Degrees of Freedom in Some Underspread MIMO Fading Channels

Raúl H. Etkin, *Student Member, IEEE,* and David N. C. Tse, *Member, IEEE*

***Abstract*—Consider a multiple-input multiple-output (MIMO) fading channel in which the fading process varies slowly over time. Assuming that neither the transmitter nor the receiver have knowledge of the fading process, do multiple transmit and receive antennas provide significant capacity improvements at high signal-to-noise ratio (SNR)? For regular fading processes, recent results show that capacity ultimately grows doubly logarithmically with the SNR independently of the number of transmit and receive antennas used. We show that for the Gauss–Markov fading process in all regimes of practical interest the use of multiple antennas provides large capacity improvements. Nonregular fading processes show completely different high-SNR behaviors due to the perfect predictability of the process from noiseless observations. We analyze the capacity of MIMO channels with nonregular fading by presenting a lower bound, which we specialize to the case of band-limited slowly varying fading processes to show that the use of multiple antennas is still highly beneficial. In both cases, regular and nonregular fading, this capacity improvement can be seen as the benefit of having multiple spatial degrees of freedom. For the Gauss–Markov fading model and all regimes of practical interest, we present a communication scheme that achieves the full number of degrees of freedom of the channel with tractable complexity. Our results for underspread Gauss–Markov and band-limited nonregular fading channels suggest that multiple antennas are useful at high SNR.**

***Index Terms*—Channel capacity, decision-oriented training, fading number, high signal-to-noise ratio (SNR), multiple antennas, noncoherent communication.**

I. Introduction

RECENT information-theoretic results suggest that in richly scattered wireless environments, systems with multiple transmit and multiple receive antennas (multiple-input multiple-output (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 independent and identically distributed (i.i.d.) across antenna pairs, and showed that at high signal-to-noise ratio (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.

Manuscript received May 23, 2003; revised November 28, 2005. The work of R. H. Etkin was supported by the National Science Foundation under Grant CCR-0118784.

The authors are with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: retkin@eecs.berkeley.edu; dtse@eecs.berkeley.edu).

Communicated by G. Caire, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2006.871595

The preceding 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? In particular, does the number of transmit and receive antennas have a significant impact in the high-SNR capacity? Is the notion of degrees of freedom still relevant ?

As a first step to answer these questions, Marzetta and Hochwald analyzed in [3] the block-fading model, where the fading matrix $\boldsymbol{H} = [h_{i,j}]$ is assumed to remain constant over the duration of a block of length T, and varies independently between blocks. The high SNR capacity of this model in the MIMO case was characterized by Zheng and Tse in [4] where it is shown that

$$C \sim M^*(1 - M^*/T) \log \mathsf{SNR}$$

where $M^* = \min\{n_r, n_t, \lfloor T/2 \rfloor\}$. One has to be careful in interpreting this result in the regime of validity of the block-fading model. At high SNR, small variations in the channel coefficients can introduce enough uncertainty to affect capacity, and this is not taken into account when one assumes a constant channel in the coherence block.

To overcome this difficulty, Liang and Veeravalli [5] introduced the time-selective block-fading model, where the fading matrix is allowed to vary within each block, and is assumed to vary independently across blocks. For the MIMO case, they obtained a capacity lower bound

$$C_{\text{lower}} = T^*(1 - T^*Q/T) \log \mathsf{SNR} + c_1 + o(1)$$

where $T^* = \min\{n_r, n_t, \lfloor T/(2Q) \rfloor\}$, Q is the rank of the covariance matrix of $\{h_{i,j}(t)\}_{t=1}^{T}$, and c_1 is a constant independent of SNR. In all cases, the block-fading model assumes independent channel realizations in different blocks. This assumption may be justified in time-interleaved or frequency-hopped systems, but may be artificial for systems that operate by transmitting a continuous symbol stream over time.

In this work, we will analyze a model in which the fading process is stationary and is allowed to vary from symbol to symbol. To answer the above questions in this context, Lapidoth and Moser [6], building on earlier work by Taricco and

Elia for memoryless fading [7], recently showed a contrasting result: that at high SNR, for regular[1] Gaussian fading processes, 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 [6] explicitly incorporates the channel variation (and the associated uncertainty) in its model, the $\log\log \mathsf{SNR}$ 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.

The $\log\log \mathsf{SNR}$ result is obtained in the context of regular fading processes. When the fading process is nonregular, it can be perfectly predicted from noiseless observations of its infinite past, and one should expect different high SNR capacity asymptotics. In fact, Lapidoth shows in [8], [9] that in the case of nonregular fading the capacity can grow as $\log\log \mathsf{SNR}$, $\log \mathsf{SNR}$, or $(\log \mathsf{SNR})^\beta$, $0 < \beta < 1$, in a single-input single-output (SISO) channel. Therefore, regular and nonregular fading processes can result in completely different high SNR capacities and need to be analyzed separately.

Typical wireless channels are *underspread*, that is, they satisfy $T_c W_c \gg 1$, where T_c is the coherence time and W_c the coherence bandwidth of the channel.[2] In the context of regular fading, the underspread condition implies that the one-step minimum mean-square error (MMSE) in the prediction of the fading process is small. In the nonregular fading case, when the process is band-limited to a bandwidth of $2f_m$ where f_m is the maximum Doppler shift, the underspread condition implies $2f_m/W_c \ll 1$.

A. Main Results

In this work, we would like to answer the questions raised in the beginning of this section for some channels with regular and nonregular fading processes. Since we are mainly interested in the influence of the number of transmit and receive antennas on the capacity, we give particular emphasis to the MIMO case.

We analyze the regular fading case in the context of a Gauss–Markov fading model. Suppose the fading process for each channel gain is first-order Gauss–Markov with a one-step MMSE prediction error in the prediction of ϵ from sample to sample. For underspread channels, ϵ is small. We propose that the capacity $C(\mathsf{SNR}, \epsilon)$ of a first-order Gauss-Markov 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: $1/\epsilon \le \mathsf{SNR} < \exp\left(\epsilon^{-\min\{n_t,n_r\}}\right)$

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

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

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

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 limited by channel uncertainty. However, when the SNR gets much larger, the doubly logarithmic 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. This observation does not apply to the third regime, where the degrees of freedom in the channel have no effect in the first-order term of capacity.

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 4.1×10^7 to 4.3×10^{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. To address the case of nonregular fading processes, we present a capacity lower bound

$$C(\mathsf{SNR}) \ge \min\{n_r, n_t\} \left(1 - \min\{n_t, n_r\}|\mathcal{D}|\right) \log \mathsf{SNR} + K(n_r, n_t)$$

for large SNR and some constant $K(n_r, n_t)$ independent of SNR, where $|\mathcal{D}|$ is the Lebesgue measure of the set of frequencies where the power spectral density (PSD) of the fading process is nonzero. For band-limited underspread channels $|\mathcal{D}|$ is typically small. This result is of particular practical importance since channel models based on the physics of the channel such as Jakes' model [10] have band-limited fading PSDs and hence are nonregular.[3] In this case, the parameter $|\mathcal{D}|$ can be associated with the Doppler spread of the channel: $|\mathcal{D}| = 2f_m/W_c$, where f_m is the maximum Doppler shift and W_c is the coherence bandwidth of the channel. This lower bound

[1]A process is called regular when the mean-square estimation error in the estimate of the present value from the (noiseless) infinite past is positive.

[2]See Section II for a definition of these parameters.

[3]Refer to Section VI and Appendix I for details regarding the assumptions required to reach this result, and to see a complete derivation.

extends to the MIMO setting the results of [8], [9], where lower and upper bounds on the capacity of a SISO channel with a peak power constraint in the input and nonregular fading are presented.

It is interesting to compare this lower bound with the results for the block-fading model of [4] and the time-selective block-fading model of [5]. Intuitively, in the block-fading case, the channel is perfectly predictable within each block, and a fraction M^*/T of the available degrees of freedom needs to be spent in channel estimation and cannot be used for communication. A similar intuition can be applied to the time-selective block-fading model, where the channel predictability in each block is controlled by the parameter Q. Channels with smaller Q are easier to predict and result in more degrees of freedom available for communication. In our case, the predictability of the channel is governed by $|\mathcal{D}|$ and the fraction of degrees of freedom lost due to channel uncertainty is given by $\min\{n_r, n_t\}|\mathcal{D}|$.

These results suggest 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.

B. Related Works

The first-order Gauss–Markov fading model was analyzed in the context of a SISO system in [11], where it is shown that as the SNR increases to infinity, Gaussian inputs result in bounded mutual information. It is also shown that for this kind of inputs, the mutual information only increases substantially with SNR when the channel estimation error variance ϵ is negligible as compared to $1/\mathsf{SNR}$. This result is a generalization of the work in [12] where i.i.d. Rayleigh-fading was assumed. [12] also shows for the SISO case that unless the channel estimation error variance is negligible with respect to $1/\mathsf{SNR}$, Gaussian inputs and scaled nearest neighbor decoding have poor performance at high SNR. Lapidoth and Moser [6] show in Theorem 4.2 that for MIMO channels with independent, stationary, and ergodic fading and noise processes with finite entropy rate, inputs with finite expected log-norm result in bounded mutual information regardless of the SNR. The effect of channel measurement errors in a SISO channel was studied in [13], where bounds on mutual information were applied to a Gauss–Markov fading model. Weingarten *et al.*[14] analyze the MIMO case with Gaussian inputs and scaled nearest neighbor decoding with some side information at the receiver. Our characterization of the first regime is consistent with these results, and is obtained in more generality, in the context of the Gauss–Markov fading model, for MIMO channels with arbitrary inputs with no specific decoding structure assumptions. Lapidoth and Moser [6] characterize the regime in which the $\log\log \mathsf{SNR}$ behavior kicks in through the notion of a *fading number* (see Section VII-B for a discussion about the fading number and its relation to the results that we present), but does not provide a tight bound for the fading number of the MIMO channel when the fading process is correlated over time. In particular, the bound that can be derived from [6] does not capture the influence of the degrees of freedom $\min\{n_r, n_t\}$ on the channel capacity. The SISO nonregular fading channel with a peak power constraint in the input was considered in [8], [9], where it is shown that

$$\lim_{\mathsf{SNR}\to\infty} C(\mathsf{SNR})/\log(\mathsf{SNR}) = 1 - |\mathcal{D}|.$$

C. Notation

We will use lower case or upper case letters to represent scalars, boldface lower case letters for vectors, upper case calligraphic letters for sets, and boldface upper case letters for matrices. For example, we write n for a scalar, $\boldsymbol{v}$ for a vector, $\mathcal{A}$ for a set, and $\boldsymbol{H}$ for a matrix. We will denote by $\boldsymbol{v}^{\mathcal{A}}$ the vector with components $v[n]$ for $n \in \mathcal{A}$, where $v[n]$ is the nth component of $\boldsymbol{v}$. 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 $\boldsymbol{H}^T$ for the transpose, $\boldsymbol{H}^*$ for the conjugate, and $\boldsymbol{H}^\dagger$ for the Hermitian (conjugate transpose) of the complex matrix $\boldsymbol{H}$ and we represent the kxk identity matrix by $\boldsymbol{I}_k$. Finally, we write $\|\boldsymbol{v}\|$ for the L_2 norm of the vector $\boldsymbol{v}$, i.e., $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v}^\dagger \boldsymbol{v}}$, and $\|\boldsymbol{H}\|_F$ for the Frobenius norm of the matrix $\boldsymbol{H}$, i.e., $\|\boldsymbol{H}\|_F = \sqrt{\mathrm{tr}(\boldsymbol{H}\boldsymbol{H}^\dagger)}$.

II. 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

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

where $\boldsymbol{y}[n] \in \mathbb{C}^{n_r}$ is the channel output, $\boldsymbol{x}[n] \in \mathbb{C}^{n_t}$ is the channel input with average power constraint

$$\sum_{n=1}^{B} E[\|\boldsymbol{x}[n]\|^2] \leq B n_t;$$

$\boldsymbol{z}[n] \sim \mathcal{CN}(0, I_{n_r})$ is circularly symmetric white complex Gaussian noise, SNR is the SNR, and $\boldsymbol{H}[n] \in \mathbb{C}^{n_r \mathrm{x} n_t}$ is the fading matrix with i.i.d. circularly symmetric complex Gaussian components of zero mean and unit variance. In Sections III–V, the time variation of the channel is modeled by a Gauss–Markov process[4]

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

where $\boldsymbol{W}[n] \in \mathbb{C}^{n_r \mathrm{x} n_t}$ 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 $\epsilon \in \mathbb{R}$, $0 \leq \epsilon \leq 1$. As $\epsilon \to 0$, we get the limiting case of a constant channel. Also, $\boldsymbol{H}[0]$ has zero mean, unit variance complex Gaussian independent components.

[4]In Section VI, we allow the fading process to be any nonregular process with a power spectral density.

We will use the channel by grouping input symbols in blocks of size B, i.e.,

$$\boldsymbol{x}^{\mathcal{N}} = (\boldsymbol{x}[1], \ldots, \boldsymbol{x}[B]) \in \mathbb{C}^{n_t \mathrm{x} B}$$

where $\mathcal{N} = \{1, \ldots, B\}$, and compute the channel capacity using the following expression:

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

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 $\boldsymbol{H}[n]$ 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[15]

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

where σ_τ is the root mean square (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 1 to 2 μs in urban environments and from 10 to 100 ns in indoor environments (see [15] 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 autocorrelation function is above 0.5 of its value at 0 then [16]

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

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

We note that the above calculations correspond to fitting the free parameter of the first-order Gauss–Markov fading process (i.e., ϵ) to some physical parameters obtained by measurement of real systems. There are different ways of fitting the model to the measurements, like adjusting ϵ to obtain a given value of the autocorrelation function of the fading process at time $n = 1$, or adjusting ϵ to obtain a given noisy prediction error for a given value of SNR. Our approach is conservative in the sense that the values of ϵ obtained by the other suggested methods are smaller, and hence the fading is easier to predict.

We point out that the Gauss–Markov assumption of the fading process is made to make the capacity analysis tractable, while capturing in the parameter ϵ the noiseless prediction error existent in all regular processes. As is shown in [8], [9] for SISO channels with nonregular fading processes, the high SNR capacity may have very different behaviors as compared to the regular case. The Gauss–Markov model, being regular, cannot be fitted to predict behaviors typical of nonregular processes. In Section VI, we analyze (through a lower bound) the capacity of the MIMO fading channel when the fading coefficients correspond to band-limited nonregular processes.

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 multiuser systems that are interference limited, the SNR is limited by the number of users that are sharing the channel.[5] In multiuser 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 30 dB.

III. The Three Regimes

The input symbols of the channel (1) are vectors in $\mathbb{C}^{n_t}$. The rank of the matrix $\boldsymbol{H}[n]$ determines how many of these dimensions are resolvable at the receiver. The assumption that $\boldsymbol{H}[n]$ is formed by i.i.d. Gaussian components implies that $\boldsymbol{H}[n]$ is full rank with probability one. If $n_r \le n_t$, the received signal is a vector in $\mathbb{C}^{n_r}$ so at most n_r of the n_t dimensions of the input vector can effectively be resolved. If, on the other hand, $n_r \ge n_t$ the information-bearing component of the received signal is contained in a subspace of dimension n_t of $\mathbb{C}^{n_r}$. Therefore, in both cases $\mathrm{rank}(\boldsymbol{H}[n]) =_{\text{a.s.}} \min\{n_t, n_r\}$ 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,

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

TABLE I

Environment	$n_t = n_r$	$\epsilon(\max)$	$\epsilon(\min)$	$\theta_1(\min)$	$\theta_1(\max)$	$\theta_2(\min)$	$\theta_2(\max)$
Indoors	1	10^{-4}	3×10^{-7}	40 dB	65 dB	4.3×10^4 dB	1.4×10^7 dB
Indoors	4	10^{-4}	3×10^{-7}	40 dB	65 dB	4.3×10^{16} dB	5.4×10^{26} dB
Urban	1	0.018	10^{-4}	17.4 dB	40 dB	241 dB	4.3×10^4 dB
Urban	4	0.018	10^{-4}	17.4 dB	40 dB	4.1×10^7 dB	4.3×10^{16} dB

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:

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

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 nonasymptotic 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 two quantities we can differentiate two 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 two 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 Section IV, 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 [6], 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 SNRs 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 SNRs 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 one 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 do not 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 decibels 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 II: see the Table I at the top of the page.[6]

From Table I 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.

We point out that the values in Table I are obtained by fitting the first-order Gauss–Markov model to the parameters of phys-

[6]Here $\theta_i(\min)$ corresponds to the maximum value of ϵ and $\theta_i(\max)$ to the minimum value of ϵ for a particular scenario.

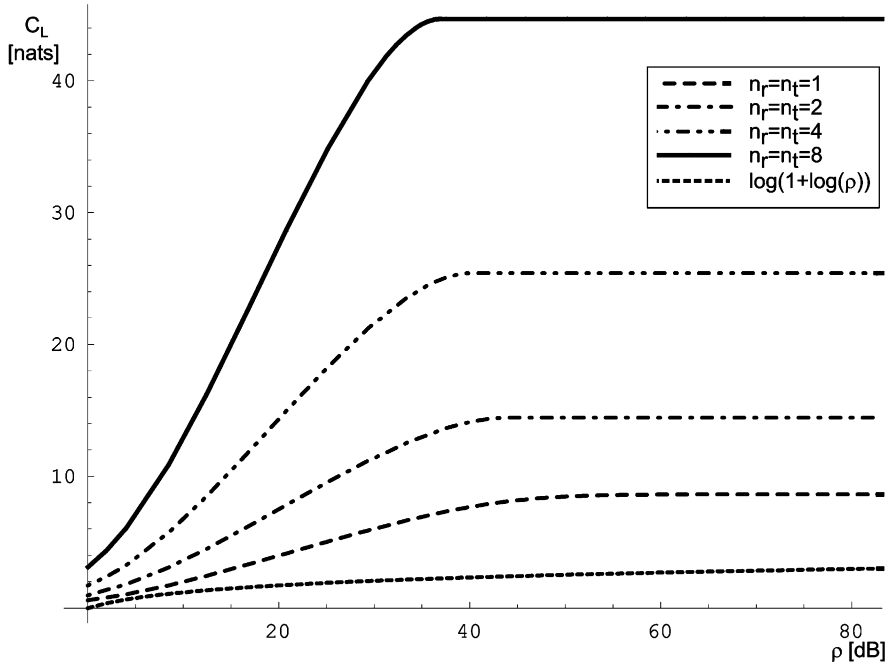

Fig. 1. Capacity lower bound as a function of the SNR (ρ) for different values of $n_t = n_r$, and $\epsilon = 10^{-4}$. The curve $\log(1 + \log \mathsf{SNR})$ is plotted for reference; it would cross the curve corresponding to $n_r = n_t = 1$ at SNR $\approx$ 43000 dB. The crossing points for the other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

ical channels. In general, these numbers are likely to change for higher order Markov processes or other regular processes.

