Degrees of Freedom in Underspread MIMO Fading Channels

Raul Etkin and David Tse
Dept. of Electrical Engineering and Computer Science
University of California, Berkeley
e-mail: {retkin,dtse}@eecs.berkeley.edu

Abstract

We analyze the capacity of a Multiple Input-Multiple Output (MIMO) fading channel in which the fading process varies slowly over time. We assume that neither the transmitter nor the receiver have knowledge of the fading process, and analyze the capacity of the channel in the large signal to noise ratio, and slow channel variation regime. In all regimes of practical interest the use of multiple antennas provides large capacity improvements. This capacity improvement can be seen as the benefit of having multiple spatial degrees of freedom. We present a communication scheme that achieves the full number of degrees of freedom of the channel with tractable complexity.

1 Introduction

Recent information theoretic results suggest that in richly scattered wireless environments, systems with multiple transmit and multiple receive antennas (MIMO systems) can have very large capacities. In particular, Foschini and Gans [1] and Telatar [2] considered a channel with n_t transmit and n_r receive antennas, with Rayleigh flat faded channel gains i.i.d. across antenna pairs, and showed that at high SNR, the capacity of this channel grows like $\min(n_t, n_r) \log \mathsf{SNR}$ for large SNR. This yields a $\min\{n_t, n_r\}$-fold increase in capacity over a channel with a single transmit and a single receive antenna. The parameter $\min\{n_t, n_r\}$ can be interpreted as the number of *degrees of freedom* (d.o.f.) of the channel: the dimension of the space over which communication can take place.

The above result assumes that the receiver can perfectly track the fading gains of the channel (so-called perfect channel state information CSI at the receiver). In high mobility applications, this may not be a reasonable assumption. Moreover, in the high SNR regime where the amount of noise is small, it is conceivable that the impact of channel uncertainty on performance is more pronounced. This leads to the question: what is the high SNR capacity of time-varying fading channels without the prior assumption of CSI? Towards this end, Lapidoth and Moser [3], building on earlier work by Taricco and Elia [4], recently showed a contrasting result: that at high SNR, the first order term of the capacity is $\log \log \mathsf{SNR}$, regardless of what the number of transmit and receive antennas is. Thus, not only does capacity grow much slower than in the case with perfect CSI, but this result also suggests that without CSI, the performance gain from having multiple antennas, if any, will only appear as a second-order effect. The increase in the number of the degrees of freedom has minimal impact.

Does this then suggest that the perfect CSI results are very fragile? Even though [3] explicitly incorporates the channel variation (and the associated uncertainty) in its model, this result is still asymptotic in the SNR and thus one has to be careful in interpreting its regime of validity. In particular, since the channel variation process is fixed while the SNR is taken to infinity, it is conceivable that the $\log\log \mathsf{SNR}$ growth only occurs when the noise level is much smaller than the amount of channel variation from one sample to the next. However, typical wireless channels are *underspread*, which means that this variation is small. Thus, one has to look at the effect of the SNR and the amount of channel variation *simultaneously* to get a more complete picture.

In this work, we would like to put forth such a picture. Suppose the fading process for each channel gain is Gauss-Markov with a one-step MMSE prediction error of ϵ from sample to sample. For underspread channels, ϵ is small. We propose that the capacity $C(\mathsf{SNR}, \epsilon)$ of an underspread MIMO fading channel (without CSI) for $\mathsf{SNR} \gg 1$ and $\epsilon \ll 1$ can be described in three regimes:

Regime 1: $\mathsf{SNR} < 1/\epsilon$

$$C(\mathsf{SNR}, \epsilon) \sim \min\{n_t, n_r\} \log \mathsf{SNR}.$$

Regime 2: $1/\epsilon \leq \mathsf{SNR} < \exp\left(\epsilon^{-\min\{n_t, n_r\}}\right)$

$$C(\mathsf{SNR}, \epsilon) \sim \min\{n_t, n_r\} \log 1/\epsilon.$$

Regime 3: $\mathsf{SNR} \geq \exp\left(\epsilon^{-\min\{n_t, n_r\}}\right)$

$$C(\mathsf{SNR}, \epsilon) \sim \log\log \mathsf{SNR}.$$

In the first regime, the channel prediction error is smaller than the inverse of the SNR. The system is noise-limited and its capacity behaves as though there is perfect CSI at the receiver. In the second regime, the SNR is now larger than the inverse of the channel prediction error and the system is now limited by channel uncertainty. However, when the SNR gets much larger, the Lapidoth-Moser regime kicks in and the system is again noise-limited, albeit with a much smaller growth rate. The important point is that in both regime 1 and 2, the capacity is proportional to the degrees of freedom in the channel.

To get a feeling of the values of SNR that separate the three regimes consider an $n_r = n_t = 4$ system. For urban environments with mobile speeds in the order of $5 - 50$ km/h, with carrier frequencies ranging from 800 Mhz to 5 Ghz the threshold between regimes 1 and 2 can range from 17.4 to 40 dB, while the threshold that separates regimes 2 and 3 can range from $4.1 \cdot 10^7$ to $4.3 \cdot 10^{16}$ dB. For indoor environments these thresholds are even larger.

In the rest of this paper, we present quantitative results to support this picture as well as quantify when regime 3 kicks in. We argue that typical wireless scenarios fall in regimes 1 and 2 but rarely in 3. This suggests that in underspread fading channels, multiple antennas provide significant gains and the concept of degrees of freedom is a useful measure of that performance gain, even without the assumption of perfect CSI.

We will use lowercase or uppercase letters to represent scalars, boldface lowercase letters for vectors, uppercase calligraphic letters for sets, and boldface uppercase letters for matrices. For example we write n for a scalar, $\mathbf{v}$ for a vector, $\mathcal{A}$ for a set and $\mathbf{H}$ for a matrix. We will denote by $\mathbf{v}^{\mathcal{A}}$ the vector with components $v[n]$ for $n \in \mathcal{A}$, where $v[n]$ is the nth component of $\mathbf{v}$. We use $h(\cdot)$ to represent differential entropy to the base e, $1(\cdot)$ for the indicator function, $\log(\cdot)$ for natural logarithm, and $\delta_{i,j}$ for the Kroneker's delta function. We write $\mathbf{H}^T$ for the transpose, $\mathbf{H}^*$ for the conjugate, and $\mathbf{H}^\dagger$ for the hermitian (conjugate transpose) of the complex matrix $\mathbf{H}$ and we represent the kxk identity matrix by $\mathbf{I}_k$. Finally we write $\|\mathbf{v}\|$ for the L_2 norm of the vector $\mathbf{v}$, i.e. $\|\mathbf{v}\| = \sqrt{\mathbf{v}^\dagger\mathbf{v}}$, and $\|\mathbf{H}\|_F$ for the Frobenius norm of the matrix $\mathbf{H}$, i.e. $\|\mathbf{H}\|_F = \sqrt{\mathrm{tr}(\mathbf{H}\mathbf{H}^\dagger)}$.

2 Channel Model

Throughout this work we will use a flat fading Rayleigh, discrete-time, baseband model. We will consider the general case when n_t transmit and n_r receive antennas are used. The channel equation is:

$$\mathbf{y}[n] = \mathbf{H}[n]\mathbf{x}[n] + \sqrt{(n_t/\mathsf{SNR})}\mathbf{z}[n] \tag{1}$$

where $\mathbf{y}[n] \in \mathcal{C}^{n_r}$ is the channel output, $\mathbf{x}[n] \in \mathcal{C}^{n_t}$ is the channel input with average power constraint $\sum_{n=1}^{B} E[\|\mathbf{x}[n]\|^2] \leq Bn_t$, $\mathbf{z}[n] \sim \mathcal{CN}(0, I_{n_r})$ is circularly symmetric white complex Gaussian noise, SNR is the signal to noise ratio, and $\mathbf{H}[n] \in \mathcal{C}^{n_r \mathrm{x} n_t}$ is the fading matrix with i.i.d. (independent and identically distributed) circularly symmetric complex Gaussian components of zero mean and unit variance. The time variation of the channel is modeled by a Gauss-Markov process:

$$\mathbf{H}[n+1] = \sqrt{1-\epsilon}\mathbf{H}[n] + \sqrt{\epsilon}\mathbf{W}[n] \tag{2}$$

where $\mathbf{W}[n] \in \mathcal{C}^{n_r \mathrm{x} n_t}$ has circularly symmetric complex Gaussian components of zero mean and unit variance, independent across rows, columns and time indices n, $0 \leq n \leq B$. The coherence time of the channel is controlled by the parameter $\epsilon \in \mathcal{R}$, $0 \leq \epsilon \leq 1$. As $\epsilon \to 0$ we get the limiting case of a constant channel. Also $\mathbf{H}[0]$ has zero mean, unit variance complex Gaussian independent components.

We will use the channel by grouping input symbols in blocks of size B, i.e. $\mathbf{x}^{\mathcal{N}} = (\mathbf{x}[1], \ldots, \mathbf{x}[B]) \in \mathcal{C}^{n_t \mathrm{x} B}$ where $\mathcal{N} = \{1, \ldots, B\}$, and compute the channel capacity using the following expression:

$$C(\mathsf{SNR}, \epsilon) = \lim_{B\to\infty} \left[\sup_{\substack{p(\mathbf{x}^{\mathcal{N}}) \\ \sum_{n=1}^{B} E[\|\mathbf{x}[n]\|^2] \leq Bn_t}} \frac{1}{B} I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}) \right] \tag{3}$$

The above maximization is performed by selecting a channel input distribution among all possible input distributions that satisfy the power constraint. Note that the channel matrix $\mathbf{H}[n]$ is unknown at both the transmitter and the receiver. No closed form expression for (3) is known.

It is useful to compute the typical values of ϵ for different applications. Since we are dealing with a flat fading channel, the signals must be restricted to a bandwidth of the order of the coherence bandwidth of the channel W_c [5]:

$$W_c \approx \frac{1}{5\sigma_\tau} \tag{4}$$

where σ_τ is the RMS delay spread of the channel. We note that it is possible to define W_c in different ways, and here we are only interested in doing an order of magnitude calculation. Measured values of σ_τ range from 1μs to 2μs in urban environments and from 10ns to 100ns in indoor environments, [5] and references therein. The bandwidth of the channel determines the sampling interval used in the discrete time model. If the passband channel has a bandwidth W_c, the baseband representation has a bandwidth $W_c/2$ and the sampling theorem allows us to take samples at a rate W_c without loss of information. The coherence time of the channel T_c represents the time over which the fading coefficients are highly correlated. If we define T_c as the time over which the auto correlation function is above 0.5 of its value at 0 then [6]:

$$T_c \approx \frac{9}{16\pi f_m} \tag{5}$$

where f_m is the maximum Doppler shift given by $f_m = v/\lambda$, where v is the mobile speed and λ is the wavelength. We can compute the autocorrelation of the process defined by (2), set it equal to $1/2$ and solve for the corresponding value of ϵ:

$$\epsilon = 1 - 2^{\frac{-2}{T_c W_c}} \simeq \frac{2 \log 2}{T_c W_c} \tag{6}$$

where the approximation is done for $T_c W_c \gg 1$. When $T_c W_c \gg 1$ the channel is said to be underspread [7], in which case we have $\epsilon \ll 1$. For indoor environments ϵ ranges from $3 \cdot 10^{-7}$ to 10^{-4} for mobile speeds of $1 - 5$km/h and carrier frequencies ranging from 800MHz to 5GHz. For slow fading outdoor environments with mobile speeds of the order of 5km/h ϵ varies from 10^{-4} to $1.8 \cdot 10^{-3}$ for the same range of carrier frequencies. Even in a fast fading scenario with a mobile speed of 50km/h and a high carrier frequency of 5GHz, ϵ is still only $1.8 \cdot 10^{-2}$. This suggests that the regime of $\epsilon \to 0$ is a natural one to look at.

The other parameter of our model that we want to consider is the SNR. We will analyze different high SNR regimes depending on how the SNR compares to $1/\epsilon$, so it is useful to have an idea of the practical values that the SNR can take. There are a number of factors that limit how large the SNR can be. In multi-user systems that are interference limited, the SNR is limited by the number of users that are sharing the channel[1] . In multi-user systems where the users are kept orthogonal in time, frequency or code, the achievable values of SNR are generally much larger. But even in point to point links there are a number of factors that limit SNR such as antenna effective noise temperature, receiver noise figure, quantization noise, etc. In practice it is difficult to achieve values of SNR much larger than 30dB.

3 The Three Regimes

The input symbols of the channel (1) are vectors in $\mathcal{C}^{n_t}$. The rank of the matrix $\mathbf{H}[n]$ determines how many of these dimensions are resolvable at the receiver. The assumption that $\mathbf{H}[n]$ is formed by i.i.d. Gaussian components implies that $\mathbf{H}[n]$ is full rank with probability one. If $n_r \leq n_t$ the received signal is a vector in $\mathcal{C}^{n_r}$ so at most n_r of the n_t dimensions of the input vector can effectively

[1] In fact when there is interference SNR is to be interpreted as signal to interference plus noise ratio.

be resolved. If on the other hand $n_r \geq n_t$ the information bearing component of the received signal is contained in a subspace of dimension n_t of $\mathcal{C}^{n_r}$. Therefore in both cases $\text{rank}(\mathbf{H}[n]) =_{a.s.} \min\{n_t, n_r\}$ determines how many d.o.f. the channel offers for conveying information. This intuitive argument based on the physics of the channel is backed up by the result [1, 2] that for large SNR the capacity of the channel (1) grows as $\min\{n_t, n_r\} \log \mathsf{SNR}$ when the channel realization is known at the receiver. Based on this we define the number of *degrees of freedom* of the channel to be $\min\{n_t, n_r\}$.

Now the following question arises: do the degrees of freedom still determine the capacity of a fading channel without the assumption of CSI? To answer this question we start by presenting an asymptotic lower bound on the capacity $C(\mathsf{SNR}, \epsilon)$, asymptotic for high SNR and small prediction error ϵ.

Theorem 1

$$\liminf_{\substack{\mathsf{SNR}\to\infty \\ \epsilon\to 0}} \{C(\mathsf{SNR}, \epsilon) - n_{min} \log [\min(\mathsf{SNR}, 1/\epsilon)] - K_1(n_r, n_t)\} \geq 0 \tag{7}$$

where $n_{min} = \min\{n_r, n_t\}$, and $K_1(n_r, n_t)$ is some constant independent of ϵ and SNR.

Proof: See Appendix B. ■

This bound and the non-asymptotic version that we present later in this section can be derived by using Gaussian inputs which are i.i.d. across time and space. We leave the proofs of these theorems for the appendices, and focus on the implications of these results.

If the receiver were to predict the channel from previous values of the channel gains, the prediction error would be ϵ. On the other hand $1/\mathsf{SNR}$ is proportional to the noise power. By comparing these 2 quantities we can differentiate 2 regimes of operation. For large SNR and small ϵ, with $\mathsf{SNR} < 1/\epsilon$, the lower bound behaves as $n_{min} \log \mathsf{SNR} + K_1(n_r, n_t)$. In this case the prediction error is negligible as compared to the noise power and the performance is similar to that of the channel with perfect CSI at the receiver. As SNR is increased beyond $1/\epsilon$ the lower bound takes a constant value (as a function of SNR) given by $n_{min} \log(1/\epsilon) + K_1(n_r, n_t)$. In this case the prediction error becomes dominant and any additional reduction in noise variance does not have a significant impact on the lower bound.

We see that $\theta_1 = 1/\epsilon$ establishes a threshold on SNR that separates 2 regimes of operation: the regime where the capacity is limited by the noise, and the regime where the capacity is limited by the channel prediction error. In both cases we see that increasing n_{min} results in improved performance. Irrespective of the source of uncertainty (prediction error or noise) n_{min} plays a significant role in the lower bound and, as will be shown later in this section, on channel capacity.

In the next section we present a communication scheme that achieves the lower bound (7), using i.i.d Gaussian inputs, interleaving, decision-oriented channel estimation and weighted minimum Euclidean distance decoding. The analysis of the performance of this scheme shows that the achievable throughput can be lower bounded by the capacity of a Gaussian fading channel with perfect CSI at the receiver and with a noise term with variance that depends linearly on SNR^{-1} and ϵ. We see that channel estimation error results in an increase in noise variance that depends linearly on ϵ. Both in regimes 1 and 2 the degrees of freedom of the channel have the same role in capacity: that of defining the dimensionality of the space over which communication takes place.

The difference between regimes 1 and 2 is the dominant noise term; additive Gaussian noise and channel estimation error respectively.

However we note that the lower bound (7) is not tight for very large SNR. As is shown in [3] channel capacity ultimately grows as $\log\log \mathsf{SNR}$. We must then define a third regime of operation. This regime corresponds to the range of SNR's for which the doubly logarithmic term has a significant influence on the high SNR capacity expansion. For this to occur the $\log\log \mathsf{SNR}$ term must be comparable to the value of the capacity, which must be at least as large as the lower bound (7). So we define a second threshold, $\theta_2 = \exp[(1/\epsilon)^{n_{min}}]$ which corresponds to the value of SNR that makes $\log\log \mathsf{SNR}$ of the same order of the lower bound. The range of SNR's corresponding to the third regime must therefore lie to the right of θ_2. In this third regime increasing the number of transmit or receive antennas beyond 1 does not produce a significant increase in C, so we cannot fully exploit any of the available degrees of freedom. Basically, communication over all of the degrees of freedom becomes limited by channel uncertainty and the additional SNR can only be exploited by conveying information in the norm of the input and output vectors. This method of conveying information does not take advantage of the increase in dimensionality, and if this is the main term of the capacity we don't observe significant gains in the addition of degrees of freedom.

It is interesting to note that θ_2 grows doubly exponentially with n_{min}, or equivalently θ_2 grows exponentially in dB with n_{min}. Thus, the range of SNR values for which the third regime is relevant increases very rapidly with the number of antennas. To get an idea of the typical values that θ_1 and θ_2 can take, we can compute them for the extreme values of ϵ, corresponding to the minimum and maximum mobile speeds and carrier frequencies, considered in Section 2 [2]:

Environment	$n_t = n_r$	ϵ (max)	ϵ (min)	θ_1 (min)	θ_1 (max)	θ_2 (min)	θ_2 (max)
Indoors	1	10^{-4}	$3 \cdot 10^{-7}$	40 dB	65 dB	$4.3 \cdot 10^{4}$ dB	$1.4 \cdot 10^{7}$ dB
Indoors	4	10^{-4}	$3 \cdot 10^{-7}$	40 dB	65 dB	$4.3 \cdot 10^{16}$ dB	$5.4 \cdot 10^{26}$ dB
Urban	1	0.018	10^{-4}	17.4 dB	40 dB	241 dB	$4.3 \cdot 10^{4}$ dB
Urban	4	0.018	10^{-4}	17.4 dB	40 dB	$4.1 \cdot 10^{7}$ dB	$4.3 \cdot 10^{16}$ dB

From this table we see that in practice, the third regime of operation requires extremely large values of SNR, values which increase very rapidly with n_{min}. In practical systems these values of SNR are never attained.

To see the effect of the increase in d.o.f. on the lower bound for specific values of ϵ and SNR we include a non-asymptotic lower bound on channel capacity derived using information theoretic principles:

Theorem 2

$$
\begin{aligned}
C(\mathsf{SNR}, \epsilon) \geq & \sum_{k=n_{max}-(n_{min}-1)}^{n_{max}} E\left[\log\left(1 + \frac{\min\{\mathsf{SNR}, \mathsf{SNR}^*\}}{n_t}\chi^2_{2k}\right)\right] + n_r n_t \log(2) \\
& - \; n_r n_t \log\left[\min(\mathsf{SNR}, \mathsf{SNR}^*)\frac{\epsilon}{n_t} + 2 - \epsilon + \sqrt{\left(\min(\mathsf{SNR}, \mathsf{SNR}^*)\frac{\epsilon}{n_t} + 2 - \epsilon\right)^2 - 4(1-\epsilon)}\right]
\end{aligned}
\quad (8)
$$

[2]Here θ_i(min) corresponds to the maximum value of ϵ and θ_i(max) to the minimum value of ϵ for a particular scenario.

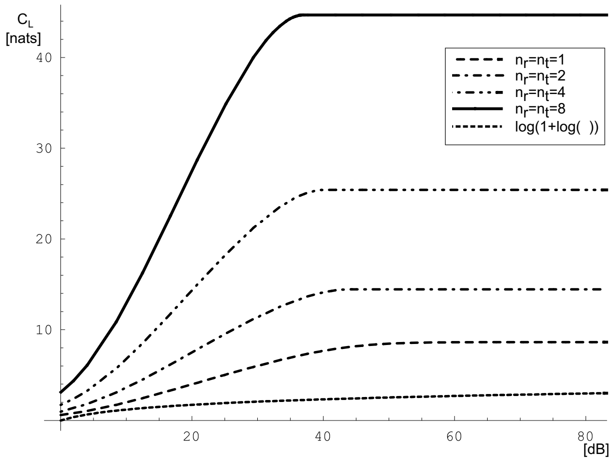

Figure 1: Capacity lower bound as a function of the SNR (ρ) for different values of $n_t = n_r$, and $\epsilon = 10^{-4}$. The curve $\log(1 + \log \mathsf{SNR})$ is plotted for reference; it would cross the curve corresponding to $n_r = n_t = 1$ at $\mathsf{SNR} \approx 43000$dB. The crossing points for the other curves occur at much larger values of SNR.

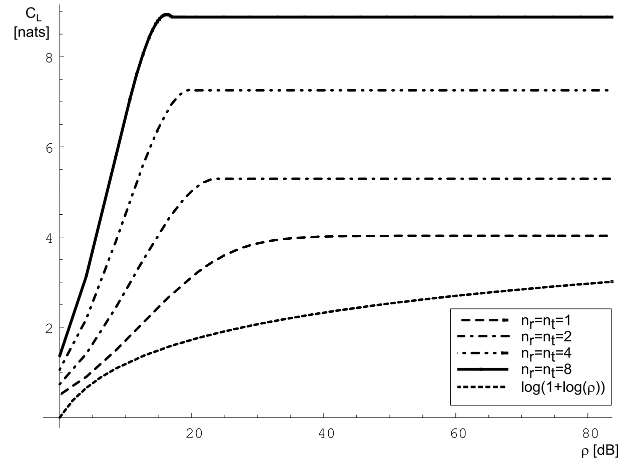

Figure 2: Capacity lower bound as a function of the SNR (ρ) for different values of $n_t = n_r$, and $\epsilon = 0.01$. The curve $\log(1 + \log \mathsf{SNR})$ is plotted for reference; it would cross the curve corresponding to $n_r = n_t = 1$ at $\mathsf{SNR} \approx 400$dB. The crossing points for the other curves occur at much larger values of SNR.

where $n_{min} = \min\{n_r, n_t\}$, $n_{max} = \max\{n_r, n_t\}$, χ^2_{2k} *is a* χ^2 *random variable with* $2k$ *degrees of freedom, and*

$$\mathsf{SNR}^* = \begin{cases} \frac{n_t}{\epsilon} \cdot \frac{n_{min}^2(2-\epsilon)+\sqrt{4n_{min}^4(1-\epsilon)+\epsilon^2 n_{min}^2 n_r^2 n_t^2}}{n_r^2 n_t^2 - n_{min}^2} & if \quad n_{max} > 1 \\ \infty & if \quad n_{max} = 1 \end{cases} \tag{9}$$

Proof: See Appendix A. ■

Figures 1 and 2 show the value of this lower bound for different number of transmit and receive antennas as a function of the SNR for $\epsilon = 10^{-4}$ and $\epsilon = 10^{-2}$ respectively. We observe that for all practical values of SNR increasing the number of degrees of freedom results in a considerable improvement.

The performance analysis of the communication scheme presented in the next section also provides a lower bound on channel capacity. Because of the bounding techniques used, the resulting lower bound is not as tight as (8) for certain values of the parameters, so we include both bounds for completeness.

Motivated by the lower bound in Theorem 1, we argued that increasing the number of degrees of freedom of the system can be exploited in the first 2 regimes. However, the lower bound gives only partial information about the capacity of the channel. For the conclusions to be fundamental we need to complement the argument with an upper bound that shows the same behavior as the lower bound for large SNR and small ϵ. Particular focus is on regime 1 and regime 2 which cover most common wireless scenarios.

We present the following asymptotic upper bound on the channel capacity.

