Capacity of Noncoherent Time-Selective Rayleigh-Fading Channels

Yingbin Liang, *Student Member, IEEE,* and Venugopal V. Veeravalli, *Senior Member, IEEE*

***Abstract*—The capacity of noncoherent time-selective Rayleigh-fading channels is studied under various models for the variations in time. The study includes both single-input and single-output (SISO) and multiple-input and multiple-output (MIMO) systems. A block-fading model is first considered where the channel changes correlatively over each block period of length T, and independently across blocks. The predictability of the channel is characterized through the rank Q of the correlation matrix of the vector of channel gains in each block. This model includes, as special cases, the standard block-fading model where the channel remains constant over block periods $(Q = 1)$, and models where the fading process has finite differential entropy rate $(Q = T)$. The capacity is initially studied for long block lengths and some straightforward but interesting asymptotes are established. For the case where Q is kept fixed as $T \to \infty$, it is shown that the noncoherent capacity converges to the coherent capacity. For the case where both $T, Q \to \infty$, with Q/T being held constant, a bound on the capacity loss due to channel unpredictability is established. The more interesting scenario of large signal-to-noise ratio (SNR) is then explored in detail. For SISO systems, useful upper and lower bounds on the large SNR asymptotic capacity are derived, and it is shown that the capacity grows logarithmically with SNR with a slope of $\frac{T-Q}{T}$, for $Q < T$. Next, in order to facilitate the analysis of MIMO systems, the rank-Q block-fading model is specialized to the case where each T-symbol block consists of Q subblocks of length L, with the channel remaining constant over each subblock and changing correlatively across subblocks. For this model, it is shown that the $\log \mathsf{SNR}$ growth behavior of the capacity is the same as that of the standard block-fading model with block length L. Finally, the SISO and MIMO channel models are generalized to allow the fading process to be correlated across blocks in a stationary and ergodic manner. It is shown that the $\log \mathsf{SNR}$ growth behavior of the capacity is not affected by the correlation across blocks.**

***Index Terms*—Correlated fading, high signal-to-noise ratio (SNR), multiple-antenna channels, multiple-input multiple-output (MIMO), wireless channels.**

I. Introduction

RECENT work on *noncoherent* fading channels, where neither the transmitter nor the receiver knows the channel states, has generated many interesting results. The capacities of these channels and the corresponding optimal input distributions have been investigated in the literature due to their theoretical and practical importance in mobile wireless communication systems. Studies of capacity provide fundamental transmission limits for such channels, and explorations of the optimal input distribution suggest practical signaling schemes.

Abou-Faycal *et al.* [1] study a single-input single-output (SISO) memoryless Rayleigh-fading channel. While they do not explicitly obtain the noncoherent capacity of this channel, they prove that the capacity-achieving distribution is discrete with a finite number of mass points. Asymptotically firm upper and lower bounds on the noncoherent capacity for this channel model are derived by Lapidoth and Moser [2], and we describe this work in more detail below. Another study of optimal input distributions for noncoherent Gaussian Markov channels is performed by Chen *et al.* in [3].

Marzetta and Hochwald [4], [5] study multiple-input multiple-output (MIMO) Rayleigh-fading channels where the components of the channel matrix fade independently. A key assumption made in that work is that the channel remains constant over blocks consisting of several symbol periods, and changes independently from block to block—we refer to this model as the *standard block-fading model*. Under the block-fading assumption, Marzetta and Hochwald characterize the capacity-achieving input distribution and suggest useful code-design criteria. Furthermore, for the special case of SISO channels, they obtain asymptotic expressions for the capacity for large block lengths and for high signal-to-noise ratio (SNR). Zheng and Tse [6] extend these SISO asymptotic results to the MIMO case, and also provide a geometric interpretation for the capacity. In all cases, the capacity of standard block-fading channels is shown to grow *logarithmically* with SNR.

Lapidoth and Moser [2] study MIMO channels without the restrictions of block fading, Rayleigh statistics, and independent channel matrix components. Furthermore, they allow for partial side information about the channel at the receiver. However, a couple of key assumptions are made in this work. The first is that the fading process is stationary on a symbol-by-symbol basis. The second is that the fading process has finite differential entropy rate, and that the mutual information rate between the fading process and the side information is finite. Under this model, they prove that the channel capacity grows only double-logarithmically in SNR at high SNR. They also evaluate the second-order term in the high SNR asymptotic expansion of the capacity (which is called the *fading number*) in some special cases, and provide corresponding fading number achieving input distributions. We note that the double-logarithmic growth in SNR of the capacity is pessimistic, and it is due to the finite entropy rate assumption on the fading process.

Manuscript received March 17, 2003; revised September 10, 2004. This work was supported by the NSF CAREER/PECASE Award CCR 00-49089, through the University of Illinois. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Lausanne, Switzerland, June/July 2002.

The authors are with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana IL 61801 USA (e-mail: yliang1@uiuc.edu; vvv@uiuc.edu).

Communicated by İ. E. Telatar, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2004.838113

In this paper, we also assume that neither the transmitter nor the receiver knows the channel state information with both knowing the channel statistics. We consider a generalization of the standard block-fading model, where the channel changes over the block period rather than remaining constant. We refer to this model as the *time-selective block-fading model*. We first consider the scenario where the fading is independent from block to block, and later we further generalize the model to allow the fading process to be correlated across blocks. As in [1], [3]–[6], we restrict our attention to Rayleigh-fading (or zero-mean complex Gaussian) channels. For MIMO channels, we make the further restriction that the components of the channel matrix fade independently. With independent blocks, our model includes as special cases the standard block-fading model studied in [4]–[6] and the memoryless fading model studied in [1]. When we allow for correlation across blocks, our model covers the one studied in [2], where the fading is stationary on a symbol-by-symbol basis, albeit under the restriction to Rayleigh fading and independent channel matrix components. We characterize the capacity of such generalized block-fading channels through bounds and asymptotic analysis.

The paper is organized as follows. In Section II, we explain our notation and introduce our time-selective block-fading models for SISO and MIMO systems. We also provide a physical justification for the model we use based on a study of the underlying continuous-time waveform channel. In Section III, we review some known results on the capacity of fading channels which are useful in our analysis. In Section IV, we analyze the SISO model in detail. We provide bounds on the capacity, and based on these bounds, we characterize the asymptotic capacity for large block lengths. Then we move on to the high-SNR regime, and characterize the capacity more precisely in this asymptotic regime. In Section V, we extend the lower bound on the capacity for the SISO case to the MIMO case, and characterize the capacity for long block lengths. To study the capacity in the high-SNR regime, we consider a special case of the MIMO model (that we term the *subblock-correlated* fading model), which is simpler to analyze and reasonably realistic. For this model, we obtain the first-order term in high-SNR expansion of the capacity. In Section VI, we generalize our time-selective block-fading models further to allow the fading process to be correlated across blocks, and study the $\log \mathsf{SNR}$ growth behavior of the capacity at high SNR. In Section VII, we summarize our results.

II. System Model and Notation

A. Notation

The following notation is used in this paper. For deterministic objects, upper case letters denote matrices, lower case letters denote scalars, and underlined lower case letters denote vectors. Random objects are identified by corresponding boldfaced letters. For example, $\boldsymbol{X}$ denotes a random matrix, X denotes the realization of $\boldsymbol{X}$, $\underline{\boldsymbol{x}}$ denotes a random vector, and $\boldsymbol{x}$ denotes a random scalar. Superscripts are used to indicate the entries of matrices. For example, the symbol $\boldsymbol{H}^{(i,j)}$ denotes the component at the ith row and jth column of the random matrix $\boldsymbol{H}$.

Although upper case letters are typically used for matrices, there are some exceptions, and these exceptions are noted explicitly in the paper. For example, T and L denote block lengths, and SNR in the formulas denotes the SNR.

Some symbols have special meanings throughout the paper. For example, γ denotes Euler's constant which is defined by

$$\gamma = -\int_0^\infty e^{-y}\log y dy \approx 0.5772\ldots$$

and $\boldsymbol{\varphi}_i$ represents a random variable with Gamma distribution $\Gamma(i,1)$, where the probability density function (pdf) of $\Gamma(\alpha,\beta)$ is given by

$$f_{\Gamma(\alpha,\beta)}(x) = \frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-\beta x}u(x)\,.$$

The unit step function is denoted by u, i.e., $u(x)=1$ for $x\ge 0$ and $u(x)=0$ for $x<0$.

The proper complex Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{CN}(\mu,\sigma^2)$, and the joint distribution of a proper complex Gaussian vector with mean $\underline{\mu}$ and covariance matrix Σ is denoted by $\mathcal{CN}(\underline{\mu},\Sigma)$.

The trace of a matrix is denoted by $\mathrm{Tr}(\cdot)$, the Hermitian conjugation and the transpose of a matrix are denoted by $(\cdot)^\dagger$ and $(\cdot)^\top$, respectively, and the Euclidean norm of a vector is denoted by $\|\cdot\|$.

Some special matrices and vectors are denoted as follows. The m-by-m identity matrix is denoted by I_m, the m-by-m matrix with all components equal to 1 is denoted by $\mathbb{1}_m$, and the vector with dimension m and all components equal to zero is denoted by $\underline{0}_m$.

The differential entropy of $\boldsymbol{x}$ is denoted by $h(\boldsymbol{x})$, and the mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$ is denoted by $I(\boldsymbol{x};\boldsymbol{y})$. The logarithmic function to the base e is denoted by $\log(\cdot)$. The differential entropy is defined to the base e as well.

B. SISO Continuous-Time Flat-Fading Model and Discrete-Time Approximation

While the information-theoretic analysis in this paper is carried out in discrete time, we first study the fading channel in continuous time (before discretization) in order to motivate our time-selective block-fading models.

Continuous-Time Flat-Fading Model: In wireless channels, the signal leaving the transmitter reaches the receiver along several paths after reflections by scatterers in the environment. Fading in wireless channels is caused by movements in the transmitter, receiver, or the scatterers. These movements cause the linear complex baseband channel connecting the transmitter and the receiver to become time varying. For movements that are of the order of a few wavelengths, the path gains and delays can be assumed to be constant and the time variations are primarily due to changes in the phases of the different paths that connect the transmitter and receiver [7]. We denote the time interval corresponding to such *small* scale variations by $\mathcal{T}_{\rm small}$. Note that $\mathcal{T}_{\rm small}$ is not to be confused with the *coherence time* of the channel, which is the time period for which the channel does not vary significantly, i.e., even the path phases are roughly constant.

Now consider the case where the transmitter and receiver are each equipped with one antenna (SISO system). If the transmitted signal $\boldsymbol{x}(t)$ is *narrowband* relative to inverse of the delay spread of the paths at the receiver, the channel is said to be frequency *flat* and the output signal $y(t)$ can be approximately written as

$$\boldsymbol{y}(t) = \boldsymbol{h}(t)\boldsymbol{x}(t) + \boldsymbol{w}(t) \tag{1}$$

where $\boldsymbol{w}(t)$ is additive white Gaussian noise, and $\boldsymbol{h}(t)$ is the *flat fading process*. We make the flat fading assumption throughout the paper.

We model $\boldsymbol{h}(t)$ as a stationary, zero-mean proper complex Gaussian (or Rayleigh-fading) process over intervals of duration $\mathcal{T}_{\text{small}}$. The zero-mean proper complex Gaussian model is based on two implicit assumptions about the scattering environment: i) the scattering is diffuse, i.e., there are a large number of roughly equally strong paths, and ii) the path phases are independent and uniformly distributed on $[0, 2\pi]$.

Stationarity follows from the assumption that the *Doppler* frequency shift along each path is constant so that the path phase varies linearly with time with an initial random phase that is uniformly distributed on $[0, 2\pi]$. In particular, the autocorrelation function of $\boldsymbol{h}(t)$ can be expressed compactly in terms of the *Doppler power density* Ψ as follows [7]:

$$R_h(\tau) = \int_{-f_{\max}}^{f_{\max}} \Psi(f) e^{j2\pi f\tau} df \,. \tag{2}$$

The function Ψ represents the distribution of path powers as a function of Doppler frequency. It is clear that Ψ is also the power spectral density of the process $\boldsymbol{h}(t)$. Note that the support of Ψ is $[-f_{\max}, f_{\max}]$, where $f_{\max}$ is the maximum Doppler frequency for all paths. If the maximum possible speed for the mobile units (terminals or scatterers) is $v_{\max}$, then $f_{\max} = 2v_{\max}/\lambda_c$, where λ_c is the carrier wavelength. The bandwidth of the process $\boldsymbol{h}(t)$ is upper-bounded by $B_d = 2f_{\max}$. We refer to B_d as the *Doppler bandwidth.*

Remark 1: While the stationarity of $\boldsymbol{h}(t)$ seems natural and convenient for analysis, it is important to treat this assumption with some caution in information-theoretic analyses where we let the time horizon for transmission go to infinity (as in [2]). The fading process $\boldsymbol{h}(t)$ is accurately modeled to be stationary only over time durations that are smaller than $\mathcal{T}_{\text{small}}$. Beyond $\mathcal{T}_{\text{small}}$, large-scale variations in the path gains and delays cause the stationary model to break down.

Note that (2) implies that $\boldsymbol{h}(t)$ is a *strictly* bandlimited process. We emphasize the bandlimitedness of $\boldsymbol{h}(t)$ is not an approximation since it follows directly from the physical limits on the speed of the mobile units. Now consider the time interval $[0, \mathcal{T}]$, where $\mathcal{T} < \mathcal{T}_{\text{small}}$. Over $[0, \mathcal{T}]$, $\boldsymbol{h}(t)$ has a Fourier series representation. The bandlimitedness of $\boldsymbol{h}(t)$ can be exploited to truncate the Fourier series to obtain the following approximate representation (in the mean-squared sense):

$$\boldsymbol{h}(t) \approx \sum_{m=-M}^{M} \hat{\boldsymbol{h}}_m \, e^{j\frac{2\pi m t}{\mathcal{T}}}, \qquad t \in [0, \mathcal{T}] \tag{3}$$

where $M = \lfloor \mathcal{T} f_{\max} \rfloor$. Furthermore, the coefficients $\{\hat{\boldsymbol{h}}_m\}$ are approximately independent zero-mean proper complex Gaussian random variables. The approximation of course improves with increasing $\mathcal{T}$.