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 nonasymptotic lower bound on channel capacity derived using information-theoretic principles.

Theorem 2: The capacity $C(\mathsf{SNR}, \epsilon)$ is lower-bounded by the expression at the bottom of the page.

Proof: See Appendix A. □

Figs. 1 and 2 show the value of this lower bound for different numbers 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 Section IV 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 two 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

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

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 V-A. □

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

where

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

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

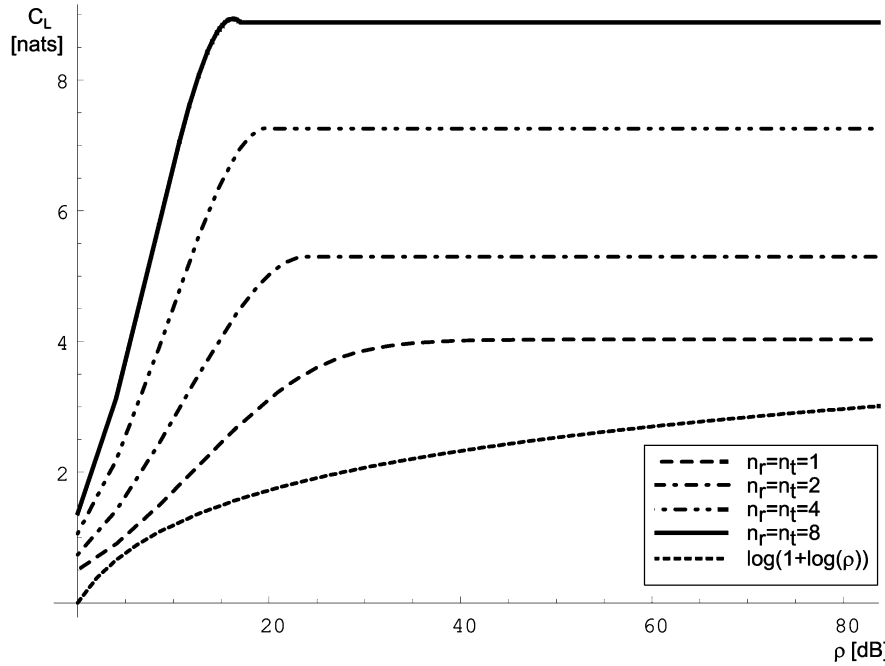

Fig. 2. Capacity lower bound as a function of the SNR (ρ) for different values of $n_t = n_r$, and $\epsilon = 0.01$. The curve $\log(1+\log \mathsf{SNR})$ is plotted for reference; it would cross the curve corresponding to $n_r = n_t = 1$ at SNR $\approx$ 400 dB. The crossing points for the other curves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

Theorem 3 together with Theorem 1 allow us to tightly characterize the behavior of channel capacity in the first two 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 $\log\log \mathsf{SNR}$ result of [6] 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

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

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 V-B. □

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 this bound becomes loose and fails to characterize the $1 \cdot \log\log \mathsf{SNR}$ behavior. In fact, by [6, Theorem 4.29 (eq. 136)] applied to the first-order Gauss–Markov fading model one can obtain

$$\limsup_{\mathsf{SNR}\to\infty} \Big[C(\mathsf{SNR},\epsilon) - n_r n_t \log(1/\epsilon) + \log(\log \mathsf{SNR}) + K_4(n_r,n_t) \Big] \leq 0 \quad (12)$$

for some constant $K_4(n_r,n_t)$ independent of SNR and ϵ. This bound is loose however in the second regime because the coefficient of $\log(1/\epsilon)$ is $n_r n_t$ instead of $n_{\min}$. This bound has recently been improved in [23], from which it is possible to show that

$$\limsup_{\mathsf{SNR}\to\infty} \Big[C(\mathsf{SNR},\epsilon) - n_r \log(1/\epsilon) + \log(\log \mathsf{SNR}) + K_5(n_r,n_t) \Big] \leq 0. \quad (13)$$

Note, however, that for $n_t < n_r$, which is the case addressed in Theorem 4, this bound is still loose in the $\log(1/\epsilon)$ term. For this reason, we need the bound of Theorem 4 to characterize capacity in the first and second regimes for $n_t < n_r$.

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 I) 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.

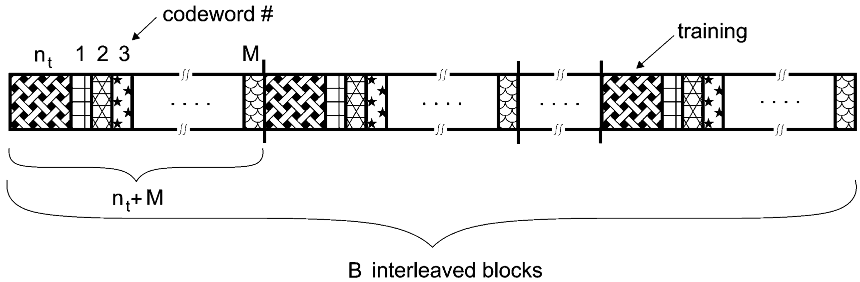

Fig. 3. Schematic representation of the transmission scheme. Symbols of the same codeword have the same shading. B is the codeword length, M the interleaver block length, and n_t the number of training symbols used in each interleaved block.

IV. 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 as follows.

- **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 nonnegligible 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 $\boldsymbol{I}_{n_t}$, 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. Fig. 3 shows a schematic representation of the transmission scheme.

Broadly speaking, the scheme consists of using the training sequences to obtain estimates of $\boldsymbol{H}[n]$ 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 $\boldsymbol{x}[n]$, $n = k(n_t+M)+n_t+1$ for $k = 0, 1, \ldots, B-1$, are known to the receiver and can be used as if they were training symbols to estimate $\boldsymbol{H}[n]$, $n = k(n_t+M)+n_t+2$ 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.

A. Channel Estimation

Consider the received vector when we send the first training symbol

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

where $\boldsymbol{h}_1[1]$ is the first column of $\boldsymbol{H}[1]$. The optimal MMSE estimate of $\hat{\boldsymbol{h}}_1[1]$ given $\boldsymbol{y}[1]$ is the conditional expectation $E\{\boldsymbol{h}_1[1]|\boldsymbol{y}[1]\}$. To simplify the analysis, we use the suboptimal estimate $\hat{\boldsymbol{h}}_1[1] = \boldsymbol{y}[1]$, 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

$$\hat{\boldsymbol{H}}[k\cdot L+n_t+1] = [\boldsymbol{y}[k\cdot L+1] \quad \boldsymbol{y}[k\cdot L+2] \quad \ldots \quad \boldsymbol{y}[k\cdot L+n_t]]$$

with

$$\boldsymbol{\Delta}[k\cdot L+n_t+1] = \boldsymbol{H}[k\cdot L+n_t+1] - \hat{\boldsymbol{H}}[k\cdot L+n_t+1]$$

as the corresponding estimation error. In order to characterize $\boldsymbol{\Delta}$, we will focus on the estimation error in $\hat{\boldsymbol{h}}_1[1]$ given by $\sqrt{n_t/\mathsf{SNR}}\boldsymbol{z}[1]$ which has covariance $\frac{n_t}{\mathsf{SNR}}\boldsymbol{I}_{n_r}$. 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 $\mathcal{CN}(\boldsymbol{0}, n_t\epsilon\boldsymbol{I}_{n_r})$ term.[7] Therefore, the overall estimation error of the first column of $\boldsymbol{H}$ at time n_t+1 is circularly symmetric complex Gaussian with covariance $n_t(\mathsf{SNR}^{-1}+\epsilon)\boldsymbol{I}_{n_r}$. Also, the estimates of the other columns of $\boldsymbol{H}$, i.e., $\hat{\boldsymbol{h}}_i[n_t+1] = \boldsymbol{y}[i]$, 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 $\boldsymbol{H}$ by $n_t(\mathsf{SNR}^{-1}+\epsilon)$. We also note that the

[7]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.

estimation error $\boldsymbol{\Delta}$ is circularly symmetric complex Gaussian with independent components.

Consider the problem of estimating $\boldsymbol{H}[n+1]$ based on a previous estimate $\hat{\boldsymbol{H}}[n]$, the currently received vector $\boldsymbol{y}[n]$, and the successfully decoded symbol $\boldsymbol{x}[n]$. 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 $\boldsymbol{x}[n]$, the estimation of the different rows of $\boldsymbol{H}[n]$ decouples into n_r independent estimation problems. This result requires the independence of the rows of $\boldsymbol{\Delta}[n_t+1]$ which was shown earlier. Therefore, without loss of generality, we will consider the estimation of the first row of $\boldsymbol{H}$, which we will denote by $\boldsymbol{h}^T$.

Assume that at time n we have an estimate $\hat{\boldsymbol{h}}[n]$ of $\boldsymbol{h}[n]$, and that the estimation error $\boldsymbol{\delta}[n] = \boldsymbol{h}[n] - \hat{\boldsymbol{h}}[n]$ has zero mean and covariance $\boldsymbol{K}[n]$. We also assume that $\boldsymbol{x}[n]$ 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

$$\hat{\boldsymbol{h}}[n+1] = E\left[\boldsymbol{h}[n+1] |\, y[n], \boldsymbol{x}[n], \hat{\boldsymbol{h}}[n]\right]$$

and the covariance matrix of the corresponding estimation error, $\boldsymbol{K}[n+1]$. We use $\boldsymbol{w}[n]^T$ to denote the first row of $\boldsymbol{W}[n]$, and consider

$$\begin{aligned}\boldsymbol{h}[n+1] &= \sqrt{1-\epsilon}\boldsymbol{h}[n] + \sqrt{\epsilon}\boldsymbol{w}[n] \\ &= \sqrt{1-\epsilon}\left(\hat{\boldsymbol{h}}[n] + \boldsymbol{\delta}[n]\right) + \sqrt{\epsilon}\boldsymbol{w}[n]\end{aligned}$$

to get the updated estimate

$$\begin{aligned}\hat{\boldsymbol{h}}[n+1] &= E\left[\boldsymbol{h}[n+1] | y[n], \boldsymbol{x}[n], \hat{\boldsymbol{h}}[n]\right] \\ &= \sqrt{1-\epsilon}\hat{\boldsymbol{h}}[n] + \sqrt{1-\epsilon}E\left[\boldsymbol{\delta}[n] | y[n], \boldsymbol{x}[n], \hat{\boldsymbol{h}}[n]\right]\end{aligned} \quad (14)$$

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

$$\begin{aligned}\boldsymbol{K}[n+1] &= \mathrm{Cov}(\boldsymbol{\delta}[n+1] | y[n], \boldsymbol{x}[n], \hat{\boldsymbol{h}}[n]) \\ &= (1-\epsilon)\mathrm{Cov}(\boldsymbol{\delta}[n] | y[n], \boldsymbol{x}[n], \hat{\boldsymbol{h}}[n]) + \epsilon \boldsymbol{I}_{n_t}\end{aligned} \quad (15)$$

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

To compute the expectation in (14) and covariance in (15), we need the joint distribution of $[y[n]\boldsymbol{\delta}[n]^T]^T$ conditioned on $\hat{\boldsymbol{h}}[n]$ and $\boldsymbol{x}[n]$. This distribution is complex Gaussian, so we only need to compute the first and second moments.[8] Letting $z[n]$ denote the first element of $\boldsymbol{z}[n]$ we can write

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

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

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

It follows that

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

and

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

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

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

We finally replace in (14) and (15) obtaining

$$\begin{aligned}\hat{\boldsymbol{h}}[n+1] &= \sqrt{1-\epsilon}\hat{\boldsymbol{h}}[n] + \sqrt{1-\epsilon} \\ &\quad \cdot \left[\boldsymbol{K}[n]\boldsymbol{x}[n]^*\left(\boldsymbol{x}[n]^T\boldsymbol{K}[n]\boldsymbol{x}[n]^* + \frac{n_t}{\mathsf{SNR}}\right)^{-1} \right. \\ &\qquad \left. \cdot \left(y[n] - \hat{\boldsymbol{h}}[n]^T\boldsymbol{x}[n]\right)\right] \\ \boldsymbol{K}[n+1] &= (1-\epsilon)\left[\boldsymbol{K}[n] - \left(\boldsymbol{x}[n]^T\boldsymbol{K}[n]\boldsymbol{x}[n]^* + \frac{n_t}{\mathsf{SNR}}\right)^{-1} \right. \\ &\qquad \left. \cdot \boldsymbol{K}[n]\boldsymbol{x}[n]^*\boldsymbol{x}[n]^T\boldsymbol{K}[n]\right] + \epsilon\boldsymbol{I}_{n_t}.\end{aligned} \quad (16)$$

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

B. 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 $\{\boldsymbol{y}[n] : n \in \mathcal{N}_m\}$ and the estimates of the channel fading matrix $\{\hat{\boldsymbol{H}}[n] : n \in \mathcal{N}_m\}$. Define $\boldsymbol{Y}_m$ to be the matrix whose columns are the received vectors corresponding to the mth codeword, and

[8]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.

$\boldsymbol{V}_k$ to be the matrix whose columns are the vectors of the form $\hat{\boldsymbol{H}}[n]\boldsymbol{x}_k[r]$ for $n \in \mathcal{N}_m$ and $r \in \{1, 2, \dots, B\}$, where k denotes the codeword index in the codebook. Then the decoding rule for the mth codeword is given by

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

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 were 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, MWEDD 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.

C. 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 $\{\boldsymbol{\Delta}[k(n_t + M) + n_t + m]\}_{k=0}^{B-1}$ and the channel estimates $\{\hat{\boldsymbol{H}}[k(n_t + M) + n_t + m]\}_{k=0}^{B-1}$ become independent over k, so the mth transmitted codeword sees an equivalent channel given by

$$\begin{aligned}\boldsymbol{y}[n] &= \hat{\boldsymbol{H}}[n]\boldsymbol{x}[n] + \boldsymbol{\Delta}[n]\boldsymbol{x}[n] + \sqrt{\frac{n_t}{\mathsf{SNR}}}\boldsymbol{z}[n] \\ &= \hat{\boldsymbol{H}}[n]\boldsymbol{x}[n] + \tilde{\boldsymbol{z}}[n] \qquad (17)\end{aligned}$$

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

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

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

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

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

Lemma 1: $E[\tilde{\boldsymbol{z}}[n]\tilde{\boldsymbol{z}}[n]^\dagger] = \sigma^2[n]\boldsymbol{I}_{n_r}$ where, for n large enough so that $E[\mathrm{tr}(\boldsymbol{K}[n])]$ 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 $\tilde{\boldsymbol{z}}[n]$, we now turn our attention to the estimate of the fading matrix $\hat{\boldsymbol{H}}[n]$ which is known at the receiver. As we argued before, conditioned on $\{\boldsymbol{x}[n]\}$, the rows of $\hat{\boldsymbol{H}}[n]$ 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 $\{\boldsymbol{x}[n]\}$. It follows that in steady state we can write

$$E\left[\hat{\boldsymbol{h}}_i^\dagger[n]\hat{\boldsymbol{h}}_j[n]\right] = \alpha^2\delta_{i,j}$$

for some constant α^2, where $\hat{\boldsymbol{h}}_i^T$ is the ith row of $\hat{\boldsymbol{H}}[n]$.

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

As $\mathsf{SNR} \to \infty$ and $\epsilon \to 0$, the estimate $\hat{\boldsymbol{H}}[n]$ converges to $\boldsymbol{H}[n]$ with probability 1, so it follows that $\hat{\boldsymbol{H}}[n]$ also converges in distribution to $\boldsymbol{H}[n]$. Therefore, in the limit as $\mathsf{SNR} \to \infty$ and $\epsilon \to 0$, $\hat{\boldsymbol{H}}[n]$ 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 (17) 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 [18] for noise not independent with the channel input.

Theorem 5: Let C denote the capacity of the channel

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

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

Consider the channel

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

where $\hat{\boldsymbol{H}}[n] \in \mathbb{C}^{n_r\times n_t}$ is the same as before, and is also known at the receiver; $\tilde{\boldsymbol{z}}[n]$ is independent over time, is uncorrelated with the input $\boldsymbol{x}[n]$, is independent of $\hat{\boldsymbol{H}}[n]$, and has the same covariance as $\boldsymbol{z}[n]$; and $\boldsymbol{x}[n] \in \mathbb{C}^{n_t}$ is independent of $\hat{\boldsymbol{H}}[n]$ and is subject to the power constraint $E[\|\boldsymbol{x}[n]\|^2] \leq n_t$. Let R be the supremum of the achievable rates in this channel when the input $\boldsymbol{x}[n]$ 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

$$R < E\left[\log\det\left(\boldsymbol{I}_{n_r} + \frac{\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^\dagger}{\sigma^2}\right)\right]$$

where the expectation is taken over the random matrix $\hat{\boldsymbol{H}}$. For large SNR and small ϵ, we can do the following approximations:

$$\begin{aligned} &E\left[\log\det\left(\boldsymbol{I}_{n_r} + \frac{\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^\dagger}{\sigma^2}\right)\right] \\ &\quad \geq E\left[\log\det\left(\frac{\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^\dagger}{\sigma^2}\right)\right] \\ &\quad = \min\{n_t, n_r\}\log\left(\frac{1}{\sigma^2}\right) + E\left[\log\det\left(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^\dagger\right)\right] \\ &\quad \approx \min\{n_t, n_r\}\log\left(\frac{1}{\sigma^2}\right) + E\left[\log\det\left(\boldsymbol{H}\boldsymbol{H}^\dagger\right)\right] \\ &\quad \sim \min\{n_t, n_r\}\log\left[\left(\mathsf{SNR}^{-1} + \epsilon\right)^{-1}\right] \end{aligned} \tag{18}$$

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 $\min\{n_r, n_t\}\log[(\mathsf{SNR}^{-1} + \epsilon)^{-1}]$ for $\mathsf{SNR} \to \infty$ and $\epsilon \to 0$. Thus, our scheme achieves the full degrees of freedom of the channel.