Theorem 3 *Let* $n_t \geq n_r$. *Then,*

$$\limsup_{\substack{\mathsf{SNR}\to\infty \\ \epsilon\to 0}} \left[C(\mathsf{SNR}, \epsilon) - \{n_{min} \log\left[\min(\mathsf{SNR}, 1/\epsilon)\right] + \log(\log \mathsf{SNR}) + K_2(n_r, n_t)\}\right] \leq 0 \tag{10}$$

where $n_{min} = \min\{n_r, n_t\}$ *and* $K_2(n_r, n_t)$ *is some constant independent of* SNR *and* ϵ.

Proof: See Section 5.1. ■

Theorem 3 together with Theorem 1 allow us to tightly characterize the behavior of channel capacity in the first 2 regimes for the case $n_t \geq n_r$. As before we focus on the scenario when $\epsilon \approx 0$ and $\mathsf{SNR} \gg 1$.

For the case when $\mathsf{SNR} < 1/\epsilon$, the above asymptotic upper bound gives

$$n_{min} \log \mathsf{SNR} + \log\log \mathsf{SNR} \approx n_{min} \log \mathsf{SNR},$$

which matches the lower bound in regime 1 (to within a constant not depending on SNR and ϵ.) For the case when $\mathsf{SNR} > 1/\epsilon$ but $\mathsf{SNR} \ll \exp(1/\epsilon)$, the above asymptotic upper bound becomes

$$n_{min} \log 1/\epsilon + \log\log \mathsf{SNR} \approx n_{min} \log 1/\epsilon,$$

which matches the lower bound in regime 2.

Finally for the case $\mathsf{SNR} > \exp(1/\epsilon)$ we can approximate

$$n_{min} \log 1/\epsilon + \log\log \mathsf{SNR} \approx \log\log \mathsf{SNR},$$

in accordance with the Lapidoth-Moser result whenever $n_t \geq n_r$.

For the case $n_t < n_r$ we present an alternative upper bound:

Theorem 4 *Let* $n_t < n_r$. *Then,*

$$\limsup_{\substack{\mathsf{SNR}\to\infty \\ \epsilon\to 0}} \left[C(\mathsf{SNR}, \epsilon) - \{n_{min} \log\left[\min(\mathsf{SNR}, 1/\epsilon)\right] + n_t \log(\log \mathsf{SNR}) + K_3(n_r, n_t)\}\right] \leq 0 \tag{11}$$

where $n_{min} = \min\{n_r, n_t\}$ *and* $K_3(n_r, n_t)$ *is some constant independent of* SNR *and* ϵ.

Proof: See Section 5.2. ■

This bound differs from (10) in that the coefficient of the $\log\log \mathsf{SNR}$ term is n_t instead of 1. Since the doubly logarithmic term is negligible compared to the other terms in regimes 1 and 2, this bound also matches the lower bound in the first two regimes. However in the third regime we conjecture that this bound becomes loose and fails to characterize the Lapidoth-Moser behavior.

In summary, regardless of the values n_r and n_t the lower and upper bounds match for regimes 1 and 2, which as we saw in the examples (cf. Table 1) are the regimes of most practical importance. In these two regimes the increase in the number of degrees of freedom (n_{min}) results in considerable capacity improvements.

4 Communication Scheme for Regimes 1 and 2

The derivation of the lower bound (7) suggests using i.i.d. circularly symmetric complex Gaussian inputs and maximum likelihood decoding. If the channel fading process is known at the receiver, the optimum decoding strategy reduces to minimum weighted Euclidean distance decoding (MWEDD). In practice the channel is unknown, and to use MWEDD we need to have an estimate of the channel. Here two problems arise:

- **Channel estimation**: if we use pilot sequences to estimate the channel, a fraction of the available degrees of freedom is wasted in training. We need to find a scheme that can estimate the channel without spending a significant amount of time in the training process.

- **Estimation error**: due to additive noise and channel variation, any estimation strategy will have a non-negligible channel estimation error. This error appears as an extra noise term in the channel equation, and this term is not independent of the input. Then, MWEDD is no longer the optimal decoding scheme and we have to prove that i.i.d. circularly symmetric complex Gaussian inputs and a mismatched decoder can still achieve the full number of degrees of freedom.

To tackle the channel estimation problem, we propose a decision-oriented training scheme. We use i.i.d. circularly symmetric complex Gaussian codewords of length B and we interleave M of them with a block interleaver. At the beginning of each interleaved block of length M we append the training sequence $\mathbf{I}_{n_t}$, i.e. we send at time $n = k(M+n_t)+m$, $k = 0, 1, \ldots, B-1$, a 1 in transmit antenna m leaving the remaining antennas silent. Figure 3 shows a schematic representation of the transmission scheme.

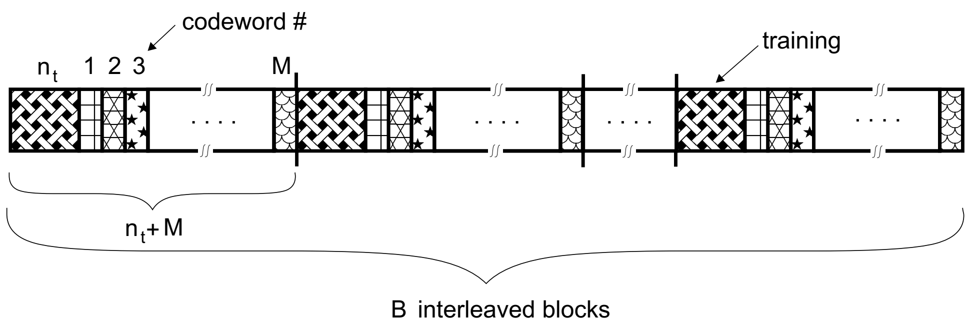

Figure 3: Schematic representation of the transmission scheme. Symbols of the same codeword have the same shading. B is the codeword length, M the interleaver block length, and n_t the number of training symbols used in each interleaved block.

Broadly speaking, the scheme consists of using the training sequences to obtain estimates of $\mathbf{H}[n]$ at times $n = k(n_t+M)+n_t+1$ for $k = 0, 1, \ldots, B-1$, and using these estimates to decode the first codeword of length B. Assuming that an appropriate coding rate is used, an arbitrarily small probability of decoding error can be achieved by choosing B large enough. Once the first block is successfully decoded, the symbols $\mathbf{x}[n]$, $n = k(n_t+M)+n_t+1$ for $k = 0, 1, \ldots, B-1$, are known to the receiver and can be used as if they were training symbols to estimate $\mathbf{H}[n]$, $n = k(n_t+M)+n_t+2$ for $k = 0, 1, \ldots, B-1$. These estimates are then used to decode the second codeword. Continuing in this way, at each step the scheme uses the previously decoded codeword to update the estimates of the fading matrix, which are then used in the decoding process of the next codeword. We can make the fraction of time spent in the transmission of the training sequences arbitrarily small by taking M large enough. We finish the description of the scheme by specifying the estimation algorithm, the decoding rule, and the transmission rate used.

4.1 Channel estimation

Consider the received vector when we send the first training symbol:

$$\mathbf{y}[1] = H[1]\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \sqrt{\frac{n_t}{\mathsf{SNR}}}\mathbf{z}[1] = \mathbf{h}_1[1] + \sqrt{\frac{n_t}{\mathsf{SNR}}}\mathbf{z}[1]$$

where $\mathbf{h}_1[1]$ is the first column of $\mathbf{H}[1]$. The optimal MMSE estimate of $\mathbf{h}_1[1]$ given $\mathbf{y}[1]$ is the conditional expectation $E\{\mathbf{h}_1[1]|\mathbf{y}[1]\}$. To simplify the analysis we use the suboptimal estimate $\hat{\mathbf{h}}_1[1] = \mathbf{y}[1]$, which is a good estimate for large SNR. In general we will estimate the channel fading matrix at time $n = k\cdot L + n_t + 1$, $(L = n_t + M)$, i.e. the time when we transmit the first information symbol of the interleaved block k, using:

$$\hat{\mathbf{H}}[k\cdot L + n_t + 1] = \begin{bmatrix} \mathbf{y}[k\cdot L+1] & \mathbf{y}[k\cdot L+2] & \cdots & \mathbf{y}[k\cdot L+n_t] \end{bmatrix}$$

with $\mathbf{\Delta}[k\cdot L + n_t + 1] = \mathbf{H}[k\cdot L + n_t + 1] - \hat{\mathbf{H}}[k\cdot L + n_t + 1]$ as the corresponding estimation error. In order to characterize $\mathbf{\Delta}$ we will focus on the estimation error in $\hat{\mathbf{h}}_1[1]$ given by $\sqrt{n_t/\mathsf{SNR}}\mathbf{z}[1]$ which has covariance $\frac{n_t}{\mathsf{SNR}}\mathbf{I}_{n_r}$. Our goal is to compute the covariance of the estimation error at time $n = n_t + 1$, i.e. the time when the first information symbol is transmitted. Due to channel variation, the estimation error at time $n_t + 1$ contains an additional independent $\mathcal{CN}(\mathbf{0}, n_t\epsilon\mathbf{I}_{n_r})$ term[3]. Therefore, the overall estimation error of the first column of $\mathbf{H}$ at time $n_t + 1$ is circularly symmetric complex Gaussian with covariance $n_t(\mathsf{SNR}^{-1} + \epsilon)\mathbf{I}_{n_r}$. Also, the estimates of the other columns of $\mathbf{H}$, i.e. $\hat{\mathbf{h}}_i[n_t+1] = \mathbf{y}[i]$, contain a smaller prediction error due to channel variation, so we can upper bound the variance of the estimation error of all the elements of $\mathbf{H}$ by $n_t(\mathsf{SNR}^{-1}+\epsilon)$. We also note that the estimation error $\mathbf{\Delta}$ is circularly symmetric complex Gaussian with independent components.

Consider the problem of estimating $\mathbf{H}[n+1]$ based on a previous estimate $\hat{\mathbf{H}}[n]$, the currently received vector $\mathbf{y}[n]$ and the successfully decoded symbol $\mathbf{x}[n]$. This is a Gaussian estimation problem, for which the optimal solution is given by the Kalman filter. We can simplify the derivation of the Kalman filter equations by noting that conditioned on $\mathbf{x}[n]$ the estimation of the different rows of $\mathbf{H}[n]$ decouples into n_r independent estimation problems. This result requires the independence of the rows of $\mathbf{\Delta}[n_t + 1]$ which was shown above. Therefore without loss of generality we will consider the estimation of the first row of $\mathbf{H}$, which we will denote by $\mathbf{h}^T$.

Assume that at time n we have an estimate $\hat{\mathbf{h}}[n]$ of $\mathbf{h}[n]$, and that the estimation error $\delta[n] = \mathbf{h}[n] - \hat{\mathbf{h}}[n]$ has zero mean and covariance $\mathbf{K}[n]$. We also assume that $\mathbf{x}[n]$ has been successfully decoded and is known to the receiver, together with the received signal in the first antenna, $y[n]$. We want to obtain $\hat{\mathbf{h}}[n+1] = E\left[\mathbf{h}[n+1]|\, y[n], \mathbf{x}[n], \hat{\mathbf{h}}[n]\right]$ and the covariance matrix of the corresponding estimation error, $\mathbf{K}[n+1]$. We use $\mathbf{w}[n]^T$ to denote the first row of $\mathbf{W}[n]$, and

[3] We have made the approximation $\sqrt{1-\epsilon} \approx 1$ when computing the variation of the channel matrix. This approximation is valid for $\epsilon \to 0$ for any finite n_t.

consider

$$\mathbf{h}[n+1] = \sqrt{1-\epsilon}\mathbf{h}[n] + \sqrt{\epsilon}\mathbf{w}[n] = \sqrt{1-\epsilon}\left(\hat{\mathbf{h}}[n] + \delta[n]\right) + \sqrt{\epsilon}\mathbf{w}[n]$$

to get the updated estimate

$$\hat{\mathbf{h}}[n+1] = E\left[\mathbf{h}[n+1]|y[n], \mathbf{x}[n], \hat{\mathbf{h}}[n]\right] = \sqrt{1-\epsilon}\hat{\mathbf{h}}[n] + \sqrt{1-\epsilon}E\left[\delta[n]|y[n], \mathbf{x}[n], \hat{\mathbf{h}}[n]\right] \tag{12}$$

where we used the fact that $\mathbf{w}[n]$ is zero mean and independent of $y[n]$, $\hat{\mathbf{h}}[n]$ and $\mathbf{x}[n]$. We also need the conditional covariance of the estimation error $\delta[n+1] = \mathbf{h}[n+1] - \hat{\mathbf{h}}[n+1]$:

$$\mathbf{K}[n+1] = \text{Cov}(\delta[n+1]|y[n], \mathbf{x}[n], \hat{\mathbf{h}}[n]) = (1-\epsilon)\text{Cov}(\delta[n]|y[n], \mathbf{x}[n], \hat{\mathbf{h}}[n]) + \epsilon\mathbf{I}_{n_t} \tag{13}$$

where we used the independence of $\mathbf{w}[n]$ with everything else.

To compute the expectation in (12) and covariance in (13) we need the joint distribution of $[y[n]\delta[n]^T]^T$ conditioned on $\hat{\mathbf{h}}[n]$ and $\mathbf{x}[n]$. This distribution is complex Gaussian, so we only need to compute the first and second moments[4]. Letting $z[n]$ denote the first element of $\mathbf{z}[n]$ we can write

$$y[n] = \hat{\mathbf{h}}[n]^T\mathbf{x}[n] + \delta[n]^T\mathbf{x}[n] + \sqrt{\frac{n_t}{\mathsf{SNR}}}z[n]$$

and noting that $z[n]$ is independent of everything else, we have that conditioned on $\hat{\mathbf{h}}[n]$ and $\mathbf{x}[n]$, $y[n]$ has mean $\hat{\mathbf{h}}[n]^T\mathbf{x}[n]$ and covariance $(\mathbf{x}[n]^T\mathbf{K}[n]\mathbf{x}[n]^* + n_t/\mathsf{SNR})$. By the orthogonality principle $\delta[n]$ is independent of $\hat{\mathbf{h}}[n]$, and is also independent of $\mathbf{x}[n]$. Therefore, conditioned on $\hat{\mathbf{h}}[n]$ and $\mathbf{x}[n]$, $\delta[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}[n])$. Finally we have,

$$\begin{aligned} E\left[y[n]\delta[n]^\dagger|\hat{\mathbf{h}}[n], \mathbf{x}[n]\right] &= E\left[\hat{\mathbf{h}}[n]^T\mathbf{x}[n]\delta[n]^\dagger + \delta[n]^T\mathbf{x}[n]\delta[n]^\dagger + \sqrt{\frac{n_t}{\mathsf{SNR}}}z[n]\delta[n]^\dagger \middle| \hat{\mathbf{h}}[n], \mathbf{x}[n]\right] \\ &= \mathbf{x}[n]^T\mathbf{K}[n] \end{aligned}$$

It follows that

$$E\left\{\left[\begin{array}{c} y[n] \\ \delta[n] \end{array}\right] \middle| \hat{\mathbf{h}}[n], \mathbf{x}[n]\right\} = \left[\begin{array}{c} \hat{\mathbf{h}}[n]^T\mathbf{x}[n] \\ \mathbf{0} \end{array}\right]$$

and

$$E\left\{\left[\begin{array}{c} y[n] \\ \delta[n] \end{array}\right]\left[\begin{array}{cc} y[n]^* & \delta[n]^\dagger \end{array}\right] \middle| \hat{\mathbf{h}}[n], \mathbf{x}[n]\right\} = \left[\begin{array}{cc} \mathbf{x}[n]^T\mathbf{K}[n]\mathbf{x}[n]^* + \frac{n_t}{\mathsf{SNR}} & \mathbf{x}[n]^T\mathbf{K}[n] \\ \mathbf{K}[n]\mathbf{x}[n]^* & \mathbf{K}[n] \end{array}\right]$$

We can now get the distribution of $\delta[n]$ conditioned on $(y[n], \hat{\mathbf{h}}[n], \mathbf{x}[n])$ which is complex Gaussian, by obtaining the corresponding moments:

$$\begin{aligned} E\left[\delta[n]|\, y[n], \hat{\mathbf{h}}[n], \mathbf{x}[n]\right] &= \mathbf{K}[n]\mathbf{x}[n]^*\left(\mathbf{x}[n]^T\mathbf{K}[n]\mathbf{x}[n]^* + \frac{n_t}{\mathsf{SNR}}\right)^{-1}\left(y[n] - \hat{\mathbf{h}}[n]^T\mathbf{x}[n]\right) \\ \text{Cov}\left[\delta[n]|\, y[n], \hat{\mathbf{h}}[n], \mathbf{x}[n]\right] &= \mathbf{K}[n] - \left(\mathbf{x}[n]^T\mathbf{K}[n]\mathbf{x}[n]^* + \frac{n_t}{\mathsf{SNR}}\right)^{-1}\mathbf{K}[n]\mathbf{x}[n]^*\mathbf{x}[n]^T\mathbf{K}[n] \end{aligned}$$

[4]The centered version of this random vector is circularly symmetric, so we only need to compute the mean vector and covariance matrix, knowing that the pseudocovariance matrix is zero.

We finally replace in (12) and (13) obtaining:

$$\hat{\mathbf{h}}[n+1] = \sqrt{1-\epsilon}\hat{\mathbf{h}}[n] + \sqrt{1-\epsilon}\left[\mathbf{K}[n]\mathbf{x}[n]^*\left(\mathbf{x}[n]^T\mathbf{K}[n]\mathbf{x}[n]^* + \frac{n_t}{\mathsf{SNR}}\right)^{-1}\left(y[n] - \hat{\mathbf{h}}[n]^T\mathbf{x}[n]\right)\right]$$

$$\mathbf{K}[n+1] = (1-\epsilon)\left[\mathbf{K}[n] - \left(\mathbf{x}[n]^T\mathbf{K}[n]\mathbf{x}[n]^* + \frac{n_t}{\mathsf{SNR}}\right)^{-1}\mathbf{K}[n]\mathbf{x}[n]^*\mathbf{x}[n]^T\mathbf{K}[n]\right] + \epsilon\mathbf{I}_{n_t} \tag{14}$$

It is interesting to note that $\mathbf{K}[n+1]$ does not depend on $y[n]$ and $\hat{\mathbf{h}}[n]$ so the estimation errors corresponding to all the rows of $\mathbf{H}[n]$ have the same covariance. However, the estimates for the different rows of $\mathbf{H}$ will be different in general.

4.2 Decoding

Define $\mathcal{N}_m = \{n : n = (n_t + M)k + n_t + m, k = 0, \ldots, B-1\}$, for $m = 1, 2, \ldots, M$, i.e. the set of times when the symbols corresponding to the mth codeword are transmitted. The decoding of the mth codeword is based on the received vectors $\{\mathbf{y}[n] : n \in \mathcal{N}_m\}$ and the estimates of the channel fading matrix $\{\hat{\mathbf{H}}[n] : n \in \mathcal{N}_m\}$. Define $\mathbf{Y}_m$ to be the matrix whose columns are the received vectors corresponding to the mth codeword, and $\mathbf{V}_k$ to be the matrix whose columns are the vectors of the form $\hat{\mathbf{H}}[n]\mathbf{x}_k[r]$ for $n \in \mathcal{N}_m$ and $r \in \{1, 2, \ldots, B\}$, where k denotes the codeword index in the codebook. Then the decoding rule for the mth codeword is given by:

$$\hat{k}_m = \arg\min_k \|\mathbf{Y}_m - \mathbf{V}_k\|_F$$

This is a weighted minimum Euclidean distance decoding rule. The vectors of the candidate codeword k are weighted by the estimates of the channel fading matrices, and then compared to the received vectors to find the one which is closest in Euclidean distance.

If the fading matrices where perfectly known at the receiver, then this decoding rule would correspond to maximum likelihood decoding, which minimizes the probability of decoding error. However, in practice there will be some estimation error in the fading matrices, and there is an additional noise term that is not Gaussian and is not independent of the channel input. In this case WMEDD is no longer optimal, but as we show next, it lets us achieve the full number of degrees of freedom of the channel in the first two regimes.

4.3 Coding rate

In this subsection we obtain a lower bound on the maximum rate achievable with our communication scheme. Once this lower bound is found, we can build a family of codebooks, each with an appropriate number of codewords so that the error probability goes to zero as the block length goes to infinity.

We start by characterizing the channel that results from our communication scheme. As $M \to \infty$, the estimation errors $\{\mathbf{\Delta}[k(n_t + M) + n_t + m]\}_{k=0}^{B-1}$ and the channel estimates $\{\hat{\mathbf{H}}[k(n_t + M) + n_t + m]\}_{k=0}^{B-1}$ become independent over k, so the mth transmitted codeword sees an equivalent channel given by:

$$\mathbf{y}[n] = \hat{\mathbf{H}}[n]\mathbf{x}[n] + \boldsymbol{\Delta}[n]\mathbf{x}[n] + \sqrt{\frac{n_t}{\mathsf{SNR}}}\mathbf{z}[n] = \hat{\mathbf{H}}[n]\mathbf{x}[n] + \tilde{\mathbf{z}}[n] \tag{15}$$

where $\tilde{\mathbf{z}}[n] = \boldsymbol{\Delta}[n]\mathbf{x}[n] + \sqrt{\frac{n_t}{\mathsf{SNR}}}\mathbf{z}[n]$ and $n \in \mathcal{N}_m$, and $\{\tilde{\mathbf{z}}[n]\}_{n\in\mathcal{N}_m}$ and $\{\hat{\mathbf{H}}[n]\}_{n\in\mathcal{N}_m}$ are i.i.d. over n. We note that the noise $\tilde{\mathbf{z}}[n]$ depends on the input $\mathbf{x}[n]$, but it is uncorrelated with it.

$$\begin{aligned} E\left[\tilde{\mathbf{z}}[n]\mathbf{x}[n]^\dagger\right] &= E\left[\boldsymbol{\Delta}[n]\mathbf{x}[n]\mathbf{x}[n]^\dagger\right] + E\left[\sqrt{\frac{n_t}{\mathsf{SNR}}}\mathbf{z}[n]\mathbf{x}[n]^\dagger\right] \\ &= E\left[\boldsymbol{\Delta}[n]\right]E\left[\mathbf{x}[n]\mathbf{x}[n]^\dagger\right] + \sqrt{\frac{n_t}{\mathsf{SNR}}}E\left[\mathbf{z}[n]\right]E\left[\mathbf{x}[n]^\dagger\right] = \mathbf{0} \end{aligned}$$

Using the independence of $\boldsymbol{\Delta}[n]$, $\mathbf{x}[n]$ and $\mathbf{z}[n]$ we can express the covariance matrix of the noise as:

$$E\left[\tilde{\mathbf{z}}[n]\tilde{\mathbf{z}}[n]^\dagger\right] = E\left[\boldsymbol{\Delta}[n]\boldsymbol{\Delta}[n]^\dagger\right] + \frac{n_t}{\mathsf{SNR}}\mathbf{I}_{n_r}$$

By computing an upper bound on $E\left[\boldsymbol{\Delta}[n]\boldsymbol{\Delta}[n]^\dagger\right]$ the following lemma gives an upper bound on $E\left[\tilde{\mathbf{z}}[n]\tilde{\mathbf{z}}[n]^\dagger\right]$.

Lemma 1 $E\left[\tilde{\mathbf{z}}[n]\tilde{\mathbf{z}}[n]^\dagger\right] = \sigma^2[n]\mathbf{I}_{n_r}$ *where, for n large enough so that $E[tr(\mathbf{K}[n])]$ reaches a steady state, $\sigma^2[n]$ is upper bounded by:*

$$\sigma^2[n] \leq \frac{n_t^2 + n_t}{\mathsf{SNR}} + n_t^2 2^{n_t}\epsilon$$

Proof: See Appendix G. ■

Having characterized the noise term $\tilde{\mathbf{z}}[n]$ we now turn our attention to the estimate of the fading matrix $\hat{\mathbf{H}}[n]$ which is known at the receiver. As we argued before, conditioned on $\{\mathbf{x}[n]\}$ the rows of $\hat{\mathbf{H}}[n]$ are independent, and by symmetry, identically distributed. They are also zero mean, so it follows that they are uncorrelated. The same holds true after removing the conditioning on $\{\mathbf{x}[n]\}$. It follows that in steady state we can write $E\left(\hat{\mathbf{h}}_i^\dagger[n]\hat{\mathbf{h}}_j[n]\right) = \alpha^2\delta_{i,j}$ for some constant α^2, where $\hat{\mathbf{h}}_i^T$ is the ith row of $\hat{\mathbf{H}}[n]$.