Remark 2: It is important to note that the approximation for $\boldsymbol{h}(t)$ given in (3) is not necessarily stationary unless we assume that the coefficients $\{\hat{\boldsymbol{h}}_m\}$ are independent.

Treating the approximation of (3) as an equality, we see that the randomness of $\boldsymbol{h}(t)$ is captured by the $2M+1$ *independent* (not necessarily identically distributed) random variables $\hat{\boldsymbol{h}}_m$, $m = -M, \ldots, 0, \ldots, M$.

Slow Fading and Discretization: We are now ready to discretize the system model given in (1). Let $\mathcal{T}_{\text{s}}$ denote the symbol period for discrete-time signaling on the channel. We make the so-called *slow fading* assumption that the symbol period $\mathcal{T}_{\text{s}}$ is much smaller than the inverse of the Doppler bandwidth B_d, i.e.,

$$B_d \mathcal{T}_{\text{s}} \ll 1 \,. \tag{4}$$

This is a reasonable assumption for most modern terrestrial wireless communication systems [8].

Under the slow fading assumption, $\boldsymbol{h}(t)$ is approximately constant over the symbol interval $\mathcal{T}_{\text{s}}$. Thus, discretizing the model of (1) via standard symbol matched filtering and sampling yields the following system model:

$$\boldsymbol{y}_k = \boldsymbol{h}_k \boldsymbol{x}_k + \boldsymbol{w}_k, \qquad k = 1, 2, \ldots \tag{5}$$

where $\{\boldsymbol{x}_k\}$ is an appropriately scaled input sequence, $\{\boldsymbol{w}_k\}$ is a sequence of independent and identically distributed (i.i.d.) zero-mean proper complex Gaussian random variables, and

$$\boldsymbol{h}_k := \boldsymbol{h}(k\mathcal{T}_{\text{s}}) = \sum_{m=-M}^{M} \hat{\boldsymbol{h}}_m \, e^{j\frac{2\pi m k \mathcal{T}_{\text{s}}}{\mathcal{T}}} \tag{6}$$

with $\{\hat{\boldsymbol{h}}_m\}$ being independent zero-mean proper complex random variables. Note that the independence of $\{\hat{\boldsymbol{h}}_m\}$ implies that the discrete-time process $\{\boldsymbol{h}_k\}$ is *stationary.*

Note that since the continuous-time model of (1) is valid over time intervals of duration less than $\mathcal{T}_{\text{small}}$, the discrete-time model is valid for blocks of symbols of block length less than $\mathcal{T}_{\text{small}}/\mathcal{T}_{\text{s}}$. This leads us to the time-selective block-fading model that we now introduce.

C. SISO Time-Selective Block-Fading Model

Consider the discrete-time system of (5) over a T symbol block, with $T < \mathcal{T}_{\text{small}}/\mathcal{T}_{\text{s}}$

$$\boldsymbol{y}_k = \sqrt{\mathsf{SNR}}\boldsymbol{h}_k \boldsymbol{x}_k + \boldsymbol{w}_k, \qquad k = 1, \ldots, T \tag{7}$$

where we normalize the system so that the channel inputs $\{\boldsymbol{x}_k\}$ have power constraint $\mathrm{E}[|\boldsymbol{x}_k|^2] \le 1$, the channel gains $\{\boldsymbol{h}_k\}$ have unit variance, and $\{\boldsymbol{w}_k\}$ is a sequence of i.i.d. $\mathcal{CN}(0,1)$ random variables. The term SNR then represents the SNR.

We collect the channel gains corresponding to one block in the vector

$$\underline{\boldsymbol{h}} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_T]^\top \,.$$

Under our Rayleigh-fading model, $\underline{\boldsymbol{h}}$ is a zero-mean proper complex Gaussian vector with covariance matrix Σ_h that has all the diagonal components equal to 1. Based on the model for $\boldsymbol{h}_k$ given in (6), the sequence $\{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots\}$ is stationary within the block, and hence Σ_h is *Toeplitz*.

The rank of Σ_h is denoted by Q and it satisfies the inequality

$$Q := \operatorname{Rank}(\Sigma_h) \leq 2M + 1 = 2\lfloor T\mathcal{T}_{\rm s} f_{\max} \rfloor + 1 \tag{8}$$

where the last equality is obtained by setting $\mathcal{T} = T\mathcal{T}_{\rm s}$ in (3).

For slow fading, $B_d\mathcal{T}_{\rm s} = 2\mathcal{T}_{\rm s} f_{\max} \ll 1$. This implies that for reasonably large T, we have $Q \ll T$, i.e., that Σ_h is highly rank deficient. In the standard block-fading model [4], [5] it is assumed that the channel gain is constant over each block (i.e., $Q = 1$), which is a good approximation for sufficiently small Doppler bandwidths.

The question that now remains is how one should model the channel variations from one block to the next. In the standard block-fading model, the channel is assumed to change in an i.i.d. manner from block to block. The independence can be justified in certain time-division or frequency-hopping systems, where the blocks are separated sufficiently in time or frequency to undergo independent fading. The independence assumption is also convenient for information-theoretic analysis as it allows us to focus on one block in studying the capacity.

Without time or frequency hopping, the channel variations from one block to the next are dictated by the long-term variations in the scattering environment. If we assume that the variations in average channel power are compensated for by other means, such as power control, it is reasonable to model the variation from block to block as stationary and ergodic.

Remark 3: The block stationary model does not imply that the fading process is stationary on a symbol-by-symbol basis as in the analysis of [2]. But as we explained earlier in Remark 1, the symbol-by-symbol stationary model is not realistic for time intervals that are larger than $\mathcal{T}_{\rm small}/\mathcal{T}_{\rm s}$ symbols anyway. For this reason, it may be more accurate to model the fading process using a block-fading model with possible correlation across blocks than it is to model it as a symbol-by-symbol stationary process.

Throughout the paper, we assume that neither the transmitter nor the receiver knows the channel state information (realizations of channel gains), but both know the channel statistics. In our analysis of the noncoherent capacity of time-selective block-fading channels, we will first assume that the fading is independent from block to block. We will consider the generalization to dependent block fading in Section VI.

D. MIMO Time-Selective Block-Fading Model

We now generalize our time-selective block-fading model to MIMO systems that employ multiple antennas at the transmitter and receiver.

The channel output $\underline{\boldsymbol{y}}_k \in \mathbb{C}^{n_r}$, corresponding to the kth symbol vector within a block of T symbol vectors, is given by

$$\underline{\boldsymbol{y}}_k = \sqrt{\frac{\mathsf{SNR}}{n_t}} \boldsymbol{H}_k \underline{\boldsymbol{x}}_k + \underline{\boldsymbol{w}}_k, \qquad k = 1, \ldots, T \tag{9}$$

where n_t and n_r denote the numbers of transmit and receive antennas, respectively, and $\{\underline{\boldsymbol{w}}_k\}_{k=1}^{\infty}$ is a sequence of i.i.d. $\mathcal{CN}(\underline{0}_{n_r}, I_{n_r})$ random vectors. The channel input $\underline{\boldsymbol{x}}_k \in \mathbb{C}^{n_t}$ has power constraint

$$\frac{1}{n_t} \mathrm{E}[\|\underline{\boldsymbol{x}}_k\|^2] \leq 1 \tag{10}$$

where $\|\underline{\boldsymbol{x}}_k\|$ denotes the Euclidean norm of vector $\underline{\boldsymbol{x}}_k$. We use $\boldsymbol{H}_k$ to denote the $n_r \times n_t$ channel gain matrix at time instant k, and $\boldsymbol{H}_k^{(i,j)}$ to denote its entry at the ith row and jth column. We assume that the scattering is sufficiently rich so that the channel gains corresponding to different antenna pairs are independent; hence, each $\boldsymbol{H}_k$ has i.i.d. entries. We then group the channel gains from time instant 1 to T corresponding to the ith receive and jth transmit antenna pair into a vector

$$\underline{\boldsymbol{h}}^{(i,j)} = \left[\boldsymbol{H}_1^{(i,j)}, \boldsymbol{H}_2^{(i,j)}, \ldots, \boldsymbol{H}_T^{(i,j)}\right]. \tag{11}$$

Under our Rayleigh-fading model, each $\underline{\boldsymbol{h}}^{(i,j)}$ is a zero-mean proper complex Gaussian vector. All the antenna pairs, of course, have identically distributed channel gains, and hence the covariance matrix of $\underline{\boldsymbol{h}}^{(i,j)}$ is the same for all pairs (i, j). We denote this covariance matrix by Σ_h. As in the SISO time-selective model, the channel gains are normalized so that all the diagonal components of Σ_h equal 1, and Q denotes the rank of the Toeplitz matrix Σ_h.

We initially assume that the channel matrix changes independently from block to block and later allow for correlation across blocks. Also, in order to facilitate the analysis of the MIMO system, we specialize the rank-Q block-fading model to the case where each T-symbol block consists of Q subblocks with length L, with the channel remaining constant over each subblock and changing correlatively across subblocks. This special case will be described in detail in Section V-B.

III. Known Results

In this section, we review some existing results on the channel capacity for a MIMO system with n_t transmit antennas and n_r receive antennas. The channel at time instants k is given by

$$\underline{\boldsymbol{y}}_k = \sqrt{\frac{\mathsf{SNR}}{n_t}} \boldsymbol{H}_k \underline{\boldsymbol{x}}_k + \underline{\boldsymbol{w}}_k \tag{12}$$

where $\boldsymbol{H}_k$ is the channel matrix at time instant k with i.i.d. $\mathcal{CN}(0, 1)$ entries.

Various channel models may follow from (12) by different assumptions about the time variations of the channel matrix sequence $\{\boldsymbol{H}_k\}$ and about the availability of channel state information. The channel capacities of some of these channel models are summarized below.

The first result concerns the coherent capacity of a MIMO system, and is derived in [9], [10].

Lemma 1: In (12), assume that the channel matrix $\boldsymbol{H}_k$ varies in a stationary and ergodic manner over time k. If the channel state information is perfectly known at the receiver only, then the coherent capacity is given by

$$C_{\rm coh}(\mathsf{SNR}) = \mathrm{E}_{\boldsymbol{H}} \log\det\left(I_{n_r} + \frac{\mathsf{SNR}}{n_t} \boldsymbol{H}\boldsymbol{H}^{\dagger}\right) \tag{13}$$

where $\boldsymbol{H}$ denotes the random matrix with i.i.d. $\mathcal{CN}(0, 1)$ entries.

In particular, the coherent capacity for a single antenna system is

$$C_{\rm coh}(\mathsf{SNR}) = \mathrm{E}_{\boldsymbol{h}} \log(1+\mathsf{SNR}|\boldsymbol{h}|^2) = \mathrm{E}_{\boldsymbol{\varphi}_1} \log(1+\mathsf{SNR}\boldsymbol{\varphi}_1) \tag{14}$$

where $\boldsymbol{h}$ is a random variable with distribution of $\mathcal{CN}(0,1)$, and $\boldsymbol{\varphi}_1$ is defined in Section II-A.

A lower bound on the capacity in (13) is given by

$$C_{\rm coh}(\mathsf{SNR}) \geq \min\{n_t, n_r\} \log \frac{\mathsf{SNR}}{n_t} + \sum_{i=|n_t-n_r|+1}^{\max\{n_t,n_r\}} \mathrm{E} \log \boldsymbol{\varphi}_i \tag{15}$$

where $\boldsymbol{\varphi}_i$ is defined in Section II-A.

The noncoherent capacity of the standard block-fading MIMO channel is computed in [6]. We summarize the main result here.

Lemma 2: In (12), assume the standard block-fading model, where the channel matrix remains constant for T-symbol blocks, and changes independently from block to block. Let $K = \min\{n_t, n_r\}$. If $T \geq K + n_r$, the noncoherent capacity is given by

$$C(\mathsf{SNR}) = K\left(1 - \frac{K}{T}\right) \log \mathsf{SNR} + c + o(1) \tag{16}$$

where c is a constant that does not depend on SNR, and $o(1)$ goes to zero as SNR goes to infinity. In the general case, including $T < K + n_r$, the capacity satisfies

$$c_1 \leq C(\mathsf{SNR}) - n^*\left(1 - \frac{n^*}{T}\right) \log \mathsf{SNR} \leq c_2. \tag{17}$$

Here, c_1 and c_2 are constants that do not depend on the SNR, and $n^* = \min\{n_t, n_r, \lfloor \frac{T}{2} \rfloor\}$. At high SNR, the optimal input must have n^* antennas transmitting signals with power much higher than the noise level.

For a MIMO channel with the fading process being stationary and ergodic on a symbol-by-symbol basis and having finite entropy rate, the noncoherent capacity is derived in [2].

