symbols in a coherence block and L is the number of different selectivity branches that the codeword experiences during transmission. Hence, in order to obtain the lower bound on the maximal achievable rate, we need to take into consideration that the channel is also changing L times during the codeword. We are interested in

$$i(\mathbb{X}^L; \mathbb{Y}^L) \triangleq \log \left(\frac{f_{\mathbb{Y}^L|\mathbb{X}^L}(\mathbb{Y}^L|\mathbb{X}^L)}{f_{\mathbb{Y}^L}(\mathbb{Y}^L)} \right) \quad (100)$$

where \mathbb{Y}^L is the sequence of L received matrices and \mathbb{X}^L is the sequence of L transmitted matrices. To evaluate (100), $\{\mathbb{X}_l\}_{l=1}^L$ are chosen to be iid according to the scaled isotropic distribution introduced in Section 4.1. This input distribution will, for each l , result in the output distribution, $f_{\mathbb{Y}}(\mathbb{Y})$, given in (93) in Section 4.2. Since the transmitted messages are iid and the channel is memoryless, we may write

$$f_{\mathbb{Y}^L}(\mathbb{Y}^L) = \prod_{l=1}^L f_{\mathbb{Y}}(\mathbb{Y}_l). \quad (101)$$

Furthermore, conditioned on \mathbb{X}_l , the columns of the output \mathbb{Y}_l for $l = 1, \dots, L$ are independent and Gaussian. Therefore, the joint pdf of all the received messages is

$$f_{\mathbb{Y}^L|\mathbb{X}^L}(\mathbb{Y}^L|\mathbb{X}^L) = \prod_{l=1}^L f_{\mathbb{Y}|\mathbb{X}}(\mathbb{Y}_l|\mathbb{X}_l) \quad (102)$$

where the pdf, $f_{\mathbb{Y}|\mathbb{X}}(\mathbb{Y}_l|\mathbb{X}_l)$, is given in (44) for each l . From (101) and (102), the total information density in (100) can now be written as

$$\begin{aligned} i(\mathbb{X}^L; \mathbb{Y}^L) &= \log \left(\prod_{l=1}^L \left(\frac{f_{\mathbb{Y}_l|\mathbb{X}_l}(\mathbb{Y}_l|\mathbb{X}_l)}{f_{\mathbb{Y}_l}(\mathbb{Y}_l)} \right) \right) \\ &= \sum_{l=1}^L i(\mathbb{X}_l; \mathbb{Y}_l) \end{aligned} \quad (103)$$

where the information density for each $l = 1, \dots, L$ is given in (95).

5.2 Rate lower bound

To compute the lower bound on $R^*(n, \epsilon)$, a bound, introduced in section 2, called Dependence Testing (DT) bound will be utilized. This way of computing the lower bound on the maximal achievable rate has already been applied successfully in the SISO case [5]. The bound states that for any input distribution, $P_{\mathbb{X}}$, there exists a code with M codewords and an average probability of error no larger than

$$\epsilon \leq \mathbb{E} \left[e^{-\{i(\mathbb{X}; \mathbb{Y}) - \log(\frac{M-1}{2})\}^+} \right] \quad (104)$$

where $i(\mathbb{X}; \mathbb{Y})$ is the information density defined in section 2. As in the upper bound on $R^*(n, \epsilon)$, a codeword spans several coherence blocks and hence, the information density in (104) need to be adjusted accordingly. This was done for the case of USTM as the input distribution in section 5.1 and it will be reused in the DT-bound. Hence, by replacing the information density in (104) with the expression in (103), everything that is needed for the DT-bound is known. Note that since M is the cardinality of the codebook, it is related to the information (measured in nats) by $M = e^k$ where k is the information that is transmitted. The lower bound on $R^*(n, \epsilon)$ is obtained by solving

$$\underline{R}(n, \epsilon) \triangleq \max \left\{ \frac{\log(M)}{n} : M \text{ does not satisfy (104)} \right\}. \quad (105)$$

6 Space-Time Block Codes

In this section, we will derive bounds on the rate of two space-time block codes, STBC, that are employed in today's communication systems. This, along with the bounds derived in the previous section, will give some insight in how close to the optimal performance one is operating. Two different setups, 2×2 and 4×4 , will be considered. In the former, the STBC will be assumed to follow the well known Alamouti scheme which achieves full diversity order of four and rate one [37]. This is the optimal scheme in terms of outage probability when the power is radiated isotropically [11, pp. 193]. In the latter case, it has been proven that there are no complex STBC that achieves full diversity [38]. Instead, we will assume an STBC that is employed in LTE-A called space-frequency block code plus frequency switched transmit diversity, SFBC+FSTD, which achieves a diversity order of eight at rate one [39, pp. 164].

As a performance metric, the outage capacity for both of the STBC's will also be presented. The outage capacity for the Alamouti scheme is given by

$$C_{\text{out},\epsilon}^{\text{Ala}} = \sup \{R : P_{\text{out}}^{\text{ALA}}(R) \leq \epsilon\} \quad (106)$$

where $P_{\text{out}}(R)$ is the outage probability defined as

$$P_{\text{out}}^{\text{ALA}}(R) = \Pr \left\{ \frac{1}{L} \sum_{l=1}^L \log \left(1 + \frac{\rho}{n_t} \text{tr} \{ \mathbb{H} \mathbb{H}^H \} \right) < R \right\} \quad (107)$$

where \mathbb{H} is a 2×2 matrix with entries iid $\mathcal{CN}(0, 1)$. The outage capacity is similar for the SFBC+FSTD scheme. In Section 6.2, it is shown that the SFBC+FSTD scheme transform a 4×4 MIMO channel with coherence time T into two parallel 2×4 MIMO channels with coherence time $T/2$. Hence, the performance of this code will be exactly the same as the performance of the Alamouti code used on a 2×4 MIMO channel with coherence time $T/2$. The outage capacity is given as

$$C_{\text{out},\epsilon}^{\text{FSTD}} = \sup \{R : P_{\text{out}}^{\text{FSTD}}(R) \leq \epsilon\} \quad (108)$$

where $P_{\text{out}}(R)^{\text{FSTD}}$ is the outage probability defined as

$$P_{\text{out}}^{\text{FSTD}}(R) = \Pr \left\{ \frac{1}{L} \sum_{l=1}^L \log \left(1 + \frac{\rho}{n_t} \text{tr} \{ \mathbb{H} \mathbb{H}^H \} \right) < R \right\} \quad (109)$$

where \mathbb{H} is a 2×4 matrix with entries iid $\mathcal{CN}(0, 1)$.

6.1 Alamouti lower bound

In this section, a finite blocklength lower bound for the Alamouti scheme will be derived. The bound is based on the DT bound that has been discussed in the foregoing sections. In Section 5.2, the lower bound was derived under the assumption of USTM, i.e. orthogonal inputs with uniform power allocation. The Alamouti scheme also utilizes orthogonal transmission with equal power allocation and is therefore a subset of the USTM family. Throughout this section, it will be assumed that $n_t = n_r = 2$ and $T/n_t \in \mathbb{N}$.

The aim from here will be to derive an expression for the information density as was done for the USTM case in 5.2. We start by constructing two orthogonal vectors that follow the Alamouti scheme. Let

$$\bar{\mathbf{X}} = \sqrt{\frac{\rho T}{n_t}} \Phi \quad (110)$$

where $\Phi \in \mathbb{C}^{T \times 1}$ and is uniformly distributed over the unit hyper-sphere in \mathbb{C}^T . From this, the second input vector is constructed according to the Alamouti scheme as

$$\tilde{\mathbf{X}} = [-\bar{X}_2^* \quad \bar{X}_1^* \quad -\bar{X}_4^* \quad \bar{X}_3^* \quad \dots]^T \quad (111)$$

and the input matrix becomes

$$\mathbb{X} = [\bar{\mathbf{X}} \quad \tilde{\mathbf{X}}] \in \mathbb{C}^{T \times 2}. \quad (112)$$

As in (85), the conditional output pdf may be expressed as

$$\begin{aligned} f_{\mathbf{Y}|\mathbb{X}}(\mathbf{Y}|\mathbb{X}) &= \frac{1}{\pi^{n_r T} \det(\mathbb{X}\mathbb{X}^H + \mathbf{I}_T)^{n_r}} e^{\text{tr}\left\{\mathbf{Y}^H \left(\frac{\rho}{n_t} \mathbb{X}\mathbb{X}^H + \mathbf{I}_T\right)^{-1} \mathbf{Y}\right\}} \\ &= \frac{1}{\pi^{n_r T} \det\left(\frac{\rho}{n_t} \mathbb{X}\mathbb{X}^H + \mathbf{I}_T\right)^{n_r}} e^{-\text{tr}\{\mathbf{Y}\mathbf{Y}^H\}} e^{c_1 \text{tr}\{\mathbf{Y}^H \underline{\Phi} \Phi^H \mathbf{Y}\}} \end{aligned} \quad (113)$$

where $c_1 = \frac{1}{1 + \frac{1}{\frac{\rho T}{n_t}}}$ and $\underline{\Phi} = \frac{1}{\sqrt{\frac{\rho T}{n_t}}} \mathbb{X}$ is unitary. The unconditional output pdf is now given as

$$f_{\mathbf{Y}}(\mathbf{Y}) = \frac{1}{\pi^{n_r T} \det\left(\frac{\rho}{n_t} \mathbb{X}\mathbb{X}^H + \mathbf{I}_T\right)^{n_r}} e^{-\text{tr}\{\mathbf{Y}\mathbf{Y}^H\}} \mathbb{E}_{\underline{\Phi}} \left[e^{c_1 \text{tr}\{\mathbf{Y}^H \underline{\Phi} \Phi^H \mathbf{Y}\}} \right]. \quad (114)$$

Let $\mathbf{y}_1, \mathbf{y}_2, \phi_1, \phi_2$ be the columns of \mathbf{Y} and $\underline{\Phi}$ respectively. Also, note that $\phi_1 \perp \phi_2$. The trace in (114) may now be written as

$$\begin{aligned} \text{tr}\{\mathbf{Y}^H \underline{\Phi} \Phi^H \mathbf{Y}\} &= \text{tr}\{\mathbf{Y}^H (\phi_1 \phi_1^H + \phi_2 \phi_2^H) \mathbf{Y}\} \\ &= \text{tr}\{\mathbf{y}_1^H (\phi_1 \phi_1^H + \phi_2 \phi_2^H) \mathbf{y}_1\} + \text{tr}\{\mathbf{y}_2^H (\phi_1 \phi_1^H + \phi_2 \phi_2^H) \mathbf{y}_2\}. \\ &= |\langle \mathbf{y}_1, \phi_1 \rangle|^2 + |\langle \mathbf{y}_2, \phi_2 \rangle|^2 + |\langle \mathbf{y}_1, \phi_2 \rangle|^2 + |\langle \mathbf{y}_2, \phi_1 \rangle|^2 \end{aligned} \quad (115)$$

Note that $\phi_2 = [-\phi_{12}^* \quad \phi_{11}^* \quad -\phi_{14}^* \quad \phi_{13}^* \quad \dots]$. We write

$$\begin{aligned} \langle \mathbf{y}_i, \phi_2 \rangle &= y_{i1}^* (-\phi_{12}^*) + y_{i2}^* (\phi_{11}^*) + y_{i3}^* (-\phi_{14}^*) + y_{i4}^* (\phi_{13}^*) \\ &= \phi_{11}^* y_{i2}^* + \phi_{12}^* (-y_{i1}^*) + \phi_{13}^* y_{i4}^* + \phi_{14}^* (-y_{i3}^*) \\ &= \langle \phi_1, \tilde{\mathbf{y}}_i \rangle \end{aligned} \quad (116)$$

where $\tilde{\mathbf{y}}_i = [y_{i2}^* \quad -y_{i1}^* \quad y_{i4}^* \quad -y_{i3}^* \quad \dots]$ for $i = 1, 2$. By utilizing (116) in (115), we write

$$\begin{aligned} \text{tr}\{\mathbf{Y}^H \underline{\Phi} \Phi^H \mathbf{Y}\} &= \text{tr}\{\mathbf{y}_1^H \phi_1 \phi_1^H \mathbf{y}_1\} + \text{tr}\{\tilde{\mathbf{y}}_1^H \phi_1 \phi_1^H \tilde{\mathbf{y}}_1\} + \text{tr}\{\mathbf{y}_2^H \phi_1 \phi_1^H \mathbf{y}_2\} + \text{tr}\{\tilde{\mathbf{y}}_2^H \phi_1 \phi_1^H \tilde{\mathbf{y}}_2\} \\ &= \text{tr}\left\{\phi_1^H \left(\tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^H\right) \phi_1\right\} \end{aligned} \quad (117)$$

where $\tilde{\mathbf{Y}} = [\mathbf{y}_1 \quad \tilde{\mathbf{y}}_1 \quad \mathbf{y}_2 \quad \tilde{\mathbf{y}}_2] \in \mathbb{C}^{T \times 4}$.