By the orthogonality principle, the estimation error $\boldsymbol{\Delta}[n]$ is independent of the estimate $\hat{\mathbf{H}}[n]$, and so is the Gaussian noise $\mathbf{z}[n]$. Therefore, $\hat{\mathbf{H}}[n]$ is independent of $\tilde{\mathbf{z}}[n]$.

As $\mathsf{SNR} \to \infty$ and $\epsilon \to 0$ the estimate $\hat{\mathbf{H}}[n]$ converges to $\mathbf{H}[n]$ with probability 1, so it follows that $\hat{\mathbf{H}}[n]$ also converges in distribution to $\mathbf{H}[n]$. Therefore, in the limit as $\mathsf{SNR} \to \infty$ and $\epsilon \to 0$, $\hat{\mathbf{H}}[n]$ converges in distribution to a circularly symmetric complex Gaussian matrix, with independent components of variance 1, in which case α^2 converges to n_t.

We now show that for a channel like (15) with MWEDD the worst case noise is circularly symmetric complex Gaussian, independent of the input and the channel fading matrix. Then, we can lower bound the maximum achievable rate by that of the additive, independent Gaussian noise channel, with Rayleigh fading known at the receiver. The following theorem extends a result of [10] for noise not independent with the channel input.

Theorem 5 *Let C denote the capacity of the channel*

$$\mathbf{y}[n] = \hat{\mathbf{H}}[n]\mathbf{x}[n] + \mathbf{z}[n]$$

where $\hat{\mathbf{H}}[n] \in \mathcal{C}^{n_r \times n_t}$ is known at the receiver, is i.i.d. over time, and $E\left(\hat{\mathbf{h}}_i^\dagger[n]\hat{\mathbf{h}}_j[n]\right) = \alpha^2\delta_{i,j}$, $i,j \in \{1,\ldots,n_r\}$, for some constant α^2, where $\hat{\mathbf{h}}_i^T$ is the ith row of $\hat{\mathbf{H}}[n]$. Also $\mathbf{z}[n] \in \mathcal{C}^{n_r}$ is $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{n_r})$ is independent over time and independent of $\hat{\mathbf{H}}[n]$; and the input $\mathbf{x}[n] \in \mathcal{C}^{n_t}$ is independent of everything else and is subject to the power constraint $E\left[\|\mathbf{x}[n]\|^2\right] \leq n_t$.

Consider the channel

$$\tilde{\mathbf{y}}[n] = \hat{\mathbf{H}}[n]\mathbf{x}[n] + \tilde{\mathbf{z}}[n]$$

where $\hat{\mathbf{H}}[n] \in \mathcal{C}^{n_r \times n_t}$ is the same as before, and is also known at the receiver; $\tilde{\mathbf{z}}[n]$ is independent over time, is uncorrelated with the input $\mathbf{x}[n]$, is independent of $\hat{\mathbf{H}}[n]$ and has the same covariance as $\mathbf{z}[n]$; and $\mathbf{x}[n] \in \mathcal{C}^{n_t}$ is independent of $\hat{\mathbf{H}}[n]$ and is subject to the power constraint $E\left[\|\mathbf{x}[n]\|^2\right] \leq n_t$. Let R be the supremum of the achievable rates in this channel when the input $\mathbf{x}[n]$ is constrained to be formed by i.i.d. $\mathcal{CN}(0,1)$ components, independent over time, and weighted minimum Euclidean distance decoding is used.

Assuming that all random vectors are jointly defined we have,

$$C \leq R$$

Proof: See Appendix H. ■

Then we can choose any rate $R < E\left[\log\det\left(\mathbf{I}_{n_r} + \frac{\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger}{\sigma^2}\right)\right]$ where the expectation is taken over the random matrix $\hat{\mathbf{H}}$. For large SNR and small ϵ we can do the following approximations:

$$\begin{aligned} E\left[\log\det\left(\mathbf{I}_{n_r} + \frac{\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger}{\sigma^2}\right)\right] &\geq E\left[\log\det\left(\frac{\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger}{\sigma^2}\right)\right] \\ &= \min\{n_t, n_r\}\log\left(\frac{1}{\sigma^2}\right) + E\left[\log\det\left(\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger\right)\right] \\ &\approx \min\{n_t, n_r\}\log\left(\frac{1}{\sigma^2}\right) + E\left[\log\det\left(\mathbf{H}\mathbf{H}^\dagger\right)\right] \\ &\sim \min\{n_t, n_r\}\log\left[\left(\mathsf{SNR}^{-1} + \epsilon\right)^{-1}\right] \end{aligned} \qquad (16)$$

where $n_r \leq n_t$ was assumed; the other case is similar. Therefore we have that the maximum achievable rate R of our scheme grows at least as fast as $\min\{n_r, n_t\}\log\left[\left(\mathsf{SNR}^{-1} + \epsilon\right)^{-1}\right]$ for $\mathsf{SNR} \to \infty$ and $\epsilon \to 0$. Thus our scheme achieves the full degrees of freedom of the channel.

5 Proofs of the Upper Bounds

In this section we will present the proofs of Theorems 3 and 4. We first prove Theorem 3 in Subsection 5.1 introducing many results that are also useful in the proof of Theorem 4. Later in Subsection 5.2 we build on the previous Subsection to obtain an upper bound for the $n_t = 1$ case, and extend the result for the $n_t < n_r$ case, proving Theorem 4.

We will introduce many changes of variables. The following table summarizes the definitions.

Variable	Definition
$\tilde{\mathbf{y}}[n]$	$\mathbf{H}[n]\mathbf{x}[n]$
$m[n]$	$\|\mathbf{x}[n]\|$
$\mathbf{d}[n]$	$\mathbf{x}[n]/\|\mathbf{x}[n]\|$
$\mathbf{c}[n]$	$\mathbf{H}[n]\mathbf{d}[n]$
$\rho_i[n]$	$\log(\|\tilde{y}_i[n]\|)$
$\phi_i[n]$	$\angle\tilde{y}_i[n]$
$r_i[n]$	$\tilde{y}_i[n]/\tilde{y}_1[n]$
$\tilde{w}_i[1]$	$(\|h_i[1]\|/h_i[1])w_i[1]$

5.1 Proof of Theorem 3

Consider first the case when $\mathsf{SNR} < 1/\epsilon$, or equivalently $1/\mathsf{SNR} > \epsilon$. This corresponds to a noise variance larger than the one-step MMSE prediction error of the channel. In this regime the capacity is mainly limited by the noise, so we can obtain a tight upper bound for the capacity by simply assuming that the channel realization $\mathbf{H}^{\mathcal{N}}$ is perfectly known at the receiver. The capacity under this assumption was explicitly computed in [2], from which it follows that $C_{\text{Known }\mathbf{H}} \approx n_{min} \log \mathsf{SNR} + K$ for some constant K independent of SNR, for large SNR. This corresponds to the bounds in Theorems 3 and 4 in regime 1 where the $\log\log \mathsf{SNR}$ term is negligible.

Next, consider the case when $\mathsf{SNR} \geq 1/\epsilon$. Channel uncertainty becomes the main capacity limiting factor when the additive noise becomes negligible as compared to the one-step MMSE prediction error of the channel. However, this does not mean that we can remove the noise from the channel equation and expect to have a tight upper bound for capacity. In fact, removing the noise corresponds to letting $\mathsf{SNR} \to \infty$, in which case $C(\mathsf{SNR}, \epsilon) \to \infty$. In [3] it is shown that for a memoryless channel it is possible upper bound the capacity of a noisy channel by the capacity of a noiseless channel with the addition of a lower bound constraint on the input. Intuitively, little information can be transmitted when the input is much smaller than the noise, so adding an appropriate lower bound on the input does not reduce capacity significantly. On the other hand, once the lower bound constraint on the input is added, the noise can be removed while obtaining a finite capacity. The following theorem generalizes the same idea for channels with memory.

Theorem 6 *Let $\mathcal{D}$ be the family of distributions:*

$$\mathcal{D} = \left\{ p(\mathbf{x}^{\mathcal{N}}) : \frac{1}{B} \sum_{n \in \mathcal{N}} E[\|\mathbf{x}[n]\|^2] \leq n_t + 1/\mathsf{SNR} \textit{ and } \|\mathbf{x}[n]\| \geq \mathsf{SNR}^{-1/2}, n \in \mathcal{N}, \textit{ with probability 1} \right\}$$

Then,

$$C(\mathsf{SNR},\epsilon) \leq (2n_r+1)\log(2) + \lim_{B\to\infty} \sup_{p(\mathbf{x}^{\mathcal{N}})\in\mathcal{D}} \frac{1}{B} I(\mathbf{x}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}})$$

Proof: See Appendix C. ■

Having a lower bound on the input norm allows us to remove the additive noise from the channel equation and still get a tight upper bound for the capacity. This is done at the expense of an additive constant, $(2n_r+1)\log(2)$, that does not modify the asymptotic behavior of the bound. From this point on, we will assume that the input satisfies

$$\begin{aligned} \|\mathbf{x}[n]\| &\geq \mathsf{SNR}^{-1/2}, n \in \mathcal{N} \\ \frac{1}{B}\sum_{n\in\mathcal{N}} E[\|\mathbf{x}[n]\|^2] &\leq n_t + \frac{1}{\mathsf{SNR}} \end{aligned} \tag{17}$$

Removing the noise does not reduce mutual information. By using the chain rule we have that $I(\mathbf{x}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}}) \leq I(\mathbf{x}^{\mathcal{N}};\tilde{\mathbf{y}}^{\mathcal{N}})$ where

$$\tilde{\mathbf{y}}[n] = \mathbf{H}[n]\mathbf{x}[n] = \mathbf{y}[n] - \sqrt{\frac{n_t}{\mathsf{SNR}}}\mathbf{z}[n],\ \forall n \in \mathcal{N} \tag{18}$$

After removing the noise we get the multiplicative channel (18).

To get some intuition about the techniques used later on we will consider the scalar $n_t = n_r = 1$ SISO case first. In the scalar case, the resulting multiplicative channel is given by:

$$\tilde{y}[n] = h[n]x[n]$$

where the input distribution is chosen from the set $\mathcal{D}$. We observe that if we directly write the mutual information $I(\mathbf{x}^{\mathcal{N}};\tilde{\mathbf{y}}^{\mathcal{N}})$ in terms of differential entropies,

$$I(\mathbf{x}^{\mathcal{N}};\tilde{\mathbf{y}}^{\mathcal{N}}) = h(\tilde{\mathbf{y}}^{\mathcal{N}}) - h(\tilde{\mathbf{y}}^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}})$$

both terms depend on the input distribution and cannot be maximized independently without obtaining an extremely loose upper bound. However we can introduce a change of variables that converts the multiplicative channel into an additive channel. For a complex number x, let $\log(x) = \log|x| + j\angle x$, where $\angle x \in (-\pi,\pi]$, and for a complex vector $\mathbf{x} = (x_1,\ldots,x_n)^T$ let $\log(\mathbf{x})$ be the vector whose components are $\log(x_i)$. After taking $\log(\cdot)$ the channel equation becomes:

$$\log(\tilde{y}[n]) = \log(x[n]) + \log(h[n])$$

and the $\log(h[n])$ term plays the role of an additive noise term, which is unknown to the receiver and is correlated over time.

With the above definition $\log(\cdot)$ is a one to one transformation that applied to the channel equation leaves the mutual information unaltered:

$$I(\mathbf{x}^{\mathcal{N}};\tilde{\mathbf{y}}^{\mathcal{N}}) = I(\log(\mathbf{x}^{\mathcal{N}});\log(\tilde{\mathbf{y}}^{\mathcal{N}})) = h(\log(\tilde{\mathbf{y}}^{\mathcal{N}})) - h(\log(\tilde{\mathbf{y}}^{\mathcal{N}})|\log(\mathbf{x}^{\mathcal{N}})) = h(\log(\tilde{\mathbf{y}}^{\mathcal{N}})) - h(\log(\mathbf{h}^{\mathcal{N}}))$$

Note that the second term of the RHS is independent of the input distribution, and in principle can be computed. To bound the term $h(\log(\tilde{\mathbf{y}}^{\mathcal{N}}))$ we can find an upper bound on the variance of the components of $\log(\tilde{\mathbf{y}}^{\mathcal{N}})$ and use the Gaussian bound assuming that all the components are independent. This is achieved in the actual proof by using triangle inequality and Jensen's inequality. The $\log\log \mathsf{SNR}$ term of the upper bound comes from this Gaussian bound.

The computation of $h(\log(\mathbf{h}^{\mathcal{N}}))$ is simplified by expressing it in terms of $h(\mathbf{h}^{\mathcal{N}})$, finding the Jacobian of the $\log(\cdot)$ transformation and using a property that relates the differential entropies of vectors related by a one to one transformation. The computation of $h(\mathbf{h}^{\mathcal{N}})$ is complicated by the correlation of the fading channel. Since we are interested in obtaining an upper bound for the capacity of the channel, we just need to obtain a lower bound on $h(\mathbf{h}^{\mathcal{N}})$. This is done by using the chain rule for differential entropies, and for each time n conditioning on $h[n-1]$. This extra conditioning does not increase the differential entropy, and each of the corresponding terms can be computed explicitly as a function of ϵ. The $\log\epsilon$ term in the upper bound comes from this lower bound on $h(\log(\mathbf{h}^{\mathcal{N}}))$.

We now consider the general vector case. In the scalar case we applied $\log(\cdot)$ to convert the multiplicative channel into an additive channel. However, in this case the channel equation is in vector form so we need to introduce a change of variables before we can take logarithms. Information gets transmitted through this channel by means of $m[n] = \|\mathbf{x}[n]\|$, the norm of the input, and $\mathbf{d}[n] = \mathbf{x}[n]/\|\mathbf{x}[n]\|$, the direction of the input. The SNR has influence only on the communication over the norm component of the input, so this transformation allows us to separate the component that depends on the SNR.

We also define $\mathbf{c}[n] = \mathbf{H}[n]\mathbf{d}[n]$ for $n \in \mathcal{N}$. Finally, using i to index the receive antenna number we define $\rho_i[n] = \log(|\tilde{y}_i[n]|)$ and $\phi_i[n] = \angle\tilde{y}_i[n]$, where $\tilde{y}_i[n]$ is the ith component of $\tilde{\mathbf{y}}[n]$, and $\pi < \phi_i[n] \leq \pi$ for $n \in \mathcal{N}$, and group all this variables into the Bn_r dimensional vectors $\rho^{\mathcal{N}}$ and $\phi^{\mathcal{N}}$. Note that this transformation includes the logarithm used in the scalar case. The following fact relates the differential entropies of random vectors defined in terms of a one to one mapping:

Fact 1 *Let the vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *be related by a one to one transformation with Jacobian* $\mathbf{J}$ *where* $J(i,j) = \partial w_i/\partial v_j$. *Then*

$$h(\mathbf{w}) = h(\mathbf{v}) + E[\log\det J]$$

We show in a lemma how to compute the mutual information in terms of these new variables:

Lemma 2

$$I(\mathbf{x}^{\mathcal{N}};\tilde{\mathbf{y}}^{\mathcal{N}}) = h(\rho^{\mathcal{N}},\phi^{\mathcal{N}}) - h(\mathbf{c}^{\mathcal{N}}|\mathbf{d}^{\mathcal{N}}) + \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} E\left[\log(|c_i[n]|^2)\right] \tag{19}$$

where $c_i[n]$ *is the* i*th component of* $\mathbf{c}[n]$.

Proof: See Appendix D. ■

Lemma 2 summarizes in one expression the result of taking $\log(\cdot)$ in the channel equation, computing the Jacobian of the one to one transformation, and simplifying the computation of the second differential entropy in the mutual information expression by inverting the $\log(\cdot)$ transformation.

We conclude the computation of the upper bound by computing or bounding each of the terms of (19). This is done in the following three lemmas:

Lemma 3

$$\sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} E\left[\log(|c_i[n]|^2)\right] = -n_r B\gamma$$

where $\gamma = 0.5772\ldots$ is Euler's constant.

Proof: To compute the double sum we need to determine the distribution of $|c_i[n]|^2$, $i = 1, \ldots, n_r$, $n \in \mathcal{N}$. Recall that $c_i[n] = \mathbf{h}_i[n]^T\mathbf{d}[n]$, where $\mathbf{h}_i[n]^T$ is the ith row of $\mathbf{H}[n]$. Since $\mathbf{d}[n]$ is a norm 1 vector and $\mathbf{h}_i[n]^T$ has i.i.d. $\mathcal{CN}(0,1)$ components, conditioned on $\mathbf{d}[n]$, $c_i[n]$ has $\mathcal{CN}(0,1)$ distribution independent of $\mathbf{d}[n]$. It follows that without conditioning $c_i[n]$ has $\mathcal{CN}(0,1)$ distribution, and therefore $|c_i[n]|^2 \sim Exp(1)$. Then we can compute $E\left[\log(|c_i[n]|^2)\right]$ explicitly:

$$E\left[\log(|c_i[n]|^2)\right] = \int_0^\infty \log(x)e^{-x}dx = -\gamma$$

independent of n and i. ■

Lemma 4

$$h(\mathbf{c}^{\mathcal{N}}|\mathbf{d}^{\mathcal{N}}) \geq n_r B\log(\pi e\epsilon)$$

Proof: We start by using the chain rule for differential entropies and lower bounding by adding extra conditioning:

$$\begin{aligned} h(\mathbf{c}^{\mathcal{N}}|\mathbf{d}^{\mathcal{N}}) &= h(\mathbf{c}[1]|\mathbf{d}^{\mathcal{N}}) + \sum_{n=2}^{B} h(\mathbf{c}[n]|\mathbf{c}[n-1],\cdots,\mathbf{c}[1],\mathbf{d}^{\mathcal{N}}) \\ &\geq h(\mathbf{c}[1]|\mathbf{H}[0],\mathbf{d}^{\mathcal{N}}) + \sum_{n=2}^{B} h(\mathbf{c}[n]|\mathbf{H}[n-1],\mathbf{c}[n-1],\cdots,\mathbf{c}[1],\mathbf{d}^{\mathcal{N}}) \end{aligned}$$

Conditioned on $(\mathbf{H}[n-1],\mathbf{d}[n])$, $\mathbf{c}[n] \sim \mathcal{CN}(\sqrt{1-\epsilon}\mathbf{H}[n-1]\mathbf{d}[n], \epsilon\mathbf{I}_{n_r})$, and it follows that $h(\mathbf{c}[1]|\mathbf{H}[0],\mathbf{d}^{\mathcal{N}}) = h(\mathbf{c}[n]|\mathbf{H}[n-1],\mathbf{c}[n-1],\cdots,\mathbf{c}[1],\mathbf{d}^{\mathcal{N}}) = n_r\log(\pi e\epsilon)$[5]. To see this we write $\mathbf{c}[n]$ as a function of $(\mathbf{H}[n-1],\mathbf{d}[n])$ for $n \in \mathcal{N}$:

$$\mathbf{c}[n] = \mathbf{H}[n]\mathbf{d}[n] = \left(\sqrt{1-\epsilon}\mathbf{H}[n-1] + \sqrt{\epsilon}\mathbf{W}[n]\right)\mathbf{d}[n] = \sqrt{1-\epsilon}\mathbf{H}[n-1]\mathbf{d}[n] + \sqrt{\epsilon}\tilde{\mathbf{w}}[n]$$

where $\tilde{\mathbf{w}}[n] = \mathbf{W}[n]\mathbf{d}[n]$ has $\mathcal{CN}(\mathbf{0},\mathbf{I}_{n_r})$ distribution because $\mathbf{W}[n]$ has i.i.d. $\mathcal{CN}(0,1)$ components and $\mathbf{d}[n]$ has norm 1. ■

[5]Note that conditioned on $(\mathbf{H}[n-1],\mathbf{d}[n])$, $\mathbf{c}[n]$ is independent of $\mathbf{c}[n-1],\cdots,\mathbf{c}[1]$ and $\mathbf{d}^{\mathcal{N}}$.

Lemma 5

$$\frac{1}{B}h(\rho^{\mathcal{N}},\phi^{\mathcal{N}}) \le \frac{1}{2}\log(2\pi e\sigma^2) + \frac{1}{2}\log(2\pi e\pi^2) + (n_r-1)[2-\log(2)] + (n_r-1)\frac{1}{2}\log\left[2\pi e(2\pi)^2\right]$$

where $\sigma^2 = \left\{\sqrt{n_t + 1/\mathsf{SNR} + 1 + [\log(\sqrt{\mathsf{SNR}})]^2} + \sqrt{1/2}\right\}^2$

Proof: See Appendix E. ∎

We replace in (19) the results of the last three lemmas to obtain the upper bound:

$$\lim_{B\to\infty}\sup_{p(\mathbf{x}^{\mathcal{N}})\in\mathcal{D}}\frac{1}{B}I(\mathbf{x}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}}) \le \frac{1}{2}\log(\sigma^2) - n_r\log(\epsilon) + K(n_r)$$

where $\mathcal{D}$ is as in Theorem 6, $\sigma^2 = \left\{\sqrt{n_t + 1/\mathsf{SNR} + 1 + [\log(\sqrt{\mathsf{SNR}})]^2} + \sqrt{1/2}\right\}^2$ and $K(n_r) = \log(2\pi e\pi) + (n_r-1)[\log(1/2)+2] + (n_r-1)\frac{1}{2}\log(2\pi e(2\pi)^2) - n_r\gamma - n_r\log(\pi e)$.

Finally we use Theorem 6 to obtain an upper bound for $C(\mathsf{SNR},\epsilon)$:

$$C(\mathsf{SNR},\epsilon) \le \frac{1}{2}\log(\sigma^2) - n_r\log(\epsilon) + K(n_r) + (2n_r+1)\log(2)$$

As $\mathsf{SNR}\to\infty$, $(1/2)\log(\sigma^2) - (\log\log\mathsf{SNR} - \log 2) \to 0$, so for $n_r \le n_t$ theorem 3 follows by taking $K_2(n_r,n_t) = 2n_r\log(2) + K(n_r)$. However, for $n_r > n_t$ this bound becomes loose in regime 2 because the coefficient of the $\log(\epsilon)$ term is n_r instead of $\min\{n_r,n_t\} = n_t$.

5.2 Proof of Theorem 4

As mentioned in the introduction of this section, many of the results of Subsection 5.1 apply to this proof as well. The main problem of the previous proof resides in Lemma 5 which is not tight for $n_r > n_t$. In fact we conjecture that the right upper bound for $\frac{1}{B}h(\rho^{\mathcal{N}},\phi^{\mathcal{N}})$ should contain a term of the form $(n_r - \min\{n_t,n_r\})\log\epsilon$ which is non zero for $n_r > n_t$. We could not obtain the right form for the general case, but we obtained a tight bound for the special case of $n_t = 1$. This is given in the following lemma.

Lemma 6 *Let* $n_t = 1$*. Then,*

$$\lim_{\substack{B\to\infty\\ \epsilon\to 0\\ \mathsf{SNR}\to\infty}}\left\{\frac{1}{B}h(\rho^{\mathcal{N}},\phi^{\mathcal{N}}) - \left[\log(2\pi e\pi) + \log(\sigma^2) + (n_r-1)\log(\epsilon) + 3.97722\right]\right\} \le 0$$

where $\sigma^2 = \left\{\sqrt{1/\mathsf{SNR} + 2 + [\log(\sqrt{\mathsf{SNR}})]^2} + \sqrt{1/2}\right\}^2$

Proof: See Appendix F. ∎

Lemma 6 together with Theorem 6, and Lemmas 2, 3 and 4 allow us to obtain a tight upper bound for the capacity of the single transmit antenna channel:

$$\limsup_{\substack{\mathsf{SNR}\to\infty\\ \epsilon\to 0}}\left[C(\mathsf{SNR},\epsilon) - \{\log\log\mathsf{SNR} + \log(1/\epsilon) + K(n_r)\}\right] \le 0 \tag{20}$$

where $K(n_r) = 2n_r \log(2) - n_r\gamma + \log(2\pi e\pi) + 3.97722$.