Lemma 3: Consider a MIMO fading channel given in (12), where $\{\boldsymbol{H}_k\}$ is stationary and ergodic (not necessarily Gaussian with possibly correlated matrix components), satisfying

$$h(\{\boldsymbol{H}_k\}) > -\infty.$$

Assume that $\{\boldsymbol{H}_k\}$ is independent of $\{\underline{\boldsymbol{w}}_k\}$ and $\{\underline{\boldsymbol{x}}_k\}$, and that

$$h(\{\underline{\boldsymbol{w}}_k\}) > -\infty.$$

Then the noncoherent capacity satisfies

$$\overline{\lim_{\mathsf{SNR}\to\infty}} \left\{ C(\mathsf{SNR}) - \log\log \mathsf{SNR} \right\} < \infty. \tag{18}$$

Moreover, define the second-order term in the high-SNR expansion of the capacity as the fading number, i.e.,

$$\chi(\{\boldsymbol{H}_k\}) = \overline{\lim_{\mathsf{SNR}\to\infty}} \left\{ C(\mathsf{SNR}) - \log\log \mathsf{SNR} \right\} \tag{19}$$

where variance of the noise is assumed to be 1. Then the following lemma provides results for the fading number.

Lemma 4: For the channel model described in Lemma 3, assume that the channel is memoryless and rotation commutative in the generalized sense (as defined in [2]), then the $\overline{\lim}$ in (19) is also a $\underline{\lim}$ and the fading number is given by

$$\chi_{\rm iid}(\boldsymbol{H}) = \log \frac{2\pi^{n_r}}{\Gamma(n_r)} - \log 2 + n_r \mathrm{E}\left[\log \|\boldsymbol{H}\underline{\hat{e}}\|^2\right] - h(\boldsymbol{H}\underline{\hat{e}}) \tag{20}$$

where $\underline{\hat{e}}$ is any deterministic unit-vector in $\mathbb{C}^{n_t}$. This fading number is achievable by inputs that can be expressed as the product of a random vector $\underline{\hat{\boldsymbol{x}}}$ and an independent circularly symmetric scalar random variable $\boldsymbol{a}$. The vector $\underline{\hat{\boldsymbol{x}}}$ is uniformly distributed on the unit n_t-sphere, and the distribution of $\boldsymbol{a}$ is such that $\log|\boldsymbol{a}|$ is uniformly distributed over the interval $[\log a_{\min}, \frac{1}{2}\log \mathsf{SNR}]$ for any $a_{\min}(\mathsf{SNR})$ satisfying

$$\lim_{\mathsf{SNR}\to\infty} a_{\min} = \infty, \quad \text{and} \quad \lim_{\mathsf{SNR}\to\infty} \frac{\log a_{\min}}{\log \mathsf{SNR}} = 0.$$

In the special case of the MIMO Rayleigh-fading channel where $n_t = n_r$, the fading number is given by

$$\chi = n_r \psi(n_r) - n_r - \log \Gamma(n_r) \tag{21}$$

where

$$\psi(n_r) = -\gamma + \sum_{j=1}^{n_r - 1} \frac{1}{j}. \tag{22}$$

For the SISO Rayleigh-fading case, the above fading number reduces to $\chi = -1 - \gamma$.

IV. Noncoherent Capacity for the SISO Time-Selective Block-Fading Model

We first study the SISO time-selective block-fading model given in Section II-C with independent fading across blocks. In Section V, we extend our results to MIMO channels. In Section VI, we generalize the results to time-selective models with correlated fading across blocks.

In the following result we provide a lower bound on the noncoherent capacity.

Proposition 1: For the SISO time-selective block-fading model given in Section II-C with independent fading across blocks, a lower bound on the capacity as a function of SNR is given by

$$\begin{aligned} C_{\rm low}(\mathsf{SNR}) &= \mathrm{E}_{\boldsymbol{\varphi}_1} \log(1+\mathsf{SNR}\boldsymbol{\varphi}_1) - \frac{Q}{T} \mathrm{E}_{\boldsymbol{\varphi}_T} \log\left(1 + \frac{\mathsf{SNR}}{Q}\boldsymbol{\varphi}_T\right) \\ &= C_{\rm coh}(\mathsf{SNR}) - \frac{Q}{T} \mathrm{E}_{\boldsymbol{\varphi}_T} \log\left(1 + \frac{\mathsf{SNR}}{Q}\boldsymbol{\varphi}_T\right) \end{aligned} \tag{23}$$

where $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_T$ have the Gamma distributions defined in Section II-A, and $C_{\rm coh}$ is the coherent capacity (see (14)).

Proof: The result follows as a special case of Proposition 5, which is a more general result for MIMO channels. See Appendix I for the proof of Proposition 5. □

The coherent capacity $C_{\rm coh}(\mathsf{SNR})$ is an obvious upper bound on the noncoherent capacity $C(\mathsf{SNR})$. Thus, we have

$$C_{\rm coh}(\mathsf{SNR}) - \frac{Q}{T} \mathrm{E}_{\boldsymbol{\varphi}_T} \log\left(1 + \frac{\mathsf{SNR}}{Q}\boldsymbol{\varphi}_T\right) \leq C(\mathsf{SNR}) \leq C_{\rm coh}(\mathsf{SNR}). \tag{24}$$

A. Long Block Length Asymptotic Capacity

To proceed further with the analysis, we resort to asymptotic analyses. We begin by studying the capacity for long block lengths, with the understanding that this asymptote needs to be treated with some caution as we point out in Section II (Remarks 1 and 3). In particular, the block length T should be less than $\mathcal{T}_{\rm small}/\mathcal{T}_{\rm s}$ in order for our block-fading model to be of practical relevance. Thus, large values of T are relevant only in scenarios where the fading is very slow compared to the symbol rate. On the other hand, the large T asymptote does provide some useful insights into the role of channel predictability in determining the noncoherent capacity.

We have the following result by letting $T \to \infty$ in (24).

Proposition 2: For the SISO time-selective block-fading model given in Section II-C with independent fading across blocks, the noncoherent capacity converges to the coherent capacity as $T \to \infty$ and Q is fixed, i.e.,

$$\lim_{T\to\infty} C(\mathsf{SNR}) = C_{\rm coh}(\mathsf{SNR}).$$

Furthermore, if $T, Q \to \infty$, with $\frac{Q}{T} = \eta$ being fixed

$$\varliminf_{T,Q\to\infty;\frac{Q}{T}=\eta} C(\mathsf{SNR}) \geq C_{\rm coh}(\mathsf{SNR}) - \eta \log\left(1 + \frac{\mathsf{SNR}}{\eta}\right). \quad (25)$$

The first part of the proposition (with fixed Q) corresponds an environment where, in addition to the fading being very slow, the scattering is very sparse so that only a finite number of effective scatterers contribute to the fading. For such a scenario, the coherent capacity is asymptotically achievable without channel state information at the receiver. This is reasonable since we can send a finite number of training symbols to accurately estimate the channel state at the receiver, and then the channel state can be assumed to be perfectly known in the remaining (infinite) channel uses.

The second part of Proposition 2 corresponds to a rich scattering scenario where the number of effective scatterers that contribute to the fading increases linearly with T. In fact, for the slow-fading model (see Sections II-B and II-C), it can be argued that Q/T approximately equals $B_d\mathcal{T}_{\rm s}$. In such a scenario, the asymptotic result suggests that the capacity remains bounded away from the coherent capacity. The capacity loss relative to the coherent capacity is captured by the second term in (25), and this loss is due to channel unpredictability.

B. High-SNR Asymptotic Capacity

The coherent capacity served as a sufficiently tight upper bound for the study of the asymptotic capacity for long block lengths. In the high-SNR regime, however, the coherent capacity is no longer a sufficiently tight upper bound. In the following, we first compute the high SNR expansion of the lower bound given in Proposition 1. Then we derive another upper bound on the capacity whose first-order term is asymptotically tight in the high-SNR regime.

The following proposition gives a high-SNR asymptotic lower bound on the capacity.

Proposition 3: The lower bound on the capacity (23) has the following asymptotic expansion in SNR:

$$C_{\rm low}(\mathsf{SNR}) = \frac{T-Q}{T}\log \mathsf{SNR} - \frac{T-Q}{T}\gamma - \frac{Q}{T}\left[\sum_{i=1}^{T-1}\frac{1}{i} - \log Q\right] + o(1) \quad (26)$$

where γ is Euler's constant as defined in Section II-A and $o(1)$ goes to zero at high SNR.

Proof: See Appendix II. □

Before we give an upper bound for the general case $1 \leq Q \leq T$, we present in the following lemma an upper bound for the special case when $Q = T$. This lemma is useful in deriving the upper bound for the general case.

Lemma 5: [1] For the SISO time-selective block-fading model defined in Section II-C with independent fading across blocks, assume $Q = T$. The capacity $C(\mathsf{SNR})$ is bounded by

$$C_{\rm iid}(\mathsf{SNR}) \leq C(\mathsf{SNR}) \leq C_{\rm iid}(\mathsf{SNR}) - \log \lambda_{\min}\{\Sigma_h\} \quad (27)$$

where $C_{\rm iid}(\mathsf{SNR})$ is the capacity of the i.i.d. Rayleigh-fading channel with channel gain having distribution of $\mathcal{CN}(0,1)$ at each time instant. In the upper bound, the symbol $\lambda_{\min}\{\Sigma_h\}$ denotes the minimal eigenvalue of Σ_h, and it can be easily seen that $\lambda_{\min}\{\Sigma_h\} \leq 1$. The upper bound has the following asymptotic expansion in SNR:

$$C(\mathsf{SNR}) \leq \log\log \mathsf{SNR} - \gamma - 1 - \log \lambda_{\min}\{\Sigma_h\} + o(1). \quad (28)$$

Proof: See Appendix III for the proof of the left inequality in (27). The right inequality in (27) follows as a special case of Lemma 6 (of Section V-B) with $L = 1$ and $Q = T$. The inequality (28) follows directly from (27) and Lemma 4. □

Lemma 5 indicates that although the channel gains in one block may have memory, as long as the covariance matrix Σ_h has full rank, the capacity increase due to channel memory remains bounded as SNR approaches infinity.

We now generalize the upper bound in (27) to the case $1 \leq Q \leq T$.

Proposition 4: Consider the SISO time-selective block-fading model given in Section II-C with independent fading across blocks. For $1 \leq Q \leq T$, an asymptotic upper bound on the capacity is given by

$$C(\mathsf{SNR}) \leq \frac{T-Q}{T}\log \mathsf{SNR} + \frac{Q}{T}\Big(\log\log \mathsf{SNR} - \gamma - 1 - \log \lambda_{\min}\{(\Sigma_h)_Q\}\Big) + o(1) \quad (29)$$

where $(\Sigma_h)_Q$ is one of the $Q \times Q$ full-rank principal submatrices of Σ_h.

Proof: See Appendix V. □

[1]A similar upper bound as in (27) is also derived by Lapidoth and Moser in [2, Lemma 4.3]. However, there are differences in the conditions required in the proofs. Our derivation does not require the fading process to be stationary, which allows the result to be extended to other scenarios such as the orthogonal frequency-division multiplexing (OFDM) model of Corollary 1, where the fading coefficients are generally not stationary across frequency tones. The proof given in [2] requires the fading process to be stationary and ergodic, but it is valid for more general fading distributions than the Rayleigh and Ricean.

The asymptotic bounds given in Propositions 3 and 4 can be combined to lead to the following result on the high SNR behavior of the capacity.

Theorem 1: Consider the SISO time-selective block-fading model given in Section II-C with independent fading across blocks. For $Q < T$, the high-SNR asymptotic capacity satisfies the relationship

$$\lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log \mathsf{SNR}} = \frac{T-Q}{T}. \tag{30}$$

It is instructive to interpret the intuition behind this result. From our assumption that the rank $Q < T$, Q random variables completely determine the channel coefficients in the entire block. Theorem 1 suggests that we have to essentially give up Q out of T channel uses in the high-SNR regime. In particular, consider the training scheme that takes Q channel uses to estimate the channel. As $\mathsf{SNR} \to \infty$, this training scheme can effectively obtain perfect estimates of the channel gains in the entire block, and remaining $T-Q$ channel uses can assume perfect channel knowledge. Our results say that such a training scheme achieves the first-order term of the high SNR expansion of the capacity. Our proof for the lower bound also suggests that using i.i.d. Gaussian inputs is another way to achieve this first-order term.

Our analysis thus far has considered only time-selective fading. In the following result, we show that frequency selectivity can also be incorporated in our approach, under the restriction of OFDM signaling [11].

Corollary 1: Consider a SISO system with Q taps in the delay spread of the channel. Let $\boldsymbol{h}_0, \boldsymbol{h}_1, \ldots, \boldsymbol{h}_{Q-1}$ denote the channel gains corresponding to the Q taps. Then the output at time k is given by

$$\boldsymbol{y}_k = \sqrt{\mathsf{SNR}} \sum_{i=0}^{Q-1} \boldsymbol{h}_i \boldsymbol{x}_{k-i} + \boldsymbol{w}_k. \tag{31}$$

Assume that $[\boldsymbol{h}_0, \boldsymbol{h}_1, \ldots, \boldsymbol{h}_{Q-1}]^\top$ is a circularly complex Gaussian random vector with mean zero and covariance matrix having full rank. Suppose we use an OFDM signaling scheme with T tones in the frequency domain $(T > Q)$. Further assume that the channel gains remain constant for the duration of one OFDM symbol and change to new independent realizations for the next OFDM symbol. Then, ignoring the loss in capacity due to the use of prefix symbols, the noncoherent capacity satisfies (30).

Proof: The discrete Fourier transform (DFT) of $\boldsymbol{h}_0$, $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{Q-1}$ is given by

$$\hat{\boldsymbol{h}}_k = \sum_{i=0}^{Q-1} \boldsymbol{h}_i e^{-j\frac{2\pi}{T}ki}, \qquad k = 0, 1, \ldots, T-1.$$

Similarly, let $\{\hat{\boldsymbol{y}}_k\}_{k=0}^{T-1}$, $\{\hat{\boldsymbol{x}}_k\}_{k=0}^{T-1}$, and $\{\hat{\boldsymbol{w}}_k\}_{k=0}^{T-1}$ be the DFTs of $\{\boldsymbol{y}_k\}_{k=0}^{T-1}$, $\{\boldsymbol{x}_k\}_{k=0}^{T-1}$, and $\{\boldsymbol{w}_k\}_{k=0}^{T-1}$, respectively.