D. Practical Considerations

Many of the elements of the scheme analyzed in this section are very useful in practice and have been considered in the literature. For example, Weingarten *et al.* analyze in [19] the performance of Gaussian codes and a more general version of the weighted minimum Euclidean distance decoding rule that we considered here and apply the results to a block-fading channel model. Baltersee *et al.* use in [20] a perfect interleaver to justify the assumption of independent channel realizations in the different coding blocks, and compute the achievable rates for an optimal decoder.

For practical systems, decoding delay is an important parameter that must be minimized. One way to reduce the interleaving delay is to use per-survivor processing [21] (PSP) to obtain reliable decisions for channel estimation before the reception of the complete codewords. A practical example of noncoherent sequence detection using PSP can be found in [22].

V. Proofs of the Upper Bounds

In this section, we will present the proofs of Theorems 3 and 4. We first prove Theorem 3 in Section V-A introducing many results that are also useful in the proof of Theorem 4. Later, in Section V-B, 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
$\tilde{\boldsymbol{y}}[n]$	$\boldsymbol{H}[n]\boldsymbol{x}[n]$
$m[n]$	$\lvert\boldsymbol{x}[n]\rvert$
$\boldsymbol{d}[n]$	$\boldsymbol{x}[n]/\lvert\boldsymbol{x}[n]\rvert$
$\boldsymbol{c}[n]$	$\boldsymbol{H}[n]\boldsymbol{d}[n]$
$\rho_i[n]$	$\log(\lvert\tilde{y}_i[n]\rvert)$
$\phi_i[n]$	$\angle\tilde{y}_i[n]$
$r_i[n]$	$\tilde{y}_i[n]/\tilde{y}_1[n]$
$\tilde{w}_i[1]$	$(\lvert h_i[1]\rvert/h_i[1])w_i[1]$

.

A. 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 $\boldsymbol{H}^{\mathcal{N}}$ is perfectly known at the receiver. The capacity under this assumption was explicitly computed in [2], from which it follows that $C_{\text{Known}\,\boldsymbol{H}} \approx n_{\min}\log\mathsf{SNR} + K$ 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 $C(\mathsf{SNR}, \epsilon) \to \infty$. It is shown in [6] that for a memoryless channel it is possible to 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

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

Then

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

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

$$\begin{aligned} \|\boldsymbol{x}[n]\| &\ge \mathsf{SNR}^{-1/2}, \qquad n\in\mathcal{N} \\ \frac{1}{B}\sum_{n\in\mathcal{N}} E[\|\boldsymbol{x}[n]\|^2] &\le n_t + \frac{1}{\mathsf{SNR}}. \end{aligned} \tag{19}$$

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

$$\tilde{\boldsymbol{y}}[n] = \boldsymbol{H}[n]\boldsymbol{x}[n] = \boldsymbol{y}[n] - \sqrt{\frac{n_t}{\mathsf{SNR}}}\boldsymbol{z}[n], \qquad \forall\, n\in\mathcal{N}. \tag{20}$$

After removing the noise we get the multiplicative channel (20).

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 $I(\boldsymbol{x}^{\mathcal{N}};\tilde{\boldsymbol{y}}^{\mathcal{N}})$ in terms of differential entropies

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

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 $\boldsymbol{x} = (x_1,\ldots,x_n)^T$ let $\log(\boldsymbol{x})$ 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

$$\begin{aligned} I(\boldsymbol{x}^{\mathcal{N}};\tilde{\boldsymbol{y}}^{\mathcal{N}}) &= I(\log(\boldsymbol{x}^{\mathcal{N}});\log(\tilde{\boldsymbol{y}}^{\mathcal{N}})) \\ &= h(\log(\tilde{\boldsymbol{y}}^{\mathcal{N}})) - h(\log(\tilde{\boldsymbol{y}}^{\mathcal{N}})|\log(\boldsymbol{x}^{\mathcal{N}})) \\ &= h(\log(\tilde{\boldsymbol{y}}^{\mathcal{N}})) - h(\log(\boldsymbol{h}^{\mathcal{N}})). \end{aligned}$$

Note that the second term of the right-hand side (RHS) is independent of the input distribution, and, in principle, can be computed. To bound the term $h(\log(\tilde{\boldsymbol{y}}^{\mathcal{N}}))$ we can find an upper bound on the variance of the components of $\log(\tilde{\boldsymbol{y}}^{\mathcal{N}})$ 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 $h(\log(\boldsymbol{h}^{\mathcal{N}}))$ is simplified by expressing it in terms of $h(\boldsymbol{h}^{\mathcal{N}})$, 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 $h(\boldsymbol{h}^{\mathcal{N}})$ 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 $h(\boldsymbol{h}^{\mathcal{N}})$. 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 $h(\log(\boldsymbol{h}^{\mathcal{N}}))$.

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 $m[n] = \|\boldsymbol{x}[n]\|$, the norm of the input, and $\boldsymbol{d}[n] = \boldsymbol{x}[n]/\|\boldsymbol{x}[n]\|$, 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 $\boldsymbol{c}[n] = \boldsymbol{H}[n]\boldsymbol{d}[n]$ 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 $\tilde{\boldsymbol{y}}[n]$, and $\pi < \phi_i[n] \le \pi$ for $n\in\mathcal{N}$, and group all this variables into the Bn_r-dimensional vectors $\boldsymbol{\rho}^{\mathcal{N}}$ and $\boldsymbol{\phi}^{\mathcal{N}}$. 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 $\boldsymbol{v}$ and $\boldsymbol{w}$ be related by a one-to-one transformation with Jacobian $\boldsymbol{J}$ where $J(i,j) = \partial w_i/\partial v_j$. Then

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

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

Lemma 2:

$$\begin{aligned} I(\boldsymbol{x}^{\mathcal{N}};\tilde{\boldsymbol{y}}^{\mathcal{N}}) &= h(\boldsymbol{\rho}^{\mathcal{N}},\boldsymbol{\phi}^{\mathcal{N}}) - h(\boldsymbol{c}^{\mathcal{N}}|\boldsymbol{d}^{\mathcal{N}}) \\ &\quad + \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} E\left[\log\left(|c_i[n]|^2\right)\right] \end{aligned} \tag{21}$$

where $c_i[n]$ is the ith component of $\boldsymbol{c}[n]$.

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 (21). This is done in the following three lemmas.

Lemma 3:

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

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 $c_i[n] = \boldsymbol{h}_i[n]^T\boldsymbol{d}[n]$, where $\boldsymbol{h}_i[n]^T$ is the ith row of $\boldsymbol{H}[n]$. Since $\boldsymbol{d}[n]$ is a norm 1 vector and $\boldsymbol{h}_i[n]^T$ has i.i.d. $\mathcal{CN}(0,1)$ components, conditioned on $\boldsymbol{d}[n]$, $c_i[n]$ has $\mathcal{CN}(0,1)$ distribution independent of $\boldsymbol{d}[n]$. It follows that without conditioning, $c_i[n]$ has $\mathcal{CN}(0,1)$ distribution, and therefore $|c_i[n]|^2 \sim \text{Exp}(1)$. Then we can compute $E\left[\log\left(|c_i[n]|^2\right)\right]$ explicitly

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

independent of n and i. □

Lemma 4:

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

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

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

Conditioned on $(\boldsymbol{H}[n-1],\boldsymbol{d}[n])$

$$\boldsymbol{c}[n] \sim \mathcal{CN}(\sqrt{1-\epsilon}\boldsymbol{H}[n-1]\boldsymbol{d}[n], \epsilon\boldsymbol{I}_{n_r})$$

and it follows that[9]

$$\begin{aligned} h(\boldsymbol{c}[1]|\boldsymbol{H}[0],\boldsymbol{d}^{\mathcal{N}}) &= h(\boldsymbol{c}[n]|\boldsymbol{H}[n-1],\boldsymbol{c}[n-1],\ldots,\boldsymbol{c}[1],\boldsymbol{d}^{\mathcal{N}}) \\ &= n_r\log(\pi e\epsilon). \end{aligned}$$

To see this, we write $\boldsymbol{c}[n]$ as a function of $(\boldsymbol{H}[n-1],\boldsymbol{d}[n])$ for $n\in\mathcal{N}$

$$\begin{aligned} \boldsymbol{c}[n] &= \boldsymbol{H}[n]\boldsymbol{d}[n] \\ &= \left(\sqrt{1-\epsilon}\boldsymbol{H}[n-1] + \sqrt{\epsilon}\boldsymbol{W}[n]\right)\boldsymbol{d}[n] \\ &= \sqrt{1-\epsilon}\boldsymbol{H}[n-1]\boldsymbol{d}[n] + \sqrt{\epsilon}\tilde{\boldsymbol{w}}[n] \end{aligned}$$

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

Lemma 5:

$$\begin{aligned} \frac{1}{B}h(\boldsymbol{\rho}^{\mathcal{N}},\boldsymbol{\phi}^{\mathcal{N}}) &\leq \frac{1}{2}\log(2\pi e\sigma^2) + \frac{1}{2}\log(2\pi e\pi^2) \\ &\quad + (n_r-1)[2-\log(2)] + (n_r-1)\frac{1}{2}\log\left[2\pi e(2\pi)^2\right] \end{aligned}$$

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. □

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

We replace in (21) the results of the last three lemmas to obtain the upper bound

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

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

$$\begin{aligned} K(n_r) &= \log(2\pi e\pi) + (n_r-1)[\log(1/2)+2] \\ &\quad + (n_r-1)\frac{1}{2}\log(2\pi e(2\pi)^2) - n_r\gamma - n_r\log(\pi e). \end{aligned}$$

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

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

As $\mathsf{SNR}\to\infty$, $(1/2)\log(\sigma^2) - (\log\log\mathsf{SNR} - \log 2) \to 0$, so for $n_r \leq n_t$, 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 $\min\{n_r,n_t\} = n_t$.

B. Proof of Theorem 4

As mentioned in the introduction of this section, many of the results of Section V-A 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 $\frac{1}{B}h(\boldsymbol{\rho}^{\mathcal{N}},\boldsymbol{\phi}^{\mathcal{N}})$ should contain a term of the form $(n_r - \min\{n_t,n_r\})\log\epsilon$ which is nonzero 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

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

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 2–4 allow us to obtain a tight upper bound for the capacity of the single transmit antenna channel

$$\limsup_{\substack{\mathsf{SNR}\to\infty\\ \epsilon\to 0}}\left[C(\mathsf{SNR},\epsilon) - \left\{\log\log\mathsf{SNR} + \log(1/\epsilon) + K(n_r)\right\}\right] \leq 0 \quad (22)$$

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

An upper bound for the capacity of the MIMO channel can be obtained by rewriting the channel (1) as the sum of the outputs of n_t independent single-input multiple-output (SIMO) ($n_t = 1$) channels as represented in Fig. 4. Let $\boldsymbol{h}_i[n]\in\mathbb{C}^{n_r}$ be the ith column of $\boldsymbol{H}[n]$, $x_i[n]$ be the ith component of $\boldsymbol{x}[n]$, and $\boldsymbol{z}^{(i)}[n] \sim \mathcal{CN}(0,I_{n_r})$, $i = 1,2,\ldots,n_t$, $n\in\mathcal{N}$, be independent

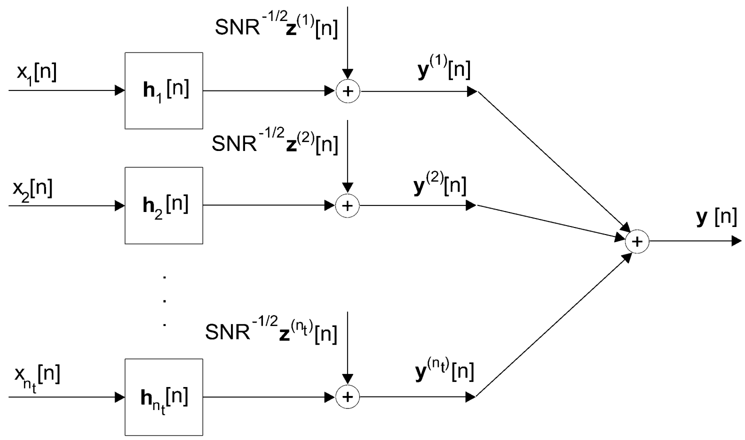

Fig. 4. Representation of the MIMO channel as the sum of the outputs of n_t independent SIMO channels.

across i and n, and independent of the channel input and fading gains. Define n_t independent SIMO channels by

$$\boldsymbol{y}^{(i)}[n] = \boldsymbol{h}_i[n]x_i[n] + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}^{(i)}[n], \qquad i = 1, 2, \ldots, n_t. \tag{23}$$

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

$$\begin{aligned}\boldsymbol{y}[n] &= \sum_{i=1}^{n_t} \boldsymbol{y}^{(i)}[n] \\ &= \sum_{i=1}^{n_t} \boldsymbol{h}_i[n]x_i[n] + \sqrt{\frac{n_t}{\mathsf{SNR}}}\frac{1}{\sqrt{n_t}}\sum_{i=1}^{n_t}\boldsymbol{z}^{(i)}[n] \\ &= H[n]\boldsymbol{x}[n] + \sqrt{\tfrac{n_t}{\mathsf{SNR}}}\boldsymbol{z}[n], \qquad n \in \mathcal{N} \end{aligned} \tag{24}$$

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

Let

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

Then for any distribution of $\boldsymbol{x}^{\mathcal{N}}$, 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

$$\begin{aligned} I(\boldsymbol{x}^{\mathcal{N}}; \boldsymbol{y}^{\mathcal{N}}) &\le I(\boldsymbol{x}^{\mathcal{N}}; \boldsymbol{y}^{(1)\mathcal{N}}, \boldsymbol{y}^{(2)\mathcal{N}}, \ldots, \boldsymbol{y}^{(n_t)\mathcal{N}}) \\ &= h(\boldsymbol{y}^{(1)\mathcal{N}}, \boldsymbol{y}^{(2)\mathcal{N}}, \ldots, \boldsymbol{y}^{(n_t)\mathcal{N}}) \\ &\quad - h(\boldsymbol{y}^{(1)\mathcal{N}}, \boldsymbol{y}^{(2)\mathcal{N}}, \ldots, \boldsymbol{y}^{(n_t)\mathcal{N}} \mid \boldsymbol{x}^{\mathcal{N}}) \\ &= h(\boldsymbol{y}^{(1)\mathcal{N}}) + \sum_{i=2}^{n_t} h(\boldsymbol{y}^{(i)\mathcal{N}} \mid \boldsymbol{y}^{(i-1)\mathcal{N}}, \ldots, \boldsymbol{y}^{(1)\mathcal{N}}) \\ &\quad - \sum_{i=1}^{n_t} h(\boldsymbol{y}^{(i)\mathcal{N}} \mid \boldsymbol{x}^{\mathcal{N}}) \\ &\le \sum_{i=1}^{n_t} h(\boldsymbol{y}^{(i)\mathcal{N}}) - \sum_{i=1}^{n_t} h(\boldsymbol{y}^{(i)\mathcal{N}} \mid \boldsymbol{x}_i^{\mathcal{N}}) \\ &= \sum_{i=1}^{n_t} I(\boldsymbol{x}_i^{\mathcal{N}}; \boldsymbol{y}^{(i)\mathcal{N}}). \end{aligned} \tag{25}$$

This upper bound can be better understood by interpreting $\boldsymbol{y}^{(i)}[n]$ 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

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

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

$$\frac{1}{B}\sum_{n\in\mathcal{N}} E\left[|x_i[n]|^2\right] \le n_t.$$

Therefore, we can bound each of the n_t terms of (25) by the bound (22) 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

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

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 $\mathsf{SNR} > 1/\epsilon$.

VI. Channels With Band-Limited Nonregular Fading Processes

A. Capacity Lower Bound

The results presented so far rely on the first-order Gauss–Markov fading model defined in Section II. The first-order Gauss–Markov fading process is regular: the mean-squared estimation error (MSE) of the present of the process using noiseless observations of the infinite past is positive.

The noiseless MSE is given by

$$\epsilon(0) = \exp\left\{\int_{-1/2}^{1/2}\log[S_h(f)]df\right\} \tag{26}$$

where $S_h(f)$ is the power spectral density of the discrete-time fading process $\{h[n]\}$. In the case of an $n_t \times n_r$ MIMO system, there are $n_r n_t$ independent such processes.

If the process is nonregular, the present value can be perfectly estimated from noiseless observations of the infinite past. In this case, $\epsilon(0) = 0$. However, if the observed samples are corrupted by i.i.d. additive Gaussian noise of variance δ^2, the resulting MSE is in general greater than zero, and is given by

$$\epsilon(\delta^2) = \exp\left\{\int_{-1/2}^{1/2} \log[S_h(f) + \delta^2] df\right\} - \delta^2. \tag{27}$$

Lapidoth in [8], [9] analyzes the capacity of SISO channels with nonregular fading processes where the input of the channel is subject to a peak power constraint. He shows that depending on how $\epsilon(\delta^2)$ goes to 0 when $\delta^2 \to 0$ the high SNR capacity of the channel can be $\log \mathsf{SNR}$, $\log\log \mathsf{SNR}$, or $(\log \mathsf{SNR})^\beta$ for $0 < \beta < 1$.

It follows that regular and nonregular fading processes lead to totally different high SNR capacity asymptotics. This motivates the question on how the degrees of freedom of a MIMO channel influence the high SNR capacity when the fading is nonregular.

To this end, in the following theorem we present a capacity lower bound that can be applied to any MIMO flat-fading channel with spatially i.i.d. circularly symmetric complex Gaussian entries in the fading matrix.

Theorem 7: The capacity $C(\mathsf{SNR})$ of a MIMO fading channel with an $n_r \times n_t$ circularly symmetric complex Gaussian fading matrix with spatially i.i.d. components of variance 1, when the fading realization is unknown to both transmitter and receiver can be lower-bounded as

$$C(\mathsf{SNR}) \geq \min\{n_r, n_t\}\left(1 - \min\{n_r, n_t\}|\mathcal{D}|\right)\log(\mathsf{SNR}) + K_6(n_r, n_t) \tag{28}$$

for some constant $K_6(n_r, n_t)$ independent of SNR, where $\mathcal{D}$ is the set of frequencies where the power spectral density $S_h(f)$ of each fading process of the fading matrix is greater than zero

$$\mathcal{D} = \left\{f : S_h(f) > 0, -\frac{1}{2} \leq f \leq \frac{1}{2}\right\}$$

and $|\mathcal{D}|$ denotes the Lebesgue measure of this set.