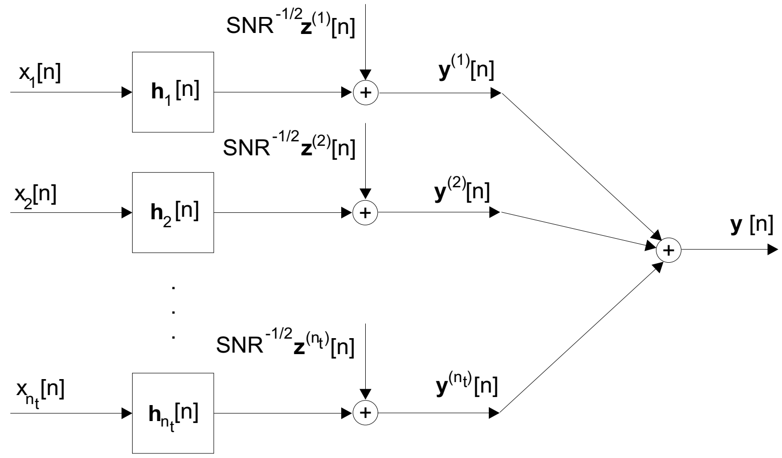

Figure 4: Representation of the MIMO channel as the sum of the outputs of n_t independent SIMO channels.

An upper bound for the capacity of the MIMO channel can be obtained by rewriting the channel equation (1) as the sum of the outputs of n_t independent SIMO ($n_t = 1$) channels as represented in Figure 4. Let $\mathbf{h}_i[n] \in C^{n_r}$ be the i^{th} column of $\mathbf{H}[n]$, $x_i[n]$ be the i^{th} component of $\mathbf{x}[n]$ and $\mathbf{z}^{(i)}[n] \sim \mathcal{CN}(0, I_{n_r}), i = 1, 2, \ldots, n_t, n \in \mathcal{N}$, be independent across i and n, and independent of the channel input and fading gains. Define n_t independent SIMO channels by:

$$\mathbf{y}^{(i)}[n] = \mathbf{h}_i[n]x_i[n] + \frac{1}{\sqrt{\mathsf{SNR}}}\mathbf{z}^{(i)}[n] \quad i = 1, 2, \ldots, n_t \tag{21}$$

Letting $\mathbf{y}[n] = \sum_{i=1}^{n_t} \mathbf{y}^{(i)}[n]$ and $\mathbf{z}[n] = \frac{1}{\sqrt{n_t}} \sum_{i=1}^{n_t} \mathbf{z}^{(i)}[n]$ we have the original channel equation expressed as a function of the individual SIMO channels

$$\begin{aligned} \mathbf{y}[n] = \sum_{i=1}^{n_t} \mathbf{y}^{(i)}[n] &= \sum_{i=1}^{n_t} \mathbf{h}_i[n]x_i[n] + \sqrt{\frac{n_t}{\mathsf{SNR}}}\frac{1}{\sqrt{n_t}}\sum_{i=1}^{n_t} \mathbf{z}^{(i)}[n] \\ &= H[n]\mathbf{x}[n] + \sqrt{\tfrac{n_t}{\mathsf{SNR}}}\mathbf{z}[n], \quad n \in \mathcal{N} \end{aligned} \tag{22}$$

where $\mathbf{z}[n] \sim \mathcal{CN}(\mathbf{0}, I_{n_r})$ is independent across n, and is independent of the channel input and fading matrix.

Let

$$\begin{aligned} \mathbf{x}^{\mathcal{N}} &= [\mathbf{x}[1]^T \mathbf{x}[2]^T \cdots \mathbf{x}[B]^T]^T \in \mathcal{C}^{Bn_t} \\ \mathbf{x}_i^{\mathcal{N}} &= [x_i[1]^T x_i[2]^T \cdots x_i[B]^T]^T \in \mathcal{C}^{B} \quad i = 1, 2 \ldots, n_t \\ \mathbf{y}^{\mathcal{N}} &= [\mathbf{y}[1]^T \mathbf{y}[2]^T \cdots \mathbf{y}[B]^T]^T \in \mathcal{C}^{Bn_r} \\ \mathbf{y}^{(i)\mathcal{N}} &= [\mathbf{y}^{(i)}[1]^T \mathbf{y}^{(i)}[2]^T \cdots \mathbf{y}^{(i)}[B]^T]^T \in \mathcal{C}^{Bn_r} \quad i = 1, 2 \ldots, n_t \end{aligned}$$

Then for any distribution of $\mathbf{x}^{\mathcal{N}}$ we can use the data processing inequality, the chain rule for differential entropies, and the fact that conditioning does not increase differential entropy to get the upper bound:

$$
\begin{aligned}
I(\mathbf{x}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}}) &\leq I(\mathbf{x}^{\mathcal{N}};\mathbf{y}^{(1)\mathcal{N}},\mathbf{y}^{(2)\mathcal{N}},\ldots,\mathbf{y}^{(n_t)\mathcal{N}}) \\
&= h(\mathbf{y}^{(1)\mathcal{N}},\mathbf{y}^{(2)\mathcal{N}},\ldots,\mathbf{y}^{(n_t)\mathcal{N}}) - h(\mathbf{y}^{(1)\mathcal{N}},\mathbf{y}^{(2)\mathcal{N}},\ldots,\mathbf{y}^{(n_t)\mathcal{N}} \mid \mathbf{x}^{\mathcal{N}}) \\
&= h(\mathbf{y}^{(1)\mathcal{N}}) + \sum_{i=2}^{n_t} h(\mathbf{y}^{(i)\mathcal{N}} \mid \mathbf{y}^{(i-1)\mathcal{N}},\ldots,\mathbf{y}^{(1)\mathcal{N}}) - \sum_{i=1}^{n_t} h(\mathbf{y}^{(i)\mathcal{N}} \mid \mathbf{x}^{\mathcal{N}}) \\
&\leq \sum_{i=1}^{n_t} h(\mathbf{y}^{(i)\mathcal{N}}) - \sum_{i=1}^{n_t} h(\mathbf{y}^{(i)\mathcal{N}} \mid \mathbf{x}_i^{\mathcal{N}}) \\
&= \sum_{i=1}^{n_t} I(\mathbf{x}_i^{\mathcal{N}};\mathbf{y}^{(i)\mathcal{N}}) \qquad (23)
\end{aligned}
$$

This upper bound can be better understood by interpreting $\mathbf{y}^{(i)}[n]$ as the received signal corresponding to the transmission in the transmit antenna i as if it could be received without interference from the signals transmitted in the other antennas.

The power constraint on the input distribution

$$
\frac{1}{B}\sum_{n\in\mathcal{N}} E\left[\|\mathbf{x}[n]\|^2\right] \leq n_t
$$

implies that the input to each of the n_t SIMO channels must satisfy

$$
\frac{1}{B}\sum_{n\in\mathcal{N}} E\left[|\mathbf{x}_i[n]|^2\right] \leq n_t
$$

Therefore we can bound each of the n_t terms of (23) by the bound (20) replacing $\mathsf{SNR} \to n_t\mathsf{SNR}$. However, for large SNR, $\log\log(n_t\mathsf{SNR}) \approx \log\log\mathsf{SNR}$ in the sense that in the limit the difference between the two sides converges to 0. As a result we obtain for the $n_r > n_t$ case the bound:

$$
\limsup_{\substack{\mathsf{SNR}\to\infty \\ \epsilon\to 0}} \left[C(\mathsf{SNR},\epsilon) - \{n_t\log\log\mathsf{SNR} + n_t\log(1/\epsilon) + K(n_r,n_t)\}\right] \leq 0
$$

where $K(n_r,n_t) = n_t[2n_r\log(2) - n_r\gamma + \log(2\pi e\pi) + 3.97722]$, which corresponds to the bound of Theorem 4 for $\mathsf{SNR} > 1/\epsilon$.

6 Conclusion

Motivated by the capacity results of the MIMO channel with perfect channel state information at the receiver we identified the parameter $\min\{n_r,n_t\}$ with the number of degrees of freedom of the channel. We then studied whether the concept of degrees of freedom could also be used in a MIMO channel where the fading matrix varies continuously and is unknown to both the receiver and the transmitter.

Towards this end we obtained lower and upper bounds for channel capacity, the upper bounds being asymptotic in $\mathsf{SNR} \to \infty$ and $\epsilon \to 0$. These bounds show different behaviors depending on the relationship between SNR^{-1} and ϵ. Depending on the relative values of SNR and ϵ we defined three regimes of operation and argued, using numerical examples, that most channels are underspread and the majority of practical systems operate in the first two regimes.

We found that in the first two regimes the lower and upper bounds match to within an additive constant, and the capacity grows as $\min\{n_r, n_t\} \log \min\{\mathsf{SNR}, \epsilon^{-1}\}$ for large SNR and small ϵ. This implies that even in the case of a continuously varying channel without the assumption of perfect CSI the concept of degrees of freedom has a great practical importance. Systems operating in the first two regimes can obtain large capacity improvements by the use of multiple antennas. This capacity improvement can be readily achieved by using a concrete communication scheme of tractable complexity.

Our upper bound is consistent with Lapidoth-Moser's result for the third regime, where capacity only grows doubly logarithmically on the SNR for large enough SNR. However, our numerical examples show that this behavior requires extremely large values of SNR to take place, and these large values are never attained in practical systems for underspread channels.

Appendix A

Proof of Theorem 2

Let $\mathbf{x}^{\mathcal{N}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_B)$. With this choice of input distribution we can write,

$$C \geq \lim_{B\to\infty} \frac{1}{B} I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}) = \lim_{B\to\infty} \frac{1}{B} \left[h(\mathbf{y}^{\mathcal{N}}) - h(\mathbf{y}^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}}) \right] \geq \lim_{B\to\infty} \frac{1}{B} \left[h(\mathbf{y}^{\mathcal{N}}|\mathbf{H}^{\mathcal{N}}) - h(\mathbf{y}^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}}) \right] \tag{24}$$

where $\mathbf{H}^{\mathcal{N}} = [\mathbf{H}[1]\mathbf{H}[2] \cdots \mathbf{H}[B]]$ and the last inequality follows from the fact that conditioning does not increase differential entropy.

We present bounds for the two terms of the right hand side of (24) in the following two lemmas.

Lemma 7 *Let* $\mathbf{x}^{\mathcal{N}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_B)$*. Then*

$$\frac{1}{B} h(\mathbf{y}^{\mathcal{N}}|\mathbf{H}^{\mathcal{N}}) \geq n_r \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + E\left[\sum_{k=n_{max}-(n_{min}-1)}^{n_{max}} \log\left(1 + \frac{\mathsf{SNR}}{n_t}\chi^2_{2k}\right)\right] \tag{25}$$

where $n_{max} = \max\{n_t, n_r\}$ *and* $n_{min} = \min\{n_t, n_r\}$*.*

Proof: Conditioned on $\mathbf{H}^{\mathcal{N}}$ the components of $\mathbf{y}^{\mathcal{N}}$ are independent across time, so we can write:

$$h(\mathbf{y}^{\mathcal{N}}|\mathbf{H}^{\mathcal{N}}) = E_{\mathbf{H}^{\mathcal{N}}}\left[\sum_{n=1}^{B} \log\det(\pi e K_{\mathbf{y}[n]|\mathbf{H}[n]})\right] = B \cdot E_{\mathbf{H}[1]}\left[\log\det(\pi e K_{\mathbf{y}[1]|\mathbf{H}[1]})\right] \tag{26}$$

where $K_{\mathbf{y}[n]|\mathbf{H}[n]} = E[\mathbf{y}[n]\mathbf{y}[n]^\dagger|\mathbf{H}[n]] = \mathbf{H}[n]\mathbf{H}[n]^\dagger + (n_t/\mathsf{SNR})\mathbf{I}_{n_r}$.

Then we can rewrite (26) as:

$$h(\mathbf{y}^{\mathcal{N}}|\mathbf{H}^{\mathcal{N}}) \quad = \quad B \cdot E_{\mathbf{H}[1]}\left\{\log\det\left[\frac{\pi e n_t}{\mathsf{SNR}}\left(\mathbf{I}_{n_r} + \frac{\mathsf{SNR}}{n_t}\mathbf{H}[1]\mathbf{H}[1]^\dagger\right)\right]\right\}$$

$$= \quad B\left\{n_r \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + E_{\mathbf{H}[1]}\left[\log\det\left(\mathbf{I}_{n_r} + \frac{\mathsf{SNR}}{n_t}\mathbf{H}[1]\mathbf{H}[1]^{\dagger}\right)\right]\right\} \tag{27}$$

For the case $n_t \geq n_r$ Foschini and Gans obtained the following lower bound [1]:

$$E_{\mathbf{H}[1]}\left[\log\det\left(\mathbf{I}_{n_r} + \frac{\mathsf{SNR}}{n_t}\mathbf{H}[1]\mathbf{H}[1]^{\dagger}\right)\right] \geq E\left[\sum_{k=n_t-(n_r-1)}^{n_t} \log\left(1 + \frac{\mathsf{SNR}}{n_t}\chi^2_{2k}\right)\right] \tag{28}$$

where χ^2_{2k} is a chi-squared random variable with $2k$ degrees of freedom.

On the other hand, if $n_t < n_r$ we can use the above lower bound in the following way:

$$\begin{aligned} E_{\mathbf{H}[1]}\left[\log\det\left(\mathbf{I}_{n_r} + \frac{\mathsf{SNR}}{n_t}\mathbf{H}[1]\mathbf{H}[1]^{\dagger}\right)\right] &= E_{\mathbf{H}[1]}\left[\log\det\left(\mathbf{I}_{n_t} + \frac{\mathsf{SNR}}{n_t}\mathbf{H}[1]^{\dagger}\mathbf{H}[1]\right)\right] \\ &\geq E\left[\sum_{k=n_r-(n_t-1)}^{n_r} \log\left(1 + \frac{\mathsf{SNR}}{n_t}\chi^2_{2k}\right)\right] \end{aligned} \tag{29}$$

The result follows from (27), (28) and (29). ■

Lemma 8 *Let* $\mathbf{x}^{\mathcal{N}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_B)$. *Then*

$$\begin{aligned} \lim_{B\to\infty} \frac{1}{B} h(\mathbf{y}^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}}) &\leq n_r \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) - n_r n_t \log(2) \\ &+ n_r n_t \log\left[\frac{\mathsf{SNR}\epsilon}{n_t} + 2 - \epsilon + \sqrt{\left(\frac{\mathsf{SNR}\epsilon}{n_t} + 2 - \epsilon\right)^2 - 4 + 4\epsilon}\right] \end{aligned} \tag{30}$$

Proof: Conditioned on $\mathbf{x}^{\mathcal{N}}$ the n_r components of $\mathbf{y}[n]$ are i.i.d. and therefore $h(\mathbf{y}^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}}) = n_r \cdot h(\mathbf{y}_1^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}})$, where $\mathbf{y}_1^{\mathcal{N}}$ is the vector of received signals in the first receive antenna for $n \in \mathcal{N}$.

Let $\tilde{\mathbf{h}} = [\mathbf{h}_1[1]^T\mathbf{h}_1[2]^T \cdots \mathbf{h}_1[B]^T]^T \in \mathcal{C}^{n_t B}$ where $\mathbf{h}_1[n]^T$ is the first row of $\mathbf{H}[n]$, $n \in \mathcal{N}$, and define $\tilde{X} \in \mathcal{C}^{Bxn_t B}$ in the following way:

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{x}[1]^T & \mathbf{0}^T & \cdots & \mathbf{0}^T \\ \mathbf{0}^T & \mathbf{x}[2]^T & \cdots & \mathbf{0}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}^T & \mathbf{0}^T & \cdots & \mathbf{x}[B]^T \end{bmatrix}$$

where $\mathbf{0}^T = [0, \cdots, 0] \in \mathcal{C}^{1xn_t}$. In the same way define $\mathbf{z}_1^{\mathcal{N}} = [z_1[1]z_1[2]\cdots z_1[B]]^T \in \mathcal{C}^{Bx1}$, where $z_1[n]$ is the first component of $\mathbf{z}[n]$, $n \in \mathcal{N}$.

Then we have

$$\mathbf{y}_1^{\mathcal{N}} = \tilde{\mathbf{X}}\tilde{\mathbf{h}} + \sqrt{\frac{n_t}{\mathsf{SNR}}}\mathbf{z}_1^{\mathcal{N}}$$

and using the Gaussianity of $\mathbf{y}_1^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}}$,

$$h(\mathbf{y}_1^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}}) = h(\mathbf{y}_1^{\mathcal{N}}|\tilde{\mathbf{X}}) = E_{\tilde{\mathbf{X}}}\left[\log\det\left(\pi e K_{\mathbf{y}_1^{\mathcal{N}}|\tilde{\mathbf{X}}}\right)\right]$$

where

$$
\begin{aligned}
K_{\mathbf{y}_1^{\mathcal{N}}|\tilde{\mathbf{X}}} &= E\left[\mathbf{y}_1^{\mathcal{N}}\mathbf{y}_1^{\mathcal{N}\dagger}|\tilde{\mathbf{X}}\right] \\
&= E\left[\tilde{\mathbf{X}}\tilde{\mathbf{h}}\tilde{\mathbf{h}}^\dagger\tilde{\mathbf{X}}^\dagger + \sqrt{(n_t/\mathsf{SNR})}\tilde{\mathbf{X}}\tilde{\mathbf{h}}\mathbf{z}_1^{\mathcal{N}\dagger} + \sqrt{(n_t/\mathsf{SNR})}\mathbf{z}_1^{\mathcal{N}}\tilde{\mathbf{h}}^\dagger\tilde{\mathbf{X}}^\dagger + (n_t/\mathsf{SNR})\mathbf{z}_1^{\mathcal{N}}\mathbf{z}_1^{\mathcal{N}\dagger}\right|\tilde{\mathbf{X}}\right] \\
&= \tilde{\mathbf{X}}K_{\tilde{\mathbf{h}}}\tilde{\mathbf{X}}^\dagger + (n_t/\mathsf{SNR})\mathbf{I}_B
\end{aligned}
$$

with $K_{\tilde{\mathbf{h}}} = E[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^\dagger]$.

We can upper bound $h(\mathbf{y}^N|\mathbf{x}^N)$ as follows:

$$
\begin{aligned}
\frac{h(\mathbf{y}^N|\mathbf{x}^N)}{n_r} &= B\log(\pi e) + E_{\tilde{\mathbf{X}}}\left[\log\det\left(\tilde{\mathbf{X}}K_{\tilde{\mathbf{h}}}\tilde{\mathbf{X}}^\dagger + \frac{n_t}{\mathsf{SNR}}\mathbf{I}_B\right)\right] \\
&= B\log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + E_{\tilde{\mathbf{X}}}\left[\log\det\left(\frac{\mathsf{SNR}}{n_t}\tilde{\mathbf{X}}K_{\tilde{\mathbf{h}}}\tilde{\mathbf{X}}^\dagger + \mathbf{I}_B\right)\right] \\
&= B\log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + E_{\tilde{\mathbf{X}}}\left[\log\det\left(\frac{\mathsf{SNR}}{n_t}\tilde{\mathbf{X}}^\dagger\tilde{\mathbf{X}}K_{\tilde{\mathbf{h}}} + \mathbf{I}_{n_tB}\right)\right] \\
&\le B\log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \log\det\left[\frac{\mathsf{SNR}}{n_t}E_{\tilde{\mathbf{X}}}\left(\tilde{\mathbf{X}}^\dagger\tilde{\mathbf{X}}\right)K_{\tilde{\mathbf{h}}} + \mathbf{I}_{n_tB}\right] \\
&= B\log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \log\det\left[\frac{\mathsf{SNR}}{n_t}K_{\tilde{\mathbf{h}}} + \mathbf{I}_{n_tB}\right] \qquad (31)
\end{aligned}
$$

where the inequality follows from Jensen's inequality and the last equality results from $E_{\tilde{\mathbf{X}}}\left(\tilde{\mathbf{X}}^\dagger\tilde{\mathbf{X}}\right) = \mathbf{I}_{n_tB}$. To upper bound the last expression we start by computing $K_{\tilde{\mathbf{h}}}$ in terms of $R_{h_{11}}[n]$, the autocorrelation function of the entry $(1,1)$ of $\mathbf{H}[n]$.

$$
\begin{aligned}
K_{\tilde{\mathbf{h}}} &= E[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^\dagger] \\
&= E\begin{bmatrix}
|h_{11}[1]|^2 & \cdots & h_{11}[1]h_{1n_t}^*[1] & h_{11}[1]h_{11}^*[2] & \cdots & h_{11}[1]h_{1n_t}^*[2] & \cdots & h_{11}[1]h_{1n_t}^*[B] \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
h_{1n_t}[1]h_{11}^*[1] & \cdots & |h_{1n_t}[1]|^2 & h_{1n_t}[1]h_{11}^*[2] & \cdots & h_{1n_t}[1]h_{1n_t}^*[2] & \cdots & h_{1n_t}[1]h_{1n_t}^*[B] \\
h_{11}[2]h_{11}^*[1] & \cdots & h_{11}[2]h_{1n_t}^*[1] & |h_{11}[2]|^2 & \cdots & h_{11}[2]h_{1n_t}^*[2] & \cdots & h_{11}[2]h_{1n_t}^*[B] \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
h_{1n_t}[2]h_{11}^*[1] & \cdots & h_{1n_t}[2]h_{1n_t}^*[1] & h_{1n_t}[2]h_{11}^*[2] & \cdots & |h_{1n_t}[2]|^2 & \cdots & h_{1n_t}[2]h_{1n_t}^*[B] \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \\
h_{1n_t}[B]h_{11}^*[1] & \cdots & h_{1n_t}[B]h_{1n_t}^*[1] & h_{1n_t}[B]h_{11}^*[2] & \cdots & h_{1n_t}[B]h_{1n_t}^*[2] & \cdots & |h_{1n_t}[B]|^2
\end{bmatrix} \\
&= \begin{bmatrix}
R_{h_{11}}[0]\mathbf{I}_{n_t} & R_{h_{11}}[1]^*\mathbf{I}_{n_t} & \cdots & R_{h_{11}}[B-1]^*\mathbf{I}_{n_t} \\
R_{h_{11}}[1]\mathbf{I}_{n_t} & R_{h_{11}}[0]\mathbf{I}_{n_t} & \cdots & R_{h_{11}}[B-2]^*\mathbf{I}_{n_t} \\
\vdots & \vdots & \ddots & \vdots \\
R_{h_{11}}[B-1]\mathbf{I}_{n_t} & R_{h_{11}}[B-2]\mathbf{I}_{n_t} & \cdots & R_{h_{11}}[0]\mathbf{I}_{n_t}
\end{bmatrix}
\end{aligned}
$$

where we have used the fact that the entries of $\mathbf{H}[n]$ are i.i.d.

To compute the right hand side of (31) we use the following known results:

Fact 2 *Let $\hat{h}[n], n \in \mathcal{Z}$, be a random process with autocorrelation function:*

$$R_{\hat{h}}[n] = \begin{cases} R_{h_{11}}[n/n_t] & if \quad n = k \cdot n_t \\ 0 & otherwise \end{cases}$$

Then the power spectral density of $\hat{h}[n]$ is given by:

$$S_{\hat{h}}(f) = S_{h_{11}}(n_t f)$$

Fact 3 *The distribution of the eigenvalues of a Toeplitz matrix converges to the Fourier transform of its rows, as the dimension of the matrix goes to infinity.*

We can use these facts for the Toeplitz matrix $K_{\tilde{\mathbf{h}}}$. Letting $\{\lambda_i : 1 \le i \le n_t B\}$ be the set of eigenvalues of $K_{\tilde{\mathbf{h}}}$ we obtain the following asymptotic bound:

$$\begin{aligned}
\lim_{B\to\infty} \frac{h(\mathbf{y}^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}})}{n_r B} &\le \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \lim_{B\to\infty} \frac{1}{B} \log\det\left[\frac{\mathsf{SNR}}{n_t} K_{\tilde{\mathbf{h}}} + \mathbf{I}_{n_t B}\right] \\
&= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \lim_{B\to\infty} \frac{1}{B} \sum_{i=1}^{n_t B} \log\left(\frac{\mathsf{SNR}}{n_t}\lambda_i + 1\right) \\
&= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_t \int_0^1 \log\left[\frac{\mathsf{SNR}}{n_t} S_{\hat{h}}(f) + 1\right] df \\
&= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_t \int_0^1 \log\left[\frac{\mathsf{SNR}}{n_t} S_{h_{11}}(n_t f) + 1\right] df \\
&= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \int_0^{n_t} \log\left[\frac{\mathsf{SNR}}{n_t} S_{h_{11}}(f') + 1\right] df' \\
&= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_t \int_0^1 \log\left[\frac{\mathsf{SNR}}{n_t} S_{h_{11}}(f) + 1\right] df
\end{aligned} \tag{32}$$

where we exploited the periodicity of $S_{h_{11}}(f)$ in the last equality.