Then, in the frequency domain, the output from the kth tone is given by

$$\hat{\boldsymbol{y}}_k = \sqrt{\mathsf{SNR}}\hat{\boldsymbol{h}}_k \hat{\boldsymbol{x}}_k + \hat{\boldsymbol{w}}_k, \qquad k = 0, 1, \ldots, T-1. \tag{32}$$

Since

$$\hat{\underline{\boldsymbol{h}}} = [\hat{\boldsymbol{h}}_0, \hat{\boldsymbol{h}}_1, \ldots, \hat{\boldsymbol{h}}_{T-1}]^\top$$

is a circularly complex Gaussian random vector with covariance matrix having rank Q, (32) is identical to (7), except that the channel gains may not have the same variances. But this does not affect the first-order high-SNR expansion term of the capacity. So the result follows from Theorem 1. □

The same approach can also be generalized to doubly selective (both time and frequency-selective) underspread fading channels, under restriction of the pulse-shaped OFDM modulation [12] or short-time Fourier signaling [13].

V. Noncoherent Capacity for MIMO Time-Selective Block-Fading Model

In the previous section we studied the asymptotic properties of the noncoherent capacity for SISO channels. We now extend this study to MIMO time-selective block-fading channels.

A. Lower Bound on Capacity and Long Block-Length Asymptotics

We first derive a lower bound on the noncoherent capacity for the MIMO time-selective block-fading model with independent fading across blocks. We then show that this lower bound is asymptotically tight for long block lengths.

Proposition 5: For the MIMO time-selective block-fading model given in Section II-D with independent fading across blocks, a lower bound on the capacity as a function of SNR is given by

$$C_{\text{low}}(\mathsf{SNR}) = \mathrm{E}_{\boldsymbol{H}} \log\det\left(I_{\nu_r} + \frac{\mathsf{SNR}}{\nu_t}\boldsymbol{H}\boldsymbol{H}^\dagger\right) - \frac{\nu_r \nu_t Q}{T} \mathrm{E}_{\boldsymbol{\varphi}_{\nu_t T}}\left[\log\left(1 + \frac{\mathsf{SNR}}{\nu_t^2 Q}\boldsymbol{\varphi}_{\nu_t T}\right)\right] \tag{33}$$

where ν_t and ν_r are the numbers of transmit and receive antennas actually used that maximize the lower bound.

Proof: See Appendix I. □

Proposition 5 suggests that in multiple-antenna systems we may not want to use all the available antennas. The reason is that although using more antennas can create more equivalent parallel channels, which contributes positively to capacity, this also introduces more unknown channel gains in one block, which contributes negatively to capacity. This point is further clarified if we consider the following example training scheme. In each block, we estimate the channel for a certain number of symbol periods and then use the remaining symbol periods coherently. When the block length is long enough, we may use all the antennas to make the number of parallel channels as large as possible. This is because after channel estimation, there are still enough symbol periods for coherent transmission in the block. The gains offered by using all possible parallel channels will offset the losses in estimating the unknown channel gains. However, when the block length is small, if we use too many antennas, the block may end before we estimate all the channel gains so that we transmit little or no information in the block.

Hence, there is a tradeoff between the number of parallel channels we want to create and the number of unknown channel gains we have to estimate. Choosing to use the optimal number of antennas will balance this tradeoff and maximize the transmission rate.

As in the SISO case, we first consider the large T asymptote to establish some interesting boundary values for the capacity. For the case where Q remains fixed as $T \to \infty$, if we set $\nu_t = n_t$ and $\nu_r = n_r$ in (33) and compare it to the simple upper bound given by coherent capacity of Lemma 1, we immediately get the asymptotic result shown in the following proposition. In this case, using all the available antennas is optimal. For the case where both $T, Q \to \infty$, with Q/T being held constant, as in the SISO case, we are able to bound the capacity loss due to channel unpredictability. Note that in this case using all the available antennas may not be optimum.

Proposition 6: For the MIMO time-selective block-fading model given in Section II-D with independent fading across blocks, where $T \to \infty$ with Q being fixed, the noncoherent capacity converges to the coherent capacity

$$\lim_{T\to\infty} C(\mathsf{SNR}) = C_{\text{coh}}(\mathsf{SNR}) = \mathrm{E}_{\boldsymbol{H}} \log\det\left(I_{n_r} + \frac{\mathsf{SNR}}{n_t}\boldsymbol{H}\boldsymbol{H}^\dagger\right).$$

Furthermore, if $T, Q \to \infty$, with $\frac{Q}{T} = \eta$ being fixed

$$\underline{\lim}_{T,Q\to\infty;\frac{Q}{T}=\eta} C(\mathsf{SNR}) \geq \max_{1\leq\nu_t\leq n_t, 1\leq\nu_r\leq n_r} \Bigg\{ \mathrm{E}_{\boldsymbol{H}} \log\det\left(I_{\nu_r} + \frac{\mathsf{SNR}}{\nu_t}\boldsymbol{H}\boldsymbol{H}^\dagger\right) - \eta\nu_t\nu_r \log\left(1 + \frac{\mathsf{SNR}}{\eta\nu_t}\right)\Bigg\}$$

where ν_t and ν_r need to be chosen to make the lower bound tightest.

B. Subblock-Correlated Model and High-SNR Asymptotic Capacity

We now move on to the more interesting high-SNR regime. It would be natural to follow the same steps as in the high SNR analysis for the SISO model. However, finding a tight upper bound for the capacity for the general model given in Section II-D appears to be difficult. Therefore, we will study a special case of the general model, *subblock-correlated model*, that is simpler to analyze and still reasonably realistic.

For the MIMO time-selective block-fading model given in Section II-D, we assume a special structure for the vector $\underline{\boldsymbol{h}}^{(i,j)}$ in (11). Suppose each T symbol block has Q subblocks with each subblock undergoing constant fading, and the fading across subblocks being correlated. The length of each subblock is L, with $L > 1$ and $QL = T$. Hence, $\underline{\boldsymbol{h}}^{(i,j)}$ can be written as

$$\underline{\boldsymbol{h}}^{(i,j)} = \Big[\underbrace{\boldsymbol{H}_1^{(i,j)}, \ldots, \boldsymbol{H}_1^{(i,j)}}_{L \text{ symbols}}, \underbrace{\boldsymbol{H}_2^{(i,j)}, \ldots, \boldsymbol{H}_2^{(i,j)}}_{L \text{ symbols}}, \ldots, \underbrace{\boldsymbol{H}_Q^{(i,j)}, \ldots, \boldsymbol{H}_Q^{(i,j)}}_{L \text{ symbols}}\Big]. \quad (34)$$

Let Σ_{cor} denote the covariance matrix of such $\underline{\boldsymbol{h}}^{(i,j)}$, and note that it is same for all (i,j) pairs as assumed before. The channel gains from each of the subblocks can be grouped into a new vector

$$\underline{\boldsymbol{b}}^{(i,j)} = \left[\boldsymbol{H}_1^{(i,j)}, \boldsymbol{H}_2^{(i,j)}, \ldots, \boldsymbol{H}_Q^{(i,j)}\right]$$

with covariance matrix Σ_b (same for all (i,j) pairs) which is assumed to have full rank Q. The relationship between Σ_{cor} and Σ_b can be expressed as $\Sigma_{\text{cor}} = \Sigma_b \otimes \mathbb{I}_L$ where $\otimes$ denotes the Kronecker product and $\mathbb{I}_L$ denotes the $L \times L$ matrix with all components equal to 1. So each subblock follows the standard block fading model, and the correlation between the subblocks is described by Σ_b. We call this model the *subblock-correlated model*. For the above subblock-correlated model, if the subblocks have independent fading coefficients, we use Σ_{ind} to denote the covariance matrix of $\underline{\boldsymbol{h}}^{(i,j)}$. Then $\Sigma_{\text{ind}} = I_Q \otimes \mathbb{I}_L$, where I_Q denotes $Q \times Q$ identity matrix. We call this model the *subblock-independent model*. Note that if the channel matrix changes independently from one block to another, the subblock-independent model is nothing but the standard block-fading model with block length equal to L.

The following lemma relates the capacities of the subblock-correlated and the subblock-independent models.

Lemma 6: Assume that the channel matrix changes independently from one block to another. Let $C_{\text{cor}}(\mathsf{SNR})$ and $C_{\text{ind}}(\mathsf{SNR})$ denote the capacities of the subblock-correlated model and subblock-independent model, respectively. Then

$$C_{\text{ind}}(\mathsf{SNR}) \leq C_{\text{cor}}(\mathsf{SNR}) \leq C_{\text{ind}}(\mathsf{SNR}) - n_r \log\lambda_{\min}\{\Sigma_b\} \quad (35)$$

where $\lambda_{\min}\{\Sigma_b\}$, which is less than 1, denotes the minimal eigenvalue of Σ_b.

Proof: The proof of the left inequality in (35) is a straightforward extension of the proof in Appendix III. See Appendix IV for the proof of the right inequality in (35). □

The above lemma implies that the subblock-correlated model with subblock length L (with independent fading across blocks) has the same first-order term in the high SNR expansion of the capacity as that of the standard block-fading model with block length L, i.e., the two channels have the same $\log \mathsf{SNR}$ growth behavior. As in the SISO case, the capacity gain due to the correlation between subblocks is bounded as SNR approaches infinity. By applying (17) in Lemma 2 for the capacity $C_{\text{ind}}(\mathsf{SNR})$ of the subblock-independent model, we obtain the following asymptotic result for the subblock-correlated model.

Corollary 2: The capacity of the MIMO subblock-correlated model given in (34) with independent fading across blocks satisfies

$$\lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log \mathsf{SNR}} = n^*_{\text{cor}}\left(1 - \frac{n^*_{\text{cor}}}{L}\right) \quad (36)$$

where $n^*_{\text{cor}} = \min\{n_t, n_r, \lfloor \frac{L}{2} \rfloor\}$.

Now we note that the high SNR expansion of the lower bound given in Proposition 5 yields an asymptotic lower bound on the capacity of the general MIMO time-selective model given in Section II-D. This lower bound is, of course, applicable for the subblock-correlated model, and it happens to be tight in the first-order term in this special case, i.e., it achieves the $\log \mathsf{SNR}$ growth behavior shown in (36).

Proposition 7: For the MIMO time-selective block-fading model given in Section II-D with independent fading across blocks

$$C_{\text{low}}(\mathsf{SNR}) = n^* \left(1 - \frac{n^* Q}{T}\right) \log \mathsf{SNR} + c_1 + o(1) \quad (37)$$

where $n^* = \min\{n_t, n_r, \lfloor \frac{T}{2Q} \rfloor\}$, and c_1 is a constant that does not depend on SNR.

In particular, for the subblock-correlated model given in (34)

$$C_{\text{low}}(\mathsf{SNR}) = n^*_{\text{cor}} \left(1 - \frac{n^*_{\text{cor}}}{L}\right) \log \mathsf{SNR} + c_1 + o(1) \quad (38)$$

where n^*_{cor} is defined in Corollary 2.

Proof: By applying Proposition 5, we get

$$C(\mathsf{SNR}) \geq \mathrm{E}_{\boldsymbol{H}} \log \det \left(I_{\nu_r} + \frac{\mathsf{SNR}}{\nu_t} \boldsymbol{H}\boldsymbol{H}^\dagger\right) - \frac{\nu_r \nu_t Q}{T} \mathrm{E}_{\boldsymbol{\varphi}_{\nu_t T}} \left[\log\left(1 + \frac{\mathsf{SNR}}{\nu_t^2 Q} \boldsymbol{\varphi}_{\nu_t T}\right)\right]. \quad (39)$$

Using (15) and Lemma 7 in Appendix II, the lower bound can be expanded as

$$C(\mathsf{SNR}) \geq \left[\min\{\nu_t, \nu_r\} - \frac{\nu_t \nu_r Q}{T}\right] \log \mathsf{SNR} + c_2 + o(1) \quad (40)$$

where c_2 is a constant that does not depend on SNR. Optimizing the first term in (40) subject to $1 \leq \nu_t \leq n_t$ and $1 \leq \nu_r \leq n_r$, we get the tightest bound shown in (37). This lower bound is achieved when $\nu_t = \nu_r = n^*$.

For the subblock-correlated model, the result in (38) follows from the fact that $T = QL$. □

The derivation of the lower bound in Proposition 5 and Proposition 7 show that $\underline{\boldsymbol{x}}_k$ with i.i.d. Gaussian components achieves the first-order term for the subblock-correlated model if n^*_{cor} transmit and receive antennas are used.

Although the exact first-order term of the high SNR expansion of the capacity is only obtained for the SISO time-selective block-fading model and the MIMO subblock-correlated model, the intuition behind these results leads us to make the following conjecture.

Conjecture 1: For the MIMO time-selective block fading model given in Section II-D with independent fading across blocks, the lower bound given in Proposition 7 is tight in the first-order high SNR expansion term of the capacity, i.e., the capacity satisfies

$$\lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log \mathsf{SNR}} = n^* \left(1 - n^* \frac{Q}{T}\right). \quad (41)$$

An explanation for this conjecture is as follows. If we use ν_t transmit and ν_r receive antennas, we may use $\nu_t Q$ symbol periods to estimate the $\nu_t \nu_r Q$ fading gains in the first Q channel matrices in each block. Then the channel matrices in the entire block can be assumed to be known, and the remaining symbol periods can be used coherently to achieve

$$\frac{\min\{\nu_t, \nu_r\}(T - \nu_t Q)}{T} \log \mathsf{SNR} = \min\{\nu_t, \nu_r\} \left(1 - \nu_t \frac{Q}{T}\right) \log \mathsf{SNR} \quad (42)$$

as the first-order term in the high SNR expansion of the noncoherent capacity. While we may not necessarily explicitly estimate the channel, Theorem 1 and Corollary 2 suggest that the training scheme achieves the first-order term in the capacity asymptotics. We now need to optimize (42) subject to $1 \leq \nu_t \leq n_t$ and $1 \leq \nu_r \leq n_r$. The solution is to choose $\nu_t = n^*$ and $\nu_t \leq \nu_r \leq n_r$, and the first-order term of the capacity is then as given in the above conjecture.