Proof: The proof of the theorem is very similar to the proof of Theorem 2 given in Appendix A. In fact, most of the steps of the proof apply to any channel with spatially i.i.d. circularly symmetric complex Gaussian fading processes, regardless of the time correlation structure.

We first note that we can use $\tilde{n}_r \leq n_r$ receive and $\tilde{n}_t \leq n_t$ transmit antennas to achieve a lower bound on capacity. We use as input to the channel $\boldsymbol{x}^{\mathcal{N}} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_B)$ and apply Lemmas 7 and 8 to obtain

$$C(\mathsf{SNR}) \geq E\left[\sum_{k=n_{\max}-(n_{\min}-1)}^{n_{\max}} \log\left(1 + \frac{\mathsf{SNR}}{\tilde{n}_t}\chi^2_{2k}\right)\right] - \tilde{n}_r\tilde{n}_t \int_0^1 \log\left[\frac{\mathsf{SNR}}{\tilde{n}_t} S_h(f) + 1\right] df \tag{29}$$

where $n_{\max} = \max\{\tilde{n}_r, \tilde{n}_t\}$, $n_{\min} = \min\{\tilde{n}_r, \tilde{n}_t\}$, and χ^2_{2k} is a χ^2 random variable with $2k$ degrees of freedom.

We now use (27) to express the last term of (29) in terms of the noisy prediction error of the fading process

$$\log\left[\epsilon(\delta^2) + \delta^2\right] = \int_0^1 \log\left[S_h(f) + \delta^2\right] df \tag{30}$$

where we used the periodicity of $S_h(f)$. It follows that

$$C(\mathsf{SNR}) \geq \sum_{k=n_{\max}-(n_{\min}-1)}^{n_{\max}} E\left[\log\left(1 + \frac{\mathsf{SNR}}{\tilde{n}_t}\chi^2_{2k}\right)\right] - \tilde{n}_r\tilde{n}_t \log\left(\frac{\mathsf{SNR}}{\tilde{n}_t}\right) - \tilde{n}_r\tilde{n}_t \log\left[\epsilon\left(\frac{\tilde{n}_t}{\mathsf{SNR}}\right) + \frac{\tilde{n}_t}{\mathsf{SNR}}\right]. \tag{31}$$

To obtain a lower bound that depends on $|\mathcal{D}|$ instead of the noisy prediction error variance $\epsilon(\cdot)$ we use (30) and apply Jensen's inequality:

$$\begin{aligned}\log\left[\epsilon(\delta^2) + \delta^2\right] &= \int_0^1 \log\left[S_h(f) + \delta^2\right] df \\ &= |\mathcal{D}| \int_{\mathcal{D}} \frac{1}{|\mathcal{D}|} \log\left[S_h(f) + \delta^2\right] df \\ &\quad + (1 - |\mathcal{D}|)\log\left(\delta^2\right) \\ &\leq |\mathcal{D}| \log\left\{\int_{\mathcal{D}} \frac{1}{|\mathcal{D}|}\left[S_h(f) + \delta^2\right] df\right\} \\ &\quad + (1 - |\mathcal{D}|)\log\left(\delta^2\right) \\ &= |\mathcal{D}| \log\left(\frac{1}{|\mathcal{D}|} + \delta^2\right) + (1 - |\mathcal{D}|)\log\left(\delta^2\right)\end{aligned} \tag{32}$$

where the last equality follows from the unit variance of the fading coefficients.

Evaluating (32) in $\delta^2 = \tilde{n}_t/\mathsf{SNR}$ and replacing in (31) we obtain

$$C(\mathsf{SNR}) \geq \sum_{k=n_{\max}-(n_{\min}-1)}^{n_{\max}} E\left[\log\left(1 + \frac{\mathsf{SNR}}{\tilde{n}_t}\chi^2_{2k}\right)\right] - \tilde{n}_r\tilde{n}_t|\mathcal{D}| \log\left(\frac{\mathsf{SNR}}{\tilde{n}_t}\right) - \tilde{n}_r\tilde{n}_t|\mathcal{D}| \log\left(\frac{1}{|\mathcal{D}|} + \frac{\tilde{n}_t}{\mathsf{SNR}}\right). \tag{33}$$

For high SNR, the first term of the lower bound (33) grows as $n_{\min} \log \mathsf{SNR}$ and the last term approaches a constant, so by taking $\tilde{n}_r = \tilde{n}_t = \min\{n_r, n_t\}$ the statement of the theorem results. □

In Appendix I, we derive the power spectral density of a discrete-time flat-fading channel using Jake's model [10]. For this derivation, we assume that the signal bandwidth W is much smaller than the delay spread of the channel T_d (flat fading assumption), and that both the input and output of the channel are sampled at a rate W.[10] Since the channel can be interpreted as a time-varying system, the signal at the output may have a bandwidth larger than the signal at the input, in which case sampling at a rate W may incur in loss of information. However, since we only obtain a lower bound in this section, the approach

[10]Note that the baseband equivalent input is band-limited to a bandwidth $W/2$.

of sampling at rate W is valid. Under these assumptions, we show that $|\mathcal{D}| = 2f_m/W$ where f_m is the maximum Doppler shift. For underspread channels $2f_m/W_c \ll 1$ and W, can be chosen large enough to obtain $|\mathcal{D}| \ll 1$ while satisfying the flat fading condition. Therefore, the number of degrees of freedom $\min\{n_r, n_t\}(1-\min\{n_r, n_r\}|\mathcal{D}|)$ is slightly reduced from the perfect CSI at the receiver case. Note that the CSI at the receiver assumption gives an upper bound to the noncoherent channel capacity. Recently, it has been shown in [23] that the pre-log, that is, the asymptotic ratio of the capacity to $\log \mathsf{SNR}$, can be upper-bounded by

$$\lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log(\mathsf{SNR})} \leq n_r(1-|\mathcal{D}|)$$

which for $n_r \leq n_t$ is at least as tight as the perfect CSI bound.

We see that for band-limited MIMO underspread channels with nonregular fading statistics, the number of degrees of freedom $\min\{n_r, n_t\}$ plays an essential role in channel capacity and provides significant capacity improvements.

B. Relation to the Capacity of the Liang–Veeravalli Model

Our results for the capacity of the MIMO underspread flat-fading channel with band-limited nonregular fading via the lower bound (28) can be compared to the results obtained in [5] for the time-selective block-fading model.

Liang and Veeravalli [5] analyze a model where the discrete time is divided into blocks of length T. In the traditional block-fading model [3], the fading matrix $\boldsymbol{H} = [h_{ij}]$ is constant within each block and varies to an independent realization in different blocks. In contrast, in the time-selective block-fading model, the fading matrix is allowed to vary within each block and varies to independent or correlated realizations in different blocks. One can define a vector $\boldsymbol{h}_{ij} = [h_{ij}[1], \ldots, h_{i,j}[T]]^T$ that groups the (i,j)th entries of the fading matrix for the different times within a block. The corresponding covariance matrix $\boldsymbol{\Sigma}_h = E[\boldsymbol{h}_{ij}\boldsymbol{h}_{ij}^\dagger]$ has rank Q. $\boldsymbol{H}$ is assumed to have spatially i.i.d. entries, and hence the distribution of $\boldsymbol{h}_{ij}$ is the same for all (i,j).

It is argued that for flat, slow fading using Jakes' model for the fading correlation $Q \ll T$ and that if Q components of $\boldsymbol{h}_{ij}$ are available, near-perfect prediction of the remaining channel gains is possible. This models the situation of a nonregular fading process where noiseless observations of the infinite past are enough to perfectly predict the future evolution of the process.

For the case of the MIMO channel they obtained the capacity lower bound

$$C_{\text{lower}} = T^*\left(1-\frac{T^*Q}{T}\right)\log \mathsf{SNR} + c_1 + o(1) \qquad (34)$$

where

$$T^* = \min\left\{n_r, n_t, \left\lfloor \frac{T}{2Q} \right\rfloor\right\}$$

and c_1 is a constant that does not depend on SNR. They conjectured that this lower bound is tight in the leading term, and gave an intuitive argument to justify the conjecture: if n_t transmit and n_r receive antennas are used, one has to lose n_tQ symbol periods to estimate the n_rn_tQ fading gains in the first Q channel matrices in each block. Then the channel matrices in the entire block are known and the remaining symbol periods can be used coherently to achieve a rate

$$\min\{n_r, n_t\}\left(1-n_t\frac{Q}{T}\right)\log \mathsf{SNR}.$$

There is an analogy between the lower bound (34) for the time-selective block-fading model and our lower bound (28) obtained for the discrete-time continuous fading model. In fact, $|\mathcal{D}|$ plays the role of $\frac{Q}{T}$ and represents the "predictability" of the channel. Channels with smaller values of $|\mathcal{D}|$ or $\frac{Q}{T}$ are easier to predict, and using the above intuitive argument, require fewer symbol times to obtain accurate predictions.

VII. Discussion and Comments

A. The Gauss–Markov Model

We used a first-order Gauss–Markov fading model to analyze the high SNR capacity of regular underspread MIMO fading channels when the fading realization is unknown to both transmitter and receiver. This model allowed us to obtain lower and upper bounds on channel capacity, which we used to define three regimes of operation.

Our intuition suggests, in a sense that needs to be made precise, that these three regimes should appear in all Gaussian MIMO channels with regular fading processes. The Gauss–Markov model captures through its parameter ϵ an essential element of all regular fading processes: the nonzero noiseless prediction error. The impossibility of perfectly predicting the fading from noiseless observations of its past, becomes the capacity limiter in the second regime, and produces the collapse of the degrees of freedom of the channel in the third regime.

We used the Gauss–Markov model fitted to physical parameters of real communications systems to quantify the boundaries between the three regimes, and observed that the appearance of the third regime required extremely large SNR values in underspread channels. These boundaries are likely to change for higher order Gauss–Markov fading models, or other regular fading processes.

B. The Fading Number

Lapidoth and Moser define in [6] the fading number χ

$$\chi(\{\boldsymbol{H}[n]\}_n) = \limsup_{\mathsf{SNR}\to\infty} \{C(\mathsf{SNR}) - \log\log \mathsf{SNR}\}.$$

Note that implicit in the definition of χ is the fact that the statistics of the fading process $\{\boldsymbol{H}[n]\}_n$ are kept fixed, while the SNR goes to infinity. In the context of the Gauss–Markov fading model, this is equivalent to keeping ϵ fixed when taking the limit.

The fading number corresponds to the second term of the high SNR capacity expansion, in the regime where the $\log\log \mathsf{SNR}$ is the leading term. That is, for sufficiently high SNR, we can write $C \approx \chi + \log(1+\log(1+\mathsf{SNR}))$. Since χ is an asymptotic parameter, it is unclear whether this approximation is reasonable for moderate values of SNR. For example, at the SNR values of the second regime where the $\log(1+\log(1+\mathsf{SNR}))$ term is still negligible the definition of χ does not allow us to judge whether $C \approx \chi$ is accurate. However, based on our results for the first-order Gauss–Markov model for the case of $n_t \geq n_r$, the expansion $C \approx \chi + \log(1+\log(1+\mathsf{SNR}))$ holds for both the second and third regimes.

Lapidoth and Moser [6] use the fading number to characterize the region where the $\log\log \mathsf{SNR}$ term dominates capacity: "at rates that significantly exceed the fading number, one should expect to square the SNR for every additional bit per channel use." This characterization of the third regime is given in terms of rates rather than SNR values.

In our approach, we compute capacity lower and upper bounds by taking joint limits when both $\mathsf{SNR} \to \infty$ and $\epsilon \to 0$. This allows us to characterize capacity not only in the third regime, but also in the second and first regimes, where the $\log\log \mathsf{SNR}$ term is still negligible. The thresholds between the three regimes that we defined for the Gauss–Markov fading model are based on these joint limits. Based on our results, it turns out that both characterizations, in terms of rates or SNR values, of the threshold that define the start of the third regime for the Gauss–Markov model are essentially the same.

For the SISO case, [8] gives also a characterization of the $\log\log \mathsf{SNR}$ regime in terms of the noisy prediction error, which is a function of the SNR. In addition, it notes that for the SISO Gauss–Markov model both characterizations, in terms of rates (through the fading number) and SNR values (through the noisy prediction error), are equivalent to our characterization of the third regime.

Our bound of Theorem 3 can also be used to obtain an upper bound on the fading number for Gauss–Markov MIMO channels when $n_t \geq n_r$. That is, for small enough ϵ and $n_t \geq n_r$, we can use Theorem 3 to write $\chi \leq n_r \log(1/\epsilon) + K_2(n_r, n_t)$. To the best of our knowledge, [6] does not provide a tight characterization of the fading number for this case of great practical interest. Specializing the results of [6] for the case of the Gauss–Markov fading model, one can obtain (12), which as was noted in Section III, is loose in the $\log(1/\epsilon)$ term. Note however that a tighter bound, i.e., (13), can be obtained from the recent results of [23]. Even though our results are obtained for the special case of Gauss–Markov fading, we consider them an important step toward the characterization of the high SNR capacity of MIMO channels with correlated regular fading.

We note that other works that appeared after the original submission of this manuscript have characterized the fading number in the cases of general SIMO fading [24] and MISO fading [23]. In addition, [25] presents new upper and lower bounds on the fading number of peak and average power constrained stationary MIMO regular fading channels. However, the problem of obtaining a tight upper bound for the fading number for the case where $n_r > n_t > 1$ remains open.

C. Nonregular Fading Processes

The high SNR capacity of SISO channels with nonregular fading processes was recently studied in [8], [9]. It is shown that depending on how the noisy prediction error of the fading process $\epsilon(\delta^2)$ varies with $\delta^2 \to 0$,[11] the capacity can show very different high SNR behaviors. In particular, it is shown that if $|\mathcal{D}| < 1$,[12] the high SNR capacity expression grows as $(1 - |\mathcal{D}|)\log \mathsf{SNR}$.

[11] δ^2 is the noise variance of the additive noise that corrupts the observations of the past of the process.

[12] As defined in Section VI, $\mathcal{D}$ is the set of frequencies where the power spectral density of the fading process is strictly positive, and $|\cdot|$ denotes Lebesgue measure.

This result is consistent with our lower bound (28) derived for MIMO channels under an average power constraint on the input, and shows that in fact our bound is tight for $n_r = n_t = 1$ for this kind of fading processes. Characterizing the high SNR capacity of nonregular stationary Gaussian MIMO fading channels requires matching upper and lower bounds on the capacity for $n_r, n_t > 1$. Unfortunately, the bounding techniques that we used to upper-bound the capacity of the MIMO channel with Gauss–Markov fading do not seem to easily extend to the case of nonregular fading. Also, the bounding techniques used in [8], [9] do not seem to extend to the MIMO case. In addition, the upper bound on the pre-log presented in [23] may be loose when $n_r > n_t$ and does not match the pre-log predicted by our lower bound. Therefore, more work needs to be done to fully characterize the capacity of MIMO band-limited nonregular fading channels.

The lower bound (28) is loose for processes where $|\mathcal{D}| = 1$. In fact, the bound may become negative in cases where the condition $|\mathcal{D}| \ll 1$ is not met. Therefore, we cannot predict $\log\log \mathsf{SNR}$ or $(\log \mathsf{SNR})^{\beta}$, $0 \leq \beta \leq 1$, behaviors of channel capacity using our bound.

However, physical fading models such as Jakes' model result in band-limited power spectral densities for the fading distributions. As we argued in Section II, the channels of typical communication systems are underspread, in which case the condition $|\mathcal{D}| \ll 1$ holds.

VIII. Conclusion

Motivated by the capacity results of the MIMO channel with perfect CSI 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.

For the case of regular fading processes, we used a first-order Gauss–Markov fading model and 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 for $n_t \geq n_r$ is consistent with the results of [6] at high SNR, 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.

We believe that channels modeled with higher order Gauss–Markov fading processes or other regular fading laws should show similar qualitative behaviors. An open question for future research is how to capture the concept of channel variations going to zero in fading processes whose statistics depend on more than one parameter. Another open problem is to determine, in this more general setting, whether capacity can still be characterized by three operational regimes and to compute the corresponding thresholds that separate them.

For the case of band-limited nonregular fading processes, where the future evolution of the fading can be perfectly predicted from noiseless observations of the infinite past, we obtained a capacity lower bound. This bound can be used to show that at high SNR at least a rate

$$\min\{n_r, n_t\}(1-\min\{n_r, n_t\}|\mathcal{D}|)\log \mathsf{SNR}$$

can be achieved with i.i.d. Gaussian inputs. For underspread channels, where the power spectral density of the fading process is band-limited $|\mathcal{D}| \ll 1$, i.e., the Lebesgue measure of the set of frequencies where the the PSD is nonzero is small. This shows that in MIMO underspread flat-fading channels with band-limited nonregular fading statistics the number of degrees of freedom $\min\{n_r, n_t\}$ provides significant capacity improvements over the SISO case.

Appendix A

Proof of Theorem 2

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

$$\begin{aligned} C &\geq \lim_{B\to\infty} \frac{1}{B} I(\boldsymbol{x}^{\mathcal{N}}; \boldsymbol{y}^{\mathcal{N}}) \\ &= \lim_{B\to\infty} \frac{1}{B}\left[h(\boldsymbol{y}^{\mathcal{N}}) - h(\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{x}^{\mathcal{N}})\right] \\ &\geq \lim_{B\to\infty} \frac{1}{B}\left[h(\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{H}^{\mathcal{N}}) - h(\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{x}^{\mathcal{N}})\right] \end{aligned} \tag{35}$$

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

We present bounds for the two terms of the RHS of (35) in the following three lemmas.