The power spectral density of $h_{11}[n]$ is obtained by taking the discrete time Fourier transform of its autocorrelation function:

$$R_{h_{11}}[n] = (1-\epsilon)^{|n|/2}$$

In this way we obtain:

$$S_{h_{11}}(f) = \mathcal{F}\{R_{h_{11}}[n]\} = \frac{\epsilon}{1 + (\sqrt{1-\epsilon})^2 - 2\sqrt{1-\epsilon}\cos(2\pi f)}$$

Letting $\alpha = \sqrt{1-\epsilon}$ and replacing in (32) we obtain:

$$\lim_{B\to\infty} \frac{h(\mathbf{y}^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}})}{n_r B} \le \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_t \int_0^1 \log\left[\frac{\mathsf{SNR}\epsilon/n_t}{1 + \alpha^2 - 2\alpha\cos(2\pi f)} + 1\right] df$$

$$= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_t \int_0^1 \log\left[\mathsf{SNR}\epsilon/n_t + 1 + \alpha^2 - 2\alpha\cos(2\pi f)\right] df$$

$$-n_t \int_0^1 \log\left[1 + \alpha^2 - 2\alpha\cos(2\pi f)\right] df$$

$$= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \frac{n_t}{\pi} \int_0^\pi \log\left[\mathsf{SNR}\epsilon/n_t + 1 + \alpha^2 - 2\alpha\cos(\omega)\right] d\omega$$

$$-\frac{n_t}{\pi} \int_0^\pi \log\left[1 + \alpha^2 - 2\alpha\cos(\omega)\right] d\omega \qquad (33)$$

The last two integrals can be computed in closed formed using the identity $\int_0^\pi \log(a \pm b\cos x)dx = \pi\log(\frac{a+\sqrt{a^2-b^2}}{2})$ valid for $a \geq b$:

$$\int_0^\pi \log\left[1 + \alpha^2 - 2\alpha\cos(\omega)\right] d\omega = 0$$

$$\int_0^\pi \log\left[\mathsf{SNR}\epsilon/n_t + 1 + \alpha^2 - 2\alpha\cos(\omega)\right] d\omega = \pi\log\left[\frac{(\mathsf{SNR}\epsilon/n_t + 1 + \alpha^2) + \sqrt{(\mathsf{SNR}\epsilon/n_t + 1 + \alpha^2)^2 - 4\alpha^2}}{2}\right]$$

Therefore we obtain:

$$\lim_{B\to\infty} \frac{h(\mathbf{y}^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}})}{B} \leq n_r \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_r n_t \log\left[(\mathsf{SNR}\epsilon/n_t + 2 - \epsilon) + \sqrt{(\mathsf{SNR}\epsilon/n_t + 2 - \epsilon)^2 - 4(1-\epsilon)}\right]$$
$$-n_r n_t \log(2)$$

■

As a result, we can use (24), (25) and (30) to get a lower bound for the capacity of the channel.

$$C(\mathsf{SNR}, \epsilon) \geq E\left[\sum_{k=n_{max}-(n_{min}-1)}^{n_{max}} \log\left(1 + \frac{\mathsf{SNR}}{n_t}\chi^2_{2k}\right)\right] + n_r n_t \log(2)$$
$$- n_r n_t \log\left[\frac{\mathsf{SNR}\epsilon}{n_t} + 2 - \epsilon + \sqrt{\left(\frac{\mathsf{SNR}\epsilon}{n_t} + 2 - \epsilon\right)^2 - 4 + 4\epsilon}\right] \qquad (34)$$

This lower bound decreases with SNR for sufficiently large values of SNR. Since the capacity $C(\mathsf{SNR}, \epsilon)$ is an increasing function of SNR, we can improve the lower bound by keeping it constant for $\mathsf{SNR} > \mathsf{SNR}^*$, where SNR^* is the value that maximizes the right hand side of (34):

$$C(\mathsf{SNR}, \epsilon) \geq E\left[\sum_{k=n_{max}-(n_{min}-1)}^{n_{max}} \log\left(1 + \frac{\min\{\mathsf{SNR}, \mathsf{SNR}^*\}}{n_t}\chi^2_{2k}\right)\right] + n_r n_t \log(2)$$

$$- \quad n_r n_t \log \left[\frac{\min\{\mathsf{SNR}, \mathsf{SNR}^*\}\epsilon}{n_t} + 2 - \epsilon + \sqrt{\left(\frac{\min\{\mathsf{SNR}, \mathsf{SNR}^*\}\epsilon}{n_t} + 2 - \epsilon \right)^2 + 4(1-\epsilon)} \right]$$

To obtain a closed form expression for SNR^* we note that as $\mathsf{SNR} \to \infty$:

$$E\left[\sum_{k=n_{max}-(n_{min}-1)}^{n_{max}} \log\left(1 + \frac{\mathsf{SNR}}{n_t}\chi^2_{2k}\right) \right] \sim n_{min} \log \mathsf{SNR}$$

in the sense that the difference between the two sides converges to a constant. This approximation allows us to maximize the RHS of (34) and obtain:

$$\mathsf{SNR}^* \approx \begin{cases} \frac{n_t}{\epsilon} \cdot \frac{n_{min}^2(2-\epsilon) + \sqrt{4n_{min}^4(1-\epsilon) + \epsilon^2 n_{min}^2 n_r^2 n_t^2}}{n_r^2 n_t^2 - n_{min}^2} & if \quad n_{max} > 1 \\ \infty & if \quad n_{max} = 1 \end{cases}$$

valid for $\mathsf{SNR} \gg 1$. If this condition is not met, the lower bound obtained using this expression for SNR^* is still valid. However, we could obtain a tighter lower bound by computing SNR^* numerically.

Appendix B

Proof of Theorem 1

We want to characterize the behavior of $C(\mathsf{SNR}, \epsilon)$ in the limit as $\mathsf{SNR} \to \infty$ and $\epsilon \to 0$. Since we are only interested in the asymptotics, we can relax the bound (8) in the following way:

$$\begin{aligned} C(\mathsf{SNR}, \epsilon) \geq & \sum_{k=n_{max}-(n_{min}-1)}^{n_{max}} E\left[\log\left(1 + \frac{\min\{\mathsf{SNR}, \mathsf{SNR}^*\}}{n_t}\chi^2_{2k}\right)\right] + n_r n_r \log(2) \\ & - n_r n_t \log\left[\min\{\mathsf{SNR}, \mathsf{SNR}^*\}\frac{\epsilon}{n_t} + 2 - \epsilon + \sqrt{\left(\min\{\mathsf{SNR}, \mathsf{SNR}^*\}\frac{\epsilon}{n_t} + 2 - \epsilon\right)^2 - 4(1-\epsilon)}\right] \\ \geq & \sum_{k=n_{max}-(n_{min}-1)}^{n_{max}} E\left[\log\left(1 + \frac{\min\{\mathsf{SNR}, \mathsf{SNR}^*\}}{n_t}\chi^2_{2k}\right)\right] \\ & - n_r n_t \log\left[\min\{\mathsf{SNR}, \mathsf{SNR}^*\}\frac{\epsilon}{n_t} + 2\right] \end{aligned} \qquad (35)$$

When $n_r = n_t = 1$, $\mathsf{SNR}^* = \infty$ and (35) reduces to:

$$\begin{aligned} C(\mathsf{SNR}, \epsilon) & \geq E\left[\log\left(1 + \mathsf{SNR}\chi^2_{2k}\right)\right] - \log\left[\mathsf{SNR}\epsilon + 2\right] \\ & = E\left[\log\left(1 + \mathsf{SNR}\chi^2_{2k}\right)\right] - \log(\mathsf{SNR}) - \log\left[\epsilon + \frac{2}{\mathsf{SNR}}\right] \\ & \geq E\left[\log\left(1 + \mathsf{SNR}\chi^2_{2k}\right)\right] - \log(\mathsf{SNR}) - \log\left[\max\left\{\epsilon, \frac{1}{\mathsf{SNR}}\right\}(1+2)\right] \\ & = E\left[\log\left(1 + \mathsf{SNR}\chi^2_{2k}\right)\right] - \log(\mathsf{SNR}) + \log\left[\min\left\{\mathsf{SNR}, \frac{1}{\epsilon}\right\}\right] - \log(3) \end{aligned}$$

The result for $n_r = n_t = 1$ follows by noting that for some constant K:

$$\lim_{\mathsf{SNR}\to\infty} E\left[\log\left(1+\mathsf{SNR}\chi^2_{2k}\right)\right] - \log(\mathsf{SNR}) = K$$

For $\max\{n_r, n_t\} > 1$, as $\epsilon \to 0$ we can approximate SNR^* by:

$$\mathsf{SNR}^* \approx \frac{n_t}{\epsilon} \cdot \frac{4}{n_r^2 n_t^2 / n_{min}^2 - 1}$$

where in the limit the difference between the two sides converges to zero. Also, for small ϵ we can lower bound the second term of (35) by a constant that does not depend on SNR or ϵ:

$$-n_r n_t \log\left[\min\{\mathsf{SNR}, \mathsf{SNR}^*\}\frac{\epsilon}{n_t} + 2\right] \geq -n_r n_t \log\left[\mathsf{SNR}^*\frac{\epsilon}{n_t} + 2\right] = -n_r n_t \log\left[\frac{4}{n_r^2 n_t^2/n_{min}^2 - 1} + 2\right]$$

As $\mathsf{SNR} \to \infty$ and $\epsilon \to 0$, $\min(\mathsf{SNR}, \mathsf{SNR}^*) \to \infty$ so it follows that:

$$\begin{aligned}\sum_{k=n_{max}-(n_{min}-1)}^{n_{max}} E\left[\log\left(1+\frac{\min\{\mathsf{SNR},\mathsf{SNR}^*\}}{n_t}\chi^2_{2k}\right)\right] &\sim n_{min}\log\left[\min\{\mathsf{SNR},\mathsf{SNR}^*\}\right] \\ &\sim n_{min}\log\left[\min\left\{\mathsf{SNR}, \frac{1}{\epsilon}\right\}\right]\end{aligned}$$

where $\sim$ indicates that in the limit the difference between the two sides converges to a constant.

Therefore the result also follows for $\min\{n_r, n_t\} > 1$.

Appendix C

Proof of Theorem 6

We start by defining a set $\mathcal{G} = \{0,1\}^B$ and a random vector $\gamma \in \mathcal{G}$, where $\gamma[n] = 1(\|\mathbf{x}[n]\| \geq \mathsf{SNR}^{-1/2})$, and use the chain rule of mutual information in two different ways:

$$\begin{aligned} I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}, \gamma) &= I(\mathbf{x}^{\mathcal{N}}; \gamma) + I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}|\gamma) \\ I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}, \gamma) &= I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}) + I(\mathbf{x}^{\mathcal{N}}; \gamma|\mathbf{y}^{\mathcal{N}}) \end{aligned}$$

From these, the non-negativity of mutual information, and the fact that γ is a discrete random vector that takes 2^B values, we obtain the following bound:

$$\begin{aligned} I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}) &= I(\mathbf{x}^{\mathcal{N}}; \gamma) + I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}|\gamma) - I(\mathbf{x}^{\mathcal{N}}; \gamma|\mathbf{y}^{\mathcal{N}}) \leq I(\mathbf{x}^{\mathcal{N}}; \gamma) + I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}|\gamma) \\ &\leq B\log 2 + I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}|\gamma) \end{aligned} \tag{36}$$

Now the problem is that of finding an upper bound for

$$I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}|\gamma) = \sum_{\mathbf{v}\in\mathcal{G}} I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}|\gamma = \mathbf{v}) P(\gamma = \mathbf{v}) \tag{37}$$

for every possible choice of the distribution of γ. We will find an upper bound for $I(\mathbf{x}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}}|\gamma=\mathbf{v})$ for $\mathbf{v}\in\mathcal{G}$ for any input distribution that satisfies the power constraint

$$\frac{1}{B}\sum_{n=1}^{B} E\left[\|\mathbf{x}[n]\|^2\middle|\gamma=\mathbf{v}\right]=\xi_{\mathbf{v}} \tag{38}$$

where the constants $\{\xi_{\mathbf{v}}\}_{\mathbf{v}\in\mathcal{G}}$ must satisfy the total power constraint

$$\sum_{\mathbf{v}\in\mathcal{G}}\xi_{\mathbf{v}}P(\gamma=\mathbf{v})\leq n_t \tag{39}$$

Define $\mathcal{A}_{\mathbf{v}}=\{n:v[n]=0\}$ and $\mathcal{B}_{\mathbf{v}}=\{n:v[n]=1\}$. For $n\in\mathcal{A}_{\mathbf{v}}$ the norm of the input is of the same order as the norm of the noise, so the information that gets through the channel during those times must be bounded. This intuitive idea is made precise in the following lemma.

Lemma 9

$$\frac{1}{B}I(\mathbf{x}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}}|\gamma=\mathbf{v})\leq 2n_r\log(2)+\frac{1}{B}I(\mathbf{x}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\gamma=\mathbf{v}) \tag{40}$$

Proof: We use the chain rule for mutual information to obtain a decomposition of $I(\mathbf{x}^{\mathcal{N}},\mathbf{y}^{\mathcal{N}}|\gamma=\mathbf{v})$ into tree terms:

$$\begin{aligned}
I(\mathbf{x}^{\mathcal{N}},\mathbf{y}^{\mathcal{N}}|\gamma=\mathbf{v}) &= I(\mathbf{x}^{\mathcal{A}_{\mathbf{v}}},\mathbf{x}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{A}_{\mathbf{v}}},\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\gamma=\mathbf{v})\\
&= I(\mathbf{x}^{\mathcal{A}_{\mathbf{v}}},\mathbf{x}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\gamma=\mathbf{v})+I(\mathbf{x}^{\mathcal{A}_{\mathbf{v}}};\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}|\mathbf{y}^{\mathcal{B}_{\mathbf{v}}},\gamma=\mathbf{v})\\
&\quad +I(\mathbf{x}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}|\mathbf{x}^{\mathcal{A}_{\mathbf{v}}},\mathbf{y}^{\mathcal{B}_{\mathbf{v}}},\gamma=\mathbf{v})
\end{aligned} \tag{41}$$

The first term can be simplified by rewriting the mutual information in terms of differential entropies, and noting that conditioned on $\mathbf{x}^{\mathcal{B}_{\mathbf{v}}}$, $\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}$ is independent of $\mathbf{x}^{\mathcal{A}_{\mathbf{v}}}$:

$$\begin{aligned}
I(\mathbf{x}^{\mathcal{A}_{\mathbf{v}}},\mathbf{x}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\gamma=\mathbf{v}) &= h(\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\gamma=\mathbf{v})-h(\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\mathbf{x}^{\mathcal{A}_{\mathbf{v}}},\mathbf{x}^{\mathcal{B}_{\mathbf{v}}},\gamma=\mathbf{v})\\
&= h(\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\gamma=\mathbf{v})-h(\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\mathbf{x}^{\mathcal{B}_{\mathbf{v}}},\gamma=\mathbf{v})=I(\mathbf{x}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\gamma=\mathbf{v})
\end{aligned}$$

The second term of (41) can be upper bounded by rewriting the mutual information in terms of differential entropies and bounding each of the individual terms:

$$\begin{aligned}
I(\mathbf{x}^{\mathcal{A}_{\mathbf{v}}};\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}|\mathbf{y}^{\mathcal{B}_{\mathbf{v}}},\gamma=\mathbf{v}) &= h(\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}|\mathbf{y}^{\mathcal{B}_{\mathbf{v}}},\gamma=\mathbf{v})-h(\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}|\mathbf{x}^{\mathcal{A}_{\mathbf{v}}},\mathbf{y}^{\mathcal{B}_{\mathbf{v}}},\gamma=\mathbf{v})\\
&\leq h(\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}|\gamma=\mathbf{v})-h(\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}|\mathbf{x}^{\mathcal{A}_{\mathbf{v}}},\mathbf{H}^{\mathcal{A}_{\mathbf{v}}},\mathbf{y}^{\mathcal{B}_{\mathbf{v}}},\gamma=\mathbf{v})\\
&= h(\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}|\gamma=\mathbf{v})-h(\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}|\mathbf{x}^{\mathcal{A}_{\mathbf{v}}},\mathbf{H}^{\mathcal{A}_{\mathbf{v}}},\gamma=\mathbf{v})
\end{aligned} \tag{42}$$

where the inequality follows from conditioning and the last equality is due to the fact that conditioned on $(\mathbf{x}^{\mathcal{A}_{\mathbf{v}}},\mathbf{H}^{\mathcal{A}_{\mathbf{v}}})$ the randomness of $\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}$ comes only from the noise $\mathbf{z}^{\mathcal{A}_{\mathbf{v}}}$ which is independent of $\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}$.

To bound the first term of (42) we define $\mathbf{d}[n] = \mathbf{x}[n]/\|\mathbf{x}[n]\|$ and compute $E[\mathbf{y}[n]\mathbf{y}[n]^\dagger|\gamma = \mathbf{v}]$ for $n \in \mathcal{A}_\mathbf{v}$:

$$\begin{aligned}
E\left[\mathbf{y}[n]\mathbf{y}[n]^\dagger \middle| \gamma = \mathbf{v}\right] &= E\left[\left(\mathbf{H}[n]\mathbf{x}[n] + \sqrt{\frac{n_t}{\mathsf{SNR}}}\mathbf{z}[n]\right)\left(\mathbf{H}[n]\mathbf{x}[n] + \sqrt{\frac{n_t}{\mathsf{SNR}}}\mathbf{z}[n]\right)^\dagger \middle| \gamma = \mathbf{v}\right] \\
&= E\left\{E\left[(\mathbf{H}[n]\mathbf{d}[n])(\mathbf{H}[n]\mathbf{d}[n])^\dagger \cdot \|\mathbf{x}[n]\|^2 \middle| \mathbf{d}[n], \gamma = \mathbf{v}\right] \middle| \gamma = \mathbf{v}\right\} \\
&\quad + \frac{n_t}{\mathsf{SNR}} E\left[\mathbf{z}[n]\mathbf{z}[n]^\dagger\right] \\
&= E\left\{E\left[\mathbf{H}[n]\mathbf{d}[n]\mathbf{d}[n]^\dagger\mathbf{H}[n]^\dagger \middle| \mathbf{d}[n], \gamma = \mathbf{v}\right] \cdot E\left[\|\mathbf{x}[n]\|^2 \middle| \mathbf{d}[n], \gamma = \mathbf{v}\right] \middle| \gamma = \mathbf{v}\right\} \\
&\quad + \frac{n_t}{\mathsf{SNR}} \mathbf{I}_{n_r} \\
&= \mathbf{I}_{n_r} E\left[\|\mathbf{x}[n]\|^2 \middle| \gamma = \mathbf{v}\right] + \frac{n_t}{\mathsf{SNR}} \mathbf{I}_{n_r}
\end{aligned}$$

where we used the conditional independence of $\mathbf{H}[n]\mathbf{d}[n]$ and $\mathbf{x}[n]$ conditioned on $\mathbf{d}[n]$ in the third equality, and the fact that $\mathbf{d}[n]$ is a unit norm vector and the entries of $\mathbf{H}[n]$ are i.i.d. $\mathcal{CN}(0,1)$ and therefore conditioned on $\mathbf{d}[n]$, $\mathbf{H}[n]\mathbf{d}[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{n_r})$ to compute $E\left[\mathbf{H}[n]\mathbf{d}[n]\mathbf{d}[n]^\dagger\mathbf{H}[n]^\dagger \mid \mathbf{d}[n], \gamma = \mathbf{v}\right]$ in the fourth equality.

For $n \in \mathcal{A}_\mathbf{v}$, $E\left[\|\mathbf{x}[n]\|^2|\gamma = \mathbf{v}\right] \le \mathsf{SNR}^{-1}$ and hence $\det\left\{E\left[\mathbf{y}[n]\mathbf{y}[n]^\dagger|\gamma = \mathbf{v}\right]\right\} \le [(n_t+1)\mathsf{SNR}^{-1}]^{n_r}$. Then using the chain rule, removing conditioning and applying the Gaussian bound we obtain:

$$h(\mathbf{y}^{\mathcal{A}_\mathbf{v}}|\gamma = \mathbf{v}) \le \sum_{n \in \mathcal{A}_\mathbf{v}} h(\mathbf{y}[n]|\gamma = \mathbf{v}) \le |\mathcal{A}_\mathbf{v}| n_r \log[\pi e (n_t + 1)\mathsf{SNR}^{-1}]$$

where $|\mathcal{A}_\mathbf{v}|$ is the cardinality of the set $\mathcal{A}_\mathbf{v}$.