Now, considering the eigenvalue decomposition $\tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^H = \mathbf{U} \Delta \mathbf{U}^H$, we may write the expectation in (114) as

$$\begin{aligned} \mathbb{E}_{\underline{\Phi}} \left[e^{c_1 \text{tr}\{\mathbf{Y}^H \underline{\Phi} \Phi^H \mathbf{Y}\}} \right] &= \mathbb{E}_{\Phi_1} \left[e^{c_1 \text{tr}\{\Phi_1^H \mathbf{U} \Delta \mathbf{U}^H \Phi_1\}} \right] \\ &= \mathbb{E}_{\Phi_1} \left[e^{c_1 \text{tr}\{\Phi_1^H \Delta \Phi_1\}} \right] \\ &= \mathbb{E}_{\Phi_1} \left[e^{\Delta \Phi_1 c_1 \Phi_1^H} \right] \end{aligned} \quad (118)$$

where, $\Delta = \text{diag}(\sigma_1, \sigma_1, \sigma_2, \sigma_2, 0, \dots, 0) \in \mathbb{C}^{T \times T}$. From (118), it can be seen that the Alamouti scheme transforms a 2×2 -MIMO channel into a 1×4 SIMO channel. If we now let $\Phi = [\Phi_1 \quad \Phi_{\perp}]$

belong to the unitary group, $\mathcal{U}(T)$, and let $\Lambda = \text{diag}(c_1, 0, \dots, 0) \in \mathbb{R}^{T \times T}$, we may write the expectation in (118) as

$$\mathbb{E}_{\Phi} \left[e^{c_1 \text{tr}\{\mathbf{Y}^H \Phi \Phi^H \mathbf{Y}\}} \right] = \frac{1}{|\mathcal{U}(T - n_t)| |\mathcal{S}(T, n_t)|} \mathbb{E}_{\Phi} \left[e^{\Delta \Phi \Lambda \Phi^H} \right] \quad (119)$$

which resembles the Itzykson-Zuber integral that has been solved in appendix C. We get

$$\begin{aligned} & \frac{1}{|\mathcal{U}(T - n_t)| |\mathcal{S}(T, n_t)|} \int_{\mathcal{U}(T)} e^{\text{tr}\{\Delta \Phi \Lambda \Phi^H\}} d\tilde{\Phi} \\ &= \frac{1}{|\mathcal{U}(T - n_t)| |\mathcal{S}(T, n_t)|} \underbrace{\frac{|\mathcal{U}(T)| \prod_{i=T-n_t+1}^T \Gamma(i)}{\prod_{i=1}^{n_t} \lambda_i^{(T-n_t)} \prod_{i=1}^{n_t-1} i! \prod_{j=1}^{T-n_r-1} j!}}_{c_{T, n_t}} \frac{1}{\prod_{i=1}^{n_r} \sigma_i^{T-n_t}} \frac{\det(\mathbb{A})}{\det(V(\Sigma))}. \end{aligned} \quad (120)$$

Where the matrix, \mathbb{A} is given in the appendix. The only difference from the scenario in appendix C is that the non-zero eigenvalues in \mathbb{A} are not distinct. To make the eigenvalues distinct, let $\Sigma = \text{diag}(\sigma_1 + \epsilon_1, \sigma_1, \sigma_2 + \epsilon_2, \sigma_2)$, denote the i :th entry of Σ as δ_i and evaluate the limit as

$$\begin{aligned} & \frac{1}{|\mathcal{U}(T - n_t)| |\mathcal{S}(T, n_t)|} \int_{\mathcal{U}(T)} e^{\text{tr}\{\Delta \Phi \Lambda \Phi^H\}} d\tilde{\Phi} = c_{T, n_t} \lim_{\epsilon_1 \rightarrow 0, \epsilon_2 \rightarrow 0} \frac{1}{\prod_{i=1}^{n_r} \delta_i^{T-n_t}} \frac{\det(\mathbb{A})}{\det(V(\Sigma))} \\ &= c_{T, n_t} \frac{1}{\prod_{i=1}^{n_r} \sigma_i^{T-n_t}} \frac{\det(\mathbb{A}_1)}{(\sigma_1 - \sigma_2)^2} \\ &\triangleq \Omega \end{aligned} \quad (121)$$

where the matrix \mathbb{A}_1 is a block matrix given as

$$\mathbb{A}_1 = \begin{pmatrix} \mathbb{A}_{11} & \mathbb{A}_{12} \\ \mathbb{A}_{13} & \mathbb{A}_{14} \end{pmatrix} \quad (122)$$

and each of the sub-matrices are given as

- $\mathbb{A}_{11} \in \mathbb{R}^{n_t \times n_r}$ where $(\mathbb{A}_{11})_{ij} = \begin{cases} e^{\lambda \sigma_j} \sigma_j^{(n_t-i)} & \text{for } i = 1, \dots, n_t & \text{if } j \text{ even} \\ e^{\lambda \sigma_j} (\lambda \sigma_j^{n_t-i} + (n_t - i) \sigma_j^{n_t-i-1}) & \text{for } i = 1, \dots, n_t & \text{if } j \text{ odd} \end{cases}$
- $\mathbb{A}_{13} \in \mathbb{R}^{(T-n_t) \times n_r}$ where $(\mathbb{A}_{13})_{ij} = \begin{cases} \sigma_j^{T-n_t-i} & \text{for } i = 1, \dots, T - n_t & \text{if } j \text{ even} \\ (T - n_t - i) \sigma_j^{T-n_t-i-1} & \text{for } i = 1, \dots, T - n_t & \text{if } j \text{ odd} \end{cases}$
- $\mathbb{A}_{14} \in \mathbb{R}^{(T-n_t) \times (T-n_r)}$ where $(\mathbb{A}_{14})_{i,j} = \begin{cases} 0 & \text{if } i \leq n_r - n_t \\ \text{diag}((T - n_r - j)!), j = 1, \dots, T - n_r & \text{if } i > n_r - n_t \end{cases}$

and $\mathbb{A}_{12} \in \mathbb{R}^{n_t \times T}$ is given as

$$\mathbb{A}_{12} = \begin{bmatrix} (\lambda^{T-n_t-1})^{(n_t-1)} & (\lambda^{T-n_t-2})^{(n_t-1)} & \dots & \lambda^{(n_t-1)} & 0 \\ (\lambda^{T-n_t-1})^{(n_t-2)} & (\lambda^{T-n_t-2})^{(n_t-2)} & \dots & \lambda^{(n_t-2)} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (\lambda^{T-n_t-1})^{(1)} & (\lambda^{T-n_t-2})^{(1)} & \dots & \lambda^{(1)} & 0 \\ (\lambda^{T-n_t-1})^{(0)} & (\lambda^{T-n_t-2})^{(0)} & \dots & \lambda^{(0)} & 0 \end{bmatrix} \quad (123)$$

Hence, the output pdf is given as

$$f_{\mathbf{Y}}(\mathbf{Y}) = \frac{\Omega e^{-\text{tr}\{\mathbf{Y}\mathbf{Y}^H\}}}{\pi^{n_r T} \det \left(\begin{smallmatrix} \rho \\ n_t \end{smallmatrix} \mathbf{X}\mathbf{X}^H + \mathbf{I}_T \right)^{n_r}}. \quad (124)$$

Finally, the information density for each selectivity branch, under $P_{\mathbb{Y}|\mathbb{X}}$, is given as

$$\begin{aligned} i^{\text{ALA}}(\mathbf{X}_l; \mathbf{Y}_l) &= \log \left(\frac{f_{\mathbb{Y}|\mathbb{X}}(\mathbf{Y}|\mathbf{X})}{f_{\mathbb{Y}}(\mathbf{Y})} \right) \\ &= \text{tr} \{ \mathbf{Y}_l \mathbf{Y}_l^H \} - \text{tr} \left\{ \mathbf{Y}_l^H \left(\frac{\rho}{n_t} \mathbf{X}_l \mathbf{X}_l^H + \mathbf{I}_T \right)^{-1} \mathbf{Y}_l \right\} - \log(\Omega) \\ &= \text{tr} \{ \mathbf{Y}_l \mathbf{Y}_l^H \} - \text{tr} \{ \mathbf{W}_l^H \mathbf{W}_l \} - \log(\Omega) \end{aligned} \quad (125)$$

where $\mathbf{W}_l \in \mathbb{C}^{T \times 2}$ with entries iid $\mathcal{CN}(0, 1)$. From the block-memoryless assumption, the total information density to be used in the DT bound is given as

$$i^{\text{ALA}}(\mathbf{X}^L; \mathbf{Y}^L) = \sum_{l=1}^L i^{\text{ALA}}(\mathbf{X}_l; \mathbf{Y}_l) \quad . \quad (126)$$

6.2 SFBC+FSTD lower bound

As mentioned in the introduction, there is no equivalent to the Alamouti scheme in the 4×4 case. Instead, we focus on the SFBC+FSTD scheme, that is used in the LTE downlink, which transmits symbols according to

$$\mathbb{X} = \begin{bmatrix} X_1 & X_2 & 0 & 0 \\ 0 & 0 & X_3 & X_4 \\ -X_2^* & X_1^* & 0 & 0 \\ 0 & 0 & -X_4^* & X_3^* \end{bmatrix}. \quad (127)$$

Note that antenna 1 and 3, together, is following the Alamouti scheme, the same goes for antenna 2 and 4.