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

$$\begin{aligned} \frac{1}{B}h(\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{H}^{\mathcal{N}}) &\geq n_r \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) \\ &\quad + E\left[\sum_{k=n_{\max}-(n_{\min}-1)}^{n_{\max}} \log\left(1+\frac{\mathsf{SNR}}{n_t}\chi^2_{2k}\right)\right] \end{aligned} \tag{36}$$

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

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

$$\begin{aligned} h(\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{H}^{\mathcal{N}}) &= E_{\boldsymbol{H}^{\mathcal{N}}}\left[\sum_{n=1}^{B}\log\det(\pi e K_{\boldsymbol{y}[n]|\boldsymbol{H}[n]})\right] \\ &= B\cdot E_{\boldsymbol{H}[1]}\left[\log\det(\pi e K_{\boldsymbol{y}[1]|\boldsymbol{H}[1]})\right] \end{aligned} \tag{37}$$

where

$$\begin{aligned} K_{\boldsymbol{y}[n]|\boldsymbol{H}[n]} &= E[\boldsymbol{y}[n]\boldsymbol{y}[n]^{\dagger}|\boldsymbol{H}[n]] \\ &= \boldsymbol{H}[n]\boldsymbol{H}[n]^{\dagger} + (n_t/\mathsf{SNR})\boldsymbol{I}_{n_r}. \end{aligned}$$

Then we can rewrite (37) as (38) at the bottom of the page.

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

$$\begin{aligned} &E_{\boldsymbol{H}[1]}\left[\log\det\left(\boldsymbol{I}_{n_r}+\frac{\mathsf{SNR}}{n_t}\boldsymbol{H}[1]\boldsymbol{H}[1]^{\dagger}\right)\right] \\ &\qquad \geq E\left[\sum_{k=n_t-(n_r-1)}^{n_t}\log\left(1+\frac{\mathsf{SNR}}{n_t}\chi^2_{2k}\right)\right] \end{aligned} \tag{39}$$

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 preceding lower bound in the following way:

$$\begin{aligned} &E_{\boldsymbol{H}[1]}\left[\log\det\left(\boldsymbol{I}_{n_r}+\frac{\mathsf{SNR}}{n_t}\boldsymbol{H}[1]\boldsymbol{H}[1]^{\dagger}\right)\right] \\ &\quad = E_{\boldsymbol{H}[1]}\left[\log\det\left(\boldsymbol{I}_{n_t}+\frac{\mathsf{SNR}}{n_t}\boldsymbol{H}[1]^{\dagger}\boldsymbol{H}[1]\right)\right] \\ &\quad \geq E\left[\sum_{k=n_r-(n_t-1)}^{n_r}\log\left(1+\frac{\mathsf{SNR}}{n_t}\chi^2_{2k}\right)\right]. \end{aligned} \tag{40}$$

The result follows from (38)–(40). □

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

$$\begin{aligned} \lim_{B\to\infty}\frac{1}{B}h(\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{x}^{\mathcal{N}}) &\leq n_r\log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) \\ &\quad - n_r n_t\int_0^1\log\left[\frac{\mathsf{SNR}}{n_t}S_{h_{11}}(f)+1\right]df. \end{aligned} \tag{41}$$

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

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

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

$$\begin{aligned} h(\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{H}^{\mathcal{N}}) &= B\cdot E_{\boldsymbol{H}[1]}\left\{\log\det\left[\frac{\pi e n_t}{\mathsf{SNR}}\left(\boldsymbol{I}_{n_r}+\frac{\mathsf{SNR}}{n_t}\boldsymbol{H}[1]\boldsymbol{H}[1]^{\dagger}\right)\right]\right\} \\ &= B\left\{n_r\log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + E_{\boldsymbol{H}[1]}\left[\log\det\left(\boldsymbol{I}_{n_r}+\frac{\mathsf{SNR}}{n_t}\boldsymbol{H}[1]\boldsymbol{H}[1]^{\dagger}\right)\right]\right\}. \end{aligned} \tag{38}$$

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

Then we have

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

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

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

where

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

with $\boldsymbol{K}_{\tilde{\boldsymbol{h}}} = E[\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^\dagger]$.

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

$$\begin{aligned} &\frac{h(\boldsymbol{y}^N | \boldsymbol{x}^N)}{n_r} \\ &= B\log(\pi e) + E_{\tilde{\boldsymbol{X}}}\left[\log\det\left(\tilde{\boldsymbol{X}}\boldsymbol{K}_{\tilde{\boldsymbol{h}}}\tilde{\boldsymbol{X}}^\dagger + \frac{n_t}{\mathsf{SNR}}\boldsymbol{I}_B\right)\right] \\ &= B\log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) \\ &\quad + E_{\tilde{\boldsymbol{X}}}\left[\log\det\left(\frac{\mathsf{SNR}}{n_t}\tilde{\boldsymbol{X}}\boldsymbol{K}_{\tilde{\boldsymbol{h}}}\tilde{\boldsymbol{X}}^\dagger + \boldsymbol{I}_B\right)\right] \\ &= B\log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) \\ &\quad + E_{\tilde{\boldsymbol{X}}}\left[\log\det\left(\frac{\mathsf{SNR}}{n_t}\tilde{\boldsymbol{X}}^\dagger\tilde{\boldsymbol{X}}\boldsymbol{K}_{\tilde{\boldsymbol{h}}} + \boldsymbol{I}_{n_tB}\right)\right] \\ &\leq B\log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) \\ &\quad + \log\det\left[\frac{\mathsf{SNR}}{n_t}E_{\tilde{\boldsymbol{X}}}\left(\tilde{\boldsymbol{X}}^\dagger\tilde{\boldsymbol{X}}\right)\boldsymbol{K}_{\tilde{\boldsymbol{h}}} + \boldsymbol{I}_{n_tB}\right] \\ &= B\log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \log\det\left[\frac{\mathsf{SNR}}{n_t}\boldsymbol{K}_{\tilde{\boldsymbol{h}}} + \boldsymbol{I}_{n_tB}\right] \qquad (42) \end{aligned}$$

where the inequality follows from Jensen's inequality and the last equality results from $E_{\tilde{\boldsymbol{X}}}(\tilde{\boldsymbol{X}}^\dagger\tilde{\boldsymbol{X}}) = \boldsymbol{I}_{n_tB}$. To upper-bound the last expression, we start by computing $\boldsymbol{K}_{\tilde{\boldsymbol{h}}}$ in terms of $R_{h_{11}}[n]$, the autocorrelation function of the entry $(1,1)$ of $\boldsymbol{H}[n]$. $\boldsymbol{K}_{\tilde{\boldsymbol{h}}}$ is given in the expression at the bottom of the page, where we have used the fact that the entries of $\boldsymbol{H}[n]$ are i.i.d.

To compute the RHS of (42), we use the following known results.

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

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

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 $K_{\tilde{\boldsymbol{h}}}$. Letting $\{\lambda_i : 1 \leq i \leq n_tB\}$ be the set of eigenvalues of $K_{\tilde{\boldsymbol{h}}}$ we obtain the following asymptotic bound:

$$\begin{aligned} &\lim_{B\to\infty}\frac{h(\boldsymbol{y}^{\mathcal{N}} | \boldsymbol{x}^{\mathcal{N}})}{n_rB} \\ &\leq \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \lim_{B\to\infty}\frac{1}{B}\log\det\left[\frac{\mathsf{SNR}}{n_t}K_{\tilde{\boldsymbol{h}}} + \boldsymbol{I}_{n_tB}\right] \\ &= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \lim_{B\to\infty}\frac{1}{B}\sum_{i=1}^{n_tB}\log\left(\frac{\mathsf{SNR}}{n_t}\lambda_i + 1\right) \\ &= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_t\int_0^1\log\left[\frac{\mathsf{SNR}}{n_t}S_{\hat{h}}(f) + 1\right]df \\ &= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_t\int_0^1\log\left[\frac{\mathsf{SNR}}{n_t}S_{h_{11}}(n_tf) + 1\right]df \\ &= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \int_0^{n_t}\log\left[\frac{\mathsf{SNR}}{n_t}S_{h_{11}}(f') + 1\right]df' \\ &= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_t\int_0^1\log\left[\frac{\mathsf{SNR}}{n_t}S_{h_{11}}(f) + 1\right]df \qquad (43) \end{aligned}$$

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

where we exploited the periodicity of $S_{h_{11}}(f)$ in the last equality. By multiplying both sides of (43) by n_r, we obtain the statement of the lemma. □

We now specialize the result of the previous lemma to the case of the first-order Gauss–Markov process.

Lemma 9: Let $\boldsymbol{x}^{\mathcal{N}} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_B)$. Then

$$\lim_{B\to\infty} \frac{1}{B} h(\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{x}^{\mathcal{N}}) \leq n_r \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) - n_r n_t \log(2) + n_r n_t \log\left[\frac{\mathsf{SNR}\epsilon}{n_t} + 2 - \epsilon + \sqrt{\left(\frac{\mathsf{SNR}\epsilon}{n_t} + 2 - \epsilon\right)^2 - 4 + 4\epsilon}\right]. \tag{44}$$

Proof: The power spectral density of $h_{11}[n]$ is obtained by taking the discrete-time Fourier transform of its autocorrelation function, which for the first-order Gauss–Markov process is given by

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

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 (43) we obtain

$$\begin{aligned}\lim_{B\to\infty} \frac{h(\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{x}^{\mathcal{N}})}{n_r B} &\leq \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_t \int_0^1 \log\left[\frac{\mathsf{SNR}\epsilon/n_t}{1+\alpha^2 - 2\alpha\cos(2\pi f)} + 1\right] df \\ &= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + n_t \int_0^1 \log\left[\mathsf{SNR}\epsilon/n_t + 1 + \alpha^2 - 2\alpha\cos(2\pi f)\right] df - n_t \int_0^1 \log\left[1+\alpha^2 - 2\alpha\cos(2\pi f)\right] df \\ &= \log\left(\frac{\pi e n_t}{\mathsf{SNR}}\right) + \frac{n_t}{\pi} \int_0^\pi \log\left[\mathsf{SNR}\epsilon/n_t + 1 + \alpha^2 - 2\alpha\cos(\omega)\right] d\omega - \frac{n_t}{\pi} \int_0^\pi \log\left[1+\alpha^2 - 2\alpha\cos(\omega)\right] d\omega. \end{aligned} \tag{45}$$

The last two integrals can be computed in closed form using the identity

$$\int_0^\pi \log(a \pm b\cos x)\, dx = \pi \log\left(\frac{a + \sqrt{a^2 - b^2}}{2}\right)$$

valid for $a \geq b$. This is done in the first set of equations at the bottom of the page. Replacing in (45) and multiplying both sides by n_r we obtain the statement of the lemma. □

As a result, we can use (35), (36), and (44) to get a lower bound for the capacity of the channel. See (46) at the bottom of the page.

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 RHS of (46). The improved lower bound is given by the last equation at the bottom of the page.

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

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

$$C(\mathsf{SNR}, \epsilon) \geq E\left[\sum_{k=n_{\max}-(n_{\min}-1)}^{n_{\max}} \log\left(1 + \frac{\mathsf{SNR}}{n_t}\chi^2_{2k}\right)\right] + n_r n_t \log(2) - n_r n_t \log\left[\frac{\mathsf{SNR}\epsilon}{n_t} + 2 - \epsilon + \sqrt{\left(\frac{\mathsf{SNR}\epsilon}{n_t} + 2 - \epsilon\right)^2 - 4 + 4\epsilon}\right]. \tag{46}$$

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

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 (46) and obtain the first equation at the bottom of the page 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) as shown in (47) at the bottom of the page.

When $n_r = n_t = 1$, $\mathsf{SNR}^* = \infty$, and (47) reduces to

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

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 (47) by a constant that does not depend on SNR or ϵ

$$\begin{aligned} -n_r n_t \log&\left[\min\{\mathsf{SNR},\mathsf{SNR}^*\}\frac{\epsilon}{n_t}+2\right] \\ &\ge -n_r n_t \log\left[\mathsf{SNR}^*\frac{\epsilon}{n_t}+2\right] \\ &= -n_r n_t \log\left[\frac{4}{n_r^2 n_t^2/n_{\min}^2 - 1}+2\right]. \end{aligned}$$

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 $\boldsymbol{\gamma} \in \mathcal{G}$, where $\gamma[n] = 1(\|\boldsymbol{x}[n]\| \ge \mathsf{SNR}^{-1/2})$, and use the chain rule of mutual information in two different ways

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

From these, the nonnegativity of mutual information, and the fact that $\boldsymbol{\gamma}$ is a discrete random vector that takes 2^B values, we obtain the following bound:

$$\begin{aligned} I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{y}^{\mathcal{N}}) &= I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{\gamma}) + I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{\gamma}) - I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{\gamma}|\boldsymbol{y}^{\mathcal{N}}) \\ &\le I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{\gamma}) + I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{\gamma}) \\ &\le B\log 2 + I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{\gamma}). \end{aligned} \tag{48}$$

Now the problem is that of finding an upper bound for

$$I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{\gamma}) = \sum_{\boldsymbol{v}\in\mathcal{G}} I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{\gamma}=\boldsymbol{v})P(\boldsymbol{\gamma}=\boldsymbol{v}) \tag{49}$$

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

$$\begin{aligned} C(\mathsf{SNR},\epsilon) \ge& \sum_{k=n_{\max}-(n_{\min}-1)}^{n_{\max}} E\left[\log\left(1+\frac{\min\{\mathsf{SNR},\mathsf{SNR}^*\}}{n_t}\chi^2_{2k}\right)\right] + n_r n_r \log(2) \\ & - n_r n_t \log\left[\min\{\mathsf{SNR},\mathsf{SNR}^*\}\frac{\epsilon}{n_t} + 2 - \epsilon + \sqrt{\left(\min\{\mathsf{SNR},\mathsf{SNR}^*\}\frac{\epsilon}{n_t}+2-\epsilon\right)^2 - 4(1-\epsilon)}\right] \\ \ge& \sum_{k=n_{\max}-(n_{\min}-1)}^{n_{\max}} E\left[\log\left(1+\frac{\min\{\mathsf{SNR},\mathsf{SNR}^*\}}{n_t}\chi^2_{2k}\right)\right] - n_r n_t \log\left[\min\{\mathsf{SNR},\mathsf{SNR}^*\}\frac{\epsilon}{n_t}+2\right]. \end{aligned} \tag{47}$$

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

$$\frac{1}{B}\sum_{n=1}^{B}E\left[\|\boldsymbol{x}[n]\|^2\middle|\,\boldsymbol{\gamma}=\boldsymbol{v}\right]=\xi_{\boldsymbol{v}} \tag{50}$$

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

$$\sum_{\boldsymbol{v}\in\mathcal{G}}\xi_{\boldsymbol{v}}P(\boldsymbol{\gamma}=\boldsymbol{v})\le n_t. \tag{51}$$

Define $\mathcal{A}_{\boldsymbol{v}}=\{n:v[n]=0\}$ and $\mathcal{B}_{\boldsymbol{v}}=\{n:v[n]=1\}$. For $n\in\mathcal{A}_{\boldsymbol{v}}$, 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 10:

$$\frac{1}{B}I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{\gamma}=\boldsymbol{v})\le 2n_r\log(2)+\frac{1}{B}I(\boldsymbol{x}^{\mathcal{B}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}}|\boldsymbol{\gamma}=\boldsymbol{v}). \tag{52}$$

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

$$\begin{aligned}I(\boldsymbol{x}^{\mathcal{N}},\boldsymbol{y}^{\mathcal{N}}|\boldsymbol{\gamma}=\boldsymbol{v})&=I(\boldsymbol{x}^{\mathcal{A}_{\boldsymbol{v}}},\boldsymbol{x}^{\mathcal{B}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}},\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}}|\boldsymbol{\gamma}=\boldsymbol{v})\\&=I(\boldsymbol{x}^{\mathcal{A}_{\boldsymbol{v}}},\boldsymbol{x}^{\mathcal{B}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}}|\boldsymbol{\gamma}=\boldsymbol{v})\\&\quad+I(\boldsymbol{x}^{\mathcal{A}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}}|\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}},\boldsymbol{\gamma}=\boldsymbol{v})\\&\quad+I(\boldsymbol{x}^{\mathcal{B}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}}|\boldsymbol{x}^{\mathcal{A}_{\boldsymbol{v}}},\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}},\boldsymbol{\gamma}=\boldsymbol{v}).\end{aligned} \tag{53}$$

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

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

The second term of (53) can be upper-bounded by rewriting the mutual information in terms of differential entropies and bounding each of the individual terms

$$\begin{aligned}I(\boldsymbol{x}^{\mathcal{A}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}}|\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}},\boldsymbol{\gamma}=\boldsymbol{v})&=h(\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}}|\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}},\boldsymbol{\gamma}=\boldsymbol{v})\\&\quad-h(\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}}|\boldsymbol{x}^{\mathcal{A}_{\boldsymbol{v}}},\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}},\boldsymbol{\gamma}=\boldsymbol{v})\\&\le h(\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}}|\boldsymbol{\gamma}=\boldsymbol{v})\\&\quad-h(\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}}|\boldsymbol{x}^{\mathcal{A}_{\boldsymbol{v}}},\boldsymbol{H}^{\mathcal{A}_{\boldsymbol{v}}},\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}},\boldsymbol{\gamma}=\boldsymbol{v})\\&=h(\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}}|\boldsymbol{\gamma}=\boldsymbol{v})\\&\quad-h(\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}}|\boldsymbol{x}^{\mathcal{A}_{\boldsymbol{v}}},\boldsymbol{H}^{\mathcal{A}_{\boldsymbol{v}}},\boldsymbol{\gamma}=\boldsymbol{v})\end{aligned} \tag{54}$$

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

To bound the first term of (54) we define $\boldsymbol{d}[n]=\boldsymbol{x}[n]/\|\boldsymbol{x}[n]\|$ and compute $E[\boldsymbol{y}[n]\boldsymbol{y}[n]^\dagger|\boldsymbol{\gamma}=\boldsymbol{v}]$ for $n\in\mathcal{A}_{\boldsymbol{v}}$ as shown in the equation at the bottom of the page. In the third equality of that equation we used the conditional independence of $\boldsymbol{H}[n]\boldsymbol{d}[n]$ and $\boldsymbol{x}[n]$ conditioned on $\boldsymbol{d}[n]$. In the fourth equality, we used the fact that $\boldsymbol{d}[n]$ is a unit norm vector and the entries of $\boldsymbol{H}[n]$ are i.i.d. $\mathcal{CN}(0,1)$ and, therefore, conditioned on $\boldsymbol{d}[n]$, $\boldsymbol{H}[n]\boldsymbol{d}[n]\sim\mathcal{CN}(\boldsymbol{0},\boldsymbol{I}_{n_r})$ to compute $E\left[\boldsymbol{H}[n]\boldsymbol{d}[n]\boldsymbol{d}[n]^\dagger\boldsymbol{H}[n]^\dagger\mid\boldsymbol{d}[n],\boldsymbol{\gamma}=\boldsymbol{v}\right]$.