The second term of (42) can be computed explicitly:

$$h(\mathbf{y}^{\mathcal{A}_\mathbf{v}}|\mathbf{x}^{\mathcal{A}_\mathbf{v}}, \mathbf{H}^{\mathcal{A}_\mathbf{v}}, \gamma = \mathbf{v}) = h\left(\sqrt{\frac{n_t}{\mathsf{SNR}}}\mathbf{z}^{\mathcal{A}_\mathbf{v}} \middle| \gamma = \mathbf{v}\right) = |\mathcal{A}_\mathbf{v}| n_r \log(\pi e n_t \mathsf{SNR}^{-1})$$

Therefore we obtain:

$$I(\mathbf{x}^{\mathcal{A}_\mathbf{v}}; \mathbf{y}^{\mathcal{A}_\mathbf{v}}|\mathbf{y}^{\mathcal{B}_\mathbf{v}}, \gamma = \mathbf{v}) \le |\mathcal{A}_\mathbf{v}| n_r \log(1 + n_t^{-1}) \le B n_r \log(2)$$

The last term of (41) can be bounded in a similar way:

$$\begin{aligned}
I(\mathbf{x}^{\mathcal{B}_\mathbf{v}}; \mathbf{y}^{\mathcal{A}_\mathbf{v}}|\mathbf{x}^{\mathcal{A}_\mathbf{v}}, \mathbf{y}^{\mathcal{B}_\mathbf{v}}, \gamma = \mathbf{v}) &= h(\mathbf{y}^{\mathcal{A}_\mathbf{v}}|\mathbf{x}^{\mathcal{A}_\mathbf{v}}, \mathbf{y}^{\mathcal{B}_\mathbf{v}}, \gamma = \mathbf{v}) - h(\mathbf{y}^{\mathcal{A}_\mathbf{v}}|\mathbf{x}^{\mathcal{A}_\mathbf{v}}, \mathbf{x}^{\mathcal{B}_\mathbf{v}}, \mathbf{y}^{\mathcal{B}_\mathbf{v}}, \gamma = \mathbf{v}) \\
&\le h(\mathbf{y}^{\mathcal{A}_\mathbf{v}}|\gamma = \mathbf{v}) - h(\mathbf{y}^{\mathcal{A}_\mathbf{v}}|\mathbf{x}^{\mathcal{A}_\mathbf{v}}, \mathbf{H}^{\mathcal{A}_\mathbf{v}}, \mathbf{x}^{\mathcal{B}_\mathbf{v}}, \mathbf{y}^{\mathcal{B}_\mathbf{v}}, \gamma = \mathbf{v}) \\
&= h(\mathbf{y}^{\mathcal{A}_\mathbf{v}}|\gamma = \mathbf{v}) - h(\mathbf{y}^{\mathcal{A}_\mathbf{v}}|\mathbf{x}^{\mathcal{A}_\mathbf{v}}, \mathbf{H}^{\mathcal{A}_\mathbf{v}}, \gamma = \mathbf{v}) \\
&\le |\mathcal{A}_\mathbf{v}| n_r \log(2) \le B n_r \log(2)
\end{aligned}$$

Replacing these 3 bounds in (41) we get the desired result. ∎

At this point our goal is to find an upper bound for (40) that depends on $\mathbf{v}$ only through $\xi_{\mathbf{v}}$. The power constraint (38) imposes a power constraint on $\mathbf{x}^{\mathcal{B}_{\mathbf{v}}}$:

$$\frac{1}{B}\sum_{n\in\mathcal{B}_{\mathbf{v}}} E\left[\|\mathbf{x}[n]\|^2\Big|\gamma=\mathbf{v}\right] \leq \frac{1}{B}\sum_{n=1}^{B} E\left[\|\mathbf{x}[n]\|^2\Big|\gamma=\mathbf{v}\right] = \xi_{\mathbf{v}}$$

Consider a random vector $\tilde{\mathbf{x}}^{\mathcal{N}}$ such that $\tilde{\mathbf{x}}^{\mathcal{B}_{\mathbf{v}}}$ has the same distribution as $\mathbf{x}^{\mathcal{B}_{\mathbf{v}}}$ conditioned on $\{\gamma=\mathbf{v}\}$, so that $I(\tilde{\mathbf{x}}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}) = I(\mathbf{x}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\gamma=\mathbf{v})$. We also require that $\|\tilde{\mathbf{x}}[n]\| \geq \mathsf{SNR}^{-1/2}$ for $n\in\mathcal{A}_{\mathbf{v}}$. Finally, since $\tilde{\mathbf{x}}^{\mathcal{B}_{\mathbf{v}}}$ must have the same total power as $\mathbf{x}^{\mathcal{B}_{\mathbf{v}}}$, we require that

$$\frac{1}{B}\sum_{n=1}^{B} E\left[\|\tilde{\mathbf{x}}[n]\|^2\right] \leq \xi_{\mathbf{v}} + \mathsf{SNR}^{-1}$$

Using the chain rule for mutual information we can write:

$$I(\tilde{\mathbf{x}}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}}) = I(\tilde{\mathbf{x}}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}) + I(\tilde{\mathbf{x}}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{A}_{\mathbf{v}}}|\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}) + I(\tilde{\mathbf{x}}^{\mathcal{A}_{\mathbf{v}}};\mathbf{y}^{\mathcal{N}}|\tilde{\mathbf{x}}^{\mathcal{B}_{\mathbf{v}}})$$

The non-negativity of mutual information implies that $I(\tilde{\mathbf{x}}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}) \leq I(\tilde{\mathbf{x}}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}})$, so defining

$$f(\mathsf{SNR},\epsilon,\xi_{\mathbf{v}}) = \lim_{B\to\infty} \sup_{\substack{p(\mathbf{x}^{\mathcal{N}}) \\ (1/B)\sum_{n\in\mathcal{N}} E[\|\mathbf{x}[n]\|^2]\leq\xi_{\mathbf{v}}+1/\mathsf{SNR} \\ \|\mathbf{x}[n]\|\geq\mathsf{SNR}^{-1/2},n\in\mathcal{N}}} \frac{1}{B} I(\mathbf{x}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}})$$

we conclude that

$$\lim_{B\to\infty}\frac{1}{B}I(\mathbf{x}^{\mathcal{B}_{\mathbf{v}}};\mathbf{y}^{\mathcal{B}_{\mathbf{v}}}|\gamma=\mathbf{v}) \leq \lim_{B\to\infty}\frac{1}{B}I(\tilde{\mathbf{x}}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}}) \leq f(\mathsf{SNR},\epsilon,\xi_{\mathbf{v}}) \tag{43}$$

In summary, putting equations (36),(37),(40) and (43) together we have:

$$\lim_{B\to\infty}\frac{1}{B}I(\mathbf{x}^{\mathcal{N}};\mathbf{y}^{\mathcal{N}}) \leq (2n_r+1)\log(2) + \sum_{\mathbf{v}\in\mathcal{G}} f(\mathsf{SNR},\epsilon,\xi_{\mathbf{v}})P(\gamma=\mathbf{v}) \tag{44}$$

Finally to upper bound the right hand side of (44) we prove that $f(\mathsf{SNR},\epsilon,\xi)$ is a concave function of ξ and use Jensen's inequality:

Lemma 10 *$f(\mathsf{SNR},\epsilon,\xi)$ is a concave function of ξ, that is, for any $\xi_1,\xi_2\geq 0$ and $\lambda\in[0,1]$:*

$$\lambda f(\mathsf{SNR},\epsilon,\xi_1) + (1-\lambda)f(\mathsf{SNR},\epsilon,\xi_2) \leq f(\mathsf{SNR},\epsilon,\lambda\xi_1+(1-\lambda)\xi_2)$$

Proof: Assume that we have two independent channels to communicate information, indexed by i, $i=1,2$. Let n_i be the block length used in channel i, and assume that $n_1/B=\lambda$, where $B=n_1+n_2$. Also let $\mathcal{N}_i=\{1,2,\ldots,n_i\}$. We require that the input signals satisfy the power constraints:

$$\sum_{n=1}^{n_i} E\left[\|\mathbf{x}_i[n]\|^2\right] \leq n_i\xi_i + n_i\mathsf{SNR}^{-1}$$

and

$$\|\mathbf{x}_i[n]\| \geq \mathsf{SNR}^{-1} \text{ for } n \in \mathcal{N}_i$$

Then by the independence of the channels we have:

$$\begin{aligned}
\frac{1}{B} I\left(\mathbf{x}_1^{\mathcal{N}_1}, \mathbf{x}_2^{\mathcal{N}_2}; \mathbf{y}_1^{\mathcal{N}_1}, \mathbf{y}_2^{\mathcal{N}_2}\right) &= \frac{1}{B}\left[h\left(\mathbf{y}_1^{\mathcal{N}_1}, \mathbf{y}_2^{\mathcal{N}_2}\right) - h\left(\mathbf{y}_1^{\mathcal{N}_1}, \mathbf{y}_2^{\mathcal{N}_2} \middle| \mathbf{x}_1^{\mathcal{N}_1}, \mathbf{x}_2^{\mathcal{N}_2}\right)\right] \\
&\leq \frac{1}{B}\left[h\left(\mathbf{y}_1^{\mathcal{N}_1}\right) + h\left(\mathbf{y}_2^{\mathcal{N}_2}\right) - h\left(\mathbf{y}_1^{\mathcal{N}_1} \middle| \mathbf{x}_1^{\mathcal{N}_1}, \mathbf{x}_2^{\mathcal{N}_2}\right) - h\left(\mathbf{y}_2^{\mathcal{N}_2} \middle| \mathbf{x}_1^{\mathcal{N}_1}, \mathbf{x}_2^{\mathcal{N}_2}\right)\right] \\
&= \frac{1}{B}\left[h\left(\mathbf{y}_1^{\mathcal{N}_1}\right) + h\left(\mathbf{y}_2^{\mathcal{N}_2}\right) - h\left(\mathbf{y}_1^{\mathcal{N}_1} \middle| \mathbf{x}_1^{\mathcal{N}_1}\right) - h\left(\mathbf{y}_2^{\mathcal{N}_2} \middle| \mathbf{x}_2^{\mathcal{N}_2}\right)\right] \\
&= \frac{1}{B}\left[I(\mathbf{x}_1^{\mathcal{N}_1}; \mathbf{y}_1^{\mathcal{N}_1}) + I(\mathbf{x}_2^{\mathcal{N}_2}; \mathbf{y}_2^{\mathcal{N}_2})\right] \\
&= \lambda \frac{1}{n_1} I(\mathbf{x}_1^{\mathcal{N}_1}; \mathbf{y}_1^{\mathcal{N}_1}) + (1-\lambda)\frac{1}{n_2} I(\mathbf{x}_2^{\mathcal{N}_2}; \mathbf{y}_2^{\mathcal{N}_2})
\end{aligned}$$

where we can achieve equality by choosing $\mathbf{x}_1^{\mathcal{N}_1}$ and $\mathbf{x}_2^{\mathcal{N}_2}$ independent. It follows that

$$\lim_{B\to\infty} \sup_{p(\mathbf{x}_1^{\mathcal{N}_1}), p(\mathbf{x}_2^{\mathcal{N}_2})} \frac{1}{B} I\left(\mathbf{x}_1^{\mathcal{N}_1}, \mathbf{x}_2^{\mathcal{N}_2}; \mathbf{y}_1^{\mathcal{N}_1}, \mathbf{y}_2^{\mathcal{N}_2}\right) = \lambda f(\mathsf{SNR}, \epsilon, \xi_1) + (1-\lambda) f(\mathsf{SNR}, \epsilon, \xi_2) \tag{45}$$

Now assume that instead of having two independent channels we have only one channel, and we use it with block size B by feeding it with the family of inputs $(\mathbf{x}_1^{\mathcal{N}_1}, \mathbf{x}_2^{\mathcal{N}_2})$ that achieves (45) as $B \to \infty$. This can be seen as having two channels with some correlation, which becomes negligible as $B \to \infty$ as long as $\epsilon > 0$. The corresponding mutual information is also given by (45). This input satisfies the power constraints:

$$\sum_{i=1}^{2} \sum_{n=1}^{n_i} E\left[\|\mathbf{x}_i[n]\|^2\right] \leq n_1\xi_1 + n_2\xi_2 + B\mathsf{SNR}^{-1} = B[\lambda\xi_1 + (1-\lambda)\xi_2] + B\mathsf{SNR}^{-1}$$

and

$$\|\mathbf{x}_i[n]\| \geq \mathsf{SNR}^{-1} \text{ for } n \in \mathcal{N}_i,\ i = 1, 2$$

Since this is just a particular choice of input distribution that satisfies the power constraints and $f(\mathsf{SNR}, \epsilon, \lambda\xi_1 + (1-\lambda)\xi_2)$ corresponds to the supremum over all such distributions we have the desired result.

■

Using Jensen's inequality in (44) we obtain:

$$\begin{aligned}
\lim_{B\to\infty} \frac{1}{B} I(\mathbf{x}^{\mathcal{N}}; \mathbf{y}^{\mathcal{N}}) &\leq (2n_r + 1)\log(2) + f\left[\mathsf{SNR}, \epsilon, \sum_{\mathbf{v}\in\mathcal{G}} \xi_{\mathbf{v}} P(\gamma = \mathbf{v})\right] \\
&\leq (2n_r + 1)\log(2) + f(\mathsf{SNR}, \epsilon, n_t)
\end{aligned}$$

where we used (39) and the fact that $f(\mathsf{SNR}, \epsilon, \xi)$ is a non-decreasing function of ξ.

Appendix D

Proof of Lemma 2

We express the mutual information in terms of differential entropies:

$$I(\mathbf{x}^{\mathcal{N}}; \tilde{\mathbf{y}}^{\mathcal{N}}) = h(\tilde{\mathbf{y}}^{\mathcal{N}}) - h(\tilde{\mathbf{y}}^{\mathcal{N}}|\mathbf{x}^{\mathcal{N}}) \tag{46}$$

To write each of the differential entropies in terms of the new variables use the chain rule for differential entropies[6] and apply fact 1 to the transformation $(\rho_i[n], \phi_i[n]) \to \tilde{y}_i$ which has Jacobian with determinant $|\tilde{y}_i|^2$:

$$\begin{aligned}
h(\tilde{\mathbf{y}}^{\mathcal{N}}) &= \sum_{n\in\mathcal{N}} h(\tilde{\mathbf{y}}[n] \mid \tilde{\mathbf{y}}[n-1], \cdots, \tilde{\mathbf{y}}[1]) \\
&= \sum_{n\in\mathcal{N}} \sum_{i=1}^{n_r} h(\tilde{y}_i[n] \mid \tilde{y}_{i-1}[n], \cdots, \tilde{y}_1[n], \tilde{\mathbf{y}}[n-1], \cdots, \tilde{\mathbf{y}}[1]) \\
&= \sum_{n\in\mathcal{N}} \sum_{i=1}^{n_r} E\left[\log(|\tilde{y}_i[n]|^2)\right] \\
&+ \sum_{n\in\mathcal{N}} \sum_{i=1}^{n_r} h(\rho_i[n], \phi_i[n] \mid \tilde{y}_{i-1}[n], \cdots, \tilde{y}_1[n], \tilde{\mathbf{y}}[n-1], \cdots, \tilde{\mathbf{y}}[1]) \\
&= \sum_{n\in\mathcal{N}} \sum_{i=1}^{n_r} E\left[\log(|\tilde{y}_i[n]|^2)\right] + h(\rho^{\mathcal{N}}, \phi^{\mathcal{N}}) \\
&= \sum_{n\in\mathcal{N}} \sum_{i=1}^{n_r} E\left[\log(|c_i[n]|^2)\right] + n_r \sum_{n\in\mathcal{N}} E\left[\log(m[n]^2)\right] + h(\rho^{\mathcal{N}}, \phi^{\mathcal{N}}) \qquad (47)
\end{aligned}$$

where the last equality follows from $\tilde{y}_i[n] = c_i[n]m[n]$.

We also rewrite the second term of (46) in terms of the new variables:

$$\begin{aligned}
h(\tilde{\mathbf{y}}^{\mathcal{N}}|\tilde{\mathbf{x}}^{\mathcal{N}}) &= h(\tilde{\mathbf{y}}^{\mathcal{N}}|\mathbf{m}^{\mathcal{N}}, \mathbf{d}^{\mathcal{N}}) \\
&= \sum_{n\in\mathcal{N}} \sum_{i=1}^{n_r} h(\tilde{y}_i[n] \mid \tilde{y}_{i-1}[n], \cdots, \tilde{y}_1[n], \tilde{\mathbf{y}}[n-1], \cdots, \tilde{\mathbf{y}}[1], \mathbf{m}^{\mathcal{N}}, \mathbf{d}^{\mathcal{N}}) \\
&= \sum_{n\in\mathcal{N}} \sum_{i=1}^{n_r} h((c_i[n]m[n]) \mid \tilde{y}_{i-1}[n], \cdots, \tilde{y}_1[n], \tilde{\mathbf{y}}[n-1], \cdots, \tilde{\mathbf{y}}[1], \mathbf{m}^{\mathcal{N}}, \mathbf{d}^{\mathcal{N}}) \\
&= \sum_{n\in\mathcal{N}} \sum_{i=1}^{n_r} h((c_i[n]m[n]) \mid c_{i-1}[n], \cdots, c_1[n], \mathbf{c}[n-1], \cdots, \mathbf{c}[1], \mathbf{m}^{\mathcal{N}}, \mathbf{d}^{\mathcal{N}}) \\
&= n_r \sum_{n\in\mathcal{N}} E\left[\log(m[n]^2)\right] \\
&+ \sum_{n\in\mathcal{N}} \sum_{i=1}^{n_r} h(c_i[n] \mid c_{i-1}[n], \cdots, c_1[n], \mathbf{c}[n-1], \cdots, \mathbf{c}[1], \mathbf{m}^{\mathcal{N}}, \mathbf{d}^{\mathcal{N}})
\end{aligned}$$

[6]We abused notation slightly in the use of the chain rule to make the expressions more readable. To be precise the first term in each sum is an unconditional differential entropy.

$$= \quad n_r \sum_{n\in\mathcal{N}} E\left[\log(m[n]^2)\right] + h(\mathbf{c}^{\mathcal{N}}|\mathbf{d}^{\mathcal{N}}) \tag{48}$$

where we used the conditional independence between $\mathbf{m}^{\mathcal{N}}$ and $\mathbf{c}^{\mathcal{N}}$ conditioned on $\mathbf{d}^{\mathcal{N}}$ in the last equality.

The result follows by subtracting (48) from (47).

Appendix E

Proof of Lemma 5

$\rho^{\mathcal{N}}$ can be thought of as having B vector components of dimension n_r. The n_r elements of the component corresponding to time n depend on the magnitude of the input $\|\mathbf{x}[n]\|$, so they are highly correlated. Eventually we will use the independent Gaussian bound to upper bound $h(\rho^{\mathcal{N}}, \phi^{\mathcal{N}})$ so we will try to remove as much correlation as possible to obtain a tight bound. For this we introduce a change of variables that has Jacobian with determinant equal to 1 and hence does not modify the differential entropy:

$$\begin{aligned} h(\rho^{\mathcal{N}}, \phi^{\mathcal{N}}) &= h(\rho_1^{\mathcal{N}}, \phi_1^{\mathcal{N}}, \rho_2^{\mathcal{N}}, \phi_2^{\mathcal{N}}, \cdots, \rho_{n_r}^{\mathcal{N}}, \phi_{n_r}^{\mathcal{N}}) \\ &= h(\rho_1^{\mathcal{N}}, \phi_1^{\mathcal{N}}, (\rho_2^{\mathcal{N}} - \rho_1^{\mathcal{N}}), (\phi_2^{\mathcal{N}} - \phi_1^{\mathcal{N}}), \cdots, (\rho_{n_r}^{\mathcal{N}} - \rho_1^{\mathcal{N}}), (\phi_{n_r}^{\mathcal{N}} - \phi_1^{\mathcal{N}})) \end{aligned}$$

This transformation eliminates the dependence on $\|\mathbf{x}[n]\|$ in all but the components corresponding to the first receive antenna. We then use the chain rule for differential entropies and remove conditioning to get the upper bound:

$$\begin{aligned} h(\rho^{\mathcal{N}}, \phi^{\mathcal{N}}) &\leq h(\rho_1^{\mathcal{N}}, \phi_1^{\mathcal{N}}) + \sum_{i=2}^{n_r} h((\rho_i^{\mathcal{N}} - \rho_1^{\mathcal{N}}), (\phi_i^{\mathcal{N}} - \phi_1^{\mathcal{N}})) \\ &\leq h(\rho_1^{\mathcal{N}}) + h(\phi_1^{\mathcal{N}}) + \sum_{i=2}^{n_r} h((\rho_i^{\mathcal{N}} - \rho_1^{\mathcal{N}})) \\ &+ \sum_{i=2}^{n_r} h((\phi_i^{\mathcal{N}} - \phi_1^{\mathcal{N}})) \end{aligned} \tag{49}$$

Most of these terms can be easily bounded by noting that $-\pi \leq \phi_1[n] \leq \pi$ and $-2\pi \leq \phi_i[n] - \phi_1[n] \leq 2\pi$ and hence using the Gaussian bound $h(\phi_1^{\mathcal{N}}) \leq \frac{B}{2}\log(2\pi e \pi^2)$ and $h((\phi_i^{\mathcal{N}} - \phi_1^{\mathcal{N}})) \leq \frac{B}{2}\log(2\pi e(2\pi)^2)$.

Using the chain rule and removing conditioning again we can write:

$$\sum_{i=2}^{n_r} h((\rho_i^{\mathcal{N}} - \rho_1^{\mathcal{N}})) \leq \sum_{n\in\mathcal{N}} \sum_{i=2}^{n_r} h((\rho_i[n] - \rho_1[n])) \tag{50}$$

Reminding the definition of $\rho_i[n]$, we can write $(\rho_i[n] - \rho_1[n]) = \frac{1}{2}\log(|c_i[n]|^2/|c_1[n]|^2)$. Conditioned on $\mathbf{d}[n]$ the independence between the rows of $\mathbf{H}[n]$ implies the independence between

$c_i[n]$ and $c_1[n]$ for $i = 2, \ldots, n_r$. Also since $\mathbf{d}[n]$ has norm 1, and the rows of $\mathbf{H}[n]$ have i.i.d. $\mathcal{CN}(0,1)$ components, it follows that $c_i[n] \sim \mathcal{CN}(0,1)$ for every possible value of $\mathbf{d}[n]$. Then, conditioned on $\mathbf{d}[n]$, $|c_i[n]|^2/|c_1[n]|^2$ is the ratio of 2 independent Exp(1) random variables, and the corresponding conditional density does not depend on $\mathbf{d}[n]$, so it follows that the unconditional distribution of $|c_i[n]|^2/|c_1[n]|^2$ is also that of the ratio of 2 independent Exp(1) random variables. The corresponding density is given by $f_T(t) = 1/(t+1)^2$ for $t \geq 0$. Introducing the change of variables $L = \log(T)$ where T has the density found before, we obtain that the density of L is $f_L(l) = \frac{e^l}{(1+e^l)^2}$ for $l \in \mathcal{R}$. Then we can compute the right hand side of (50) explicitly:

$$\begin{aligned} \sum_{n=\in\mathcal{N}} \sum_{i=2}^{n_r} h((\rho_i[n] - \rho_1[n])) &= (n_r - 1)B \left\{ \log\left(\frac{1}{2}\right) + \int_{-\infty}^{\infty} \frac{e^l}{(1+e^l)^2} \log\left[\frac{(1+e^l)^2}{e^l}\right] dl \right\} \\ &= (n_r - 1)B[\log(1/2) + 2] \end{aligned} \tag{51}$$

It only remains to find an upper bound for the first term of (49), $h(\rho_1^{\mathcal{N}})$. For this we use the chain rule for differential entropies, remove conditioning and use the Gaussian upper bound for the differential entropy of a random variable with a given second order moment:

$$h(\rho_1^{\mathcal{N}}) \leq \sum_{n\in\mathcal{N}} h(\rho_1[n]) \leq \sum_{n\in\mathcal{N}} \frac{1}{2}\log(2\pi e\sigma_n^2) \leq \frac{B}{2}\log\left(2\pi e \frac{1}{B}\sum_{n\in\mathcal{N}} \sigma_n^2\right) \tag{52}$$

where we used Jensen's inequality and σ_n^2 is some upper bound for $E[(\rho_1[n])^2]$.

We use the triangle inequality to calculate some upper bound σ_n^2:

$$\begin{aligned} \rho_1[n] &= \log|c_1[n]| + \log(m[n]) \\ \Rightarrow \sqrt{E[(\rho_1[n])^2]} &\leq \sqrt{E[(\log|c_1[n]|)^2]} + \sqrt{E[(\log m[n])^2]} \end{aligned} \tag{53}$$

The second term of (53) can be bounded by conditioning on the event $\Psi = \{m[n] > 1\}$ and noting that on Ψ, $(\log m[n])^2 \leq [m[n]-1]^2$:

$$\begin{aligned} E[(\log m[n])^2] &= E[(\log m[n])^2|\Psi]P(\Psi) + E[(\log m[n])^2|\Psi^c]P(\Psi^c) \\ &\leq E[(m[n]-1)^2|\Psi]P(\Psi) + [\log(1/\sqrt{\mathsf{SNR}})]^2 \cdot 1 \\ &\leq E[m[n]^2|\Psi]P(\Psi) + 1 + [\log(\sqrt{\mathsf{SNR}})]^2 \\ \theta_n^2 &\stackrel{\text{def}}{=} E[\|\mathbf{x}[n]\|^2] = E[m[n]^2] \geq E[m[n]^2|\Psi]P(\Psi) \\ \Rightarrow E[(\log m[n])^2] &\leq \theta_n^2 + 1 + [\log(\sqrt{\mathsf{SNR}})]^2 \end{aligned}$$

Noting that $c_1[n] \sim \mathcal{CN}(0,1)$ the first term of (53) can be computed by numerical integration:

$$E[(\log|c_1[n]|)^2] = E\left\{\frac{1}{4}\left[\log(|c_1[n]|^2)\right]^2\right\} = \frac{1}{4}\int_0^\infty [\log(x)]^2 e^{-x} dx = 0.494528\ldots < 1/2$$

As a result, we obtain

$$\sigma_n^2 = \left\{\sqrt{\theta_n^2 + 1 + [\log(\sqrt{\mathsf{SNR}})]^2} + \sqrt{1/2}\right\}^2 = f(\theta_n^2)$$

It is easy to check that $f(\theta_n^2)$ is an increasing concave function of θ_n^2, so we can use Jensen's inequality to upper bound $\frac{1}{B}\sum_{n\in\mathcal{N}}\sigma_n^2$:

$$\begin{aligned}\frac{1}{B}\sum_{n\in\mathcal{N}}\sigma_n^2 &= \frac{1}{B}\sum_{n\in\mathcal{N}} f(\theta_n^2) \le f\left(\frac{1}{B}\sum_{n\in\mathcal{N}}\theta_n^2\right)\\ &\le f\left(n_t + \frac{1}{\mathsf{SNR}}\right) = \left\{\sqrt{n_t + \frac{1}{\mathsf{SNR}} + 1 + [\log(\sqrt{\mathsf{SNR}})]^2} + \sqrt{1/2}\right\}^2\end{aligned}$$

Here we used the fact that $\frac{1}{B}\sum_{n\in\mathcal{N}}\theta_n^2 = \frac{1}{B}\sum_{n\in\mathcal{N}} E[\|\mathbf{x}[n]\|^2] \le n_t + \frac{1}{\mathsf{SNR}}$ by the power constraint (17).

Therefore, we can upper bound (52) by:

$$h(\rho_1^{\mathcal{N}}) \le \frac{B}{2}\log\left(2\pi e\sigma^2\right)$$

where $\sigma^2 = \left(\sqrt{n_t + \frac{1}{\mathsf{SNR}} + 1 + [\log(\sqrt{\mathsf{SNR}})]^2} + \sqrt{1/2}\right)^2$.