VI. Generalization to Dependent Block Fading

Thus far, we have considered time-selective models with independent fading across blocks. In the following, we allow for correlation across blocks in these models and show that the main results in the previous sections remain unchanged.

A. SISO Case

Consider the SISO time-selective block-fading model given in Section II-C. Let $\underline{\boldsymbol{h}}_n$ indicate the vector of fading coefficients corresponding to the nth block. In Section IV, we have considered the case where the fading process is independent across blocks, i.e., the sequence of vectors $\{\underline{\boldsymbol{h}}_n\}_{n=1,2,\ldots}$ is an i.i.d. vector random process. We now consider a more general case where $\{\underline{\boldsymbol{h}}_n\}_{n=1,2,\ldots}$ is allowed to be a stationary ergodic vector process such that the channel is stationary and ergodic across blocks (see [14, Ch. 9.3] for a precise definition). Stationarity of course implies that $\underline{\boldsymbol{h}}_n$ has the same marginal distribution for each n, and as described in Section II-C, $\underline{\boldsymbol{h}}_n$ is zero-mean proper complex Gaussian with covariance matrix Σ_h having rank Q. To be consistent with our previous nomenclature, we still refer to the above model as a time-selective block-fading model, but with correlated fading across blocks.

We now further assume that the correlation between blocks is such that the rank of the vector of channel coefficients corresponding to N blocks is NQ, i.e., for any n, perfect prediction of $\underline{\boldsymbol{h}}_n$ is not possible from $\underline{\boldsymbol{h}}_1, \ldots, \underline{\boldsymbol{h}}_{n-1}$. Note that this is consistent with the physical model of (6). This predictability assumption can be expressed more precisely in the following way. Without loss of generality, assume that the vector of the first Q fading coefficients in a block has a full-rank covariance matrix. We then group the first Q fading coefficients from each of the first N blocks into one vector

$$\underline{\boldsymbol{h}}_{[1,N]} = \Big[\underbrace{\boldsymbol{h}_1^{(1)}, \boldsymbol{h}_1^{(2)}, \ldots, \boldsymbol{h}_1^{(Q)}}_{Q \text{ symbols from } \underline{\boldsymbol{h}}_1}, \underbrace{\boldsymbol{h}_2^{(1)}, \boldsymbol{h}_2^{(2)}, \ldots, \boldsymbol{h}_2^{(Q)}}_{Q \text{ symbols from } \underline{\boldsymbol{h}}_2}, \ldots, \underbrace{\boldsymbol{h}_N^{(1)}, \boldsymbol{h}_N^{(2)}, \ldots, \boldsymbol{h}_N^{(Q)}}_{Q \text{ symbols from } \underline{\boldsymbol{h}}_N}\Big]$$

where $\boldsymbol{h}_n^{(k)}$ indicates the kth fading coefficient in the nth block. Let Σ_N be the covariance matrix of $\underline{\boldsymbol{h}}_{[1,N]}$. Then the predictability assumption stated above is expressed compactly by

$$\lambda^* := \varliminf_{N\to\infty} \lambda_{\min}\{\Sigma_N\} > 0. \tag{43}$$

It can be readily seen that the SISO time-selective model defined in Section II-C with independent fading across blocks satisfies (43). We now give an example of the model with correlated fading across blocks. Assume that the sequence $\{\underline{\boldsymbol{h}}_n\}_{n=1,2,\ldots}$ is a vector Gaussian Markov process that evolves as

$$\underline{\boldsymbol{h}}_n = \alpha \underline{\boldsymbol{h}}_{n-1} + \sqrt{1-\alpha^2}\underline{\boldsymbol{z}}_n, \qquad n = 2,3,\ldots \tag{44}$$

where $\alpha \in (0,1)$. The initial random vector $\underline{\boldsymbol{h}}_1$ has the distribution $\mathcal{CN}(\underline{0}, \Sigma_h)$, and the sequence $\{\underline{\boldsymbol{z}}_n\}_{n=1,2,\ldots}$ is i.i.d. $\mathcal{CN}(\underline{0}, \Sigma_h)$. We also assume that the sequence $\{\underline{\boldsymbol{h}}_n\}_{n=1,2,\ldots}$ and the sequence $\{\underline{\boldsymbol{z}}_n\}_{n=1,2,\ldots}$ are statistically independent.

Let $(\Sigma_h)_Q$ denote the covariance matrix of the vector corresponding to the first Q fading coefficients in a block. Then it is easy to show that

$$\Sigma_N = \Sigma_\alpha^N \otimes (\Sigma_h)_Q \tag{45}$$

where

$$\Sigma_\alpha^N = \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{N-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{N-2} \\ \alpha^2 & \alpha & 1 & \cdots & \alpha^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{N-1} & \alpha^{N-2} & \alpha^{N-3} & \cdots & 1 \end{bmatrix}. \tag{46}$$

The minimum eigenvalue of Σ_N then satisfies

$$\begin{aligned} \lambda^* &:= \varliminf_{N\to\infty} \lambda_{\min}\{\Sigma_\alpha^N \otimes (\Sigma_h)_Q\} \\ &\geq \varliminf_{N\to\infty} \lambda_{\min}\{\Sigma_\alpha^N\} \quad \lambda_{\min}\{(\Sigma_h)_Q\} \\ &\geq \frac{1-\alpha}{1+\alpha} \quad \lambda_{\min}\{(\Sigma_h)_Q\} > 0. \end{aligned}$$

We further note that this vector Gaussian Markov channel is ergodic because it satisfies the asymptotically memoryless condition [14, Ch. 9.4]. This channel provides an example of a time-selective channel with correlated fading across blocks that satisfies the assumption given in (43).

We now consider the capacity of this class of channels (time-selective channels with correlated fading across blocks). Since these channels are stationary and ergodic on the block-by-block basis, the general formula on the capacity (normalized by the symbol period) is given by [14, Ch. 12.4] and [15]

$$\lim_{N\to\infty} \frac{1}{NT} \max_{p(\underline{\boldsymbol{x}}_{[1,N]})} I(\underline{\boldsymbol{y}}_{[1,N]}; \underline{\boldsymbol{x}}_{[1,N]}) \tag{47}$$

with $\underline{\boldsymbol{x}}_{[1,N]}$ and $\underline{\boldsymbol{y}}_{[1,N]}$ denoting the inputs and outputs of the first N blocks, respectively.

To lower-bound the capacity, we can still use i.i.d. Gaussian input symbols and get the same results as in Propositions 1 and 3. For the upper bound, we can use the approach in Section IV-B to get the same results as in Proposition 4, except that $\lambda_{\min}\{(\Sigma_h)_Q\}$ in (29) is replaced by λ^*. Combining these lower and upper bounds, we obtain the following result.

Proposition 8: Consider the SISO time-selective block-fading model with correlated fading across blocks. Assume (43) is satisfied and that $Q < T$. Then Theorem 1 still holds, i.e.,

$$\lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log \mathsf{SNR}} = \frac{T-Q}{T}. \tag{48}$$

Therefore, the $\log \mathsf{SNR}$ growth behavior of the SISO time-selective model with correlated fading across blocks is the same as that of the model with independent fading across blocks, provided that (43) is satisfied.

B. MIMO Case

Consider the MIMO subblock-correlated model defined in (34). We have considered the case where the fading process changes correlatively from one subblock to another, and changes independently across blocks. We now consider a more general case, where the fading process is also correlated across blocks. Interestingly, after this generalization, the boundaries between blocks are indistinguishable (statistically) from the boundaries between subblocks. We therefore refer to the subblocks as blocks, and the model is one where the channel matrix remains constant within blocks, and changes correlatively from one block to another. Hence, this model is nothing but the standard block-fading model with correlation across blocks.

We now let $\boldsymbol{H}_n$ denote the channel matrix of the nth block. We assume that the sequence of matrices $\{\boldsymbol{H}_n\}_{n=1,2,\ldots}$ is a stationary and ergodic process. Define the vector $\underline{\boldsymbol{h}}_{[1,N]}^{(i,j)}$ as

$$\underline{\boldsymbol{h}}_{[1,N]}^{(i,j)} = \left[\boldsymbol{H}_1^{(i,j)}, \boldsymbol{H}_2^{(i,j)}, \ldots, \boldsymbol{H}_N^{(i,j)}\right] \tag{49}$$

where $\boldsymbol{H}_n^{(i,j)}$ denotes the fading coefficient associated with the ith receive and jth transmit antenna in block n. Let Σ_N denote the covariance matrix of $\underline{\boldsymbol{h}}_{[1,N]}^{(i,j)}$, and assume that it is the same for all (i,j) pairs. As in the SISO case, we assume that the matrix Σ_N satisfies

$$\lambda^* := \varliminf_{N\to\infty} \lambda_{\min}\{\Sigma_N\} > 0. \tag{50}$$

The following proposition follows from the same reasoning as in the SISO case.

Proposition 9: Consider the MIMO block-fading model with the fading being constant within blocks and correlated across blocks. Assume that (50) is satisfied and that the block length $L > 1$. Then Corollary 2 still holds, i.e.,

$$\lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log \mathsf{SNR}} = n_{\text{cor}}^* \left(1 - \frac{n_{\text{cor}}^*}{L}\right) \tag{51}$$

where $n_{\text{cor}}^* = \min\{n_t, n_r, \lfloor \frac{L}{2} \rfloor\}$.

VII. Conclusion

For noncoherent channels, channel memory represented by the correlation of fading gains over time plays a key role in determining the capacity. In this paper, we studied time-selective block-fading channels, where channel memory is characterized by the rank Q of the covariance matrix Σ_h of the fading gains over one block.

For the SISO time-selective block-fading model we proved that at high SNR the noncoherent capacity grows logarithmically with SNR, with a slope of $\frac{T-Q}{T}$. This result explicitly shows that channel memory affects the capacity through Q. For channels with Σ_h having full rank $(Q = T)$, correlated channels may have higher capacity than the i.i.d. channel. We showed that the difference between the capacities is bounded by a constant that is determined by the minimal eigenvalue of Σ_h. This result suggests that channels with Σ_h having the same rank $(Q < T)$ may differ in the second-order high SNR expansion term of the capacity and this term depends on more detailed characteristics of Σ_h.

It is interesting to compare our capacity result with a recent result by Lapidoth [16] on the noncoherent capacity for a discrete-time stationary and ergodic fading channel without the restriction to block fading. From Section II-C, we know that block rank Q is approximately equal to $B_d\mathcal{T}_sT$, where B_d is the Doppler bandwidth of the fading process and $\mathcal{T}_s$ is the symbol period. Furthermore, $\mathcal{T}_s$ is approximately the inverse of the signaling bandwidth, which we denote by W. Now the $\log \mathsf{SNR}$ growth behavior of the capacity in Theorem 1 can be written as

$$\lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log \mathsf{SNR}} = \frac{T-Q}{T} \approx \frac{W - B_d}{W} \tag{52}$$

which is the ratio of the length of the frequency band where the spectral density of the fading process is null to the total bandwidth. This is consistent with the capacity result of (50) given in [16] for the stationary ergodic fading channel. Note that in our time-selective block-fading model, the blocks of fading gains are a discretized version of a time slice of the continuous stationary fading process. Each block retains the "limiting predictability" of the original fading process, which determines the high-SNR behavior of the capacity. This is why the $\log \mathsf{SNR}$ growth behavior of the two channels are consistent.

We also generalized our model to include MIMO systems and introduced a subblock-correlated model as a special case. We pointed out that as long as Σ_b has full rank, i.e., we do not have perfect prediction across subblocks, the $\log \mathsf{SNR}$ growth behavior is the same as that of the standard block fading with subblocks having independent fading. Furthermore, we showed for the SISO time-selective and MIMO subblock-correlated models that correlation across blocks does not affect the dominant term ($\log \mathsf{SNR}$ term) of the capacity, as long as perfect prediction is not possible across blocks.

In this paper, we have focussed on the first-order term of the high SNR expansion of the capacity. As future work, the bounds on the capacity can be further refined in order to obtain the second-order expansion term. It would be interesting to see how this term relates to the correlation matrix Σ_h. It is also of interest to study MIMO channel models where the elements of the channel matrix are correlated.

Appendix I
Proof of Proposition 5

We derive a lower bound on capacity by using ν_t transmit antennas and ν_r receive antennas. The lower bound can then be tightened by maximizing over $1 \le \nu_t \le n_t$ and $1 \le \nu_r \le n_r$.