Now, let $\mathbf{\Phi}_1 = [\Phi_{11} \ \Phi_{12} \ \Phi_{13} \ \Phi_{14} \ \dots] \in \mathbb{C}^{T/2}$ and $\mathbf{\Phi}_2 = [\Phi_{21} \ \Phi_{22} \ \Phi_{23} \ \Phi_{24} \ \dots] \in \mathbb{C}^{T/2}$ be independent and uniformly distributed over the unit sphere in $\mathbb{C}^{T/2}$. Also define $\tilde{\mathbf{\Phi}}_1 = [-\Phi_{12}^* \ \Phi_{11}^* \ -\Phi_{14}^* \ \Phi_{13}^* \ \dots]$ and $\tilde{\mathbf{\Phi}}_2 = [-\Phi_{22}^* \ \Phi_{21}^* \ -\Phi_{24}^* \ \Phi_{23}^* \ \dots]$. Let $T/n_t \in \mathbb{N}$ and let the complete codeword be given as

$$\mathbb{X} = \sqrt{T} \underbrace{\begin{bmatrix} \Phi_{11} & \Phi_{12} & 0 & 0 & \Phi_{13} & \Phi_{14} & 0 & 0 & \dots \\ 0 & 0 & \Phi_{21} & \Phi_{22} & 0 & 0 & \Phi_{23} & \Phi_{24} & \dots \\ -\Phi_{12}^* & \Phi_{11}^* & 0 & 0 & -\Phi_{14}^* & \Phi_{13}^* & 0 & 0 & \dots \\ 0 & 0 & -\Phi_{22}^* & \Phi_{21}^* & 0 & 0 & -\Phi_{24}^* & \Phi_{23}^* & \dots \end{bmatrix}^T}_{\triangleq \Phi} \quad (128)$$

The output pdf is given in (114) and as in the previous section, we need to compute

$$\begin{aligned} \text{tr} \{ \mathbf{Y}^H \Phi \Phi^H \mathbf{Y} \} &= \sum_{i=1}^4 \text{tr} \left\{ \mathbf{y}_i^H \left(\phi_1 \phi_1^H + \phi_2 \phi_2^H + \tilde{\phi}_1 \tilde{\phi}_1^H + \tilde{\phi}_2 \tilde{\phi}_2^H \right) \mathbf{y}_i \right\} \\ &= \sum_{i=1}^4 \left(|\mathbf{y}_i, \phi_1|^2 + |\mathbf{y}_i, \phi_2|^2 + |\mathbf{y}_i, \tilde{\phi}_1|^2 + |\mathbf{y}_i, \tilde{\phi}_2|^2 \right) \end{aligned} \quad (129)$$

By performing similar steps as in (116), (128) can be written as

$$\begin{aligned} \text{tr} \{ \mathbf{Y}^H \Phi \Phi^H \mathbf{Y} \} &= \sum_{i=1}^4 \left(\text{tr} \left\{ (\mathbf{y}_i^{\text{odd}})^H (\phi_1 \phi_1^H) \mathbf{y}_i^{\text{odd}} \right\} + \text{tr} \left\{ (\tilde{\mathbf{y}}_i^{\text{odd}})^H (\phi_1 \phi_1^H) \tilde{\mathbf{y}}_i^{\text{odd}} \right\} \right. \\ &\quad \left. + \text{tr} \left\{ (\mathbf{y}_i^{\text{even}})^H (\phi_2 \phi_2^H) \mathbf{y}_i^{\text{even}} \right\} + \text{tr} \left\{ (\tilde{\mathbf{y}}_i^{\text{even}})^H (\phi_2 \phi_2^H) \tilde{\mathbf{y}}_i^{\text{even}} \right\} \right) \end{aligned} \quad (130)$$

where

$$\begin{aligned}
\mathbf{y}_i^{\text{odd}} &= [y_{i1} \ y_{i2} \ y_{i5} \ y_{i6} \ y_{i9} \ y_{i10} \ \cdots]^\top \in \mathbb{C}^{T/2} \\
\mathbf{y}_i^{\text{even}} &= [y_{i3} \ y_{i4} \ y_{i7} \ y_{i8} \ y_{i11} \ y_{i12} \ \cdots]^\top \in \mathbb{C}^{T/2} \\
\tilde{\mathbf{y}}_i^{\text{odd}} &= [y_{i2}^* \ -y_{i1}^* \ y_{i6}^* \ -y_{i5}^* \ y_{i10}^* \ -y_{i11}^* \ \cdots]^\top \in \mathbb{C}^{T/2} \\
\tilde{\mathbf{y}}_i^{\text{even}} &= [y_{i4}^* \ -y_{i3}^* \ y_{i8}^* \ -y_{i7}^* \ y_{i12}^* \ -y_{i11}^* \ \cdots]^\top \in \mathbb{C}^{T/2}
\end{aligned} \tag{131}$$

By collecting the received vectors, we can express (130) as the sum of two traces according to

$$\text{tr} \{ \mathbf{Y}^H \boldsymbol{\Phi} \boldsymbol{\Phi}^H \mathbf{Y} \} = \text{tr} \left\{ (\mathbf{Y}^{\text{odd}})^H (\boldsymbol{\phi}_1 \boldsymbol{\phi}_1^H) \mathbf{Y}^{\text{odd}} \right\} + \text{tr} \left\{ (\mathbf{Y}^{\text{even}})^H (\boldsymbol{\phi}_2 \boldsymbol{\phi}_2^H) \mathbf{Y}^{\text{even}} \right\} \tag{132}$$

where $\mathbf{Y}^{\text{odd}} = [\mathbf{y}_1^{\text{odd}} \ \tilde{\mathbf{y}}_1^{\text{odd}} \ \mathbf{y}_2^{\text{odd}} \ \tilde{\mathbf{y}}_2^{\text{odd}} \ \mathbf{y}_3^{\text{odd}} \ \tilde{\mathbf{y}}_3^{\text{odd}} \ \mathbf{y}_4^{\text{odd}} \ \tilde{\mathbf{y}}_4^{\text{odd}}]$ and similarly for \mathbf{Y}^{even} . By the independency between $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$, the expectation in (114) may now be expressed as

$$\mathbb{E}_{\boldsymbol{\Phi}} \left[e^{\text{tr} \{ c_1 \mathbf{Y}^H \boldsymbol{\Phi} \boldsymbol{\Phi}^H \mathbf{Y} \}} \right] = \mathbb{E}_{\boldsymbol{\Phi}_1} \left[e^{c_1 \text{tr} \{ (\mathbf{Y}^{\text{odd}})^H (\boldsymbol{\phi}_1 \boldsymbol{\phi}_1^H) \mathbf{Y}^{\text{odd}} \}} \right] \mathbb{E}_{\boldsymbol{\Phi}_2} \left[e^{c_1 \text{tr} \{ (\mathbf{Y}^{\text{even}})^H (\boldsymbol{\phi}_2 \boldsymbol{\phi}_2^H) \mathbf{Y}^{\text{even}} \}} \right] \tag{133}$$

where $c_1 = \frac{1}{1 + \frac{1}{\frac{\rho T}{n_t}}}$. Hence, the output pdf factorizes into two independent parts, each of which corresponds to the output of a 2×4 noncoherent MIMO channel with coherence time $T/2$. Denote the eigenvalues of $(\mathbf{Y}^{\text{odd}} (\mathbf{Y}^{\text{odd}})^H)$ by $\Lambda^{\text{odd}} = \text{diag} (\lambda_1^{\text{odd}}, \lambda_1^{\text{odd}}, \lambda_2^{\text{odd}}, \lambda_2^{\text{odd}}, \lambda_3^{\text{odd}}, \lambda_3^{\text{odd}}, \lambda_4^{\text{odd}}, \lambda_4^{\text{odd}})$ and similarly for $\mathbf{Y}^{\text{even}} (\mathbf{Y}^{\text{even}})^H$. From here, each of the expectations in (133) can be evaluated analogously as in the Alamouti case in the previous section with the difference that $n_t = 2$, $n_r = 4$ and $L = n / (T/2)$.

The first expectation on the RHS in (133) results in

$$\mathbb{E}_{\boldsymbol{\Phi}_1} \left[e^{c_1 \text{tr} \{ (\mathbf{Y}^{\text{odd}})^H (\boldsymbol{\phi}_1 \boldsymbol{\phi}_1^H) \mathbf{Y}^{\text{odd}} \}} \right] \triangleq \Omega^{\text{odd}} \tag{134}$$

where Ω^{odd} is given in (121) and similarly for the second expectation. Hence, the output pdf is given by

$$f_{\mathbf{Y}}(\mathbf{Y}) = \frac{\Omega^{\text{odd}} \Omega^{\text{even}} e^{-\text{tr} \{ \mathbf{Y} \mathbf{Y}^H \}}}{\pi^{n_r T} \det \left(\frac{\rho}{n_t} \mathbf{X} \mathbf{X}^H + \mathbf{I}_T \right)^{n_r}} \tag{135}$$

and the information density for each selectivity branch, under $P_{\mathbf{Y}|\mathbf{X}}$, is given as

$$\begin{aligned}
i^{\text{FSTD}}(\mathbf{X}_l; \mathbf{Y}_l) &= \log \left(\frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{Y})} \right) \\
&= \text{tr} \{ \mathbf{Y}_l \mathbf{Y}_l^H \} - \text{tr} \left\{ \mathbf{Y}_l^H \left(\frac{\rho}{n_t} \mathbf{X}_l \mathbf{X}_l^H + \mathbf{I}_T \right)^{-1} \mathbf{Y}_l \right\} - \log (\Omega^{\text{odd}} \Omega^{\text{even}}) \\
&= \text{tr} \{ \mathbf{Y}_l \mathbf{Y}_l^H \} - \text{tr} \{ \mathbf{W}_l^H \mathbf{W}_l \} - \log (\Omega^{\text{odd}} \Omega^{\text{even}})
\end{aligned} \tag{136}$$

From the block-memoryless assumption, the total information density to be used in the DT bound is given as

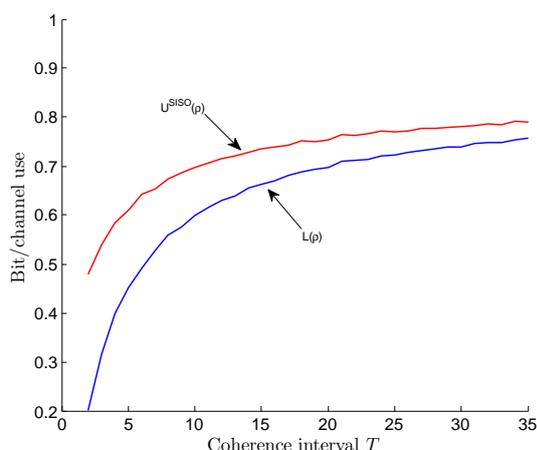
$$i^{\text{FSTD}}(\mathbf{X}^L; \mathbf{Y}^L) = \sum_{l=1}^L i^{\text{FSTD}}(\mathbf{X}_l; \mathbf{Y}_l) \tag{137}$$

7 Results

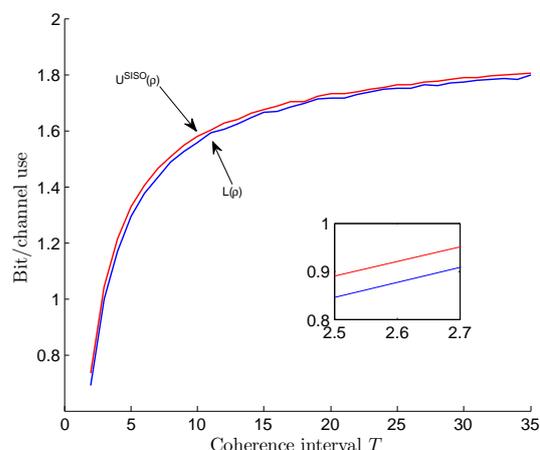
The aim of this chapter is to graphically present the bounds derived in chapter 4, 5 and 6. The figures will only be motivated, for a thorough analysis, see chapter 8.

7.1 Tightness of capacity bounds

In this section, the capacity bounds derived in Section 4.1 and 4.2 are illustrated. The bounds are plotted for SNR's of 0 dB and 6 dB. The bounds are illustrated for coherence blocks, T in the interval $[n_r + n_t, 35]$. This is the most interesting regime since, with larger T , estimating the channel becomes less penalized and the bounds converges to the coherent capacity. In figure 1, the bounds are shown for a SISO setup. Figure 2 shows the bounds for a 2x2 MIMO setup and figure 3 illustrates the bounds for a 3x3 MIMO setup.

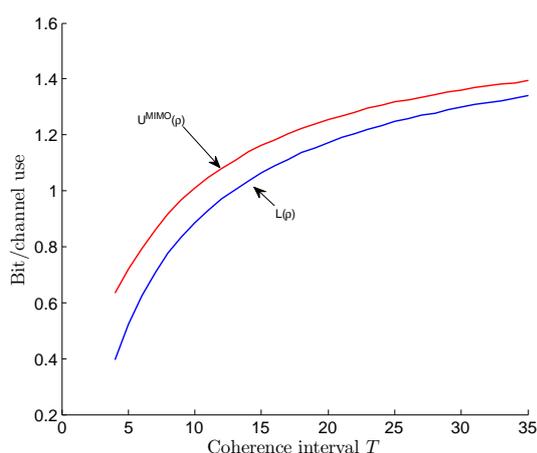


(a) Capacity bounds for SISO setup at 0 dB.