Replacing all the bounds that we found in (49) we obtain the statement of the lemma.

Appendix F

Proof of Lemma 6

As in the proof of lemma 5 we introduce a change of variables whose Jacobian has determinant 1 and hence does not modify the differential entropy, and apply the chain rule removing conditioning:

$$\begin{aligned}h(\rho^{\mathcal{N}},\phi^{\mathcal{N}}) &= h(\rho_1^{\mathcal{N}},\phi_1^{\mathcal{N}},\cdots,\rho_{n_r}^{\mathcal{N}},\phi_{n_r}^{\mathcal{N}})\\ &= h(\rho_1^{\mathcal{N}},\phi_1^{\mathcal{N}},(\rho_2^{\mathcal{N}}-\rho_1^{\mathcal{N}}),(\phi_2^{\mathcal{N}}-\phi_1^{\mathcal{N}}),\cdots,(\rho_{n_r}^{\mathcal{N}}-\rho_1^{\mathcal{N}}),(\phi_{n_r}^{\mathcal{N}}-\phi_1^{\mathcal{N}}))\\ &\le h(\rho_1^{\mathcal{N}}) + h(\phi_1^{\mathcal{N}}) + h((\rho_2^{\mathcal{N}}-\rho_1^{\mathcal{N}}),(\phi_2^{\mathcal{N}}-\phi_1^{\mathcal{N}}),\cdots,(\rho_{n_r}^{\mathcal{N}}-\rho_1^{\mathcal{N}}),(\phi_{n_r}^{\mathcal{N}}-\phi_1^{\mathcal{N}}))\end{aligned} \quad (54)$$

The first two terms of (54) can be bounded in the same way as the corresponding terms that appeared in the proof of lemma 5[7] by simply taking $n_t = 1$:

$$h(\rho_1^{\mathcal{N}}) + h(\phi_1^{\mathcal{N}}) \le \frac{B}{2}\log\left(2\pi e\sigma^2\right) + \frac{B}{2}\log(2\pi e\pi^2) \quad (55)$$

where $\sigma^2 = \left(\sqrt{\frac{1}{\mathsf{SNR}} + 2 + [\log(\sqrt{\mathsf{SNR}})]^2} + \sqrt{1/2}\right)^2$.

Recalling previous definitions we can express $\rho_i[n]-\rho_1[n]$ and $\phi_i[n]-\phi_1[n]$ in terms of $\tilde{y}_i[n]/\tilde{y}_1[n]$ for $n\in\mathcal{N}$ and $i = 2,\ldots,n_r$:

$$\rho_i[n]-\rho_1[n] = \log(|\tilde{y}_i[n]|) - \log(|\tilde{y}_1[n]|) = \log\left(\left|\frac{\tilde{y}_i[n]}{\tilde{y}_1[n]}\right|\right)$$

[7] See Appendix E for the details.

$$\phi_i[n] - \phi_1[n] = \angle \tilde{y}_i[n] - \angle \tilde{y}_1[n] = \angle\left(\frac{\tilde{y}_i[n]}{\tilde{y}_1[n]}\right)$$

To compute the third term of (54) we invert the transformation $(\log(|v|), \angle v) \to v$ which has Jacobian with determinant $|v|^2$, define $r_i[n] = \tilde{y}_i[n]/\tilde{y}_1[n] = h_i[n]/h_1[n]$, where $h_i[n]$ is the ith entry of $\mathbf{H}[n]$, and apply fact 1:

$$\begin{aligned} h((\rho_2^{\mathcal{N}} - \rho_1^{\mathcal{N}}), (\phi_2^{\mathcal{N}} - \phi_1^{\mathcal{N}}), \cdots, (\rho_{n_r}^{\mathcal{N}} - \rho_1^{\mathcal{N}}), (\phi_{n_r}^{\mathcal{N}} - \phi_1^{\mathcal{N}})) &= h(\mathbf{r}_2^{\mathcal{N}}, \cdots, \mathbf{r}_{n_r}^{\mathcal{N}}) - \sum_{n \in \mathcal{N}} \sum_{i=2}^{n_r} E\left[\log\left(|r_i[n]|^2\right)\right] \\ &= h(\mathbf{r}_2^{\mathcal{N}}, \cdots, \mathbf{r}_{n_r}^{\mathcal{N}}) \end{aligned} \quad (56)$$

where the last equality results from noting that $E\left[\log\left(|r_i[n]|^2\right)\right] = E\left[\log\left(|h_i[n]|^2\right)\right] - E\left[\log\left(|h_1[n]|^2\right)\right] = 0$ since $h_i[n]$ and $h_1[n]$ have the same distribution for $n \in \mathcal{N}$ and $i = 2, \ldots, n_r$, and each of the expectations in the subtraction is finite.

We can upper bound (56) by using chain rule, removing conditioning, and noting that $\{r_i[n]\}_{i=2}^{n_r}$ are identically distributed:

$$h(\mathbf{r}_2^{\mathcal{N}}, \cdots, \mathbf{r}_{n_r}^{\mathcal{N}}) \le \sum_{i=2}^{n_r} h(\mathbf{r}_i^{\mathcal{N}}) = (n_r - 1) h(\mathbf{r}_2^{\mathcal{N}}) \quad (57)$$

We provide an asymptotic upper bound for $h(\mathbf{r}_2^{\mathcal{N}})$ in the limit as $\epsilon \to 0$:

Lemma 11

$$\lim_{\substack{\epsilon \to 0 \\ B \to \infty}} \left\{\frac{1}{B} h(\mathbf{r}_2^{\mathcal{N}}) - \log(\epsilon) - 3.97722\right\} \le 0$$

Proof: We will provide a heuristic proof that can be formalized by proving a technical issue of convergence of differential entropies.

We upper bound $h(\mathbf{r}_2^{\mathcal{N}})$ using the chain rule for differential entropies and removing conditioning:

$$\begin{aligned} h(\mathbf{r}_2^{\mathcal{N}}) &= h(r_2[1]) + \sum_{n=2}^{B} h(r_2[n] | r_2[n-1], \cdots, r_2[1]) \\ &\le h(r_2[1]) + \sum_{n=2}^{B} h(r_2[n] | r_2[n-1]) \\ &= h(r_2[1]) + (B-1) h(r_2[2] | r_2[1]) \end{aligned} \quad (58)$$

Recalling equation (2) for the Gauss-Markov process $\{\mathbf{H}[n]\}_{n \in \mathcal{N}}$, and letting $w_i[1]$ be the ith component of $\mathbf{W}[1]$ we can write:

$$r_2[2] = \frac{\sqrt{1-\epsilon} h_2[1] + \sqrt{\epsilon} w_2[1]}{\sqrt{1-\epsilon} h_1[1] + \sqrt{\epsilon} w_1[1]} = \frac{h_2[1]}{h_1[1]} \frac{\left(1 + \frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}} \frac{\tilde{w}_2[1]}{|h_2[1]|}\right)}{\left(1 + \frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}} \frac{\tilde{w}_1[1]}{|h_1[1]|}\right)} \quad (59)$$

$$\approx \quad r_2[1]\left(1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{\tilde{w}_2[1]}{|h_2[1]|}\right)\left(1-\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{\tilde{w}_1[1]}{|h_1[1]|}\right)$$

$$\approx \quad r_2[1]\left[1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\left(\frac{\tilde{w}_2[1]}{|h_2[1]|}-\frac{\tilde{w}_1[1]}{|h_1[1]|}\right)\right]$$

$$= \quad r_2[1]\left[1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{1}{|h_2[1]|}\left(\tilde{w}_2[1]-|r_2[1]|\tilde{w}_1[1]\right)\right] \tag{60}$$

where $\approx$ means that difference between the two sides, conditioned on $r_2[1]$, goes to zero with probability 1 as $\epsilon \to 0$. In (59) we define $\tilde{w}_i[1] = (|h_i[1]|/h_i[1])w_i[1]$, $i = 1, 2$, and note that the circular symmetry of $w_i[1]$ and the independence of $h_1[1]$ and $h_2[1]$ imply $\tilde{w}_i[1] \sim \mathcal{CN}(0, 1)$ independent across i.

We want to use the approximation (60) to approximate $h(r_2[2]|r_2[1])$ for small ϵ. Under some technical conditions that we will not verify here, convergence in distribution of random variables implies the convergence of the corresponding differential entropies. Assuming that these conditions hold for this particular case we can write as $\epsilon \to 0$:

$$h(r_2[2]|r_2[1]) \quad \approx \quad h\left(r_2[1]\left[1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{1}{|h_2[1]|}\left(\tilde{w}_2[1]-|r_2[1]|\tilde{w}_1[1]\right)\right]\middle| r_2[1]\right)$$

$$= \quad E\left[\log(|r_2[1]|^2)\right] + h\left(1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{1}{|h_2[1]|}\left(\tilde{w}_2[1]-|r_2[1]|\tilde{w}_1[1]\right)\middle| r_2[1]\right)$$

$$= \quad \log\left(\frac{\epsilon}{1-\epsilon}\right) + h\left(\frac{1}{|h_2[1]|}\left(\tilde{w}_2[1]-|r_2[1]|\tilde{w}_1[1]\right)\middle| r_2[1]\right) \tag{61}$$

where we used fact 1 in the second and third lines, and as before, $E\left[\log(|r_2[1]|^2)\right] = 0$.

To compute the second term of (61) we note that the real and imaginary parts of $\frac{1}{|h_2[1]|}\left(\tilde{w}_2[1]-|r_2[1]|\tilde{w}_1[1]\right)$ are i.i.d. (with and without conditioning on $r_2[1]$), and they depend on $r_2[1]$ only through its magnitude. Then we have:

$$h\left(\frac{1}{|h_2[1]|}\left(\tilde{w}_2[1]-|r_2[1]|\tilde{w}_1[1]\right)\middle| r_2[1]\right) = 2\cdot h\left(\frac{1}{|h_2[1]|}\left(\Re\{\tilde{w}_2[1]\}-|r_2[1]|\Re\{\tilde{w}_1[1]\}\right)\middle| |r_2[1]|\right)$$

To compute the last differential entropy let $V = \frac{1}{|h_2[1]|}\left(\Re\{\tilde{w}_2[1]\}-|r_2[1]|\Re\{\tilde{w}_1[1]\}\right)$ and $W = |r_2[1]|$. Then it can be shown by doing a change of variables that

$$f_{V|W}(v|w) \quad = \quad \frac{3w(w^2+1)^4}{4\left[(1+w^2)^2+v^2w^2\right]^{5/2}} \tag{62}$$

$$f_W(w) \quad = \quad \frac{2w}{(w^2+1)^2} \tag{63}$$

for $v \in (-\infty, \infty)$ and $w \in (0, \infty)$. Then,

$$h(V|W) \quad = \quad -\int_0^\infty f_W(w)\int_{-\infty}^\infty f_{V|W}(v|w)\log[f_{V|W}(v|w)]dvdw = 1.98861... \tag{64}$$

where the last integral was solved numerically.

As a result we have that for small ϵ:

$$h(r_2[2] \mid r_2[1]) \approx \log(\epsilon) + 3.97722$$

As long as $h(r_2[1])$ is finite, the first term of (58) becomes negligible compared to the second term as $B \to \infty$. In fact one can check that $h(r_2[1]) \leq 2 - \log(2) + (1/2)\log[2\pi e \pi^2]$, so that the first term of (58) vanishes when dividing by B and letting $B \to \infty$. Therefore dividing the RHS of (58) by B and letting $B \to \infty$ we obtain the statement of the lemma. ■

Using this lemma in (57), and (55) in (54) we conclude the proof.

Appendix G

Proof of Lemma 1

To compute $E\left[\mathbf{\Delta}[n]\mathbf{\Delta}[n]^\dagger\right]$ we condition on the previously transmitted symbols $\{\mathbf{x}[i]\}_{(i<n)}$ and recall a remark that we made previously: the estimation of the different rows of $\mathbf{H}$ decouples into n_r independent and identical problems, and hence the estimation errors of the different rows, conditioned on $\{\mathbf{x}[i]\}_{(i<n)}$ are independent. We also noted that the estimation error covariances corresponding to different rows are identical. Then,

$$E\left[\mathbf{\Delta}[n]\mathbf{\Delta}[n]^\dagger\right] = E\left\{E\left[\mathbf{\Delta}[n]\mathbf{\Delta}[n]^\dagger \middle| \mathbf{x}[i] : i < n\right]\right\} = E\left[\mathrm{tr}(\mathbf{K}[n])\right]\mathbf{I}_{n_r}$$

where $\mathbf{K}[n]$ is the conditional covariance of the first row of $\mathbf{\Delta}[n]$, conditioned on $\{\mathbf{x}[i]\}_{(i<n)}$.

Using (14) we could compute $E\left[\mathrm{tr}(\mathbf{K}[n])\right]$ numerically and recursively, starting with $\mathbf{K}[n_t+1] = n_t(\mathsf{SNR}^{-1}+\epsilon)\mathbf{I}_{n_t}$. However, we can obtain an explicit expression by noting that (14) eventually reaches a steady state, and the transient behavior becomes irrelevant in the limit as $M \to \infty$. Taking traces and expectations in (14) we have:

$$\begin{aligned} E\left[\mathrm{tr}(\mathbf{K}[n+1])\right] \quad &= \quad (1-\epsilon)\left\{E\left[\mathrm{tr}(\mathbf{K}[n])\right] - E\left[\left(\mathbf{x}[n]^T\mathbf{K}[n]\mathbf{x}[n]^* + \frac{n_t}{\mathsf{SNR}}\right)^{-1}\mathrm{tr}\left(\mathbf{K}[n]\mathbf{x}[n]^*\mathbf{x}[n]^T\mathbf{K}[n]\right)\right]\right\} \\ &\quad +n_t\epsilon \end{aligned} \tag{65}$$

We will upper bound the RHS of the above equality with an expression that depends on $E[\mathrm{tr}(\mathbf{K}[n])]$, and then find the maximum value of $E[\mathrm{tr}(\mathbf{K}[n])]$ that satisfies the inequality in steady state. Let λ_i by the ith eigenvalue of $\mathbf{K}[n]$, and $\lambda_{max} = \max_i\{\lambda_i\}$. Then

$$\left(\mathbf{x}[n]^T\mathbf{K}[n]\mathbf{x}[n]^* + \frac{n_t}{\mathsf{SNR}}\right)^{-1} \geq \left(\lambda_{max}|\mathbf{x}[n]\|^2 + \frac{n_t}{\mathsf{SNR}}\right)^{-1}$$

and letting $\mathbf{K}[n] = \mathbf{Q\Lambda Q}^\dagger$ be the eigenvalue decomposition of $\mathbf{K}[n]$ we have

$$\begin{aligned} \mathrm{tr}\left(\mathbf{K}[n]\mathbf{x}[n]^*\mathbf{x}[n]^T\mathbf{K}[n]\right) \quad &= \quad \mathrm{tr}\left(\mathbf{Q\Lambda Q}^\dagger\mathbf{x}[n]^*\mathbf{x}[n]^T\mathbf{Q\Lambda Q}^\dagger\right) = \mathrm{tr}\left(\mathbf{\Lambda\Lambda}\tilde{\mathbf{x}}[n]^*\tilde{\mathbf{x}}[n]^T\right) \\ &= \quad \sum_{i=1}^{n_t}\lambda_i^2|\tilde{x}_i[n]|^2 \geq \lambda_{max}^2|\tilde{x}_i[n]|^2 \end{aligned}$$

for some $i \in \{1, \ldots, n_t\}$ where $\tilde{\mathbf{x}}[n] = \mathbf{Q}^\dagger \mathbf{x}[n]$ and $\tilde{x}_i[n]$ is the ith component of $\tilde{\mathbf{x}}[n]$.

Noting that $\{\tilde{x}_i[n]\}_{i=1}^{n_t}$ have the same distribution as $x_1[n]$ and are independent of λ_{max} we can upper bound the RHS of (65) as follows:

$$E\left[\mathrm{tr}(\mathbf{K}[n+1])\right] \leq (1-\epsilon)\left\{E\left[\mathrm{tr}(\mathbf{K}[n])\right] - E\left[\frac{\lambda_{max}|x_1[n]|^2}{\|\mathbf{x}[n]\|^2 + \frac{n_t}{\lambda_{max}\mathsf{SNR}}}\right]\right\} + n_t\epsilon \tag{66}$$

We bound $E\left[\frac{\lambda_{max}|x_1[n]|^2}{\|\mathbf{x}[n]\|^2 + \frac{n_t}{\lambda_{max}\mathsf{SNR}}}\right]$ using the following lemma.

Lemma 12

$$E\left[\frac{\lambda_{max}|x_1[n]|^2}{\|\mathbf{x}[n]\|^2 + \frac{n_t}{\lambda_{max}\mathsf{SNR}}}\right] \geq \frac{1}{2^{n_t}}\left[\frac{E[tr(\mathbf{K}[n])]}{n_t} - \frac{n_t}{\mathsf{SNR}}\right]$$

Proof: The cases $n_t = 1$, $n_t = 2$, and $n_t \geq 3$ require slightly different proofs, so we have to deal with each case separately. In the three cases we will use the inequalities: $\mathrm{tr}(\mathbf{K}[n]) = \sum_{i=1}^{n_t} \lambda_i \leq n_t \lambda_{max}$ and $e^{-x} \geq 1 - x$.

Consider first the case $n_t = 1$. In this case $\|\mathbf{x}[n]\|^2 = |x_1[n]|^2 \sim \mathrm{Exp}(1)$. Since $\mathbf{K}[n]$ is independent of $\mathbf{x}[n]$, conditioned on $\mathbf{K}[n]$, $|x_1[n]|^2$ is also Exp(1). Also $\mathbf{K}[n] = \lambda_{max}$ which is a scalar. Then we have:

$$\begin{aligned}
E\left[\frac{\lambda_{max}|x_1[n]|^2}{|x_1[n]|^2 + \frac{1}{\lambda_{max}\mathsf{SNR}}}\right] &= E\left\{E\left[\left.\frac{\lambda_{max}|x_1[n]|^2}{|x_1[n]|^2 + \frac{1}{\lambda_{max}\mathsf{SNR}}}\right|\mathbf{K}[n]\right]\right\} \\
&= E\left[\lambda_{max}\int_0^\infty \frac{x}{x + \frac{1}{\lambda_{max}\mathsf{SNR}}} e^{-x} dx\right] \\
&\geq E\left[\lambda_{max}\int_{\frac{1}{\lambda_{max}\mathsf{SNR}}}^\infty \frac{x}{x + \frac{1}{\lambda_{max}\mathsf{SNR}}} e^{-x} dx\right] \\
&\geq E\left[\lambda_{max}\int_{\frac{1}{\lambda_{max}\mathsf{SNR}}}^\infty \frac{1}{2} e^{-x} dx\right] = E\left[\frac{\lambda_{max}}{2}\exp\left(-\frac{1}{\lambda_{max}\mathsf{SNR}}\right)\right] \\
&\geq E\left[\frac{\lambda_{max}}{2}\left(1 - \frac{1}{\lambda_{max}\mathsf{SNR}}\right)\right] = \frac{1}{2}E(\lambda_{max}) - \frac{1}{2\mathsf{SNR}} \\
&= \frac{1}{2}\left[E(\mathrm{tr}(\mathbf{K}[n])) - \frac{1}{\mathsf{SNR}}\right]
\end{aligned}$$

We next consider the case $n_t = 2$. In this case $\|\mathbf{x}[n]\|^2 = |x_1[n]|^2 + |x_2[n]|^2$, where $|x_1[n]|^2$ and $|x_2[n]|^2$ are independent of $\mathbf{K}[n]$ and are i.i.d. Exp(1). Then we have:

$$\begin{aligned}
E\left[\frac{\lambda_{max}|x_1[n]|^2}{\|\mathbf{x}[n]\|^2 + \frac{2}{\lambda_{max}\mathsf{SNR}}}\right] &= E\left\{E\left[\left.\frac{\lambda_{max}|x_1[n]|^2}{|x_1[n]|^2 + |x_2[n]|^2 + \frac{2}{\lambda_{max}\mathsf{SNR}}}\right|\mathbf{K}[n], |x_2[n]|^2\right]\right\} \\
&= E\left[\lambda_{max}\int_0^\infty \left(\frac{x}{x + |x_2[n]|^2 + \frac{2}{\lambda_{max}\mathsf{SNR}}}\right) e^{-x} dx\right]
\end{aligned}$$

$$
\begin{aligned}
&\geq E\left[\lambda_{max}\int_{\left(|x_2[n]|^2+\frac{2}{\lambda_{max}\mathsf{SNR}}\right)}^{\infty}\left(\frac{x}{x+|x_2[n]|^2+\frac{2}{\lambda_{max}\mathsf{SNR}}}\right)e^{-x}dx\right]\\
&\geq E\left[\lambda_{max}\int_{\left(|x_2[n]|^2+\frac{2}{\lambda_{max}\mathsf{SNR}}\right)}^{\infty}\frac{1}{2}e^{-x}dx\right]\\
&= E\left\{\frac{\lambda_{max}}{2}\exp\left[-\left(|x_2[n]|^2+\frac{2}{\lambda_{max}\mathsf{SNR}}\right)\right]\right\}\\
&= E\left[\frac{\lambda_{max}}{2}\exp\left(-\frac{2}{\lambda_{max}\mathsf{SNR}}\right)\int_0^{\infty}e^{-2x}dx\right]\\
&\geq E\left[\frac{\lambda_{max}}{4}\left(1-\frac{2}{\lambda_{max}\mathsf{SNR}}\right)\right]\\
&\geq \frac{1}{2^2}\left\{\frac{E\left[\mathrm{tr}(\mathbf{K}[n])\right]}{2}-\frac{2}{\mathsf{SNR}}\right\}
\end{aligned}
$$

We finally consider the case $n_t \geq 3$. In this case $\|\mathbf{x}[n]\|^2 = |x_1[n]|^2 + \sum_{i=2}^{n_t}|x_i[n]|^2$, where $\sum_{i=2}^{n_t}|x_i[n]|^2$ is independent of $|x_1[n]|^2$ and $\mathbf{K}[n]$ and has a Gamma$(n_t-1,1)$ distribution. Then we have:

$$
\begin{aligned}
E\left[\frac{\lambda_{max}|x_1[n]|^2}{\|\mathbf{x}[n]\|^2+\frac{n_t}{\lambda_{max}\mathsf{SNR}}}\right] &= E\left\{E\left[\frac{\lambda_{max}|x_1[n]|^2}{\sum_{i=2}^{n_t}|x_i[n]|^2+|x_1[n]|^2+\frac{n_t}{\lambda_{max}\mathsf{SNR}}}\middle|\mathbf{K}[n],|x_1[n]|^2\right]\right\}\\
&= E\left[\lambda_{max}|x_1[n]|^2\int_0^{\infty}\left(\frac{1}{x+|x_1[n]|^2+\frac{n_t}{\lambda_{max}\mathsf{SNR}}}\right)\frac{x^{n_t-2}}{(n_t-2)!}e^{-x}dx\right]\\
&\geq E\left[\frac{\lambda_{max}|x_1[n]|^2}{(n_t-2)!}\int_{\left(|x_1[n]|^2+\frac{n_t}{\lambda_{max}\mathsf{SNR}}\right)}^{\infty}\frac{x}{x+|x_1[n]|^2+\frac{n_t}{\lambda_{max}\mathsf{SNR}}}x^{n_t-3}e^{-x}dx\right]\\
&\geq E\left[\frac{\lambda_{max}|x_1[n]|^2}{(n_t-2)!}\left(|x_1[n]|^2+\frac{n_t}{\lambda_{max}\mathsf{SNR}}\right)^{n_t-3}\int_{\left(|x_1[n]|^2+\frac{n_t}{\lambda_{max}\mathsf{SNR}}\right)}^{\infty}\frac{1}{2}e^{-x}dx\right]\\
&= E\left[\frac{\lambda_{max}|x_1[n]|^2}{2(n_t-2)!}\left(|x_1[n]|^2+\frac{n_t}{\lambda_{max}\mathsf{SNR}}\right)^{n_t-3}\exp\left[-\left(|x_1[n]|^2+\frac{n_t}{\lambda_{max}\mathsf{SNR}}\right)\right]\right]\\
&\geq E\left[\frac{\lambda_{max}}{2(n_t-2)!}\exp\left(-\frac{n_t}{\lambda_{max}\mathsf{SNR}}\right)\left(|x_1[n]|^2\right)^{n_t-2}\exp\left(-|x_1[n]|^2\right)\right]\\
&= E\left[\frac{\lambda_{max}}{2(n_t-2)!}\exp\left(-\frac{n_t}{\lambda_{max}\mathsf{SNR}}\right)\int_0^{\infty}x^{n_t-2}e^{-2x}dx\right]\\
&= E\left[\frac{\lambda_{max}}{2(n_t-2)!}\exp\left(-\frac{n_t}{\lambda_{max}\mathsf{SNR}}\right)\frac{1}{2^{n_t-1}}\int_0^{\infty}y^{n_t-2}e^{-y}dy\right]\\
&= E\left[\frac{\lambda_{max}}{2^{n_t}}\exp\left(-\frac{n_t}{\lambda_{max}\mathsf{SNR}}\right)\right] \geq E\left[\frac{\lambda_{max}}{2^{n_t}}\left(1-\frac{n_t}{\lambda_{max}\mathsf{SNR}}\right)\right]\\
&\geq \frac{1}{2^{n_t}}\left[\frac{E[\mathrm{tr}(\mathbf{K}[n])]}{n_t}-\frac{n_t}{\mathsf{SNR}}\right]
\end{aligned}
$$