For $n\in\mathcal{A}_{\boldsymbol{v}}$, $E\left[\|\boldsymbol{x}[n]\|^2|\boldsymbol{\gamma}=\boldsymbol{v}\right]\le\mathsf{SNR}^{-1}$ and hence,

$$\det\left\{E\left[\boldsymbol{y}[n]\boldsymbol{y}[n]^\dagger|\boldsymbol{\gamma}=\boldsymbol{v}\right]\right\}\le[(n_t+1)\mathsf{SNR}^{-1}]^{n_r}.$$

Then using the chain rule, removing conditioning and applying the Gaussian bound we obtain

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

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

The second term of (54) can be computed explicitly

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

Therefore, we obtain

$$\begin{aligned}I(\boldsymbol{x}^{\mathcal{A}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{A}_{\boldsymbol{v}}}|\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}},\boldsymbol{\gamma}=\boldsymbol{v})&\le|\mathcal{A}_{\boldsymbol{v}}|n_r\log(1+n_t^{-1})\\&\le Bn_r\log(2).\end{aligned}$$

The last term of (53) can be bounded in a similar way

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

Replacing these three bounds in (53) we get the desired result. □

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

At this point, our goal is to find an upper bound for (52) that depends on $\boldsymbol{v}$ only through $\xi_{\boldsymbol{v}}$. The power constraint (50) imposes a power constraint on $\boldsymbol{x}^{\mathcal{B}_{\boldsymbol{v}}}$

$$\frac{1}{B}\sum_{n\in\mathcal{B}_{\boldsymbol{v}}} E\left[\|\boldsymbol{x}[n]\|^2 \middle| \boldsymbol{\gamma}=\boldsymbol{v}\right] \le \frac{1}{B}\sum_{n=1}^{B} E\left[\|\boldsymbol{x}[n]\|^2 \middle| \boldsymbol{\gamma}=\boldsymbol{v}\right] = \xi_{\boldsymbol{v}}.$$

Consider a random vector $\tilde{\boldsymbol{x}}^{\mathcal{N}}$ such that $\tilde{\boldsymbol{x}}^{\mathcal{B}_{\boldsymbol{v}}}$ has the same distribution as $\boldsymbol{x}^{\mathcal{B}_{\boldsymbol{v}}}$ conditioned on $\{\boldsymbol{\gamma}=\boldsymbol{v}\}$ so that

$$I(\tilde{\boldsymbol{x}}^{\mathcal{B}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}}) = I(\boldsymbol{x}^{\mathcal{B}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}}|\boldsymbol{\gamma}=\boldsymbol{v}).$$

We also require that $\|\tilde{\boldsymbol{x}}[n]\| \ge \mathsf{SNR}^{-1/2}$ for $n \in \mathcal{A}_{\boldsymbol{v}}$. Finally, since $\tilde{\boldsymbol{x}}^{\mathcal{B}_{\boldsymbol{v}}}$ must have the same total power as $\boldsymbol{x}^{\mathcal{B}_{\boldsymbol{v}}}$, we require that

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

Using the chain rule for mutual information we can write

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

The nonnegativity of mutual information implies that $I(\tilde{\boldsymbol{x}}^{\mathcal{B}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}}) \le I(\tilde{\boldsymbol{x}}^{\mathcal{N}};\boldsymbol{y}^{\mathcal{N}})$, so defining

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

we conclude that

$$\lim_{B\to\infty}\frac{1}{B}I(\boldsymbol{x}^{\mathcal{B}_{\boldsymbol{v}}};\boldsymbol{y}^{\mathcal{B}_{\boldsymbol{v}}}|\boldsymbol{\gamma}=\boldsymbol{v}) \le \lim_{B\to\infty}\frac{1}{B}I(\tilde{\boldsymbol{x}}^{\mathcal{N}};\boldsymbol{y}^{\mathcal{N}}) \le f(\mathsf{SNR},\epsilon,\xi_{\boldsymbol{v}}). \quad (55)$$

In summary, putting (48), (49), (52), and (55) together, we have

$$\lim_{B\to\infty}\frac{1}{B}I(\boldsymbol{x}^{\mathcal{N}};\boldsymbol{y}^{\mathcal{N}}) \le (2n_r+1)\log(2) + \sum_{\boldsymbol{v}\in\mathcal{G}} f(\mathsf{SNR},\epsilon,\xi_{\boldsymbol{v}})P(\boldsymbol{\gamma}=\boldsymbol{v}). \quad (56)$$

Finally, to upper-bound the RHS of (56), we prove that $f(\mathsf{SNR},\epsilon,\xi)$ is a concave function of ξ and use Jensen's inequality.

Lemma 11: $f(\mathsf{SNR},\epsilon,\xi)$ is a concave function of ξ, that is, for any $\xi_1,\xi_2\ge 0$ and $\lambda\in[0,1]$

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

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 $\mathcal{N}_i=\{1,2,\dots,n_i\}$. We require that the input signals satisfy the power constraints

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

and

$$\|\boldsymbol{x}_i[n]\| \ge \mathsf{SNR}^{-1}, \qquad \text{for } n\in\mathcal{N}_i.$$

Then by the independence of the channels, we have

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

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

$$\lim_{B\to\infty}\sup_{p\left(\boldsymbol{x}_1^{\mathcal{N}_1}\right),p\left(\boldsymbol{x}_2^{\mathcal{N}_2}\right)}\frac{1}{B}I\left(\boldsymbol{x}_1^{\mathcal{N}_1},\boldsymbol{x}_2^{\mathcal{N}_2};\boldsymbol{y}_1^{\mathcal{N}_1},\boldsymbol{y}_2^{\mathcal{N}_2}\right) = \lambda f(\mathsf{SNR},\epsilon,\xi_1) + (1-\lambda)f(\mathsf{SNR},\epsilon,\xi_2). \quad (57)$$

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 $(\boldsymbol{x}_1^{\mathcal{N}_1},\boldsymbol{x}_2^{\mathcal{N}_2})$ that achieves (57) 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 (57). This input satisfies the power constraints

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

and

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

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 (56) we obtain

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

where we used (51) and the fact that $f(\mathsf{SNR},\epsilon,\xi)$ is a nondecreasing function of ξ.

APPENDIX D

Proof of Lemma 2

We express the mutual information in terms of differential entropies

$$I(\boldsymbol{x}^{\mathcal{N}};\tilde{\boldsymbol{y}}^{\mathcal{N}}) = h(\tilde{\boldsymbol{y}}^{\mathcal{N}}) - h(\tilde{\boldsymbol{y}}^{\mathcal{N}}|\boldsymbol{x}^{\mathcal{N}}). \tag{58}$$

To write each of the differential entropies in terms of the new variables use the chain rule for differential entropies and apply Fact 1 to the transformation $(\rho_i[n], \phi_i[n]) \to \tilde{y}_i$ which has Jacobian with determinant $|\tilde{y}_i|^2$ to get (59) at the bottom of the page,[13] where the last equality follows from $\tilde{y}_i[n] = c_i[n]m[n]$.

We also rewrite the second term of (58) in terms of the new variables to get (60) also at the bottom of this page, where we used the conditional independence between $\boldsymbol{m}^{\mathcal{N}}$ and $\boldsymbol{c}^{\mathcal{N}}$ conditioned on $\boldsymbol{d}^{\mathcal{N}}$ in the last equality.

The result follows by subtracting (60) from (59).

APPENDIX E

Proof of Lemma 5

$\boldsymbol{\rho}^{\mathcal{N}}$ 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 $\|\boldsymbol{x}[n]\|$, so they are highly correlated. Eventually, we will use the independent Gaussian bound to upper-bound $h(\boldsymbol{\rho}^{\mathcal{N}},\boldsymbol{\phi}^{\mathcal{N}})$ 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

$$\begin{aligned} h\left(\boldsymbol{\rho}^{\mathcal{N}},\boldsymbol{\phi}^{\mathcal{N}}\right) &= h\left(\boldsymbol{\rho}_1^{\mathcal{N}},\boldsymbol{\phi}_1^{\mathcal{N}},\boldsymbol{\rho}_2^{\mathcal{N}},\boldsymbol{\phi}_2^{\mathcal{N}},\ldots,\boldsymbol{\rho}_{n_r}^{\mathcal{N}},\boldsymbol{\phi}_{n_r}^{\mathcal{N}}\right) \\ &= h\Big(\boldsymbol{\rho}_1^{\mathcal{N}},\boldsymbol{\phi}_1^{\mathcal{N}},\left(\boldsymbol{\rho}_2^{\mathcal{N}}-\boldsymbol{\rho}_1^{\mathcal{N}}\right),\left(\boldsymbol{\phi}_2^{\mathcal{N}}-\boldsymbol{\phi}_1^{\mathcal{N}}\right),\ldots, \\ &\qquad \left(\boldsymbol{\rho}_{n_r}^{\mathcal{N}}-\boldsymbol{\rho}_1^{\mathcal{N}}\right),\left(\boldsymbol{\phi}_{n_r}^{\mathcal{N}}-\boldsymbol{\phi}_1^{\mathcal{N}}\right)\Big). \end{aligned}$$

This transformation eliminates the dependence on $\|\boldsymbol{x}[n]\|$ 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

$$\begin{aligned} h\left(\boldsymbol{\rho}^{\mathcal{N}},\boldsymbol{\phi}^{\mathcal{N}}\right) &\le h\left(\boldsymbol{\rho}_1^{\mathcal{N}},\boldsymbol{\phi}_1^{\mathcal{N}}\right) \\ &\quad + \sum_{i=2}^{n_r} h\left(\left(\boldsymbol{\rho}_i^{\mathcal{N}}-\boldsymbol{\rho}_1^{\mathcal{N}}\right),\left(\boldsymbol{\phi}_i^{\mathcal{N}}-\boldsymbol{\phi}_1^{\mathcal{N}}\right)\right) \\ &\le h\left(\boldsymbol{\rho}_1^{\mathcal{N}}\right) + h\left(\boldsymbol{\phi}_1^{\mathcal{N}}\right) + \sum_{i=2}^{n_r} h\left(\left(\boldsymbol{\rho}_i^{\mathcal{N}}-\boldsymbol{\rho}_1^{\mathcal{N}}\right)\right) \\ &\quad + \sum_{i=2}^{n_r} h\left(\left(\boldsymbol{\phi}_i^{\mathcal{N}}-\boldsymbol{\phi}_1^{\mathcal{N}}\right)\right). \end{aligned} \tag{61}$$

[13]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.

$$\begin{aligned} h(\tilde{\boldsymbol{y}}^{\mathcal{N}}) &= \sum_{n\in\mathcal{N}} h(\tilde{\boldsymbol{y}}[n] \mid \tilde{\boldsymbol{y}}[n-1],\ldots,\tilde{\boldsymbol{y}}[1]) \\ &= \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} h(\tilde{y}_i[n] \mid \tilde{y}_{i-1}[n],\ldots,\tilde{y}_1[n],\tilde{\boldsymbol{y}}[n-1],\ldots,\tilde{\boldsymbol{y}}[1]) \\ &= \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} E\left[\log(|\tilde{y}_i[n]|^2)\right] + \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} h(\rho_i[n],\phi_i[n] \mid \tilde{y}_{i-1}[n],\ldots,\tilde{y}_1[n],\tilde{\boldsymbol{y}}[n-1],\ldots,\tilde{\boldsymbol{y}}[1]) \\ &= \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} E\left[\log(|\tilde{y}_i[n]|^2)\right] + h(\boldsymbol{\rho}^{\mathcal{N}},\boldsymbol{\phi}^{\mathcal{N}}) \\ &= \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} E\left[\log(|c_i[n]|^2)\right] + n_r\sum_{n\in\mathcal{N}} E\left[\log(m[n]^2)\right] + h(\boldsymbol{\rho}^{\mathcal{N}},\boldsymbol{\phi}^{\mathcal{N}}) \end{aligned} \tag{59}$$

$$\begin{aligned} h(\tilde{\boldsymbol{y}}^{\mathcal{N}}|\tilde{\boldsymbol{x}}^{\mathcal{N}}) &= h(\tilde{\boldsymbol{y}}^{\mathcal{N}}|\boldsymbol{m}^{\mathcal{N}},\boldsymbol{d}^{\mathcal{N}}) \\ &= \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} h(\tilde{y}_i[n] \mid \tilde{y}_{i-1}[n],\ldots,\tilde{y}_1[n],\tilde{\boldsymbol{y}}[n-1],\ldots,\tilde{\boldsymbol{y}}[1],\boldsymbol{m}^{\mathcal{N}},\boldsymbol{d}^{\mathcal{N}}) \\ &= \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} h((c_i[n]m[n]) \mid \tilde{y}_{i-1}[n],\ldots,\tilde{y}_1[n],\tilde{\boldsymbol{y}}[n-1],\ldots,\tilde{\boldsymbol{y}}[1],\boldsymbol{m}^{\mathcal{N}},\boldsymbol{d}^{\mathcal{N}}) \\ &= \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} h((c_i[n]m[n]) \mid c_{i-1}[n],\ldots,c_1[n],\boldsymbol{c}[n-1],\ldots,\boldsymbol{c}[1],\boldsymbol{m}^{\mathcal{N}},\boldsymbol{d}^{\mathcal{N}}) \\ &= n_r\sum_{n\in\mathcal{N}} E\left[\log(m[n]^2)\right] + \sum_{n\in\mathcal{N}}\sum_{i=1}^{n_r} h(c_i[n] \mid c_{i-1}[n],\ldots,c_1[n],\boldsymbol{c}[n-1],\ldots,\boldsymbol{c}[1],\boldsymbol{m}^{\mathcal{N}},\boldsymbol{d}^{\mathcal{N}}) \\ &= n_r\sum_{n\in\mathcal{N}} E\left[\log(m[n]^2)\right] + h(\boldsymbol{c}^{\mathcal{N}}|\boldsymbol{d}^{\mathcal{N}}) \end{aligned} \tag{60}$$

Most of these terms can be easily bounded by noting that $-\pi \le \phi_1[n] \le \pi$ and $-2\pi \le \phi_i[n] - \phi_1[n] \le 2\pi$ and hence using the Gaussian bound

$$h(\boldsymbol{\phi}_1^{\mathcal{N}}) \le \frac{B}{2}\log(2\pi e \pi^2)$$

and

$$h\left(\left(\boldsymbol{\phi}_i^{\mathcal{N}} - \boldsymbol{\phi}_1^{\mathcal{N}}\right)\right) \le \frac{B}{2}\log(2\pi e(2\pi)^2).$$

Using the chain rule and removing conditioning again we can write

$$\sum_{i=2}^{n_r} h\left(\left(\boldsymbol{\rho}_i^{\mathcal{N}} - \boldsymbol{\rho}_1^{\mathcal{N}}\right)\right) \le \sum_{n\in\mathcal{N}}\sum_{i=2}^{n_r} h((\rho_i[n]-\rho_1[n])). \tag{62}$$

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 $\boldsymbol{d}[n]$, the independence between the rows of $\boldsymbol{H}[n]$ implies the independence between $c_i[n]$ and $c_1[n]$ for $i = 2,\ldots,n_r$. Also, since $\boldsymbol{d}[n]$ has norm 1, and the rows of $\boldsymbol{H}[n]$ 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 $\boldsymbol{d}[n]$. Then, conditioned on $\boldsymbol{d}[n]$, $|c_i[n]|^2/|c_1[n]|^2$ is the ratio of two independent $\mathrm{Exp}(1)$ random variables, and the corresponding conditional density does not depend on $\boldsymbol{d}[n]$, so it follows that the unconditional distribution of $|c_i[n]|^2/|c_1[n]|^2$ is also that of the ratio of two independent $\mathrm{Exp}(1)$ random variables. The corresponding density is given by $f_T(t) = 1/(t+1)^2$ for $t \ge 0$. 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 $l \in \mathbb{R}$. Then we can compute the RHS of (62) explicitly as shown in (63) at the bottom of the page.

It only remains to find an upper bound for the first term of (61), $h\left(\boldsymbol{\rho}_1^{\mathcal{N}}\right)$. 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

$$\begin{aligned} h\left(\boldsymbol{\rho}_1^{\mathcal{N}}\right) &\le \sum_{n\in\mathcal{N}} h(\rho_1[n]) \\ &\le \sum_{n\in\mathcal{N}} \frac{1}{2}\log(2\pi e\sigma_n^2) \\ &\le \frac{B}{2}\log\left(2\pi e\frac{1}{B}\sum_{n\in\mathcal{N}}\sigma_n^2\right) \end{aligned} \tag{64}$$

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

$$\begin{aligned} &\rho_1[n] = \log|c_1[n]| + \log(m[n]) \\ &\Rightarrow \sqrt{E[(\rho_1[n])^2]} \le \sqrt{E[(\log|c_1[n]|)^2]} + \sqrt{E[(\log m[n])^2]}. \end{aligned} \tag{65}$$

The second term of (65) can be bounded by conditioning on the event $\Psi = \{m[n] > 1\}$ and noting that on Ψ, $(\log m[n])^2 \le [m[n]-1]^2$

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

Noting that $c_1[n] \sim \mathcal{CN}(0,1)$, the first term of (65) can be computed by numerical integration

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

As a result, we obtain

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

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

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

Here we used the fact that

$$\frac{1}{B}\sum_{n\in\mathcal{N}}\theta_n^2 = \frac{1}{B}\sum_{n\in\mathcal{N}} E[\|\boldsymbol{x}[n]\|^2] \le n_t + \frac{1}{\mathsf{SNR}}$$

by the power constraint (19).