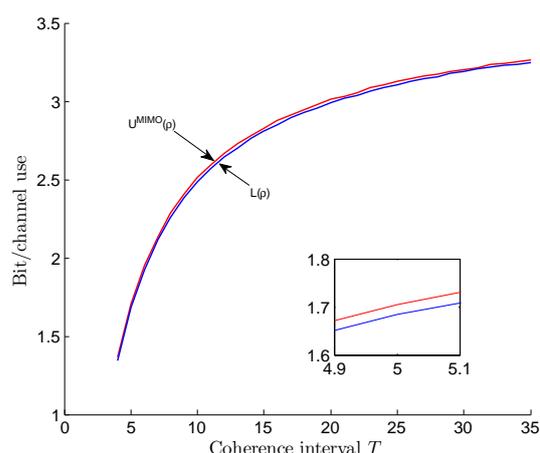


(b) Capacity bounds for a SISO setup at 6 dB.

Figure 1: Capacity bounds for SISO.

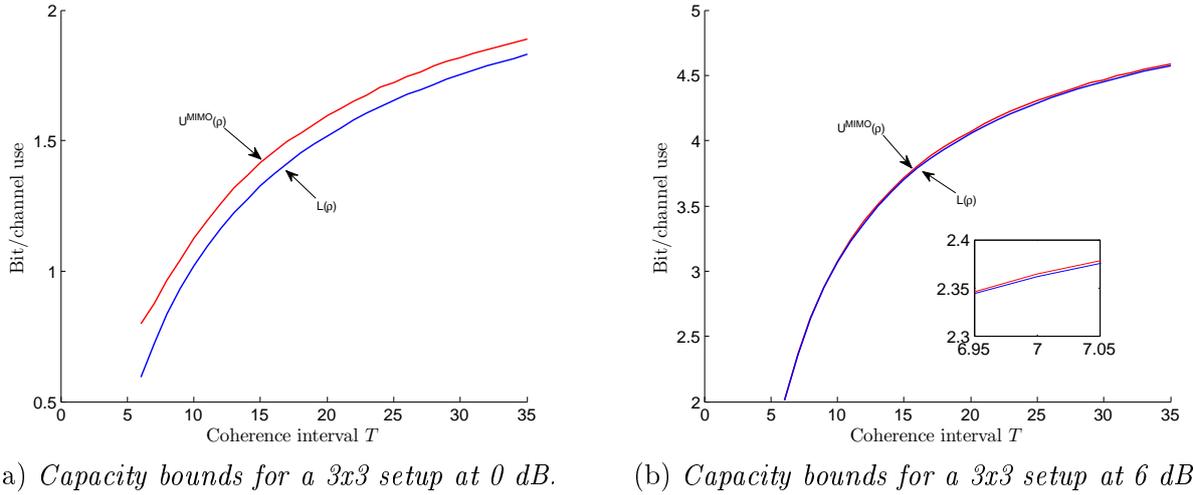


(a) Capacity bounds for a 2x2 setup at 0 dB.



(b) Capacity bounds for a 2x2 setup at 6 dB.

Figure 2: Capacity bounds for 2x2 MIMO.

Figure 3: Capacity bounds for 3×3 MIMO.

7.2 Tightness of Rate Bounds

We will, in this section, illustrate the characteristics of the maximal achievable rate, $R^*(n, \epsilon)$, in the finite blocklength regime. The considered scenarios are the supposed to reflect the two cases of using small- and large blocklengths respectively. The scenario of short blocks consists of $n_{\text{Small}} = 168$ bits while the large messages are chosen to consist of $n_{\text{Large}} = 10000$ bits. Since today's communication systems rather chooses an operation point in terms of block error probability instead of aiming for making it as low as possible, we will consider $\epsilon = 10^{-3}$ [18]. This is a realistic operation point in for example traffic applications [40, pp. 70]. The illustrations are made for setups with two and three antennas at both the receiver and the transmitter at an SNR of 6 dB.

The coherent capacity, the blue curve, is included in all of the figures as a measure of how close one can communicate to the channel's fundamental limit. Also, as discussed in Section 2, for slow-fading scenarios i.e., large T (small L), the ϵ -outage capacity is the relevant performance metric. Therefore, it is also included, the black dashed curve, to give a feeling for how close to the fundamental limit of communication the maximal achievable rate is performing in the slow-fading regime.

The bounds on the maximal achievable rate as a function of the coherence time is shown for small blocklengths, n_{Small} , in figure 4a and 4b and for large blocklengths, n_{Large} , in figure 5a and 5b.

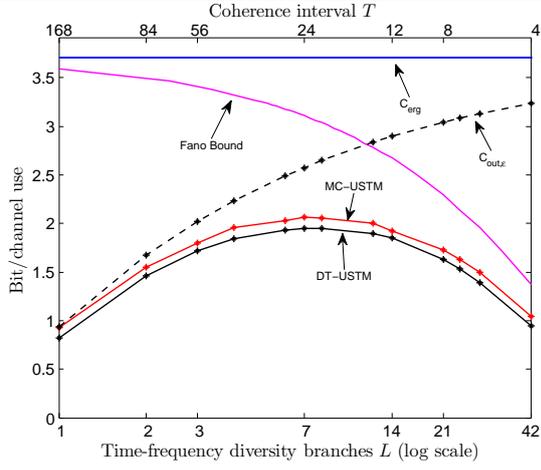
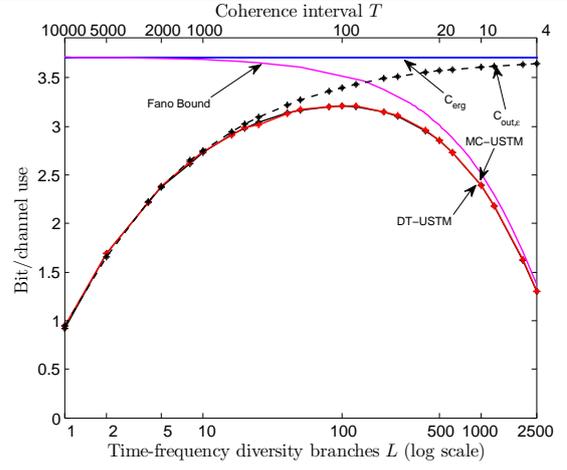
(a) 2×2 -MIMO at 6 dB and $n = 168$.(b) 2×2 -MIMO at 6 dB and $n = 10000$.

Figure 4: Bounds for the maximal achievable rate

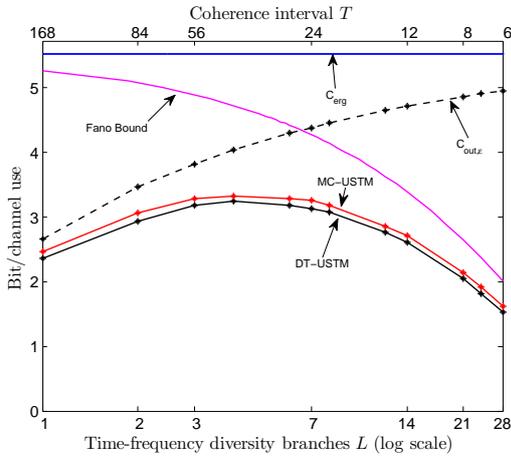
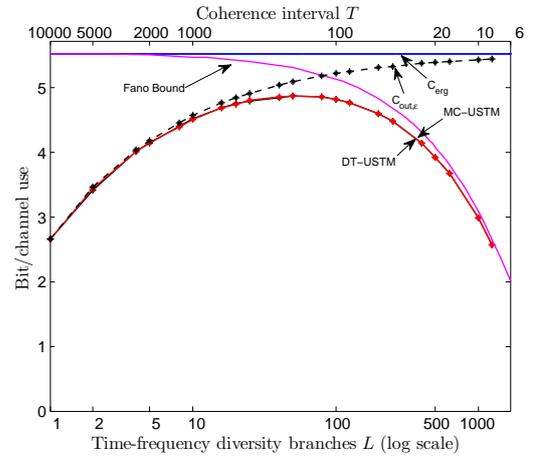
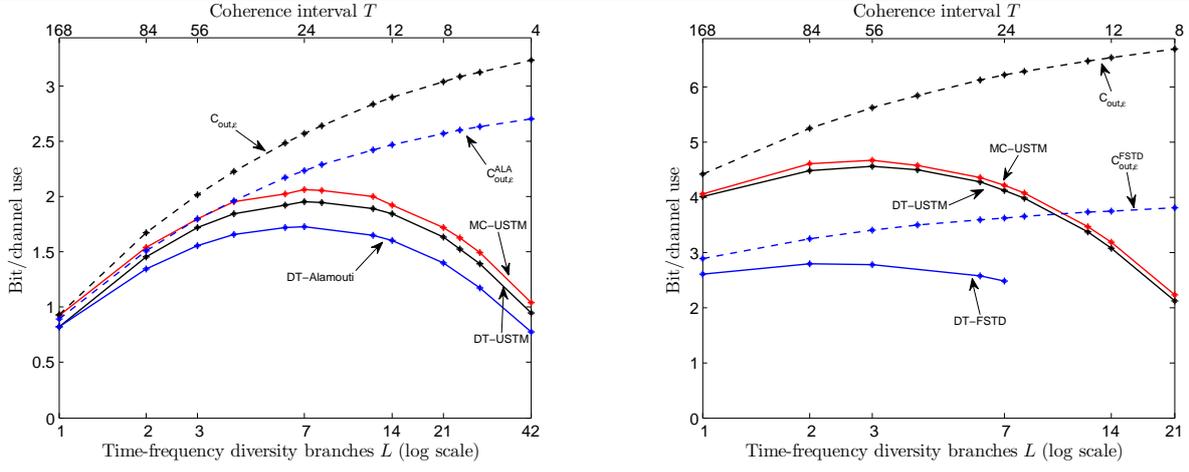
(a) 3×3 -MIMO at 6 dB and $n = 168$.(b) 3×3 -MIMO at 6 dB and $n = 10000$.

Figure 5: Bounds for the maximal achievable rate

7.3 STBC performance

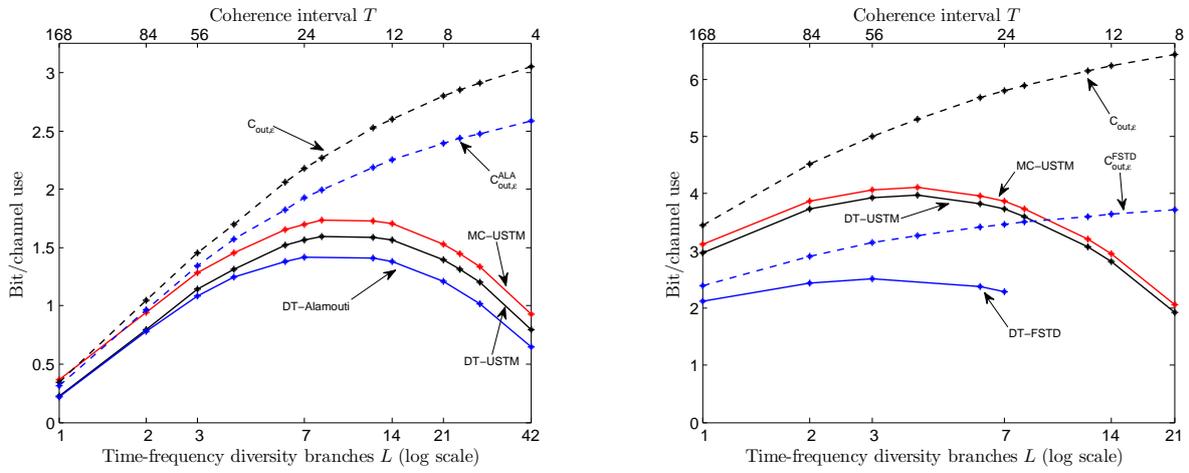
The results presented in this section emphasize the performance of diversity exploiting STBC's that are in use today under the same setup as in [18]. Specifically, the blocklength is equal to 168 bits (an LTE resource block) and the SNR is fixed to 6 dB. The STBC lower bounds that are illustrated are the ones introduced in Section 6. Two scenarios are illustrated; control signaling and ultra reliable communication. The former is appropriate for the exchange of short packages and the error rate is set to $\epsilon = 10^{-3}$. Ultra reliable communication is relevant for transmission of critical information e.g., in traffic-safety applications and the error rate is set to $\epsilon = 10^{-5}$ [40].