■

Replacing in (66) we obtain:

$$
E\left[\mathrm{tr}(\mathbf{K}[n+1])\right] \leq (1-\epsilon)\left\{E\left[\mathrm{tr}(\mathbf{K}[n])\right]\left(1-\frac{1}{n_t 2^{n_t}}\right)+\frac{n_t}{2^{n_t}\mathsf{SNR}}\right\}+n_t\epsilon
$$

which can be solved in steady state to get an upper bound for $E[\mathrm{tr}(\mathbf{K}[n])]$:

$$E[\mathrm{tr}(\mathbf{K}[n])] \leq \left[\frac{n_t(1-\epsilon)}{2^{n_t}\mathsf{SNR}} + n_t\epsilon\right]\frac{n_t 2^{n_t}}{n_t 2^{n_t}\epsilon + 1 - \epsilon} \leq n_t^2\left(\frac{1}{\mathsf{SNR}} + 2^{n_t}\epsilon\right)$$

Finally we have $E\left[\tilde{\mathbf{z}}[n]\tilde{\mathbf{z}}[n]^\dagger\right] = \sigma^2[n]\mathbf{I}_{n_r}$ where for large[8] n

$$\sigma^2[n] \leq \frac{n_t^2 + n_t}{\mathsf{SNR}} + n_t^2 2^{n_t}\epsilon$$

Appendix H

Proof of Theorem 5

We start by specifying a randomly generated codebook with block length B. The codebook $\mathcal{C} = \{\mathbf{X}_1, \cdots, \mathbf{X}_{2^{nR}}\}$, $\mathbf{X}_i \in \mathcal{C}^{n_t \times B}$, $1 \leq i \leq 2^{nR}$, is randomly generated by choosing the $n_t \cdot B$ components of the 2^{nR} codewords independently from a $\mathcal{CN}(0,1)$ distribution. We will transmit $\mathbf{X}_i(k,n)$ from transmit antenna k at time n to send message i. Let $\mathbf{v}_i[n] = \hat{\mathbf{H}}[n]\mathbf{x}_i[n]$, where $\mathbf{x}_i[n]$ is the nth column of $\mathbf{X}_i$. Let $\mathbf{V}_i$ be the matrix whose nth column is $\mathbf{v}_i[n]$. In a similar way define $\tilde{\mathbf{Y}}$ as the matrix whose nth column is $\tilde{\mathbf{y}}[n]$, and $\tilde{\mathbf{Z}}$ as the matrix whose nth column is $\tilde{\mathbf{z}}[n]$. Then we can rewrite the channel equation as:

$$\tilde{\mathbf{Y}} = \mathbf{V}_i + \tilde{\mathbf{Z}}$$

where $\tilde{\mathbf{Z}} \in \mathcal{C}^{n_r \times B}$ is the channel noise received at the n_r receive antennas in times from 1 through n, and is formed by uncorrelated across antennas and independent across time components of equal variance σ^2, also uncorrelated with the components of $\mathbf{X}_i$ and independent of the fading matrices $\{\hat{\mathbf{H}}[n]\}$.

The decoding procedure is to assume that message i was transmitted whenever codeword i has the smallest weighted Euclidean distance to the received matrix $\tilde{\mathbf{Y}}$, i.e.

$$\hat{m} = \arg\min_j \left\|\tilde{\mathbf{Y}} - \mathbf{V}_j\right\|_F$$

where $\hat{m}$ is the decoded message.

The average probability of decoding error, averaged over the randomly generated codewords is independent of the codeword being sent, and as a result we can assume WLOG that codeword 1 was transmitted. Let $\bar{p}$ be the average probability of decoding error, averaged over all codebooks and codewords, and let $\bar{p}(1)$ be the average probability of error when codeword 1 is sent. Then letting $\mathcal{N} = \{1, \ldots, B\}$ we have

$$\bar{p} = \bar{p}(1) = E_{\mathcal{C},\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}}\left[\Pr\left(Error \middle| \mathcal{C}, \tilde{\mathbf{Z}}, \hat{\mathbf{H}}^{\mathcal{N}}, m=1\right)\right] = 1 - E_{\mathcal{C},\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}}\left[\Pr\left(\overline{Error}\middle| \mathcal{C}, \tilde{\mathbf{Z}}, \hat{\mathbf{H}}^{\mathcal{N}}, m=1\right)\right] \quad (67)$$

[8]Here "large" refers to an n large enough so that $E[\mathrm{tr}(\mathbf{K}[n])]$ reaches a steady state.

where m is the transmitted codeword and the expectation is taken over the random codebook $\mathcal{C}$, the noise matrix $\tilde{\mathbf{Z}}$ and the fading matrices $\{\hat{\mathbf{H}}[n]\}$.

There is no decoding error when codeword 1 is the closest codeword in weighted Euclidean distance to the received matrix $\tilde{\mathbf{Y}}$ and as a result

$$\begin{aligned}\Pr(\overline{Error}\big|\,\mathcal{C},\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}},m=1) &= 1\left[\bigcap_{i=2}^{2^{nR}}\left(\left\|\tilde{\mathbf{Y}}-\mathbf{V}_1\right\|_F<\left\|\tilde{\mathbf{Y}}-\mathbf{V}_i\right\|_F\right)\right] \\ &= \prod_{i=2}^{2^{nR}} 1\left(\left\|\tilde{\mathbf{Z}}\right\|_F<\left\|\mathbf{V}_1+\tilde{\mathbf{Z}}-\mathbf{V}_i\right\|_F\right)\end{aligned}$$

Noting that the codewords $\mathbf{X}_2,\cdots,\mathbf{X}_{2^{nR}}$ are chosen i.i.d. and independently of $\tilde{\mathbf{Z}}$, $\{\hat{\mathbf{H}}[n]\}$ and $\mathbf{X}_1$, we can rewrite (67) as

$$\begin{aligned}\bar{p} &= 1-E_{\mathbf{X}_1,\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}}\left\{E_{\mathbf{X}_2,\cdots,\mathbf{X}_{2^{nR}}|\mathbf{X}_1,\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}}\left[\prod_{i=2}^{2^{nR}}1\left(\left\|\tilde{\mathbf{Z}}\right\|_F<\left\|\mathbf{V}_1+\tilde{\mathbf{Z}}-\mathbf{V}_i\right\|_F\right)\Big|\,\mathbf{X}_1,\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}\right]\right\} \\ &= 1-E_{\mathbf{X}_1,\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}}\left\{E_{\mathbf{X}_2,\cdots,\mathbf{X}_{2^{nR}}}\left[\prod_{i=2}^{2^{nR}}1\left(\left\|\tilde{\mathbf{Z}}\right\|_F<\left\|\mathbf{V}_1+\tilde{\mathbf{Z}}-\mathbf{V}_i\right\|_F\right)\right]\right\} \\ &= 1-E_{\mathbf{X}_1,\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}}\left\{\left[E_{\mathbf{X}_2}\left[1\left(\left\|\tilde{\mathbf{Z}}\right\|_F<\left\|\mathbf{V}_1+\tilde{\mathbf{Z}}-\mathbf{V}_2\right\|_F\right)\right]\right]^{2^{nR}-1}\right\} \\ &= 1-E_{\mathbf{X}_1,\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}}\left[P_{\bar{e}}(\tilde{\mathbf{Y}},\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}})\right]\end{aligned}$$

where $\tilde{\mathbf{Y}}=\mathbf{V}_1+\tilde{\mathbf{Z}}$ and $P_{\bar{e}}(\tilde{\mathbf{Y}},\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}})=\left\{E_{\mathbf{X}_2}\left[1\left(\left\|\tilde{\mathbf{Z}}\right\|_F<\left\|\mathbf{V}_1+\tilde{\mathbf{Z}}-\mathbf{V}_2\right\|_F\right)\right]\right\}^{2^{nR}-1}$.

We will upper bound the average probability of decoding error $\bar{p}$ by the one corresponding to a Gaussian channel, relying on the fact that this probability only depends asymptotically on the second order statistics of the random variables involved. We will use the following lemma.

Lemma 13 *Let* $\mathbf{Y}$,$\tilde{\mathbf{Y}}$,$\mathbf{Z}$ *and* $\tilde{\mathbf{Z}}$ *be any matrices in* $\mathcal{C}^{n_r\times B}$ *that satisfy the inequality* $\left\|\mathbf{Y}-\tilde{\mathbf{Y}}\right\|_F\leq\|\mathbf{Z}\|_F-\left\|\tilde{\mathbf{Z}}\right\|_F$. *Then* $P_{\bar{e}}\left(\mathbf{Y},\mathbf{Z},\hat{\mathbf{H}}^{\mathcal{N}}\right)\leq P_{\bar{e}}\left(\tilde{\mathbf{Y}},\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}\right)$.

Proof:

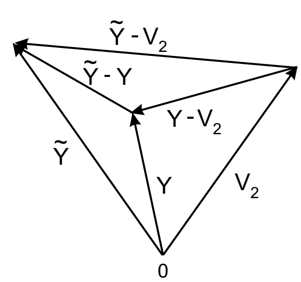

Figure 5: Two dimensional representation of $\mathbf{Y}$, $\tilde{\mathbf{Y}}$ and $\mathbf{V}_2$

From figure 5 we see that $(\tilde{\mathbf{Y}} - \mathbf{V}_2)$, $(\mathbf{Y} - \mathbf{V}_2)$ and $(\tilde{\mathbf{Y}} - \mathbf{Y})$ form a triangle and as a result

$$\left\|\tilde{\mathbf{Y}} - \mathbf{V}_2\right\|_F \geq \|\mathbf{Y} - \mathbf{V}_2\|_F - \left\|\mathbf{Y} - \tilde{\mathbf{Y}}\right\|_F \tag{68}$$

From equation (68) we have that if $\|\mathbf{Y} - \mathbf{V}_2\|_F > \|\mathbf{Z}\|_F$ it follows that

$$\left\|\tilde{\mathbf{Y}} - \mathbf{V}_2\right\|_F > \|\mathbf{Z}\|_F - \left\|\mathbf{Y} - \tilde{\mathbf{Y}}\right\|_F$$

and from the hypothesis of the lemma

$$\left\|\tilde{\mathbf{Y}} - \mathbf{V}_2\right\|_F > \|\mathbf{Z}\|_F - \|\mathbf{Z}\|_F + \left\|\tilde{\mathbf{Z}}\right\|_F = \left\|\tilde{\mathbf{Z}}\right\|_F$$

As a result, under the hypothesis of the lemma

$$1\left[\left\|\tilde{\mathbf{Z}}\right\|_F < \left\|\tilde{\mathbf{Y}} - \mathbf{V}_2\right\|_F\right] \geq 1\left[\|\mathbf{Z}\|_F < \|\mathbf{Y} - \mathbf{V}_2\|_F\right]$$

and we have that $P_{\bar{e}}(\tilde{\mathbf{Y}}, \tilde{\mathbf{Z}}, \hat{\mathbf{H}}^{\mathcal{N}}) \geq P_{\bar{e}}(\mathbf{Y}, \mathbf{Z}, \hat{\mathbf{H}}^{\mathcal{N}})$. ■

Consider 2 different channels, one with additive white circularly symmetric complex Gaussian noise of variance $\sigma^2 + \epsilon$ for each component, $\epsilon > 0$, independent of the input signal denoted by $\mathbf{Z} \in \mathcal{C}^{n_r \times B}$, and another one with noise of variance σ^2 per component, denoted by $\tilde{\mathbf{Z}} \in \mathcal{C}^{n_r \times B}$ in which $\tilde{\mathbf{Z}}(i,j)$ is independent of $\tilde{\mathbf{Z}}(h,k)$ for $j \neq k$ and uncorrelated with $\tilde{\mathbf{Z}}(h,j)$ for $i \neq h$. Since in both cases the noise is independent across columns the Strong Law of Large Numbers lets us write $\frac{1}{B}\mathbf{Z}\mathbf{Z}^\dagger \overset{a.s.}{\to} (\sigma^2 + \epsilon)\mathbf{I}_{n_r}$ and $\frac{1}{B}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\dagger \overset{a.s.}{\to} \sigma^2\mathbf{I}_{n_r}$. Also because the codewords are generated with independent Gaussian components of variance 1 and $\hat{\mathbf{H}}[n]$ is independent over time and independent of the input, and $E\left(\hat{\mathbf{h}}_i^\dagger[n]\hat{\mathbf{h}}_j[n]\right) = \alpha^2\delta_{i,j}$, by the SLLN we have that $\frac{1}{B}\mathbf{X}_i\mathbf{X}_i^\dagger \overset{a.s.}{\to} \mathbf{I}_{n_t}$ and $\frac{1}{B}\mathbf{V}_i\mathbf{V}_i^\dagger \overset{a.s.}{\to} \alpha^2\mathbf{I}_{n_r}$ Finally noting that the noise is uncorrelated with the input signal and is independent of the fading matrices $\{\hat{\mathbf{H}}[n]\}$, the SLLN implies that $\frac{1}{B}\mathbf{Y}\mathbf{Y}^\dagger \overset{a.s.}{\to} = (\alpha^2 + \sigma^2 + \epsilon)\mathbf{I}_{n_r}$ and $\frac{1}{B}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\dagger \overset{a.s.}{\to} = (\alpha^2 + \sigma^2)\mathbf{I}_{n_r}$.

Let

$$\mathcal{I} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathcal{C}^{n_r \times B}$$

We can express $\mathbf{Y}$ and $\tilde{\mathbf{Y}}$ by their singular value decompositions as follows: $\mathbf{Y} = \sqrt{B}\mathbf{U}\mathbf{\Lambda}^{1/2}\mathcal{I}\mathbf{V}^\dagger$ and $\tilde{\mathbf{Y}} = \sqrt{B}\tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}^{1/2}\mathcal{I}\tilde{\mathbf{V}}^\dagger$ where $\mathbf{U}$, $\tilde{\mathbf{U}}$, $\mathbf{V}$, and $\tilde{\mathbf{V}}$ are unitary matrices, and $\mathbf{\Lambda}$ and $\tilde{\mathbf{\Lambda}}$ are diagonal matrices with real, non negative diagonal entries. Then $\frac{1}{B}\mathbf{Y}\mathbf{Y}^\dagger = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$ and $\frac{1}{B}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\dagger = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{U}}^\dagger$. From the previous results we have that as $B \to \infty$, $\mathbf{\Lambda} \to (\alpha^2 + \sigma^2 + \epsilon)\mathbf{I}_{n_r}$ and $\tilde{\mathbf{\Lambda}} \to (\alpha^2 + \sigma^2)\mathbf{I}_{n_r}$.

We want to verify the hypothesis of Lemma 13 in the limit as $B \to \infty$. We first compute $\frac{1}{B}\|\tilde{\mathbf{Y}} - \mathbf{Y}\|_F^2$ as $B \to \infty$:

$$\lim_{B\to\infty} \frac{1}{B}\left\|\tilde{\mathbf{Y}} - \mathbf{Y}\right\|_F^2 = \lim_{B\to\infty} \frac{1}{B}\text{tr}\left[\left(\tilde{\mathbf{Y}} - \mathbf{Y}\right)\left(\tilde{\mathbf{Y}} - \mathbf{Y}\right)^\dagger\right]$$

$$
\begin{aligned}
&= \lim_{B\to\infty} \frac{1}{B}\text{tr}\left[\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\dagger - \tilde{\mathbf{Y}}\mathbf{Y}^\dagger - \mathbf{Y}\tilde{\mathbf{Y}}^\dagger + \mathbf{Y}\mathbf{Y}^\dagger\right] \\
&= n_r(\alpha^2+\sigma^2) - \lim_{B\to\infty} \text{tr}\left[\tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}^{1/2}\mathcal{I}\tilde{\mathbf{V}}^\dagger\mathbf{V}\mathcal{I}^\dagger\mathbf{\Lambda}^{1/2}\mathbf{U}^\dagger\right] \\
&\quad - \lim_{B\to\infty} \text{tr}\left[\mathbf{U}\mathbf{\Lambda}^{1/2}\mathcal{I}\mathbf{V}^\dagger\tilde{\mathbf{V}}\mathcal{I}^\dagger\tilde{\mathbf{\Lambda}}^{1/2}\tilde{\mathbf{U}}^\dagger\right] + n_r(\alpha^2+\sigma^2+\epsilon) \\
&= n_r(\alpha^2+\sigma^2) - \lim_{B\to\infty} \text{tr}\left[\tilde{\mathbf{\Lambda}}^{1/2}\mathbf{\Lambda}^{1/2}\right] \\
&\quad - \lim_{B\to\infty} \text{tr}\left[\mathbf{\Lambda}^{1/2}\tilde{\mathbf{\Lambda}}^{1/2}\right] + n_r(\alpha^2+\sigma^2+\epsilon) \\
&= n_r(\alpha^2+\sigma^2) - 2n_r\sqrt{(\alpha^2+\sigma^2)(\alpha^2+\sigma^2+\epsilon)} + n_r(\alpha^2+\sigma^2+\epsilon) \\
&= n_r\left(\sqrt{\alpha^2+\sigma^2+\epsilon} - \sqrt{\alpha^2+\sigma^2}\right)^2
\end{aligned}
$$

Also from the SLLN we have that with probability 1

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{\sqrt{B}}\left(\|\mathbf{Z}\|_F - \left\|\tilde{\mathbf{Z}}\right\|_F\right) &= \sqrt{\text{tr}\left[(\sigma^2+\epsilon)\,\mathbf{I}_{n_r}\right]} - \sqrt{\text{tr}\left[\sigma^2\mathbf{I}_{n_r}\right]} \\
&= \sqrt{n_r}\left(\sqrt{\sigma^2+\epsilon} - \sqrt{\sigma^2}\right)
\end{aligned}
$$

To check that for $B\to\infty$, $\|\tilde{\mathbf{Y}}-\mathbf{Y}\|_F \le \|\mathbf{Z}\|_F - \left\|\tilde{\mathbf{Z}}\right\|_F$ we have to verify that $f(\alpha^2) = \sqrt{\sigma^2+\epsilon} - \sqrt{\sigma^2} - \sqrt{\alpha^2+\sigma^2+\epsilon} + \sqrt{\alpha^2+\sigma^2} > 0$ for all $\alpha^2 > 0$. This follows from the fact that $f(0) = 0$ and $f(\alpha^2)$ is strictly increasing for $\alpha^2 > 0$.

$$
f'(\alpha^2) = -\frac{1}{2\sqrt{\alpha^2+\sigma^2+\epsilon}} + \frac{1}{2\sqrt{\alpha^2+\sigma^2}} > 0
$$

for $\alpha^2 > 0$.

Therefore with probability one

$$
\lim_{B\to\infty}\left[\|\mathbf{Z}\|_F - \left\|\tilde{\mathbf{Z}}\right\|_F - \left\|\tilde{\mathbf{Y}}-\mathbf{Y}\right\|_F\right] > 0 \tag{69}
$$

We can now upper bound the average probability of error $\bar{p}$ with the one corresponding to the Gaussian channel as follows:

$$
\lim_{B\to\infty}\bar{p} = 1 - \lim_{n\to\infty} E_{\mathbf{X}_1,\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}}\left[P_{\bar{e}}\left(\tilde{\mathbf{Y}},\tilde{\mathbf{Z}},\hat{\mathbf{H}}^{\mathcal{N}}\right)\right] = 1 - \lim_{B\to\infty} E_\omega\left\{P_{\bar{e}}\left[\tilde{\mathbf{Y}}(\omega),\tilde{\mathbf{Z}}(\omega),\hat{\mathbf{H}}^{\mathcal{N}}(\omega)\right]\right\}
$$

Since $P_{\bar{e}}[\tilde{\mathbf{Y}}(\omega),\tilde{\mathbf{Z}}(\omega),\hat{\mathbf{H}}^{\mathcal{N}}(\omega)]$ is an average probability, $0 \le P_{\bar{e}}[\tilde{\mathbf{Y}}(\omega),\tilde{\mathbf{Z}}(\omega),\hat{\mathbf{H}}^{\mathcal{N}}(\omega)] \le 1$ and we can use the Dominated Convergence Theorem to exchange the limit with the expectation:

$$
\lim_{B\to\infty}\bar{p} = 1 - E_\omega\left\{\lim_{B\to\infty} P_{\bar{e}}[\tilde{\mathbf{Y}}(\omega),\tilde{\mathbf{Z}}(\omega),\hat{\mathbf{H}}^{\mathcal{N}}(\omega)]\right\} \le 1 - E_\omega\left\{\lim_{B\to\infty} P_{\bar{e}}[\mathbf{Y}(\omega),\mathbf{Z}(\omega),\hat{\mathbf{H}}^{\mathcal{N}}(\omega)]\right\}
$$

where the last inequality follows from (69) and Lemma 13.

Finally $1 - E_\omega \left\{ \lim_{B\to\infty} P_{\bar{e}} \left[\mathbf{Y}(\omega), \mathbf{Z}(\omega), \hat{\mathbf{H}}^{\mathcal{N}}(\omega) \right] \right\}$ is the limit for infinite block length of the average probability of decoding error for a Gaussian channel with fading matrix known to the receiver, in which random Gaussian codebooks and weighted minimum Euclidean distance decoding are used. We can use the Channel Coding Theorem to conclude that this probability goes to 0 as $B \to \infty$ for all rates $R' < E\left[\log\det\left(\mathbf{I}_{n_r} + \frac{\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger}{\sigma^2+\epsilon}\right)\right]$ [2]. Since ϵ is arbitrarily small we have that all rates below $C = E\left[\log\det\left(\mathbf{I}_{n_r} + \frac{\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger}{\sigma^2}\right)\right]$ are achievable. Therefore, the supremum of all achievable rates R', R, must be at least as large as C.

References

[1] G. Foschini, M. Gans, *On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas*, Wireless Personal Communications, No. 6, pp. 311-335, 1998.

[2] E. Telatar, *Capacity of Multi-Antenna Gaussian Channels*, European Trans. Telecommun., Vol. 10, No. 6, pp. 585-595, 1999.

[3] A. Lapidoth, S. Moser, *Capacity Bounds Via Duality with Applications to Multi-Antenna Systems on Flat Fading channels*, Preprint, June 2002.

[4] G. Taricco and M. Elia, *Capacity of Fading Channels with no Side Information*, Electronics Letters, Vol. 33, pp. 1368-1370, July 1997.

[5] T. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, 2nd Edition, 2001.

[6] R. Steele, *Mobile Radio Communications*, IEEE Press, 1994.

[7] R. Kennedy, *Fading Dispersive Communication Channels*, John Wiley & Sons, 1969.

[8] B. Hochwald, W. Sweldens, *Differential Unitary Space-time Modulation*, IEEE Transactions on Communications, Vol. 48, No. 12, pp. 2041-2052, 2000.

[9] L. Zheng and D. Tse, *Communicating on the Grassmann Manifold: A Geometric Approach to the Non-coherent Multiple Antenna Channel*, IEEE Transactions on Information Theory, vol. 48(2), February 2002, pp. 359-383.

[10] A. Lapidoth, *Nearest Neighbor Decoding for Additive Non-Gaussian Noise Channels*, IEEE Transactions on Information Theory, vol. 42, No. 5, 1996, pp. 1520-1529.