The output vectors and input vectors in one block can be grouped into two matrices and denoted by

$$\boldsymbol{Y} = [\underline{\boldsymbol{y}}_1 \quad \underline{\boldsymbol{y}}_2 \quad \cdots \quad \underline{\boldsymbol{y}}_T] \quad \text{and} \quad \boldsymbol{X} = [\underline{\boldsymbol{x}}_1 \quad \underline{\boldsymbol{x}}_2 \quad \cdots \quad \underline{\boldsymbol{x}}_T]$$

respectively. The channel matrices are denoted by

$$\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_T \in \mathbb{C}^{\nu_r \times \nu_t}.$$

We have the following general bound on the mutual information:

$$\begin{aligned} I(\boldsymbol{Y};\boldsymbol{X}) &= I(\boldsymbol{Y};\boldsymbol{X},\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_T) \\ &\quad - I(\boldsymbol{Y};\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_T|\boldsymbol{X}) \\ &\ge I(\boldsymbol{Y};\boldsymbol{X}|\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_T) \\ &\quad - I(\boldsymbol{Y};\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_T|\boldsymbol{X}). \end{aligned}$$

Let $\underline{\boldsymbol{x}}^g$ denote the input vector when it has Gaussian distribution $\mathcal{CN}(\underline{0}_{\nu_t}, I_{\nu_t})$ with i.i.d. components. Now assume that the sequence of input vectors over the block is i.i.d. with each vector being $\mathcal{CN}(\underline{0}_{\nu_t}, I_{\nu_t})$, and define $\boldsymbol{X}_g = [\underline{\boldsymbol{x}}_1^g \quad \underline{\boldsymbol{x}}_2^g \quad \cdots \quad \underline{\boldsymbol{x}}_T^g]$. We then get the following lower bound on the capacity per channel use:

$$\begin{aligned} C &\ge \frac{1}{T} I(\boldsymbol{Y};\boldsymbol{X}_g) \\ &\ge \frac{1}{T} I(\boldsymbol{Y};\boldsymbol{X}_g|\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_T) \\ &\quad - \frac{1}{T} I(\boldsymbol{Y};\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_T|\boldsymbol{X}_g). \end{aligned} \tag{53}$$

The first term can be shown to be

$$\begin{aligned} &\frac{1}{T} I(\boldsymbol{Y};\boldsymbol{X}_g|\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_T) \\ &= \frac{1}{T}\mathrm{E}_{\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_T}\left[\sum_{k=1}^{T} I\left(\underline{\boldsymbol{y}}_k;\underline{\boldsymbol{x}}_k^g\middle|\boldsymbol{H}_k\right)\right] \\ &= \frac{1}{T}\mathrm{E}_{\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_T}\left[\sum_{k=1}^{T} \log\det\left(I_{\nu_r} + \frac{\mathsf{SNR}}{\nu_t}\boldsymbol{H}_k\boldsymbol{H}_k^\dagger\right)\right] \\ &= \mathrm{E}_{\boldsymbol{H}} \log\det\left(I_{\nu_r} + \frac{\mathsf{SNR}}{\nu_t}\boldsymbol{H}\boldsymbol{H}^\dagger\right) \end{aligned} \tag{54}$$

where $\boldsymbol{H}$ is a $\nu_r \times \nu_t$ matrix with i.i.d. $\mathcal{CN}(0,1)$ entries.

In the following, we use $\boldsymbol{x}^{(i)}$ to denote the ith component of the vector $\underline{\boldsymbol{x}}$.

Conditioned on the input matrix X in one block, the output from the ith antenna at time instant k is given by

$$\boldsymbol{y}_k^{(i)} = \sqrt{\frac{\mathsf{SNR}}{\nu_t}} \sum_{j=1}^{\nu_t} \boldsymbol{H}_k^{(i,j)} x_k^{(j)} + \boldsymbol{w}_k^{(i)}. \tag{55}$$

It is easy to see that $\boldsymbol{y}_k^{(i)}$ is complex circularly Gaussian with zero mean. The correlation between $\boldsymbol{y}_k^{(i)}$ and $\boldsymbol{y}_{k'}^{(i')}$ can be computed by

$$\begin{aligned} &\mathrm{E}\left[\boldsymbol{y}_k^{(i)} \boldsymbol{y}_{k'}^{(i')^*}\right] \\ &= \mathrm{E}\left(\sqrt{\frac{\mathsf{SNR}}{\nu_t}} \sum_{j=1}^{\nu_t} \boldsymbol{H}_k^{(i,j)} x_k^{(j)} + \boldsymbol{w}_k^{(i)}\right) \end{aligned}$$

$$\times \left(\sqrt{\frac{\mathsf{SNR}}{\nu_t}} \sum_{j'=1}^{\nu_t} \boldsymbol{H}_{k'}^{(i',j')^*} x_{k'}^{(j')^*} + \boldsymbol{w}_{k'}^{(i')^*} \right)$$
$$= \frac{\mathsf{SNR}}{\nu_t} \sum_{j=1}^{\nu_t} \sum_{j'=1}^{\nu_t} \mathrm{E}\left(\boldsymbol{H}_k^{(i,j)} \boldsymbol{H}_{k'}^{(i',j')^*} \right) x_k^{(j)} x_{k'}^{(j')^*} + \delta_{ii'}\delta_{kk'}$$
$$= \frac{\mathsf{SNR}}{\nu_t} \left[\sum_{j=1}^{\nu_t} \Sigma_h^{(k,k')} x_k^{(j)} x_{k'}^{(j)^*} + \delta_{kk'} \right] \delta_{ii'}$$

where $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$. Hence, the outputs from different antennas are independent, and the second term in (53) becomes

$$\frac{1}{T} I\left(\boldsymbol{Y}; \boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_T | \boldsymbol{X}_g\right)$$
$$= \frac{1}{T} \mathrm{E}_{\boldsymbol{X}_g} \Big[h(\boldsymbol{Y}|\boldsymbol{X}_g = X) - h(\boldsymbol{Y}|\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_T, \boldsymbol{X}_g{=}X) \Big]$$
$$= \frac{\nu_r}{T} \mathrm{E}_{\boldsymbol{X}_g} \left[\log[(\pi e)^T \det R_{\boldsymbol{y}}] - \log[(\pi e)^T] \right]$$
$$= \frac{\nu_r}{T} \mathrm{E}_{\boldsymbol{X}_g} \left[\log\left(\det R_{\boldsymbol{y}}\right) \right]$$

where $R_{\boldsymbol{y}}$ is defined as

$$R_{\boldsymbol{y}} = \frac{\mathsf{SNR}}{\nu_t} \sum_{j=1}^{\nu_t} P_j \Sigma_h P_j^\dagger + I_T \tag{56}$$

with P_j being a $T \times T$ diagonal matrix

$$P_j = \mathrm{diag}(x_1^{(j)}, x_2^{(j)}, \ldots, x_T^{(j)}). \tag{57}$$

Let $D := \sum_{j=1}^{\nu_t} P_j \Sigma_h P_j^\dagger$. Then

$$\mathrm{Rank}(D) = \mathrm{Rank}\left(\sum_{j=1}^{\nu_t} P_j \Sigma_h P_j^\dagger \right) \leq \sum_{j=1}^{\nu_t} \mathrm{Rank}\left(P_j \Sigma_h P_j^\dagger \right) \leq \sum_{j=1}^{\nu_t} \mathrm{Rank}\left(\Sigma_h\right) = \nu_t Q.$$

The trace of the matrix D satisfies

$$\mathrm{Tr}(D) = \mathrm{Tr}\left(\sum_{j=1}^{\nu_t} P_j \Sigma_h P_j^\dagger \right) = \sum_{j=1}^{\nu_t} \mathrm{Tr}\left(P_j \Sigma_h P_j^\dagger \right) = \sum_{j=1}^{\nu_t} \sum_{k=1}^{T} |x_k^{(j)}|^2.$$

Note that the matrix D can have at most $\nu_t Q$ nonzero eigenvalues. Without loss of optimality, we assume $\nu_t Q \leq T$, because $\nu_t Q > T$ leads to a looser lower bound. Let λ_i, $i = 1, \ldots, \nu_t Q$, denote the $\nu_t Q$ largest eigenvalues of D (some of them could be zero). Clearly, $\lambda_i \geq 0$ for all i. Then

$$\mathrm{Tr}(D) = \sum_{i=1}^{\nu_t Q} \lambda_i = \sum_{j=1}^{\nu_t} \sum_{k=1}^{T} |x_k^{(j)}|^2. \tag{58}$$

Hence,

$$\frac{1}{T} I(\boldsymbol{Y}; \boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_T | \boldsymbol{X}_g = X)$$
$$= \frac{\nu_r}{T} \log \prod_{i=1}^{\nu_t Q} \left[1 + \lambda_i \frac{\mathsf{SNR}}{\nu_t} \right]$$
$$= \frac{\nu_r}{T} \sum_{i=1}^{\nu_t Q} \log \left[1 + \lambda_i \frac{\mathsf{SNR}}{\nu_t} \right]$$
$$\leq \frac{\nu_r}{T} \sum_{i=1}^{\nu_t Q} \log \left[1 + \frac{\mathsf{SNR}}{\nu_t} \frac{1}{\nu_t Q} \sum_{j=1}^{\nu_t} \sum_{k=1}^{T} |x_k^{(j)}|^2 \right]$$
$$= \frac{\nu_r \nu_t Q}{T} \log \left[1 + \frac{\mathsf{SNR}}{\nu_t^2 Q} \sum_{j=1}^{\nu_t} \sum_{k=1}^{T} |x_k^{(j)}|^2 \right] \tag{59}$$

where for the above inequality we used the Lagrange Multiplier Rule to get maximum value under the constraint (58) and the constraints $\lambda_i \geq 0$ for $i = 1, \ldots, \nu_t Q$.

Since $\boldsymbol{x}_k^{(j)}$ are i.i.d. $\mathcal{CN}(0,1)$, $\sum_{j=1}^{\nu_t} \sum_{k=1}^{T} |\boldsymbol{x}_k^{(j)}|^2$ has the distribution $\Gamma(\nu_t T, 1)$ and is denoted by $\boldsymbol{\varphi}_{\nu_t T}$.

Then (59) can be expressed as

$$\frac{1}{T} I\left(\boldsymbol{Y}; \boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_T | \boldsymbol{X}_g\right) \leq \frac{\nu_r \nu_t Q}{T} \mathrm{E}_{\boldsymbol{\varphi}_{\nu_t T}} \left[\log\left(1 + \frac{\mathsf{SNR}}{\nu_t^2 Q} \boldsymbol{\varphi}_{\nu_t T} \right) \right]. \tag{60}$$

Plugging (54) and (60) into (53), we get the lower bound on the capacity shown in Proposition 5.

APPENDIX II
PROOF OF PROPOSITION 3

We first present two lemmas that will be used in the proof of Proposition 3.

Lemma 7: Let $\boldsymbol{\varphi}_k$ denote a random variable with distribution $\Gamma(k,1)$ and $k \geq 1$. Then

$$\mathrm{E}[\log(1 + \mathsf{SNR}\boldsymbol{\varphi}_k)] = \log \mathsf{SNR} + \mathrm{E}\log\boldsymbol{\varphi}_k + o(1) \tag{61}$$

where o(1) goes to zero as SNR goes to infinity.

Proof:

$$\mathrm{E}[\log(1 + \mathsf{SNR}\boldsymbol{\varphi}_k)]$$
$$= \int_0^\infty \frac{x^{k-1}e^{-x}}{\Gamma(k)} \log\left(1 + \mathsf{SNR}x\right) dx$$
$$= \int_0^\infty \frac{x^{k-1}e^{-x}}{\Gamma(k)} \left[\log \mathsf{SNR} + \log\left(\frac{1}{\mathsf{SNR}} + x \right) dx \right]$$
$$= \log \mathsf{SNR} + \int_0^\infty \frac{x^{k-1}e^{-x}}{\Gamma(k)} \log\left(\frac{1}{\mathsf{SNR}} + x \right) dx$$
$$= \log \mathsf{SNR} + \int_{\frac{1}{\mathsf{SNR}}}^\infty \frac{\left(y - \frac{1}{\mathsf{SNR}}\right)^{k-1} e^{-y + \frac{1}{\mathsf{SNR}}}}{\Gamma(k)} \log y dy$$
$$= \log \mathsf{SNR} + \int_0^\infty \frac{y^{k-1}e^{-y}}{\Gamma(k)} \log y dy - \int_0^{\frac{1}{\mathsf{SNR}}} \frac{y^{k-1}e^{-y}}{\Gamma(k)} \log y dy + O\left(\frac{1}{\mathsf{SNR}}\right)$$
$$= \log \mathsf{SNR} + \mathrm{E}\log\boldsymbol{\varphi}_k - \int_0^{\frac{1}{\mathsf{SNR}}} \frac{y^{k-1}e^{-y}}{\Gamma(k)} \log y dy + O\left(\frac{1}{\mathsf{SNR}}\right). \tag{62}$$

We may bound the third term in the preceding equation in the following way:

$$\begin{aligned}\left|\int_0^{\frac{1}{\mathsf{SNR}}} \frac{y^{k-1}e^{-y}}{\Gamma(k)} \log y dy\right| &\le \left(\frac{1}{\mathsf{SNR}}\right)^{k-1} \left|\int_0^{\frac{1}{\mathsf{SNR}}} \frac{e^{-y}}{\Gamma(k)} \log y dy\right| \\ &\le O\left(\frac{\log \mathsf{SNR}}{\mathsf{SNR}^k}\right).\end{aligned}$$

Hence, the last two terms in (62) together can be expressed by $o(1)$ and the desired result follows. □

Lemma 8: For $k \ge 1$

$$\Phi_k := \mathrm{E}\log \boldsymbol{\varphi}_k = -\gamma + \sum_{i=1}^{k-1} \frac{1}{i} \tag{63}$$

where γ is Euler's constant.

The proof of Lemma 8 follows from elementary calculus steps.