In figure 6, the control signaling scenario is illustrated while the scenario utilizing ultra reliable communication is shown in figure 7. The upper- and lower bounds on the maximal achievable rate from Section 5 is presented as a solid red and black curve respectively. The STBC that is considered in each scenario and setup, is shown as the solid blue curve. The outage capacity for for each signaling scheme (USTM and the STBC) is also shown in the figures as a black and blue dashed curve respectively.



(a) Bounds for a 2×2 setup with control signaling. (b) Bounds for a 4×4 setup with control signaling.

Figure 6: Bounds for control signaling.



(a) Bounds for a 2×2 setup with ultra reliable communication. (b) Bounds for a 4×4 setup with ultra reliable communication.

Figure 7: Bounds for ultra reliable communication.

8 Discussion

In this thesis, bounds in order to characterize the maximal achievable rate in the finite block-length regime has been presented. The first upper bound that was presented is based on the channel capacity which is not known and had to be bounded. These are illustrated in Section 7.1 for several antenna setups and SNR. These bounds are based on the assumption that USTM is used as the input distribution, which is justified by it being the capacity achieving distribution in the high SNR regime. This can be seen in the figures of Section 7.1 by comparing a fixed setup at different SNR, i.e., the upper and lower bound approaches each other for higher SNR. However, it can be seen that for SNR as low as 0 dB, the bounds are still tight. If one is to go much lower than this, other assumptions on the input distributions may yield better results. A further remark is that the bounds become tighter for an increased number of antennas.

The illustrations are all plotted for coherence blocks in the interval of $[n_t + n_r, 35]$. The reason for this is that the fast-fading regime, small T , is the most interesting. Namely, for small T , the impact of the non-coherency becomes significant; the channel varies to fast and the channel estimation occupies a lot of the information exchange which decreases the rate. If one instead increase the coherence block, the curves will converge towards the coherent capacity. This is due to the slow variations of the channel and hence, the occupied bits for channel estimation becomes insignificant.

The rate bounds are presented in Section 7.2. There are two capacity curves included, ergodic- and outage capacity. The ergodic capacity is included just to give an upper bound of how fast one would possibly be able to communicate. The outage capacity is included as a performance metric to be compared to the maximal achievable rate. The magenta colored curve is the upper bound that is based on the non-coherent channel capacity. This curve is a general upper bound on the maximal achievable rate. The second upper bound, based on the meta converse bound in Section 5.1, is not as general since it assumes USTM as the input distribution. This distribution includes all orthogonal signaling with uniform power allocation across the antennas and is a realistic bound for initial transmission.

The illustrations are shown for 2×2 - and 3×3 -MIMO for an SNR equal to 6 dB and an error target $\epsilon = 10^{-3}$. The left column displays the performance in the short blocklength regime, $n = 168$ bits, while the right column does the same for large blocklengths, $n = 10000$.

From the figures, it is clear that the maximal achievable rate, $R^*(n, \epsilon)$ is not monotonic in the coherence block, T , but there is an optimal coherence block T^* (or equivalently, an optimal number $L^* = n/T^*$ of time-frequency selectivity) that maximizes it for a specific n , ϵ , SNR and antenna setup. For $T \leq T^*$ (or $L \geq L^*$) the channel is fast fading and the cost of estimating the channel overcomes the gain due to time-frequency selectivity. On the other hand, when $T \geq T^*$, the channel is slow fading and the limited time-frequency selectivity decreases the performance. A similar observation was reported in [5] for the single-antenna case.

If the outage capacity would be considered as the performance metric in the finite block-length regime, it can be seen that it provides an accurate approximation of $R^*(n, \epsilon)$ only for large T , i.e., when the fading channel is essentially constant over the duration of the packet (quasi-static scenario). In the fast fading regime, the Fano's based bound provides a fairly accurate approximation of $R^*(n, \epsilon)$ only for very short T . By comparing the two columns, we note that the outage capacity and Fano's based bound becomes better approximations of $R^*(n, \epsilon)$ when the blocklength increases. This comes as no surprise since the maximal achievable rate goes towards the channel capacity by increasing the blocklength. Since the two curves are on top of each other, the USTM input distribution is nearly optimal which is expected since we have already seen that it is nearly optimal at SNR equal to 6 dB in Section 7.1. Also, the gap

between the Meta-converse and the DT bound is small even for short messages.

In the last Section, 7.3, $R^*(n, \epsilon)$ is compared to STBC's that are in use today. For 2×2 -MIMO, the Alamouti scheme, with diversity order 4 and no rate gain, is used for comparison. As no generalization of the Alamouti scheme exists beyond 2×2 configuration [38], we consider instead the SFBC+FSTD scheme, which provides diversity gain 8 and no rate gain, that is used in LTE.

In the figures, the MC upper bound and the DT lower bound is again depicted. The outage capacity is again included as a performance metric. Furthermore, the two additional curves in blue corresponds to the specific STBC. The solid blue curve is the DT lower bound for the corresponding scheme and the dashed blue curve corresponds to the outage capacity for the scheme. The left column illustrates the performance of a control signaling scenario, i.e., $\epsilon = 10^{-3}$ while the right column illustrates the performance of ultra reliable communication with $\epsilon = 10^{-5}$. All of the figures in this section are plotted for $n = 168$ and SNR equal to 6 dB.

If outage capacity is used as the performance metric, then a diversity scheme such as the Alamouti code is nearly optimal when the channel provides limited selectivity in time and frequency ($L \approx 1$). However, if the channel provides significant time-frequency selectivity, then one should use the antennas in multiplexing mode. For example, in figure 6a, for the case $L = 14$, the gap between the Alamouti scheme and the outage optimal scheme is about 0.5 bit/channel use and the gap increases as L grows larger. This observation is one of the key contributions in [18]. In the other regime, where L is large and T is small, the cost of learning the channel becomes significant. Therefore, large multiplexing gains are not feasible and the gap to optimality of the Alamouti scheme decreases. For example, by comparing the DT-USTM lower bound to the DT-Alamouti, the throughput reduction due to the use of the Alamouti scheme is about 0 bit/channel use for $L = 1$; 0.25 bit/channel use for $L = 12$ and 0.18 bit/channel use for $L = 42$.

A similar comparison is shown in figure 6b for 4×4 -MIMO. The SFBC+FSTD scheme provides diversity gain 8 and no multiplexing gain. As shown in the figure, the gap between the MC upper bound and the DT lower bound is small, allowing for a precise characterization of $R^*(n, \epsilon)$. In contrast, the gap between the DT-USTM and the DT-FSTD lower bound is large, suggesting that using all 8 antennas to provide diversity gain is suboptimal also when the time-frequency selectivity is limited.

In figure 7a and 7a, we consider a scenario of ultra reliable communication, $\epsilon = 10^{-5}$. Compared to the case of control signaling, $\epsilon = 10^{-3}$, the gap between the optimal schemes and the diversity-based schemes (Alamouti for 2×2 -MIMO and SFBC+FSTD for 4×4 -MIMO) gets smaller. This comes as no surprise, as the higher reliability requirement makes the exploitation of transmit diversity advantageous.

9 Conclusion

This thesis has analyzed the fundamental limits of the Rayleigh block-fading channel under an average-power constraint. For the channel capacity, an upper and a lower bound has been presented and proved to be tight for SNR down to 0 dB. The upper bound presented is new while the lower bound is the same as in [9] but more numerically stable. The maximal achievable rate has been characterized in terms of bounds; two upper- and one lower bound. One of the upper bounds, the one based on the channel capacity, and the lower bound is general while the second upper bound is valid for orthogonal and initial transmission, i.e., uniform power allocation across antennas. The maximal achievable rate has been compared to common performance metrics of today, such as outage capacity, and it has been concluded that in some regimes, these metrics are accurate while in other, they are not. Furthermore, the maximal achievable rate has been compared to diversity-exploiting schemes that are in use today. It was shown that neither diversity- nor multiplexing-exploiting schemes are universally optimal for a 2×2 configuration while in a 4×4 setup, the results suggests that all 8 antennas should be used for multiplexing.

A Appendix A

Starting from (57), the only term left to evaluate is

$$\mathbb{E}_{P_{\mathbf{X}}}\left[\mathbb{E}_{P_{\mathbf{Y}|\mathbf{X}}}\left[\log(\det(\mathbf{Y}^H\mathbf{Y}))\mid\mathbb{X}\right]\right]. \quad (\text{A.1})$$

Using the decomposition in (52) and the concavity of the logarithm, we use Jensen's inequality [19, Thm. 2.6.2] to write (A.1) as

$$\mathbb{E}_{P_{\mathbf{X}}}\left[\text{tr}\left\{\mathbb{E}_{P_{Z_1|\mathbf{X}}}\left[\mathbb{E}_{P_{Z_2|\mathbf{X}}}\left[\text{logm}\left(\mathbf{Z}_1^H\tilde{\mathbf{D}}\mathbf{Z}_1 + \mathbf{Z}_2^H\mathbf{Z}_2\right)\right]\right]\right\}\right] \leq \mathbb{E}_{P_{\mathbf{X}}}\left[\text{logm}\left(\mathbb{E}_{P_{Z_1|\mathbf{X}}}\left[\mathbb{E}_{P_{Z_2|\mathbf{X}}}\left[\mathbf{Z}_1^H\tilde{\mathbf{D}}\mathbf{Z}_1 + \mathbf{Z}_2^H\mathbf{Z}_2\right]\right]\right)\right] \quad (\text{A.2})$$

Since each element in the decomposition, κ_{ij} , is a sum of the product of scaled iid complex standard Gaussian random variables, the expectation will be zero for all κ_{ij} , $i \neq j$. Hence, the expectation of the decomposition on the right hand side of (A.2) is a diagonal matrix and the matrix logarithm can be thought of as taking the scalar logarithm of each entry on the diagonal which follows the same pdf as in (63) [33]. The expectation of κ_{ii} can be obtained as

$$\begin{aligned} \mathbb{E}_{P_{\kappa_{ii}|\mathbf{X}}}\left[\kappa_{ii}\right] &= \int_0^\infty y f_{\kappa_{ii}}(y) dy \\ &= \prod_{i=1}^{n_t} \left(\frac{1}{\xi_i}\right) \sum_{k=0}^\infty \frac{\delta_k}{\Gamma(T+k)} \int_0^\infty y^{T+k+1-1} e^{-y} dy \\ &= \prod_{i=1}^{n_t} \left(\frac{1}{\xi_i}\right) \sum_{k=0}^\infty \delta_k \frac{\Gamma(T+k+1)}{\Gamma(T+k)} \\ &= \prod_{i=1}^{n_t} \left(\frac{1}{\xi_i}\right) \sum_{k=0}^\infty (k+T)\delta_k. \end{aligned} \quad (\text{A.3})$$