Therefore, we can upper-bound (64) by

$$h\left(\boldsymbol{\rho}_1^{\mathcal{N}}\right) \le \frac{B}{2}\log\left(2\pi e\sigma^2\right)$$

where

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

$$\begin{aligned} \sum_{n=\in\mathcal{N}}\sum_{i=2}^{n_r} h((\rho_i[n]-\rho_1[n])) &= (n_r-1)B\left\{\log\left(\frac{1}{2}\right) + \int_{-\infty}^{\infty}\frac{e^l}{(1+e^l)^2}\log\left[\frac{(1+e^l)^2}{e^l}\right]dl\right\} \\ &= (n_r-1)B[\log(1/2)+2]. \end{aligned} \tag{63}$$

Replacing all the bounds that we found in (61) 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

$$\begin{aligned} h\left(\boldsymbol{\rho}^{\mathcal{N}},\boldsymbol{\phi}^{\mathcal{N}}\right) &= h\left(\boldsymbol{\rho}_1^{\mathcal{N}},\boldsymbol{\phi}_1^{\mathcal{N}},\ldots,\boldsymbol{\rho}_{n_r}^{\mathcal{N}},\boldsymbol{\phi}_{n_r}^{\mathcal{N}}\right) \\ &= h\Big(\boldsymbol{\rho}_1^{\mathcal{N}},\boldsymbol{\phi}_1^{\mathcal{N}},\left(\boldsymbol{\rho}_2^{\mathcal{N}}-\boldsymbol{\rho}_1^{\mathcal{N}}\right),\left(\boldsymbol{\phi}_2^{\mathcal{N}}-\boldsymbol{\phi}_1^{\mathcal{N}}\right),\ldots, \\ &\qquad \left(\boldsymbol{\rho}_{n_r}^{\mathcal{N}}-\boldsymbol{\rho}_1^{\mathcal{N}}\right),\left(\boldsymbol{\phi}_{n_r}^{\mathcal{N}}-\boldsymbol{\phi}_1^{\mathcal{N}}\right)\Big) \\ &\le h\left(\boldsymbol{\rho}_1^{\mathcal{N}}\right)+h\left(\boldsymbol{\phi}_1^{\mathcal{N}}\right) \\ &\quad + h\Big(\left(\boldsymbol{\rho}_2^{\mathcal{N}}-\boldsymbol{\rho}_1^{\mathcal{N}}\right),\left(\boldsymbol{\phi}_2^{\mathcal{N}}-\boldsymbol{\phi}_1^{\mathcal{N}}\right),\ldots, \\ &\qquad \left(\boldsymbol{\rho}_{n_r}^{\mathcal{N}}-\boldsymbol{\rho}_1^{\mathcal{N}}\right),\left(\boldsymbol{\phi}_{n_r}^{\mathcal{N}}-\boldsymbol{\phi}_1^{\mathcal{N}}\right)\Big). \end{aligned} \tag{66}$$

The first two terms of (66) can be bounded in the same way as the corresponding terms that appeared in the proof of Lemma 5[14] by simply taking $n_t = 1$

$$h\left(\boldsymbol{\rho}_1^{\mathcal{N}}\right)+h\left(\boldsymbol{\phi}_1^{\mathcal{N}}\right) \le \frac{B}{2}\log\left(2\pi e\sigma^2\right)+\frac{B}{2}\log(2\pi e\pi^2) \tag{67}$$

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 $i=2,\ldots,n_r$

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

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

To compute the third term of (66), 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 $\boldsymbol{H}[n]$, and apply Fact 1

$$\begin{aligned} &h\Big(\left(\boldsymbol{\rho}_2^{\mathcal{N}}-\boldsymbol{\rho}_1^{\mathcal{N}}\right),\left(\boldsymbol{\phi}_2^{\mathcal{N}}-\boldsymbol{\phi}_1^{\mathcal{N}}\right),\ldots,\left(\boldsymbol{\rho}_{n_r}^{\mathcal{N}}-\boldsymbol{\rho}_1^{\mathcal{N}}\right),\left(\boldsymbol{\phi}_{n_r}^{\mathcal{N}}-\boldsymbol{\phi}_1^{\mathcal{N}}\right)\Big) \\ &\quad = h\left(\boldsymbol{r}_2^{\mathcal{N}},\ldots,\boldsymbol{r}_{n_r}^{\mathcal{N}}\right)-\sum_{n\in\mathcal{N}}\sum_{i=2}^{n_r}E\left[\log\left(|r_i[n]|^2\right)\right] \\ &\quad = h\left(\boldsymbol{r}_2^{\mathcal{N}},\ldots,\boldsymbol{r}_{n_r}^{\mathcal{N}}\right) \end{aligned} \tag{68}$$

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.

[14]See Appendix E for the details.

We can upper-bound (68) by using chain rule, removing conditioning, and noting that $\{r_i[n]\}_{i=2}^{n_r}$ are identically distributed

$$h\left(\boldsymbol{r}_2^{\mathcal{N}},\ldots,\boldsymbol{r}_{n_r}^{\mathcal{N}}\right) \le \sum_{i=2}^{n_r} h\left(\boldsymbol{r}_i^{\mathcal{N}}\right) = (n_r-1)h\left(\boldsymbol{r}_2^{\mathcal{N}}\right). \tag{69}$$

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

Lemma 12:

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

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

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

$$\begin{aligned} h\left(\boldsymbol{r}_2^{\mathcal{N}}\right) &= h(r_2[1])+\sum_{n=2}^{B}h(r_2[n]|r_2[n-1],\ldots,r_2[1]) \\ &\le h(r_2[1])+\sum_{n=2}^{B}h(r_2[n]|r_2[n-1]) \\ &= h(r_2[1])+(B-1)h(r_2[2]|r_2[1]). \end{aligned} \tag{70}$$

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

$$r_2[2] = \frac{\sqrt{1-\epsilon}h_2[1]+\sqrt{\epsilon}w_2[1]}{\sqrt{1-\epsilon}h_1[1]+\sqrt{\epsilon}w_1[1]} = \frac{h_2[1]}{h_1[1]}\frac{\left(1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{\tilde{w}_2[1]}{|h_2[1]|}\right)}{\left(1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{\tilde{w}_1[1]}{|h_1[1]|}\right)} \tag{71}$$

$$\begin{aligned} &\approx r_2[1]\left(1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{\tilde{w}_2[1]}{|h_2[1]|}\right)\left(1-\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{\tilde{w}_1[1]}{|h_1[1]|}\right) \\ &\approx r_2[1]\left[1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\left(\frac{\tilde{w}_2[1]}{|h_2[1]|}-\frac{\tilde{w}_1[1]}{|h_1[1]|}\right)\right] \\ &= r_2[1]\left[1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{1}{|h_2[1]|}\left(\tilde{w}_2[1]-|r_2[1]|\tilde{w}_1[1]\right)\right] \end{aligned} \tag{72}$$

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 (71), 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 (72) 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 (73) at the top of the following page as $\epsilon\to 0$, 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 (73), 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

$$\begin{aligned} h(r_2[2]|r_2[1]) &\approx h\left(r_2[1]\left[1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{1}{|h_2[1]|}\left(\tilde{w}_2[1]-|r_2[1]|\tilde{w}_1[1]\right)\right]\middle| r_2[1]\right) \\ &= E\left[\log(|r_2[1]|^2)\right]+h\left(1+\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\frac{1}{|h_2[1]|}\left(\tilde{w}_2[1]-|r_2[1]|\tilde{w}_1[1]\right)\middle| r_2[1]\right) \\ &= \log\left(\frac{\epsilon}{1-\epsilon}\right)+h\left(\frac{1}{|h_2[1]|}\left(\tilde{w}_2[1]-|r_2[1]|\tilde{w}_1[1]\right)\middle| r_2[1]\right) \end{aligned} \tag{73}$$

and without conditioning on $r_2[1]$), and they depend on $r_2[1]$ only through its magnitude. Then we have

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

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

$$f_{V|W}(v|w)=\frac{3w(w^2+1)^4}{4\left[(1+w^2)^2+v^2w^2\right]^{5/2}} \tag{74}$$

$$f_W(w)=\frac{2w}{(w^2+1)^2} \tag{75}$$

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

$$\begin{aligned} h(V|W) &= -\int_0^\infty f_W(w) \\ &\quad\cdot\int_{-\infty}^\infty f_{V|W}(v|w)\log[f_{V|W}(v|w)]dvdw \\ &= 1.98861\ldots \end{aligned} \tag{76}$$

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 (70) becomes negligible compared to the second term as $B\to\infty$. In fact, one can check that $h(r_2[1])\le 2-\log(2)+(1/2)\log[2\pi e\pi^2]$, so that the first term of (70) vanishes when dividing by B and letting $B\to\infty$. Therefore, dividing the RHS of (70) by B and letting $B\to\infty$, we obtain the statement of the lemma.

Using this lemma in (69), and (67) in (66) we conclude the proof. □

APPENDIX G

Proof of Lemma 1

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

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

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

Using (16), we could compute $E\left[\mathrm{tr}(\boldsymbol{K}[n])\right]$ numerically and recursively, starting with $\boldsymbol{K}[n_t+1]=n_t(\mathsf{SNR}^{-1}+\epsilon)\boldsymbol{I}_{n_t}$. However, we can obtain an explicit expression by noting that (16) eventually reaches a steady state, and the transient behavior becomes irrelevant in the limit as $M\to\infty$. Taking traces and expectations in (16) we get (77) at the bottom of the page.

We will upper-bound the RHS of equality (77) with an expression that depends on $E[\mathrm{tr}(\boldsymbol{K}[n])]$, and then find the maximum value of $E[\mathrm{tr}(\boldsymbol{K}[n])]$ that satisfies the bound in steady state. Let λ_i be the ith eigenvalue of $\boldsymbol{K}[n]$, and $\lambda_{\max}=\max_i\{\lambda_i\}$. Then

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

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

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

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

$$E\left[\mathrm{tr}(\boldsymbol{K}[n+1])\right]=(1-\epsilon)\left\{E\left[\mathrm{tr}(\boldsymbol{K}[n])\right]-E\left[\left(\boldsymbol{x}[n]^T\boldsymbol{K}[n]\boldsymbol{x}[n]^*+\frac{n_t}{\mathsf{SNR}}\right)^{-1}\mathrm{tr}\left(\boldsymbol{K}[n]\boldsymbol{x}[n]^*\boldsymbol{x}[n]^T\boldsymbol{K}[n]\right)\right]\right\}+n_t\epsilon. \tag{77}$$

Noting that $\{\tilde{x}_i[n]\}_{i=1}^{n_t}$ have the same distribution as $x_1[n]$ and are independent of $\lambda_{\max}$ we can upper-bound the RHS of (77) as follows:

$$E\left[\mathrm{tr}(\boldsymbol{K}[n+1])\right] \leq (1-\epsilon)\left\{E\left[\mathrm{tr}(\boldsymbol{K}[n])\right] - E\left[\frac{\lambda_{\max}|x_1[n]|^2}{\|\boldsymbol{x}[n]\|^2+\frac{n_t}{\lambda_{\max}\mathsf{SNR}}}\right]\right\} + n_t\epsilon. \quad (78)$$

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

Lemma 13:

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

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: $\mathrm{tr}(\boldsymbol{K}[n]) = \sum_{i=1}^{n_t}\lambda_i \leq n_t\lambda_{\max}$ and $e^{-x} \geq 1-x$.

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

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

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

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

Replacing in (78) we obtain

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

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

$$\begin{aligned}
E[\mathrm{tr}(\boldsymbol{K}[n])] &\leq \left[\frac{n_t(1-\epsilon)}{2^{n_t}\mathsf{SNR}} + n_t\epsilon\right]\frac{n_t 2^{n_t}}{n_t 2^{n_t}\epsilon + 1 - \epsilon}\\
&\leq n_t^2\left(\frac{1}{\mathsf{SNR}} + 2^{n_t}\epsilon\right).
\end{aligned}$$

$$\begin{aligned}
E\left[\frac{\lambda_{\max}|x_1[n]|^2}{\|\boldsymbol{x}[n]\|^2+\frac{2}{\lambda_{\max}\mathsf{SNR}}}\right] &= E\left\{E\left[\left.\frac{\lambda_{\max}|x_1[n]|^2}{|x_1[n]|^2+|x_2[n]|^2+\frac{2}{\lambda_{\max}\mathsf{SNR}}}\right|\boldsymbol{K}[n], |x_2[n]|^2\right]\right\}\\
&= E\left[\lambda_{\max}\int_0^\infty\left(\frac{x}{x+|x_2[n]|^2+\frac{2}{\lambda_{\max}\mathsf{SNR}}}\right)e^{-x}dx\right]\\
&\geq E\left[\lambda_{\max}\int_{\left(|x_2[n]|^2+\frac{2}{\lambda_{\max}\mathsf{SNR}}\right)}^\infty\left(\frac{x}{x+|x_2[n]|^2+\frac{2}{\lambda_{\max}\mathsf{SNR}}}\right)e^{-x}dx\right]\\
&\geq E\left[\lambda_{\max}\int_{\left(|x_2[n]|^2+\frac{2}{\lambda_{\max}\mathsf{SNR}}\right)}^\infty\frac{1}{2}e^{-x}dx\right]\\
&= E\left\{\frac{\lambda_{\max}}{2}\exp\left[-\left(|x_2[n]|^2+\frac{2}{\lambda_{\max}\mathsf{SNR}}\right)\right]\right\}\\
&= E\left[\frac{\lambda_{\max}}{2}\exp\left(-\frac{2}{\lambda_{\max}\mathsf{SNR}}\right)\right]\int_0^\infty e^{-2x}dx\\
&\geq E\left[\frac{\lambda_{\max}}{4}\left(1-\frac{2}{\lambda_{\max}\mathsf{SNR}}\right)\right]\\
&\geq \frac{1}{2^2}\left\{\frac{E\left[\mathrm{tr}(\boldsymbol{K}[n])\right]}{2} - \frac{2}{\mathsf{SNR}}\right\}.
\end{aligned}$$

Finally, we have $E\left[\tilde{\boldsymbol{z}}[n]\tilde{\boldsymbol{z}}[n]^\dagger\right] = \sigma^2[n]\boldsymbol{I}_{n_r}$ where for large[15] 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 $\mathcal{C} = \{\boldsymbol{X}_1,\ldots,\boldsymbol{X}_{2^{nR}}\}$, $\boldsymbol{X}_i \in \mathbb{C}^{n_t\times B}$, $1 \leq i \leq 2^{nR}$, 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 $\boldsymbol{X}_i(k,n)$ from transmit antenna k at time n to send message i. Let $\boldsymbol{v}_i[n] = \hat{\boldsymbol{H}}[n]\boldsymbol{x}_i[n]$, where $\boldsymbol{x}_i[n]$ is the nth column of $\boldsymbol{X}_i$. Let $\boldsymbol{V}_i$ be the matrix whose nth column is $\boldsymbol{v}_i[n]$. In a similar way, define $\tilde{\boldsymbol{Y}}$ as the matrix whose nth column is $\tilde{\boldsymbol{y}}[n]$, and $\tilde{\boldsymbol{Z}}$ as the matrix whose nth column is $\tilde{\boldsymbol{z}}[n]$. Then we can rewrite the channel equation as

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

where $\tilde{\boldsymbol{Z}} \in \mathbb{C}^{n_r\times B}$ 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 $\boldsymbol{X}_i$ and independent of the fading matrices $\{\hat{\boldsymbol{H}}[n]\}$.

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

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

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

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 without loss of generality 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

$$\begin{aligned}\bar{p} &= \bar{p}(1) = E_{\mathcal{C},\tilde{\boldsymbol{Z}},\hat{\boldsymbol{H}}^{\mathcal{N}}}\left[\Pr\left(\text{Error}\middle|\mathcal{C},\tilde{\boldsymbol{Z}},\hat{\boldsymbol{H}}^{\mathcal{N}},m=1\right)\right]\\ &= 1 - E_{\mathcal{C},\tilde{\boldsymbol{Z}},\hat{\boldsymbol{H}}^{\mathcal{N}}}\left[\Pr\left(\overline{\text{Error}}\middle|\mathcal{C},\tilde{\boldsymbol{Z}},\hat{\boldsymbol{H}}^{\mathcal{N}},m=1\right)\right] \qquad (79)\end{aligned}$$

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

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

$$\Pr(\overline{\text{Error}}|\,\mathcal{C},\tilde{\boldsymbol{Z}},\hat{\boldsymbol{H}}^{\mathcal{N}},m=1) = 1\left[\bigcap_{i=2}^{2^{nR}}\left(\left\|\tilde{\boldsymbol{Y}}-\boldsymbol{V}_1\right\|_F < \left\|\tilde{\boldsymbol{Y}}-\boldsymbol{V}_i\right\|_F\right)\right]$$

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

$$= \prod_{i=2}^{2^{nR}} 1\left(\left\|\tilde{\boldsymbol{Z}}\right\|_F < \left\|\boldsymbol{V}_1 + \tilde{\boldsymbol{Z}} - \boldsymbol{V}_i\right\|_F\right).$$

Noting that the codewords $\boldsymbol{X}_2, \ldots, \boldsymbol{X}_{2^{nR}}$ are chosen i.i.d. and independently of $\tilde{\boldsymbol{Z}}$, $\{\hat{\boldsymbol{H}}[n]\}$, and $\boldsymbol{X}_1$, we can rewrite (79) as the equation at the bottom of the page, where $\tilde{\boldsymbol{Y}} = \boldsymbol{V}_1 + \tilde{\boldsymbol{Z}}$ and

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

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 14: Let $\boldsymbol{Y}$, $\tilde{\boldsymbol{Y}}$, $\boldsymbol{Z}$, and $\tilde{\boldsymbol{Z}}$ be any matrices in $\mathbb{C}^{n_r \times B}$ that satisfy the inequality

$$\left\|\boldsymbol{Y} - \tilde{\boldsymbol{Y}}\right\|_F \le \|\boldsymbol{Z}\|_F - \left\|\tilde{\boldsymbol{Z}}\right\|_F.$$

Then

$$P_{\bar{e}}\left(\boldsymbol{Y}, \boldsymbol{Z}, \hat{\boldsymbol{H}}^{\mathcal{N}}\right) \le P_{\bar{e}}\left(\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{Z}}, \hat{\boldsymbol{H}}^{\mathcal{N}}\right).$$

Proof: From Fig. 5, we see that $(\tilde{\boldsymbol{Y}} - \boldsymbol{V}_2)$, $(\boldsymbol{Y} - \boldsymbol{V}_2)$, and $(\tilde{\boldsymbol{Y}} - \boldsymbol{Y})$ form a triangle and as a result

$$\left\|\tilde{\boldsymbol{Y}} - \boldsymbol{V}_2\right\|_F \ge \|\boldsymbol{Y} - \boldsymbol{V}_2\|_F - \left\|\boldsymbol{Y} - \tilde{\boldsymbol{Y}}\right\|_F. \quad (80)$$

From (80) we have that if $\|\boldsymbol{Y} - \boldsymbol{V}_2\|_F > \|\boldsymbol{Z}\|_F$, it follows that

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

and from the hypothesis of the lemma

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

As a result, under the hypothesis of the lemma

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

and we have that $P_{\bar{e}}(\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{Z}}, \hat{\boldsymbol{H}}^{\mathcal{N}}) \ge P_{\bar{e}}(\boldsymbol{Y}, \boldsymbol{Z}, \hat{\boldsymbol{H}}^{\mathcal{N}})$. □

Consider two different channels, one with additive white circularly symmetric complex Gaussian noise of variance $\sigma^2 + \epsilon$, $\epsilon > 0$ for each component, independent of the input signal denoted by $\boldsymbol{Z} \in \mathbb{C}^{n_r \times B}$, and another one with noise of variance

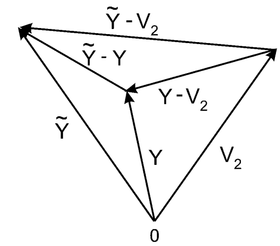

Fig. 5. Two-dimensional representation of $\boldsymbol{Y}$, $\tilde{\boldsymbol{Y}}$, and $\boldsymbol{V}_2$.