Now we are ready to prove Proposition 3. Using Proposition 1, and Lemmas 7 and 8, we obtain

$$\begin{aligned}C_{\text{low}}(\mathsf{SNR}) &= \mathrm{E}_{\boldsymbol{\varphi}_1} \log(1+\mathsf{SNR}\boldsymbol{\varphi}_1) \\ &\quad - \frac{Q}{T}\mathrm{E}_{\boldsymbol{\varphi}_T} \log\left(1+\frac{\mathsf{SNR}}{Q}\boldsymbol{\varphi}_T\right) \\ &= \log \mathsf{SNR} - \gamma + o(1) \\ &\quad - \frac{Q}{T}\left[\log\frac{\mathsf{SNR}}{Q} - \gamma + \sum_{i=1}^{T-1}\frac{1}{i}\right] + o(1) \\ &= \frac{T-Q}{T}\log \mathsf{SNR} - \frac{T-Q}{T}\gamma \\ &\quad - \frac{Q}{T}\left[\sum_{i=1}^{T-1}\frac{1}{i} - \log Q\right] + o(1).\end{aligned}$$

APPENDIX III
PROOF OF THE LEFT INEQUALITY IN (27) OF LEMMA 5

The inequality essentially says that the channel with memory has better performance than the memoryless channel provided that the two channels have the same marginal distributions for fading and noise. Although comparing the mutual informations of the two channels is a straightforward way to execute the proof, manipulating the chain rules of mutual information and entropy could overshadow the insight in the inequality. The following proof uses a codebook base argument and is more insightful.

Consider a sequence of codebooks with fixed rate, denoted by $\{\mathcal{C}(n)\}$, with n equal to the codeword length. For notational simplicity, we do not necessarily explicitly write the index n. A code $\mathcal{C}$ achieves rate R on a channel if it has rate R, and when $\mathcal{C}$ is applied to the channel, the average error probability goes to zero as n goes to infinity.

The left inequality in (27) of Lemma 5 is true if we can show that the following statement is true. For any code $\mathcal{C}$ that achieves rate R on i.i.d. fading channel, there exists a corresponding code $\mathcal{C}'$ that achieves the same rate R on the correlated fading channel. Note that $Q = T$ throughout this proof.

For a code $\mathcal{C}$ achieving rate R on the i.i.d. fading channel, let $\mathcal{C}' = \underbrace{\mathcal{C}\times\mathcal{C}\times\cdots\times\mathcal{C}}_{T}$, i.e.,

$$\mathcal{C}' = \{c' : c' = c_1c_2\cdots c_T, \quad c_i \in \mathcal{C} \text{ for } 1 \le i \le T\}.$$

Then the code $\mathcal{C}'$ has length nT and rate R. We want to prove that the code $\mathcal{C}'$ achieves rate R on the correlated fading channel. Given a codeword $c' = c_1c_2\cdots c_T \in \mathcal{C}'$, we apply it to n consecutive blocks of the correlated fading channel in the following way. We use c_1 for the first time instant of each of the n blocks, use c_2 for the second time instant of each block, and so on, for all $c_i, 1 \le i \le T$. Since the fading gains in different blocks are independent, each c_i experiences an i.i.d. fading channel. The receiver then separately decodes each of $c_1, c_2, \ldots, c_T$ according to the decision region used when $\mathcal{C}$ is applied to the i.i.d. fading channel. So although the codeword $c' = c_1c_2\cdots c_T$ experiences correlated fading, the decoder does not use the correlation information. It remains to show that the average error probability when using $\mathcal{C}'$ tends to zero as the codeword length nT goes to infinity.

Assuming that the codewords are equally likely to be sent through the channel, the average error probability satisfies

$$\begin{aligned}\bar{p}_e^{\text{cor}} &= \frac{1}{|\mathcal{C}'|}\sum_{c\in\mathcal{C}'} P(\text{error} \mid c \text{ is sent}) \\ &\le \frac{1}{|\mathcal{C}|^T}\sum_{c_1,\ldots,c_T\in\mathcal{C}}\Big[P(\hat{c}_1 \ne c_1 \mid c_1 \text{ is sent}) \\ &\quad + P(\hat{c}_2 \ne c_2 \mid c_2 \text{ is sent}) \\ &\quad + \cdots + P(\hat{c}_T \ne c_T \mid c_T \text{ is sent})\Big] \\ &= \frac{1}{|\mathcal{C}|^T}|\mathcal{C}|^{T-1}\Big[\sum_{c_1\in\mathcal{C}} P(\hat{c}_1 \ne c_1 \mid c_1 \text{ is sent}) \\ &\quad + \sum_{c_2\in\mathcal{C}} P(\hat{c}_2 \ne c_2 \mid c_2 \text{ is sent}) \\ &\quad + \cdots + \sum_{c_T\in\mathcal{C}} P(\hat{c}_T \ne c_T \mid c_T \text{ is sent})\Big] \\ &= \frac{1}{|\mathcal{C}|}\sum_{c_1\in\mathcal{C}} P(\hat{c}_1 \ne c_1 \mid c_1 \text{ is sent}) \\ &\quad + \frac{1}{|\mathcal{C}|}\sum_{c_2\in\mathcal{C}} P(\hat{c}_2 \ne c_2 \mid c_2 \text{ is sent}) \\ &\quad + \cdots + \frac{1}{|\mathcal{C}|}\sum_{c_T\in\mathcal{C}} P(\hat{c}_T \ne c_T \mid c_T \text{ is sent}) \\ &= T\,\bar{p}_e^{\text{iid}}\end{aligned}$$

where $\hat{c}_i$ denotes the decoded codeword. Since $\mathcal{C}$ achieves rate R on the i.i.d. fading channel, the average error probability $\bar{p}_e^{\text{iid}}$ goes to zero as length n goes to infinity. Therefore, $\bar{p}_e^{\text{cor}}$ goes to zero as the length nT of $\mathcal{C}'$ goes to infinity. Therefore, code $\mathcal{C}'$ achieves rate R on the correlated fading channel.

APPENDIX IV
PROOF OF THE RIGHT INEQUALITY OF LEMMA 6

In the following proof, we use the letters without tilde to denote the objects for the subblock-correlated model, and use the corresponding letters with tilde to denote the objects for the subblock-independent model.

Define the maximum mutual information between the inputs and outputs for the two channel models as

$$I_{\max} := \max_{p(\underline{\boldsymbol{x}}_1,\underline{\boldsymbol{x}}_2,\ldots,\underline{\boldsymbol{x}}_T)} I(\underline{\boldsymbol{x}}_1,\underline{\boldsymbol{x}}_2,\ldots,\underline{\boldsymbol{x}}_T;\underline{\boldsymbol{y}}_1,\underline{\boldsymbol{y}}_2,\ldots,\underline{\boldsymbol{y}}_T)$$
$$\tilde{I}_{\max} := \max_{p(\tilde{\underline{\boldsymbol{x}}}_1,\tilde{\underline{\boldsymbol{x}}}_2,\ldots,\tilde{\underline{\boldsymbol{x}}}_T)} I(\tilde{\underline{\boldsymbol{x}}}_1,\tilde{\underline{\boldsymbol{x}}}_2,\ldots,\tilde{\underline{\boldsymbol{x}}}_T;\tilde{\underline{\boldsymbol{y}}}_1,\tilde{\underline{\boldsymbol{y}}}_2,\ldots,\tilde{\underline{\boldsymbol{y}}}_T).$$

Then, the right inequality in (35) that we want to prove becomes

$$I_{\max} \le \tilde{I}_{\max} - n_r T \log \lambda_{\min}\{\Sigma_b\}. \tag{64}$$

The proof follows by bounding the difference between the mutual informations of the two channels, using appropriate matrix inequalities.

We first define some index sets. For $1 \le k \le Q$

$$\mathcal{A}_k := \{i : (k-1)L + 1 \le i \le kL\}. \tag{65}$$

Each $\mathcal{A}_k$ contains the time indexes corresponding to the kth subblock in the subblock-correlated channel.

For any given probability distribution $p(\underline{\boldsymbol{x}}_1,\underline{\boldsymbol{x}}_2,\ldots,\underline{\boldsymbol{x}}_T)$ satisfying the power constraint (10), there is a corresponding probability distribution

$$\tilde{p}_{\tilde{\underline{\boldsymbol{x}}}_1,\tilde{\underline{\boldsymbol{x}}}_2,\ldots,\tilde{\underline{\boldsymbol{x}}}_T}(\underline{\boldsymbol{x}}_1,\underline{\boldsymbol{x}}_2,\ldots,\underline{\boldsymbol{x}}_T) = \prod_{k=1}^{Q} p(\underline{\boldsymbol{x}}_i, i \in \mathcal{A}_k) \tag{66}$$

where

$$p(\underline{\boldsymbol{x}}_i, i \in \mathcal{A}_k) = \int p(\underline{\boldsymbol{x}}_1,\underline{\boldsymbol{x}}_2,\ldots,\underline{\boldsymbol{x}}_T) \prod_{j \notin \mathcal{A}_k} d\underline{\boldsymbol{x}}_j$$

is the marginal distribution of $\{\underline{\boldsymbol{x}}_i, i \in \mathcal{A}_k\}$. Clearly, $\tilde{p}_{\tilde{\underline{\boldsymbol{x}}}_1,\tilde{\underline{\boldsymbol{x}}}_2,\ldots,\tilde{\underline{\boldsymbol{x}}}_T}(\underline{\boldsymbol{x}}_1,\underline{\boldsymbol{x}}_2,\ldots,\underline{\boldsymbol{x}}_T)$ also satisfies the power constraint (10).

We use $p(\underline{\boldsymbol{x}}_1,\underline{\boldsymbol{x}}_2,\ldots,\underline{\boldsymbol{x}}_T)$ as the input distribution for the subblock-correlated channel and $\tilde{p}_{\tilde{\underline{\boldsymbol{x}}}_1,\tilde{\underline{\boldsymbol{x}}}_2,\ldots,\tilde{\underline{\boldsymbol{x}}}_T}(\underline{\boldsymbol{x}}_1,\underline{\boldsymbol{x}}_2,\ldots,\underline{\boldsymbol{x}}_T)$ for the subblock-independent channel. By the definition of the two channel models, we have the following relationship for the output distributions of the two channels:

$$\tilde{p}_{\tilde{\underline{\boldsymbol{y}}}_1,\tilde{\underline{\boldsymbol{y}}}_2,\ldots,\tilde{\underline{\boldsymbol{y}}}_T}(\underline{\boldsymbol{y}}_i, i \in \mathcal{A}_k) = p(\underline{\boldsymbol{y}}_i, i \in \mathcal{A}_k) \tag{67}$$

where

$$p(\underline{\boldsymbol{y}}_i, i \in \mathcal{A}_k) = \int p(\underline{\boldsymbol{y}}_1,\underline{\boldsymbol{y}}_2,\ldots,\underline{\boldsymbol{y}}_T) \prod_{j \notin \mathcal{A}_k} d\underline{\boldsymbol{y}}_j$$

is the joint distribution of $\{\underline{\boldsymbol{y}}_i, i \in \mathcal{A}_k\}$.

Since $\{\tilde{\underline{\boldsymbol{y}}}_i, i \in \mathcal{A}_1\}, \{\tilde{\underline{\boldsymbol{y}}}_i, i \in \mathcal{A}_2\}, \ldots, \{\tilde{\underline{\boldsymbol{y}}}_i, i \in \mathcal{A}_Q\}$ are independent, we have

$$\begin{aligned} h(\underline{\boldsymbol{y}}_1,\underline{\boldsymbol{y}}_2,\ldots,\underline{\boldsymbol{y}}_T) &\le \sum_{k=1}^{Q} h(\underline{\boldsymbol{y}}_i, i \in \mathcal{A}_k) = \sum_{k=1}^{Q} h(\tilde{\underline{\boldsymbol{y}}}_i, i \in \mathcal{A}_k) \\ &= h(\tilde{\underline{\boldsymbol{y}}}_1,\tilde{\underline{\boldsymbol{y}}}_2,\ldots,\tilde{\underline{\boldsymbol{y}}}_T). \end{aligned} \tag{68}$$

To bound the difference of the mutual informations of the two channels, we need to bound the following entropy difference. For notational simplicity, we use $\underline{\boldsymbol{x}}_{[m,n]}$ to denote the sequence of vectors $\{\underline{\boldsymbol{x}}_m, \underline{\boldsymbol{x}}_{m+1}, \ldots, \underline{\boldsymbol{x}}_n\}$. Then

$$\begin{aligned} -h&\big(\underline{\boldsymbol{y}}_{[1,T]}\big|\underline{\boldsymbol{x}}_{[1,T]}\big) + h\big(\tilde{\underline{\boldsymbol{y}}}_{[1,T]}\big|\tilde{\underline{\boldsymbol{x}}}_{[1,T]}\big) \\ &= \int d\underline{\boldsymbol{x}}_{[1,T]}\, p\big(\underline{\boldsymbol{x}}_{[1,T]}\big) \\ &\quad \int d\underline{\boldsymbol{y}}_{[1,T]}\, p\big(\underline{\boldsymbol{y}}_{[1,T]}\big|\underline{\boldsymbol{x}}_{[1,T]}\big) \log p\big(\underline{\boldsymbol{y}}_{[1,T]}\big|\underline{\boldsymbol{x}}_{[1,T]}\big) \\ &\quad - \int d\underline{\boldsymbol{x}}_{[1,T]}\, \tilde{p}\big(\underline{\boldsymbol{x}}_{[1,T]}\big) \\ &\quad \int d\underline{\boldsymbol{y}}_{[1,T]}\, \tilde{p}\big(\underline{\boldsymbol{y}}_{[1,T]}\big|\underline{\boldsymbol{x}}_{[1,T]}\big) \log \tilde{p}\big(\underline{\boldsymbol{y}}_{[1,T]}\big|\underline{\boldsymbol{x}}_{[1,T]}\big) \\ &= -n_r \int d\underline{\boldsymbol{x}}_{[1,T]} p\big(\underline{\boldsymbol{x}}_{[1,T]}\big) \\ &\quad \log\left[(\pi e)^T \det\left(I_T + \frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t} \boldsymbol{P}_j \Sigma_{\text{cor}} \boldsymbol{P}_j^{\dagger}\right)\right] \\ &\quad + n_r \int d\underline{\boldsymbol{x}}_{[1,T]} \tilde{p}\big(\underline{\boldsymbol{x}}_{[1,T]}\big) \\ &\quad \log\left[(\pi e)^T \det\left(I_T + \frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t} \boldsymbol{P}_j \Sigma_{\text{ind}} \boldsymbol{P}_j^{\dagger}\right)\right] \\ &= n_r \int d\underline{\boldsymbol{x}}_{[1,T]} p\big(\underline{\boldsymbol{x}}_{[1,T]}\big) \\ &\quad \log \frac{\det\left(I_T + \frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t} \boldsymbol{P}_j \left(I_Q \otimes \mathbb{I}_L\right) \boldsymbol{P}_j^{\dagger}\right)}{\det\left(I_T + \frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t} \boldsymbol{P}_j \left(\Sigma_b \otimes \mathbb{I}_L\right) \boldsymbol{P}_j^{\dagger}\right)} \end{aligned} \tag{69}$$

where $\boldsymbol{P}_j$ is as defined in Appendix I. The matrices $\Sigma_{\text{cor}} = \Sigma_b \otimes \mathbb{I}_L$ and $\Sigma_{\text{ind}} = I_Q \otimes \mathbb{I}_L$ are covariance matrices of fading gain vectors within one block for the subblock-correlated channel and subblock-independent channel, respectively.