Hence, the original expression in (A.1) can be bounded as

$$\begin{aligned} \mathbb{E}_{P_{\mathbf{X}}}\left[\mathbb{E}_{P_{\mathbf{Y}|\mathbf{X}}}\left[\log(\det(\mathbf{Y}^H\mathbf{Y}))\mid\mathbb{X}\right]\right] &\leq \mathbb{E}_{P_{\mathbf{X}}}\left[\text{tr}\left\{\log\left(\prod_{j=1}^{n_t} \left(\frac{1}{\xi_j}\right) \sum_{k=0}^\infty (k+T)\delta_k l_{n_r}\right)\right\}\right] \\ &= \mathbb{E}_{P_{\mathbf{X}}}\left[\sum_{i=1}^{n_r} \log\left(\prod_{j=1}^{n_t} \left(\frac{1}{\xi_j}\right) \sum_{k=0}^\infty (k+T)\delta_k\right)\right] \\ &= \mathbb{E}_{P_{\mathbf{X}}}\left[n_r \log\left(\prod_{j=1}^{n_t} \left(\frac{1}{1 + \frac{\rho}{n_t} \|\mathbf{X}_i\|^2}\right) \sum_{k=0}^\infty (k+T)\delta_k\right)\right]. \end{aligned} \quad (\text{A.4})$$

Substituting (A.4) into the bound on the mutual information in (57), we obtain

$$\begin{aligned}
I(\mathbb{X}; \mathbb{Y}) &\leq \log \left(\frac{\beta n_r^2 \Gamma_{n_r}(n_r)}{e^{n_r T} \Gamma_{n_r}(T)} \right) + \frac{n_r \mathbb{E}_{P_{\mathbb{X}}} \left[\sum_{i=1}^{n_r} \frac{\rho}{n_r} \|\mathbf{X}_i\|^2 \right] + T - n_r}{\beta} \\
&\quad + (T - n_r) \mathbb{E}_{P_{\mathbb{X}}} \left[n_r \log \left(\prod_{i=1}^{n_r} \left(\frac{1}{1 + \frac{\rho}{n_r} \|\mathbf{X}_i\|^2} \right) \sum_{k=0}^{\infty} (k + T) \delta_k \right) \right] \\
&\quad - n_r \sum_{i=1}^{n_r} \mathbb{E}_{P_{\mathbb{X}}} \left[\log \left(1 + \frac{\rho \|\mathbf{X}_i\|^2}{n_r} \right) \right] \\
&= \underbrace{\log \left(\frac{\beta n_r^2 \Gamma_{n_r}(n_r)}{e^{n_r T} \Gamma_{n_r}(T)} \right) + \frac{T - n_r}{\beta}}_{c^{\text{MIMO}}(\rho)} + n_r \mathbb{E}_{P_{\mathbb{X}}} \left[\frac{1}{\beta} \sum_{i=1}^{n_r} \left(1 + \frac{\rho}{n_r} \|\mathbf{X}_i\|^2 \right) \right. \\
&\quad \left. + (T - n_r) \log \left(\prod_{i=1}^{n_r} \left(\frac{1}{1 + \frac{\rho}{n_r} \|\mathbf{X}_i\|^2} \right) \sum_{k=0}^{\infty} (k + T) \delta_k \right) - \sum_{i=1}^{n_r} \log \left(1 + \frac{\rho}{n_r} \|\mathbf{X}_i\|^2 \right) \right]. \tag{A.5}
\end{aligned}$$

For simplicity, we define $f^{\text{MIMO}}(\mathbb{X}) : \mathbb{R}^{T \times n_t} \rightarrow \mathbb{R}$ as

$$\begin{aligned}
f_{\text{Jensen}}^{\text{MIMO}}(\mathbb{X}) &\triangleq \frac{1}{\beta} \sum_{i=1}^{n_r} \left(1 + \frac{\rho}{n_r} \|\mathbf{X}_i\|^2 \right) + (T - n_r) \log \left(\prod_{i=1}^{n_r} \left(\frac{1}{1 + \frac{\rho}{n_r} \|\mathbf{X}_i\|^2} \right) \sum_{k=0}^{\infty} (k + T) \delta_k \right) \\
&\quad - \sum_{i=1}^{n_r} \log \left(1 + \frac{\rho}{n_r} \|\mathbf{X}_i\|^2 \right). \tag{A.6}
\end{aligned}$$

Hence, the mutual information is bounded as

$$I(\mathbb{X}; \mathbb{Y}) \leq c^{\text{MIMO}}(\rho) + n_r \mathbb{E}_{P_{\mathbb{X}}} [f_{\text{Jensen}}^{\text{MIMO}}(\mathbb{X})]. \tag{A.7}$$

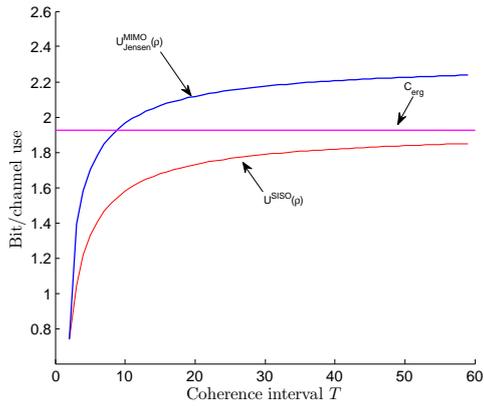
Finally, substituting (A.7) in (23) results in the following capacity upper bound

$$\begin{aligned}
C(\rho) &\leq \frac{1}{T} \inf_{\lambda \geq 0} \sup_{P_{\mathbb{X}}} \left\{ c^{\text{MIMO}}(\rho) + n_r \mathbb{E}_{P_{\mathbb{X}}} [f^{\text{MIMO}}(\mathbb{X})] + \lambda (T n_r - \mathbb{E}_{P_{\mathbb{X}}} [\text{tr} \{ \mathbf{X} \mathbf{X}^H \}]) \right\} \\
&= \frac{c^{\text{MIMO}}(\rho)}{T} - \frac{n_r}{T} \inf_{\lambda \geq 0} \sup_{P_{\mathbb{X}}} \left\{ \mathbb{E}_{P_{\mathbb{X}}} \left[f_{\text{Jensen}}^{\text{MIMO}}(\mathbb{X}) + \lambda \left(T - \frac{1}{n_r} \text{tr} \{ \mathbb{X} \mathbb{X}^H \} \right) \right] \right\} \\
&= \frac{c^{\text{MIMO}}(\rho)}{T} - \frac{n_r}{T} \inf_{\lambda \geq 0} \sup_{P_{\mathbb{X}}} \left\{ \mathbb{E}_{P_{\mathbb{X}}} \left[f_{\text{Jensen}}^{\text{MIMO}}(\mathbb{X}) + \lambda \left(T - \frac{1}{n_r} \sum_{i=1}^{n_r} \|\mathbf{X}_i\|^2 \right) \right] \right\} \tag{A.8} \\
&\leq \frac{c^{\text{MIMO}}(\rho)}{T} - \frac{n_r}{T} \inf_{\lambda \geq 0} \sup_{\|\mathbf{x}_i\|, i=1, \dots, n_r} \left\{ f_{\text{Jensen}}^{\text{MIMO}}(\mathbb{X}) + \lambda \left(T - \frac{1}{n_r} \sum_{i=1}^{n_r} \|\mathbf{X}_i\|^2 \right) \right\} \\
&\triangleq U_{\text{Jensen}}^{\text{MIMO}}(\rho)
\end{aligned}$$

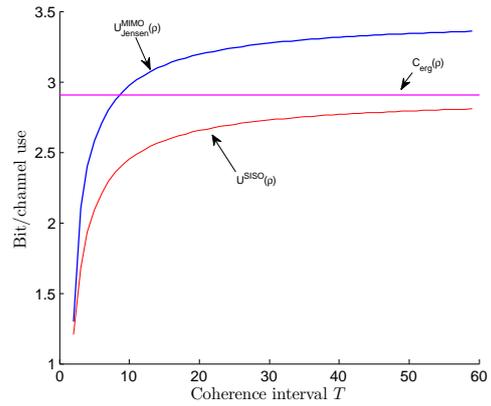
The last inequality can be understood from the fact that the expectation of a function of a rv is never larger than the maximum value taken by the function over the support of the rv.

In figure 8, the upper bounds in (68) and (A.8) are compared at two different, relatively large, SNR in a SISO setup, note that the results for MIMO are also valid for SISO. Even though both of the expressions are valid upper bounds of the capacity, (68) seem to be much tighter. The culprit for the bad performance of (A.8) is the usage of Jensen's inequality in

(A.2). In figure 9, (A.8) is plotted with the upper bound obtained by the result from (84). It can be seen that the upper bound in (A.8) is very loose also for MIMO.

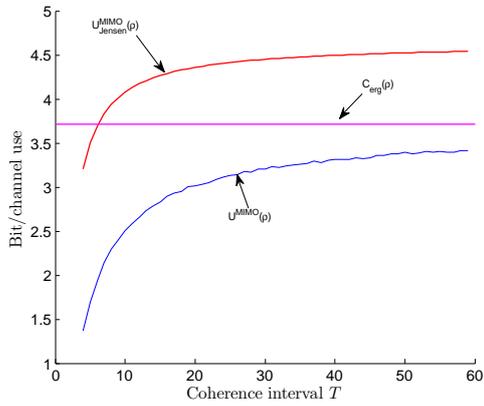


(a) Upper bound at 6 dB

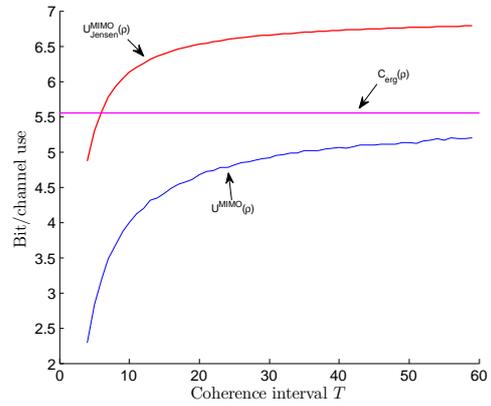


(b) Upper bound at 10 dB

Figure 8: Comparison of (68) and (A.8) for a SISO setup.



(a) Upper bound at 6 dB



(b) Upper bound at 10 dB

Figure 9: Comparison of upper bounds using (A.8) and (84).

B Appendix B

In this appendix, we evaluate the limit in the expression for the moment generating function

$$M(t) = \frac{1}{\prod_{l=1}^{n_r} l!} \int_{\xi_1, \dots, \xi_{n_r}} \left(\prod_{i=1}^{n_r} \xi_i^t \right) \left(\lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \frac{\det(\Gamma)}{\det(\Omega)} \right) \det(\Xi) d\xi_1, \dots, d\xi_{n_r} \quad (\text{B.1})$$

The denominator in (B.1) is the determinant of the Vandermonde matrix in (71). It may be written as [32, ch. 4.3]

$$\lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \det(\Omega) = \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \prod_{1 \leq i < j \leq T} \lambda_j - \lambda_i \quad (\text{B.2})$$

Now, consider taking the limit of ϵ_j for $j = T$, then we have

$$\begin{aligned} & \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \prod_{1 \leq i < j \leq T-1} (\lambda_j - \lambda_i) \prod_{1 \leq i \leq T-1} (\lambda_T - \lambda_i) \\ &= \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_{T-1} \rightarrow 0} \prod_{1 \leq i < j \leq T-1} (\lambda_j - \lambda_i) \prod_{1 \leq i \leq T-1} (1 - \lambda_i). \end{aligned} \quad (\text{B.3})$$