σ^2 per component, denoted by $\tilde{\boldsymbol{Z}} \in \mathbb{C}^{n_r \times B}$ in which $\tilde{\boldsymbol{Z}}(i,j)$ is independent of $\tilde{\boldsymbol{Z}}(h,k)$ for $j \ne k$ and uncorrelated with $\tilde{\boldsymbol{Z}}(h,j)$ for $i \ne h$. Since in both cases, the noise is independent across columns the Strong Law of Large Numbers (SLLN) lets us write

$$\frac{1}{B}\boldsymbol{Z}\boldsymbol{Z}^\dagger \xrightarrow{\text{a.s.}} (\sigma^2 + \epsilon)\boldsymbol{I}_{n_r} \quad \text{and} \quad \frac{1}{B}\tilde{\boldsymbol{Z}}\tilde{\boldsymbol{Z}}^\dagger \xrightarrow{\text{a.s.}} \sigma^2 \boldsymbol{I}_{n_r}.$$

Also, because the codewords are generated with independent Gaussian components of variance 1 and $\hat{\boldsymbol{H}}[n]$ is independent over time and independent of the input, and $E(\hat{\boldsymbol{h}}_i^\dagger[n]\hat{\boldsymbol{h}}_j[n]) = \alpha^2\delta_{i,j}$, by the SLLN we have that

$$\frac{1}{B}\boldsymbol{X}_i\boldsymbol{X}_i^\dagger \xrightarrow{\text{a.s.}} \boldsymbol{I}_{n_t} \quad \text{and} \quad \frac{1}{B}\boldsymbol{V}_i\boldsymbol{V}_i^\dagger \xrightarrow{\text{a.s.}} \alpha^2 \boldsymbol{I}_{n_r}.$$

Finally, noting that the noise is uncorrelated with the input signal and is independent of the fading matrices $\{\hat{\boldsymbol{H}}[n]\}$, the SLLN implies that

$$\frac{1}{B}\boldsymbol{Y}\boldsymbol{Y}^\dagger \xrightarrow{\text{a.s.}} (\alpha^2 + \sigma^2 + \epsilon)\boldsymbol{I}_{n_r}$$

and

$$\frac{1}{B}\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^\dagger \xrightarrow{\text{a.s.}} (\alpha^2 + \sigma^2)\boldsymbol{I}_{n_r}.$$

Let

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & 0 & \ldots & 0 \end{bmatrix} \in \mathbb{C}^{n_r \times B}.$$

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

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

We want to verify the hypothesis of Lemma 14 in the limit as $B \to \infty$. We first compute $\frac{1}{B}\|\tilde{\boldsymbol{Y}} - \boldsymbol{Y}\|_F^2$ as $B \to \infty$

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

Also, from the SLLN we have that with probability 1

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

To check that for $B \to \infty$, $\|\tilde{\boldsymbol{Y}} - \boldsymbol{Y}\|_F \le \|\boldsymbol{Z}\|_F - \left\|\tilde{\boldsymbol{Z}}\right\|_F$ 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

$$\lim_{B\to\infty}\left[\|\boldsymbol{Z}\|_F - \left\|\tilde{\boldsymbol{Z}}\right\|_F - \left\|\tilde{\boldsymbol{Y}} - \boldsymbol{Y}\right\|_F\right] > 0. \tag{81}$$

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

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

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

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

where the last inequality follows from (81) and Lemma 14.

Finally

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

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 [2]

$$R' < E\left[\log\det\left(\boldsymbol{I}_{n_r} + \frac{\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^\dagger}{\sigma^2+\epsilon}\right)\right].$$

Since ϵ is arbitrarily small we have that all rates below

$$C = E\left[\log\det\left(\boldsymbol{I}_{n_r} + \frac{\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^\dagger}{\sigma^2}\right)\right]$$

are achievable. Therefore, the supremum of all achievable rates R', R, must be at least as large as C.

Appendix I

Jake's Model for a Discrete-Time Flat-Fading Channel

This is a popular statistical model for flat fading. The transmitter is fixed, the mobile receiver is moving at speed v, and the transmitted signal is scattered by stationary objects around the mobile. There are K paths, the ith path arriving at an angle $\theta_i = 2\pi i/K, i = 0, \ldots, K-1$, with respect to the direction of motion. K is assumed to be large. The scattered path arriving at the mobile at the angle θ has a delay of $\tau_\theta(t)$ and a time-invariant gain a_θ, and the input/output relationship is given by

$$y(t) = \sum_{i=0}^{K-1} a_{\theta_i} x[t - \tau_{\theta_i}(t)]. \tag{82}$$

The most general version of the model allows the received power distribution $p(\theta)$ and the antenna gain pattern $\alpha(\theta)$ to be arbitrary functions of the angle θ, but the most common scenario assumes uniform power distribution and isotropic antenna gain pattern, i.e., the amplitudes $a_\theta = a/\sqrt{K}$ for all angles θ. This models the situation when the scatterers are located in a ring around the mobile (see Fig. 6). We scale the amplitude of each path by $\sqrt{K}$ so that the total received energy along all paths is a^2. For large K, the received energy along each path is a small fraction of the total energy.

This input/output relationship corresponds to that of a linear time-varying system. Let $x_b(t)$ and $y_b(t)$ be the baseband representations of $x(t)$ and $y(t)$. Then we can rewrite (82) as

$$\begin{aligned}\Re\left\{y_b(t)e^{j2\pi f_c t}\right\} &= \sum_{i=0}^{K-1} a_{\theta_i}\Re\left\{x_b(t-\tau_{\theta_i}(t))e^{j2\pi f_c(t-\tau_{\theta_i}(t))}\right\} \\ &= \Re\left\{\left[\sum_{i=0}^{K-1} a_{\theta_i} x_b(t-\tau_{\theta_i}(t))e^{-j2\pi f_c \tau_{\theta_i}(t)}\right]e^{j2\pi f_c t}\right\}\end{aligned}$$

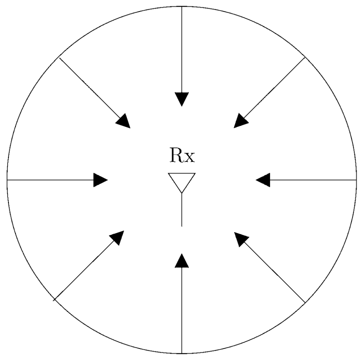

Fig. 6. The one-ring model.

where f_c is the carrier frequency.[16] Similarly, one can write

$$\Im\left\{y_b(t)e^{j2\pi f_c t}\right\} = \Im\left\{\left[\sum_{i=0}^{K-1} a_{\theta_i} x_b(t-\tau_{\theta_i}(t))e^{-j2\pi f_c \tau_{\theta_i}(t)}\right] e^{j2\pi f_c t}\right\}.$$

Hence, the baseband equivalent channel is

$$y_b(t) = \sum_{i=0}^{K-1} a_{\theta_i} e^{-j2\pi f_c \tau_{\theta_i}(t)} x_b(t-\tau_{\theta_i}(t)). \quad (83)$$

Assuming that the input signal $x(t)$ is band-limited to W, the baseband equivalent input is then limited to a bandwidth $W/2$ and can be represented using the sampling theorem as

$$x_b(t) = \sum_n x[n]\mathrm{sinc}(Wt-n)$$

where $x[n] = x_b(n/W)$ and $\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$. Using (83), the baseband output is given by

$$y_b(t) = \sum_n x[n] \sum_{i=0}^{K-1} a_{\theta_i} e^{-j2\pi f_c \tau_{\theta_i}(t)} \mathrm{sinc}(Wt - W\tau_{\theta_i}(t) - n).$$

The sampled outputs at multiples of $1/W$, $y[m] = y_b(m/W)$, are then given by

$$y[m] = \sum_n x[n] \sum_{i=0}^{K-1} a_{\theta_i} e^{-j2\pi f_c \tau_{\theta_i}(m/W)} \cdot \mathrm{sinc}(m-n-W\tau_{\theta_i}(m/W))$$

and by letting $l = m-n$ we obtain

$$y[m] = \sum_l x[m-l] \sum_{i=0}^{K-1} a_{\theta_i} e^{-j2\pi f_c \tau_{\theta_i}(m/W)} \cdot \mathrm{sinc}(l-W\tau_{\theta_i}(m/W)). \quad (84)$$

By defining

$$h_l[m] = \sum_{i=0}^{K-1} a_{\theta_i} e^{-j2\pi f_c \tau_{\theta_i}(m/W)} \mathrm{sinc}(l-W\tau_{\theta_i}(m/W))$$

[16] We denote by $\Re\{\cdot\}$ the real part operator, and by $\Im\{\cdot\}$ the imaginary part operator.

equation (84) can be written in the simple from

$$y[m] = \sum_l h_l[m]x[m-l]. \quad (85)$$

We will argue that if the communication bandwidth W is much smaller than the reciprocal of the delay spread T_d, the baseband channel can be represented by a single tap at each time. $T_d \ll 1/W$ means that most of the paths arrive in an interval much smaller than the sample time of the signal $x_b(t)$. Since the signal remains approximately constant over the interval T_d, we can extract $x_b(t-\tau_\theta(t))$ out of the sum in (83) obtaining

$$y_b(t) = x_b(t-\tau_{\theta_0}(t)) \sum_{i=0}^{K-1} a_{\theta_i} e^{-j2\pi f_c \tau_{\theta_i}(t)}. \quad (86)$$

We see from (86) that the input baseband signal $x_b(t)$ is delayed by $\tau_{\theta_0}(t)$ and multiplied by a factor $\sum_{i=0}^{K-1} a_{\theta_i} e^{-j2\pi f_c \tau_{\theta_i}(t)}$. This means that in (85), only one tap $h_l[m]$ will be nonzero at a given time. The value of l corresponding to the nonzero tap will vary with t (and m) due to the varying delay $\tau_{\theta_0}(t)$, but this variation occurs in a time scale several orders of magnitude larger than the symbol time scale. Without loss of generality we will assume that in the time scale of interest the nonzero tap is the first tap $h_0[m]$, whose value is

$$h_0[m] = \sum_{i=0}^{K-1} a_{\theta_i} e^{-j2\pi f_c \tau_{\theta_i}(m/W)}.$$

Note that in the time scale of interest $\mathrm{sinc}(-W\tau_{\theta_i}(m/W)) \approx 1$.

To derive an expression for $\tau_\theta(t)$ as a function of $\tau_\theta(0)$ we note that the projection of the velocity vector $\boldsymbol{v}$ onto the direction of the path at angle θ has magnitude $v_\theta = v\cos(\theta)$. The distance traveled by the mobile in the direction θ in time t is $v_\theta t$, which is the reduction in the distance between transmitter and receiver for the path at angle θ. It follows that

$$\tau_\theta(t) = \tau_\theta(0) - \frac{v\cos(\theta)}{c}t$$

where c is the speed of light.

In order to characterize the process $\{h_0[m]\}$, we will further assume that $\{a_{\theta_i}\}_i$ is independent of $\{\tau_{\theta_i}(0)\}_i$ and that the phase of the path at angle θ_i at time 0

$$2\pi f_c \tau_{\theta_i}(0) \bmod 2\pi$$

is uniformly distributed in $[0, 2\pi]$, and i.i.d. for all i. With these assumptions the mean of the process $\{h_0[m]\}$ is

$$\begin{aligned} E(h[m]) &= E\left\{\sum_{i=0}^{K-1} a_{\theta_i} e^{-j2\pi f_c \left[\tau_{\theta_i}(0) - \frac{vm\cos(\theta_i)}{cW}\right]}\right\} \\ &= \sum_{i=0}^{K-1} E[a_{\theta_i}]E\left[e^{-j2\pi f_c \tau_{\theta_i}(0)}\right] e^{j2\pi f_c \frac{vm\cos(\theta_i)}{cW}]} \\ &= 0 \end{aligned} \quad (87)$$

and the autocorrelation function is

$$R[m,m+n] = E\left\{ \sum_{i=0}^{K-1} \sum_{s=0}^{K-1} a_{\theta_i}^* a_{\theta_s} e^{j2\pi f_c \left[\tau_{\theta_i}(0) - \frac{vm\cos(\theta_i)}{cW} \right]} \cdot e^{-j2\pi f_c \left[\tau_{\theta_s}(0) - \frac{v(m+n)\cos(\theta_s)}{cW} \right]} \right\}$$
$$= \sum_{i=0}^{K-1} \sum_{s=0}^{K-1} E\left\{ a_{\theta_i}^* a_{\theta_s} e^{j2\pi f_c \left[\tau_{\theta_i}(0) - \tau_{\theta_s}(0) \right]} \right\} \cdot e^{j2\pi f_c \frac{v}{c} \left[\frac{m+n}{W} \cos(\theta_s) - \frac{m}{W} \cos(\theta_i) \right]}$$
$$= \sum_{i=0}^{K-1} E\left[|a_{\theta_i}|^2 \right] e^{j2\pi f_c \frac{vn}{cW} \cos(\theta_i)}. \tag{88}$$

Assuming $a_{\theta_i} = a/\sqrt{K}$ for all i, in the limit of $K \to \infty$ the sum in (88) becomes an integral and we can write

$$R[m,m+n] = \int_0^{2\pi} a^2 e^{j2\pi f_c \frac{vn}{cW} \cos(\theta)} dt.$$

We see that the process $\{h[m]\}_m$ is wide-sense stationary with autocorrelation function $R_0[n]$ given by

$$R_0[n] = 2a^2 \pi J_0 \left(n2\pi f_m / W \right) \tag{89}$$

where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind

$$J_0(x) = \frac{1}{\pi} \int_0^{\pi} e^{jx\cos\theta} d\theta$$

and $f_m = f_c v/c$ is the maximum Doppler shift. The power spectral density $S(f)$, defined on $[-1/2, 1/2]$, is given by

$$S(f) = \begin{cases} \frac{2a^2 W}{f_m \sqrt{1-(fW/f_m)^2}}, & -f_m/W \le f \le f_m/W \\ 0, & \text{else.} \end{cases} \tag{90}$$

This can be verified by computing the inverse discrete-time Fourier transform of (90) which is given by (89). Note that $S(f)$ is zero for $|f| > f_m/W$.

In addition, $\{h_0[m]\}$ is the sum of many small independent contributions, one for each angle. By the Central Limit Theorem, it is reasonable to model the process as Gaussian, so in fact the process is stationary, zero mean, with autocorrelation and power spectral density functions as given in (89) and (90).

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor Giuseppe Caire and the anonymous reviewers for their thoughtful comments and valuable suggestions which helped us improve the initial submitted version of this manuscript.

REFERENCES

[1] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, no. 6, pp. 311–335, 1998.

[2] İ.E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[3] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[4] L. Zheng and D. N. C. Tse, "Communicating on the Grassmann manifold: A geometric approach to the noncoherent multiple antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[5] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.

[6] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[7] G. Taricco and M. Elia, "Capacity of fading channels with no side information," *Electron. Lett.*, vol. 33, pp. 1368–1370, Jul. 1997.

[8] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proc. 41st Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003.

[9] ——, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.

[10] W. C. Jakes, *Microwave Mobile Communications*. Piscataway, NJ: IEEE Press, 1993.

[11] R. R. Chen, B. Hajek, R. Koetter, and U. Madhow, "On fixed input distributions for noncoherent communication over high-SNR Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3390–3396, Dec. 2004.

[12] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need perfect side information be?," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[13] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[14] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "Gaussian codes and the scaled nearest neighbor decoder in fading multi-antenna channels," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 253.

[15] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[16] R. Steele, *Mobile Radio Communications*. Piscataway, NJ: IEEE Press, 1994.

[17] R. Kennedy, *Fading Dispersive Communication Channels*. New York: Wiley, 1969.

[18] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.

[19] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "Gaussian codes and the weighted nearest neighbor decoding in fading multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1665–1686, Aug. 2004.

[20] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Select. Areas Commun.*, vol. 19, no. 12, pp. 2358–2368, Dec. 2001.

[21] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing: A General approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, no. 2, pp. 354–364, Feb. 1995.

[22] G. Colavolpe and R. Raheli, "Non-coherent sequence detection of M-ary PSK," in *Proc. IEEE Int. Conf. Communications*, vol. 1, Montreal, QC, Canada, Jun. 1997, pp. 21–25.

[23] T. Koch, "On the asymptotic capacity of multiple-input single-output fading channels with memory," Master's thesis, ETH, Zurich, Switzerland, 2004.

[24] A. Lapidoth and S. Moser, "The fading number of SIMO fading channels with memory," in *Proc. IEEE 2004 Int. Symp. Information Theory and Its Applications*, Parma, Italy, Oct. 2004, pp. 287–292.

[25] T. Koch and A. Lapidoth, "Degrees of freedom in noncoherent stationary MIMO fading channels," in *Proc. 2005 Winters Shool on Coding and Information Theory*, Bratislava, Slovakia, Feb. 2005, pp. 91–97.