We note the following property of matrices [17].

Let A, B be n-by-n Hermitian matrices, and assume that A is positive definite and that B is positive semidefinite. Then

$$\det(A+B) \ge \det(A). \tag{70}$$

Let $\lambda_{\min} \le 1$ denote the smallest eigenvalue of Σ_b. Then

$$\begin{aligned} \det&\left(I_T + \frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t} \boldsymbol{P}_j \left(\Sigma_b \otimes \mathbb{I}_L\right) \boldsymbol{P}_j^{\dagger}\right) \\ &= \det\Bigg(I_T + \frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t} \boldsymbol{P}_j \left(\Sigma_b \otimes \mathbb{I}_L\right) \boldsymbol{P}_j^{\dagger} \\ &\quad - \frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t} \boldsymbol{P}_j \left(\lambda_{\min} I_Q \otimes \mathbb{I}_L\right) \boldsymbol{P}_j^{\dagger} \\ &\quad + \frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t} \boldsymbol{P}_j \left(\lambda_{\min} I_Q \otimes \mathbb{I}_L\right) \boldsymbol{P}_j^{\dagger}\Bigg) \end{aligned}$$

$$= \det\left(I_T + \frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t}\lambda_{\min}\boldsymbol{P}_j\left(I_Q\otimes \mathbb{I}_L\right)\boldsymbol{P}_j^\dagger + \frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t}\boldsymbol{P}_j\left[\Sigma_b\otimes\mathbb{I}_L-\lambda_{\min}I_Q\otimes\mathbb{I}_L\right]\boldsymbol{P}_j^\dagger\right)$$
$$\geq \det\left(I_T+\frac{\mathsf{SNR}}{n_t}\lambda_{\min}\sum_{j=1}^{n_t}\boldsymbol{P}_j\left(I_Q\otimes\mathbb{I}_L\right)\boldsymbol{P}_j^\dagger\right)$$

where the last step used (70) and the fact that the matrix

$$\sum_{j=1}^{n_t}\boldsymbol{P}_j\left[\Sigma_b\otimes\mathbb{I}_L-\lambda_{\min}I_Q\otimes\mathbb{I}_L\right]\boldsymbol{P}_j^\dagger$$

is positive semidefinite.

Applying the above result to (69), we get

$$\begin{aligned}
-h&\left(\underline{\boldsymbol{y}}_{[1,T]}\middle|\underline{\boldsymbol{x}}_{[1,T]}\right)+h\left(\underline{\tilde{\boldsymbol{y}}}_{[1,T]}\middle|\underline{\tilde{\boldsymbol{x}}}_{[1,T]}\right)\\
&\leq n_r\int d\underline{\boldsymbol{x}}_{[1,T]}p\left(\underline{\boldsymbol{x}}_{[1,T]}\right)\log\frac{\det\left(I_T+\frac{\mathsf{SNR}}{n_t}\sum_{j=1}^{n_t}\boldsymbol{P}_j\left(I_Q\otimes\mathbb{I}_L\right)\boldsymbol{P}_j^\dagger\right)}{\det\left(I_T+\frac{\mathsf{SNR}}{n_t}\lambda_{\min}\sum_{j=1}^{n_t}\boldsymbol{P}_j\left(I_Q\otimes\mathbb{I}_L\right)\boldsymbol{P}_j^\dagger\right)}\\
&=n_r\int d\underline{\boldsymbol{x}}_{[1,T]}p\left(\underline{\boldsymbol{x}}_{[1,T]}\right)\log\frac{\det\left(\lambda_{\min}I_T+\frac{\mathsf{SNR}}{n_t}\lambda_{\min}\sum_{j=1}^{n_t}\boldsymbol{P}_j\left(I_Q\otimes\mathbb{I}_L\right)\boldsymbol{P}_j^\dagger\right)}{\lambda_{\min}^T\det\left(I_T+\frac{\mathsf{SNR}}{n_t}\lambda_{\min}\sum_{j=1}^{n_t}\boldsymbol{P}_j\left(I_Q\otimes\mathbb{I}_L\right)\boldsymbol{P}_j^\dagger\right)}\\
&\leq n_r\int d\underline{\boldsymbol{x}}_{[1,T]}p\left(\underline{\boldsymbol{x}}_{[1,T]}\right)\log\frac{\det\left(I_T+\frac{\mathsf{SNR}}{n_t}\lambda_{\min}\sum_{j=1}^{n_t}\boldsymbol{P}_j\left(I_Q\otimes\mathbb{I}_L\right)\boldsymbol{P}_j^\dagger\right)}{\lambda_{\min}^T\det\left(I_T+\frac{\mathsf{SNR}}{n_t}\lambda_{\min}\sum_{j=1}^{n_t}\boldsymbol{P}_j\left(I_Q\otimes\mathbb{I}_L\right)\boldsymbol{P}_j^\dagger\right)}\\
&=-n_rT\log\lambda_{\min}
\end{aligned}\tag{71}$$

where the second inequality follows from (70) and $0<\lambda_{\min}\leq 1$.

Combining (68) and (71), we can bound the difference of the mutual informations by

$$\begin{aligned}
I\left(\underline{\boldsymbol{y}}_{[1,T]};\underline{\boldsymbol{x}}_{[1,T]}\right)&-I\left(\underline{\tilde{\boldsymbol{y}}}_{[1,T]};\underline{\tilde{\boldsymbol{x}}}_{[1,T]}\right)\\
&=h\left(\underline{\boldsymbol{y}}_{[1,T]}\right)-h\left(\underline{\boldsymbol{y}}_{[1,T]}\middle|\underline{\boldsymbol{x}}_{[1,T]}\right)\\
&\quad-h\left(\underline{\tilde{\boldsymbol{y}}}_{[1,T]}\right)+h\left(\underline{\tilde{\boldsymbol{y}}}_{[1,T]}\middle|\underline{\tilde{\boldsymbol{x}}}_{[1,T]}\right)\\
&\leq -n_rT\log\lambda_{\min}.
\end{aligned}$$

Now, let $\underline{\boldsymbol{x}}_{[1,T]}^*$ denote the input that achieves maximum $I(\underline{\boldsymbol{y}}_{[1,T]};\underline{\boldsymbol{x}}_{[1,T]})$ of the subblock-correlated channel, and let $\underline{\tilde{\boldsymbol{x}}}_{[1,T]}^*$ denote the corresponding input for the subblock-independent channel with distribution defined by (66). Then

$$\begin{aligned}
I_{\max}&=I(\underline{\boldsymbol{y}}_{[1,T]};\underline{\boldsymbol{x}}_{[1,T]}^*)\\
&\leq I(\underline{\tilde{\boldsymbol{y}}}_{[1,T]};\underline{\tilde{\boldsymbol{x}}}_{[1,T]}^*)-n_rT\log\lambda_{\min}\\
&\leq \tilde{I}_{\max}-n_rT\log\lambda_{\min}.
\end{aligned}\tag{72}$$

APPENDIX V
PROOF OF PROPOSITION 4

The covariance matrix Σ_h is positive semidefinite with rank $Q\leq T$, and so it has a $Q\times Q$ positive-definite principal submatrix denoted by $(\Sigma_h)_Q$ [17]. Without loss of generality, suppose $(\Sigma_h)_Q$ is the covariance matrix of the first Q components of the vector $\underline{\boldsymbol{h}}$.

In the following, we use $\boldsymbol{x}_{[m,n]}$ to denote the sequence of components $\{\boldsymbol{x}_m,\boldsymbol{x}_{m+1},\ldots,\boldsymbol{x}_n\}$. Then

$$\begin{aligned}
I\left(\boldsymbol{y}_{[1,T]};\boldsymbol{x}_{[1,T]}\right)&=I\left(\boldsymbol{y}_{[1,Q]};\boldsymbol{x}_{[1,T]}\right)+I\left(\boldsymbol{y}_{[Q+1,T]};\boldsymbol{x}_{[1,T]}\middle|\boldsymbol{y}_{[1,Q]}\right)\\
&=I\left(\boldsymbol{y}_{[1,Q]};\boldsymbol{x}_{[1,Q]}\right)+I\left(\boldsymbol{y}_{[1,Q]};\boldsymbol{x}_{[Q+1,T]}\middle|\boldsymbol{x}_{[1,Q]}\right)\\
&\quad+I\left(\boldsymbol{y}_{[Q+1,T]};\boldsymbol{x}_{[1,T]}\middle|\boldsymbol{y}_{[1,Q]}\right).
\end{aligned}\tag{73}$$

By Lemma 5, the first term can be bounded as

$$I\left(\boldsymbol{y}_{[1,Q]};\boldsymbol{x}_{[1,Q]}\right)\leq Q\Big(\log\log\mathsf{SNR}-\gamma-1-\log\lambda_{\min}\{(\Sigma_h)_Q\}\Big)+o(1).\tag{74}$$

The second term in (73) is zero because conditioned on the first Q inputs, the first Q outputs only depend on the fading coefficients and noise, which are independent of the remaining $T-Q$ inputs.

The third term in (73) can be bounded by

$$\begin{aligned}
I&\left(\boldsymbol{y}_{[Q+1,T]};\boldsymbol{x}_{[1,T]}\middle|\boldsymbol{y}_{[1,Q]}\right)\\
&=h\left(\boldsymbol{y}_{[Q+1,T]}\middle|\boldsymbol{y}_{[1,Q]}\right)-h\left(\boldsymbol{y}_{[Q+1,T]}\middle|\boldsymbol{x}_{[1,T]},\boldsymbol{y}_{[1,Q]}\right)\\
&\leq h\left(\boldsymbol{y}_{[Q+1,T]}\right)-h\left(\boldsymbol{y}_{[Q+1,T]}\middle|\boldsymbol{h}_{[Q+1,T]},\boldsymbol{x}_{[1,T]},\boldsymbol{y}_{[1,Q]}\right)\\
&=h\left(\boldsymbol{y}_{[Q+1,T]}\right)-h\left(\boldsymbol{w}_{[Q+1,T]}\right)\\
&\leq (T-Q)\log[\pi e(1+\mathsf{SNR})]-(T-Q)\log(\pi e)\\
&=(T-Q)\log\mathsf{SNR}+o(1)
\end{aligned}\tag{75}$$

where we used the fact that $\mathrm{E}\left[|\boldsymbol{y}_k|^2\right]\leq 1+\mathsf{SNR}$.

Combining (74) and (75), and exploiting the fact that the upper bound does not depend on the input distribution of $\boldsymbol{x}_{[1,T]}$, we have

$$\begin{aligned}
C(\mathsf{SNR})&=\frac{1}{T}\max_{p(\boldsymbol{x})}I\left(\boldsymbol{y}_{[1,T]};\boldsymbol{x}_{[1,T]}\right)\\
&\leq\frac{T-Q}{T}\log\mathsf{SNR}+\frac{Q}{T}\Big(\log\log\mathsf{SNR}-\gamma-1-\log\lambda_{\min}\{(\Sigma_h)_Q\}\Big)+o(1).
\end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank Dr. Amos Lapidoth of ETH, Dr. Ashok Mantravadi of Qualcomm, and Dr. Lizhong Zheng of

MIT for the helpful discussions. They also would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

[1] I. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1290–1301, May 2001.

[2] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2426–2467, Oct. 2003.

[3] R. Chen, B. Hajek, R. Koetter, and U. Madhow, "On fixed input distributions for noncoherent communication over high SNR Rayleigh fading channels," *IEEE Trans. Inform. Theory*, to be published.

[4] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.

[5] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[6] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

[7] V. V. Veeravalli and A. Sayeed, *Wideband Wireless Channels: Statistical Modeling, Analysis and Simulation*, to be published.

[8] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA: Kluwer Academic, 2001.

[9] G. J. Foschini and M. Gans, "On limits of wireless communications in fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

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

[11] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.

[12] H. Bölcskei, R. Koetter, and S. Mallik, "Coding and modulation for underspread fading channels," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, June/July 2002, p. 358.

[13] K. Liu, T. Kadous, and A. M. Sayeed, "Performance of orthogonal short-time Fourier signaling over doubly dispersive channels," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, June/July 2002, p. 357.

[14] R. M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.

[15] R. L. Dobrushin, "General formulation of Shannon's main theorem in information theory," *Amer. Math. Soc. Translations*, vol. 33, pp. 323–438, 1963.

[16] A. Lapidoth, "On the asymptotic capacity of fading channels," *IEEE Trans. Inform. Theory*, submitted for publication.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.