Do the same for $j = T - 1$. We get

$$\begin{aligned} & \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_{T-1} \rightarrow 0} \prod_{1 \leq i < j \leq T-2} (\lambda_j - \lambda_i) \prod_{1 \leq i \leq T-2} (\lambda_{T-1} - \lambda_i)(1 - \lambda_{T-1}) \prod_{1 \leq i \leq T-2} (1 - \lambda_i) \\ & \stackrel{(a)}{=} \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_{T-2} \rightarrow 0} \prod_{1 \leq i < j \leq T-2} (\lambda_j - \lambda_i) \prod_{1 \leq i \leq T-2} (1 - \lambda_i) \left(\lim_{\epsilon_{T-1} \rightarrow 0} \prod_{1 \leq i \leq T-2} -(\lambda_{T-1} - \lambda_i) \right) \\ &= \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_{T-2} \rightarrow 0} \prod_{1 \leq i < j \leq T-2} (\lambda_j - \lambda_i) \prod_{1 \leq i \leq T-2} (1 - \lambda_i) \left(\prod_{1 \leq i \leq T-2} -(1 - \lambda_i) \right) \\ &= \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_{T-2} \rightarrow 0} \prod_{1 \leq i < j \leq T-2} (\lambda_j - \lambda_i) \prod_{1 \leq i \leq T-2} -(1 - \lambda_i)^2 \end{aligned} \quad (\text{B.4})$$

where (a) follows from differentiating once w.r.t. λ_{T-1} . Letting $j = T - 2$ gives

$$\begin{aligned} & \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_{T-2} \rightarrow 0} \prod_{1 \leq i < j \leq T-3} (\lambda_j - \lambda_i) \prod_{1 \leq i \leq T-3} (\lambda_{T-2} - \lambda_i)(1 - \lambda_{T-2})^2 \prod_{1 \leq i \leq T-3} (1 - \lambda_i) \\ & \stackrel{(a)}{=} \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_{T-3} \rightarrow 0} \prod_{1 \leq i < j \leq T-3} (\lambda_j - \lambda_i) \prod_{1 \leq i \leq T-3} (1 - \lambda_i)^2 \left(\lim_{\epsilon_{T-2} \rightarrow 0} \prod_{1 \leq i \leq T-3} 2(\lambda_{T-2} - \lambda_i) \right) \\ &= \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_{T-3} \rightarrow 0} \prod_{1 \leq i < j \leq T-2} (\lambda_j - \lambda_i) \prod_{1 \leq i \leq T-3} (1 - \lambda_i)^2 \left(\prod_{1 \leq i \leq T-2} 2(1 - \lambda_i) \right) \\ &= \lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_{T-3} \rightarrow 0} \prod_{1 \leq i < j \leq T-3} (\lambda_j - \lambda_i) \prod_{1 \leq i \leq T-3} 2(1 - \lambda_i)^3 \end{aligned} \quad (\text{B.5})$$

where (a) follows from differentiating twice w.r.t. λ_{T-2} . Continuing this for $j = T-3, \dots, n_r+1$ gives us

$$\lim_{\epsilon_{n_t+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \det(\Omega) = (-1)^{T-n_r-1} \prod_{1 \leq i < j \leq n_r} (\lambda_j - \lambda_i) \prod_{i=1}^{n_r} (1 - \lambda_i)^{T-n_r} \prod_{i=1}^{T-n_r-1} i! \quad (\text{B.6})$$

We go through similar steps for the numerator in (B.1). The task is to make sure that the matrix, \mathbf{C} have distinct rows. We will take the limit by applying L'hospital's rule to the $T - n_r$ lower rows, starting with the T :th row. Note that for this row, we may just take the limit as it is without differentiating. For the $T - 1$ row we need to differentiate once, for the $T - 2$ row, we need to differentiate twice and so on. After the limit is carried out Γ' is given as

$$\Gamma' = \begin{bmatrix} \Upsilon' & \mathbf{D}' \\ \mathbf{C}' & \mathbf{B}' \end{bmatrix} \quad (\text{B.7})$$

where

- $(\Upsilon')_{k,l} = \lambda_k^{l-1}$, $k = 1, \dots, n_r$ and $l = 1, \dots, T - n_r$
- $(\mathbf{D}')_{k,l} = \lambda_k^{T-n_r-1} e^{-\xi_l/\lambda_k}$, $k = 1, \dots, n_r$ and $l = 1, \dots, n_r$
- $(\mathbf{C}')_{k,l} = \frac{d^{T-k-n_r} \lambda_k^{l-1}}{d\lambda^{T-k-n_r}}$, $k = 1, \dots, T - n_r$ and $l = 1, \dots, T - n_r$
- $(\mathbf{B}')_{k,l} = e^{-\xi_l} \left(\sum_{i=0}^{T-k-n_r-1} \binom{T-k-n_r}{k} \xi_l^i \frac{\Gamma(T-n_r-i)}{\Gamma(k+2n_r)} + \xi_l^{T-k-n_r} \right)$, $k = 1, \dots, T - n_r$ and $l = 1, \dots, n_r$.

This results in a moment generating function given as

$$\begin{aligned} M(t) &= \frac{1}{\underbrace{(-1)^{T-n_r-1} \prod_{l=1}^{n_r} l! \prod_{1 \leq i < j \leq n_r} (\lambda_j - \lambda_i) \prod_{i=1}^{n_r} (1 - \lambda_i)^{T-n_r} \prod_{i=1}^{T-n_r-1} i!}_{\tau}} \\ &\quad \times \int_{\xi_1, \dots, \xi_{n_r}} \det(\Gamma') \det(\text{diag}(\xi_1^t, \dots, \xi_{n_r}^t) \Xi) d\xi_1, \dots, d\xi_{n_r} \\ &\stackrel{(a)}{=} (-1)^{\frac{(T-1)(T+4)}{2}} \tau \det(\mathbf{C}') \int_{\xi_1, \dots, \xi_{n_r}} \det(\mathbf{D}' - \Upsilon'(\mathbf{C}')^{-1}\mathbf{B}') \det(\text{diag}(\xi_1^t, \dots, \xi_{n_r}^t) \Xi) d\xi_1, \dots, d\xi_{n_r} \\ &\stackrel{(b)}{=} (-1)^{\frac{(T-1)(T+4)}{2}} \tau \prod_{i=1}^{T-n_r-1} i! \int_{\xi_1, \dots, \xi_{n_r}} \det(\mathbf{D}' - \Upsilon'(\mathbf{C}')^{-1}\mathbf{B}') \det(\text{diag}(\xi_1^t, \dots, \xi_{n_r}^t) \Xi) d\xi_1, \dots, d\xi_{n_r} \\ &\stackrel{(c)}{=} (-1)^{\frac{(T-1)(T+4)}{2}} \tau \prod_{i=1}^{T-n_r-1} i! \det(\mathbf{A}) \end{aligned} \quad (\text{B.8})$$

where in (a) we have used an identity for determinants of block-matrices [32, pp. 475] and that

$$\begin{aligned} \det \left(\begin{bmatrix} \Upsilon' & \mathbf{D}' \\ \mathbf{C}' & \mathbf{B}' \end{bmatrix} \right) &= (-1)^{\frac{(T-1)(T+4)}{2}} \det \left(\begin{bmatrix} \Upsilon' & \mathbf{D}' \\ \mathbf{C}' & \mathbf{B}' \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{I}_{T-n_r} \\ \mathbf{I}_{n_r} & \mathbf{0} \end{bmatrix} \right) \\ &= (-1)^{\frac{(T-1)(T+4)}{2}} \det \left(\begin{bmatrix} \mathbf{D}' & \Upsilon' \\ \mathbf{B}' & \mathbf{C}' \end{bmatrix} \right). \end{aligned} \quad (\text{B.9})$$

In (b) we observed that \mathbf{C}' is a lower triangular matrix and its determinant is given as $\det(\mathbf{C}') = \prod_{i=1}^{T-n_r-1} i!$. In (c) we used [34, Lemma 1] where \mathbf{A} is given as

$$\begin{aligned}
 (\mathbf{A})_{k,l} &= \int_0^\infty \xi^{t+l-1} \left((\mathbf{D}')_{k,l} - \sum_{d=1}^{T-n_r} \sum_{q=1}^{T-n_r} (\mathbf{C}'^{-1})_{d,q} (\Upsilon')_{k,d} (\mathbf{B})_{q,l} \right) d\xi \\
 &= \int_0^\infty \xi^{t+l-1} \left(\lambda_k^{T-n_r-1} e^{-\xi/\lambda_k} - \sum_{d=1}^{T-n_r} \sum_{q=1}^{T-n_r} (\mathbf{C}'^{-1})_{d,q} \lambda_k^{d-1} \right. \\
 &\quad \left. \times e^{-\xi} \left(\sum_{i=0}^{T-q-n_r-1} \binom{T-q-n_r}{q} \xi^i \frac{\Gamma(T-n_r-i)}{\Gamma(q+2n_r)} + \xi^{T-q-n_r} \right) \right) d\xi \\
 &= \lambda_k^{T-n_r-1+t+l-1} \int_0^\infty \left(\frac{\xi}{\lambda_k} \right)^{t+l-1} e^{-\xi/\lambda_k} d\xi \\
 &\quad - \sum_{d=1}^{T-n_r} \sum_{q=1}^{T-n_r} (\mathbf{C}'^{-1})_{d,q} \lambda_k^{d-1} \sum_{i=0}^{T-q-n_r-1} \binom{T-q-n_r}{q} \frac{\Gamma(T-n_r-i)}{\Gamma(q+2n_r)} \int_0^\infty \xi^{i+t+l-1} e^{-\xi} d\xi \\
 &\quad - \int_0^\infty \xi^{T-q-n_r+t+l-1} e^{-\xi} d\xi \\
 &\stackrel{(d)}{=} \lambda_k^{T-n_r-1+t+l-1} \Gamma(t+l) - \sum_{d=1}^{T-n_r} \sum_{q=1}^{T-n_r} (\mathbf{C}'^{-1})_{d,q} \lambda_k^{d-1} \sum_{i=0}^{T-q-n_r-1} \binom{T-q-n_r}{q} \frac{\Gamma(T-n_r-i)}{\Gamma(q+2n_r)} \Gamma(i+t+l) \\
 &\quad - \Gamma(T-q-n_r+t+l).
 \end{aligned} \tag{B.10}$$

where in (d) we have used the definition of the gamma function and that $\Gamma(l) = (l-1)!$ for $l \in \mathbb{Z}$.

C Appendix C

In this appendix, the Itzykson-Zuber integral is evaluated for the case of non-distinct eigenvalues. The integral is given as

$$\int_{\mathcal{U}(T)} e^{\text{tr}\{\Delta\Phi\Lambda\Phi^H\}} d\Phi \quad (\text{C.1})$$

where $\Phi \in \mathcal{U}(T)$. Let $\Delta \triangleq \text{diag}([\sigma_1 \cdots \sigma_{n_r} \quad 0 \cdots 0]^T) \in \mathbb{C}^{T \times T}$ and $\Lambda \triangleq \text{diag}([\lambda \cdots \lambda \quad 0 \cdots 0]^T) \in \mathbb{C}^{T \times T}$. The Itzykson-Zuber integral is only valid for Δ and Λ with distinct eigenvalues. Therefore, we let $\Delta_\epsilon \triangleq \text{diag}([\sigma_1^2 \cdots \sigma_{n_r}^2 \quad \epsilon_{n_r+1} \cdots \epsilon_T]^T)$ and $\Lambda'_\epsilon \triangleq \text{diag}([\lambda + \epsilon'_1 \cdots \lambda + \epsilon'_{n_t} \quad \epsilon'_{n_t+1} \cdots \epsilon'_T]^T)$, also let $n_r \geq n_t$. Let δ_i and λ'_i denote the i :th element in Δ_ϵ and Λ'_ϵ respectively. Now, we write (C.1) as

$$\begin{aligned} \int_{\mathcal{U}(T)} e^{\text{tr}\{\Delta\Phi\Lambda\Phi^H\}} d\tilde{\Phi} &= \lim_{\epsilon_{n_r+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \lim_{\epsilon'_1 \rightarrow 0, \dots, \epsilon'_{n_t} \rightarrow 0} \lim_{\epsilon'_{n_t+1} \rightarrow 0, \dots, \epsilon'_T \rightarrow 0} \int_{\mathcal{U}(T)} e^{\text{tr}\{\Delta_\epsilon \tilde{\Phi} \Lambda'_\epsilon \tilde{\Phi}^H\}} d\tilde{\Phi} \\ &= \lim_{\epsilon_{n_r+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \lim_{\epsilon'_1 \rightarrow 0, \dots, \epsilon'_{n_t} \rightarrow 0} \lim_{\epsilon'_{n_t+1} \rightarrow 0, \dots, \epsilon'_T \rightarrow 0} |\mathcal{U}(T)| \prod_{i=1}^T \Gamma(i) \frac{\det(\Delta)}{\det(V(\Delta_\epsilon)) \det(V(\Lambda'_\epsilon))} \\ &= \lim_{\epsilon_{n_r+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \lim_{\epsilon'_1 \rightarrow 0, \dots, \epsilon'_{n_t} \rightarrow 0} \frac{|\mathcal{U}(T)| \prod_{i=1}^T \Gamma(i)}{\prod_{i=1}^{n_t} (\lambda'_i)^{(T-n_t)} \prod_{i=1}^{n_t-1} i! \det(V(\Delta_\epsilon)) \det(V(\text{diag}(\epsilon'_1, \dots, \epsilon'_{n_t})))} \det(\mathbb{A}_1) \\ &= \lim_{\epsilon_{n_r+1} \rightarrow 0, \dots, \epsilon_T \rightarrow 0} \frac{|\mathcal{U}(T)| \prod_{i=T-n_t+1}^T \Gamma(i)}{\prod_{i=1}^{n_t} (\lambda_i)^{(T-n_t)} \prod_{i=1}^{n_t-1} i! \det(V(\Delta_\epsilon))} \det(\mathbb{A}_2) \end{aligned} \quad (\text{C.2})$$

Here $\mathbf{A} = (e^{\lambda_i \delta_j})_{i,j} \in \mathbb{R}^{T \times T}$. The matrix \mathbf{A}_1 is given by $(\mathbf{A}_1)_{i,j} = (e^{\lambda_i \delta_j})$, $i = 1, \dots, n_t, j = 1, \dots, T$ and $(\mathbf{A}_1)_{i,j} = \delta_j^{T-i}$, $i = n_t+1, \dots, T, j = 1, \dots, T$. The matrix \mathbf{A}_2 is given by $(\mathbf{A}_2)_{i,j} = (\delta_j^{n_t-i} e^{\lambda \delta_j})$, $i = 1, \dots, n_t, j = 1, \dots, T$ and $(\mathbf{A}_2)_{i,j} = \delta_j^{T-i}$, $i = n_t+1, \dots, T, j = 1, \dots, T$. The two limits are carried out by iterative use of L'hopital's rule similarly as in appendix B. The last inequality is also carried out by the use of L'hopital's rule and we get

$$\int_{\mathcal{U}(T)} e^{\text{tr}\{\Delta\Phi\Lambda\Phi^H\}} d\tilde{\Phi} = \frac{|\mathcal{U}(T)| \prod_{i=T-n_t+1}^T \Gamma(i)}{\prod_{i=1}^{n_t} \lambda_i^{(T-n_t)} \prod_{i=1}^{n_t-1} i!} \frac{1}{\prod_{j=1}^{T-n_r-1} j! \prod_{i=1}^{n_r} \sigma_i^{T-n_t}} \frac{\det(\mathbb{A}_3)}{\det(V(\Sigma))}. \quad (\text{C.3})$$

The matrix \mathbb{A}_3 is a block matrix given as

$$\mathbb{A}_3 = \begin{pmatrix} \mathbb{A}_{31} & \mathbb{A}_{32} \\ \mathbb{A}_{33} & \mathbb{A}_{34} \end{pmatrix} \quad (\text{C.4})$$

where

- $\mathbb{A}_{31} \in \mathbb{R}^{n_t \times n_r}$ where $(\mathbb{A}_{31})_{ij} = e^{\lambda \sigma_j} \sigma_j^{(n_t-i)}$ for $i = 1, \dots, n_t$ and $j = 1, \dots, n_r$.
- $\mathbb{A}_{33} \in \mathbb{R}^{(T-n_t) \times n_r}$ with $(\mathbb{A}_{33})_{ij} = \sigma_j^{(T-n_t-i)}$ for $i = 1, \dots, T - n_t$ and $j = 1, \dots, n_r$.
- $\mathbb{A}_{34} \in \mathbb{R}^{(T-n_t) \times (T-n_r)}$ where $(\mathbb{A}_{34})_{i,j} = \begin{cases} 0 & \text{if } i \leq n_r - n_t \\ \text{diag}((T - n_r - j)!), j = 1, \dots, T - n_r & \text{if } i > n_r - n_t \end{cases}$

and $\mathbb{A}_{32} \in \mathbb{R}^{n_t \times T}$ is given as

$$\mathbb{A}_{32} = \begin{bmatrix} (\lambda^{T-n_t-1})^{(n_t-1)} & (\lambda^{T-n_t-2})^{(n_t-1)} & \dots & \lambda^{(n_t-1)} & 0 \\ (\lambda^{T-n_t-1})^{(n_t-2)} & (\lambda^{T-n_t-2})^{(n_t-2)} & \dots & \lambda^{(n_t-2)} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (\lambda^{T-n_t-1})^{(1)} & (\lambda^{T-n_t-2})^{(1)} & \dots & \lambda^{(1)} & 0 \\ (\lambda^{T-n_t-1})^{(0)} & (\lambda^{T-n_t-2})^{(0)} & \dots & \lambda^{(0)} & 0 \end{bmatrix} \quad (\text{C.5})$$

where $(\lambda^a)^{(b)} \triangleq \frac{d^b \lambda^a}{d\lambda^b}$. In this thesis, a scenario of special interest is the square case, i.e., $n_t = n_r$, for which the determinant of \mathbb{A}_3 may be computed with the identity [32, pp. 475]

$$\begin{aligned} \det(\mathbb{A}_3) &= \det(\mathbf{A}_{34}) \det(\mathbb{A}_{31} - \mathbf{A}_{32} \mathbf{A}_{34}^{-1} \mathbb{A}_{33}) \\ &= \det \left(\left[\sigma_k^{(n_t-l)} \tilde{\gamma}(T+l-2n_t, \lambda \sigma_k^2) \right]_{1 \leq l, k \leq n_t} \right) e^{\lambda \text{tr}\{\boldsymbol{\Sigma}^2\}} \prod_{j=1}^{T-n_t-1} j!. \end{aligned} \quad (\text{C.6})$$

where $\tilde{\gamma}(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt$ denotes the regularized incomplete gamma function.

References

- [1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, October 1948.
- [2] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, pp. 2307–2359, May 2010.
- [3] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static SIMO fading channels at finite blocklength," *IEEE International Symposium on Information Theory*, vol. 31, pp. 117–132, February 2013.
- [4] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Transactions on Information Theory*, vol. 60, pp. 4232–4265, July 2014.
- [5] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Diversity versus channel knowledge at finite blocklength," *Information Theory Workshop (ITW)*, vol. 2012 IEEE, pp. 572–576, September 2012.
- [6] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons Inc., 1968.
- [7] W. E. Ryan and S. Lin, *Channel Codes classical and modern*. The Edinburgh building, Cambridge CB2 8RU, UK: Cambridge Univ. Press, 2009.
- [8] F. Rusek, A. Lozano, and N. Jindal, "Mutual information of iid complex gaussian signals on block rayleigh-faded channels," *IEEE Transactions on Information Theory*, vol. 58, pp. 331–340, January 2012.
- [9] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: the received signal density in closed form," *IEEE Transactions on Information Theory*, vol. 48, pp. 1473–1484, June 2002.
- [10] G. Alfano, C.-F. Chiasserini, M. Nordio, and S. Zhou, "Output statistics of MIMO channels with general input distribution," in *2013 IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp. 829–833, July 2013.
- [11] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. The Edinburgh building, Cambridge, UK: Cambridge Univ. Press, 2005.
- [12] A. Lapidoth and S. S. Shamai, "Fading channels: How perfect need 'perfect side information' be?," *IEEE Transactions on Information Theory*, vol. 48, pp. 1118–1134, May 2002.
- [13] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, pp. 139–157, January 1999.
- [14] R. H. Clarke, "A statistical theory of mobile-radio reception," tech. rep., Alcatel-lucent, 1968.
- [15] A. Lozano, "Interplay of spectral efficiency, power and Doppler spectrum for reference-signal-assisted wireless communication," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 5020–5028, December 2008.
- [16] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.
- [17] L. Zheng, *Diversity-Multiplexing Tradeoff: A Comprehensive View of Multiple Antenna Systems*. PhD thesis, University of California at Berkeley, Fall 2002.

- [18] A. Lozano and N. Jindal, "Transmit diversity vs. spatial multiplexing in modern MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 186–197, January 2010.
- [19] T. M. Cover and J. A. Thomas, *Elements of information theory*. 111 River Street, Hoboken, New Jersey: Wiley, second ed., 2006.
- [20] J. N. Laneman, "On the distribution of mutual information," in *Proc. Workshop on Information Theory and its Applications (ITA)*, (San Diego, CA), February 2006.
- [21] Í. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, November 1999.
- [22] H. Shin and M. Z. Win, "Gallager's exponent for MIMO channels: A reliability-rate trade-off," *IEEE Trans. Commun.*, vol. 57, pp. 972–984, April 2009.
- [23] I. Abou-Faycal and B. M. Hochwald, "Coding requirements for multiple-antenna channels with unknown Rayleigh fading," tech. rep., Bell Labs Technical Memorandum, March 1999.
- [24] S. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Upper Saddle River, New Jersey, USA: Prentice Hall, 1998.
- [25] L. Zheng and D. N. C. Tse, "Communication on the grassman manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, pp. 359–383, February 2002.
- [26] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 117–132, February 2013.
- [27] M. Katz and S. S. Shamai, "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 2257–2270, October 2004.
- [28] A. Martinez, "Spectral efficiency of optical direct detection," *Journal of the optical society of America B*, vol. 24, pp. 739–749, April 2007.
- [29] A. Edelman, *Eigenvalues and Condition Numbers of Random Matrices*. PhD thesis, Massachusetts Institute of Technology, May 1989.
- [30] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. The Edinburgh building, Cambridge CB2 8RU, UK: Cambridge Univ. Press, 2011.
- [31] S. L. Miller and D. G. Childers, *Probability and random processes with applications to signal processing and communications*. 200 Wheeler Road, Burlington, MA, USA: Elsevier, 2004.
- [32] C. D. Meyer, *Matrix analysis and applied linear algebra*. 3600 University City Science Center, Philadelphia, PA, USA: Society for Industrial and applied mathematics, 2000.
- [33] P. G. Moschopoulos, "The distribution of the sum of independent gamma random variables," *Annals of the Institute of Statistical Mathematics*, vol. 37, pp. 541–544, June 1985.
- [34] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multiantenna communication," *IEEE Transactions on Information Theory*, vol. 51, pp. 4134–4151, December 2005.
- [35] C. Itzykson and J.-B. Zuber, "The planar approximation II," *Journal of Mathematical Physics*, vol. 21, pp. 411–421, March 1980.
- [36] Y. Polyanskiy, "Scalar coherent fading channel: dispersion analysis," in *IEEE International Symposium on Information Theory Proceedings*, pp. 2959–2963, July 2011.
- [37] S. M. Alamouti, "A simple transmit diversity scheme," *IEEE Journal on selected Areas in*

- Communications*, vol. 16, pp. 1451–1458, October 1998.
- [38] V. Tarokh and H. Jafarkhani, “Space-time block codes from orthogonal designs,” *IEEE Transactions on Information Theory*, vol. 45, pp. 1456–1467, July 1999.
- [39] E. Dahlman, S. Parkvall, and J. Sköld, *4G LT /LTE-Advanced for Mobile Broadband*. The Boulevard, Langford Lane, Kidlington, Oxford: Elsevier, 2011.
- [40] METIS project, Deliverable D1.1, “Scenarios, requirements and KPIs for 5G mobile and wireless system,” tech. rep., April 2013.