Capacity of Underspread Noncoherent WSSUS Fading Channels under Peak Signal Constraints

Giuseppe Durisi, Helmut Bölcskei
Communication Technology Laboratory
ETH Zurich, 8092 Zurich, Switzerland
E-mail: {gdurisi,boelcskei}@nari.ee.ethz.ch

Shlomo Shamai (Shitz)
Technion, Israel Institute of Technology
32000 Haifa, Israel
E-mail: sshlomo@ee.technion.ac.il

***Abstract*— We characterize the capacity of the general class of noncoherent underspread wide-sense stationary uncorrelated scattering (WSSUS) time-frequency-selective Rayleigh fading channels, under peak constraints in time and frequency and in time only. Capacity upper and lower bounds are found which are explicit in the channel's scattering function and allow to identify the capacity-maximizing bandwidth for a given scattering function and a given peak-to-average power ratio.**

I. Introduction

A well-known result in information theory states that, in the infinite-bandwidth limit, the capacity of a time-frequency (TF) selective, wide-sense stationary uncorrelated scattering (WSSUS) [1] fading channel[1] equals the capacity of an additive white Gaussian noise (AWGN) channel with the same receive signal-to-noise ratio (SNR); capacity is achieved by codebooks that are "peaky" in both time and frequency [2], [3]. It is also well-known that the AWGN channel capacity cannot be achieved, in the infinite-bandwidth limit, if a peak constraint is imposed [4], [5]. In particular, as detailed below, different forms of peak constraints lead to different infinite-bandwidth capacity behavior.

Peak constraint in time: A closed-form expression for the capacity of a WSSUS TF-selective fading channel under a peak constraint in time is not available in the literature. The rate achievable by frequency-shift-keying (FSK), in the infinite-bandwidth limit, has been shown by Viterbi [6] to be given by

$$R_{\text{FSK},\infty} = \underbrace{\int_{-\infty}^{\infty} \frac{P}{N_0} S(\nu)\mathrm{d}\nu}_{C_{\text{AWGN},\infty}} - \int_{-\infty}^{\infty} \log\left(1+\frac{P}{N_0}S(\nu)\right)\mathrm{d}\nu \tag{1}$$

where P is the received power, N_0 stands for the one-sided noise spectral density and $S(\nu)$ is the power-Doppler profile [1] of the TF-selective fading channel process. The infinite-bandwidth rate achieved by FSK constitutes a lower bound on the infinite-bandwidth capacity of the channel and equals the infinite-bandwidth capacity $C_{\text{AWGN},\infty}$ of an AWGN channel (with the same receive SNR) minus a penalty term resulting from the peak constraint. In the absence of a peak constraint, the second term on the right-hand-side (RHS) of (1) can be made as small as desired by transmitting the FSK codewords with arbitrarily low duty cycle, and AWGN channel capacity can be achieved in the infinite-bandwidth limit [2, Sec. 8.6].

Expressions similar to (1) have been found in [7] for the capacity per unit energy and in [8], [9] for the infinite-bandwidth capacity of time-selective, frequency-flat Rayleigh fading channels under a peak constraint in time.

Peak constraint in time and frequency: From the results in [4], [5] it follows that, under a peak constraint in time and frequency, the capacity of a TF-selective WSSUS fading channel goes to zero as the bandwidth approaches infinity. An upper bound on the rate achievable by transmitting constant modulus symbols over a TF-selective WSSUS *underspread* [10] fading channel has been obtained in [11]. This upper bound is explicit in the channel's scattering function and hints at the existence of a scattering-function-dependent, capacity-maximizing bandwidth.

Contributions: We consider the general class of underspread [10] continuous-time WSSUS TF-selective Rayleigh fading channels under peak constraints (i) in time and (ii) in time and frequency.[2] Our specific contributions are summarized as follows:

- *Peak constraint in time:* We recover Viterbi's infinite-bandwidth lower bound (1) using an alternative proof.
- *Peak constraint in time and frequency:* We generalize the results in [11], which dealt exclusively with constant modulus inputs, by providing upper and lower bounds on capacity. Both bounds are explicit in the channel's scattering function and reveal the existence of an optimum, capacity-maximizing bandwidth, which depends on the scattering function and the peak-to-average power ratio.

Proof techniques: Our entire analysis is built on the fact that underspread channels have a well-structured set of TF-localized approximate eigenfunctions [10]. The main proof techniques used in the paper are based on the relation between mutual information and minimum mean-square error (MMSE) discovered in [12], on a generalization of Szegö's theorem on the asymptotic eigenvalue distribution of Toeplitz matrices [13] to the case of

This work was supported in part by the Swiss *Kommission für Technologie und Innovation (KTI)* under grant 6715.2 ENS-ES, by the European Commission as part of the Integrated Project PULSERS Phase II under contract FP6-027142, and by the EC FP6 Network of Excellence NEWCOM.

[1]Throughout the paper, whenever we speak of fading channels, we shall refer to (without explicitly stating it) the "noncoherent" case where neither transmitter nor receiver have access to channel state information (CSI) but both are aware of the channel law.

[2]The peak constraints, made precise in Section II, will be imposed in the channel's eigenspace rather than on the continuous-time transmit signal.

block-Toeplitz matrices with Toeplitz blocks[3] [14], and on a property of the information divergence of orthogonal signaling schemes first presented in a different context in [15].

Notation: Uppercase boldface letters denote matrices and lowercase boldface letters designate vectors. The superscripts ^{T}, ^{H} and * stand for transpose, conjugate transpose and element-wise conjugation, respectively. $\det(\mathbf{X})$ denotes the determinant of the matrix $\mathbf{X}$, $\mathbf{I}$ stands for the identity matrix of appropriate size, and $\operatorname{diag}\{\mathbf{x}\}$ denotes a diagonal square matrix having the elements of the vector $\mathbf{x}$ on its main diagonal. $\mathbb{E}\{\cdot\}$ is the expectation operator. $\delta(t)$ stands for the Dirac delta function, and $\delta_{i,j} = 1$ if $i = j$ and 0 else. All logarithms are to the base e. If a random variable (RV) x has distribution $\mathcal{P}_x$, we write $x \sim \mathcal{P}_x$. Finally, $\mathcal{CN}(0, \sigma^2)$ denotes a circularly symmetric complex Gaussian RV with variance σ^2.

II. Channel and System Model

We consider an ergodic WSSUS TF-selective Rayleigh fading channel $\mathbb{H}$ with input-output relation

$$r(t) = (\mathbb{H}x)(t) = \int_\tau h(t, t-\tau)\, x(t-\tau) \mathrm{d}\tau \qquad (2)$$

where the impulse response $h(t, t')$ is a two-dimensional zero-mean complex Gaussian random process. The time-varying transfer function of the channel is defined as [1]

$$L_\mathbb{H}(t, f) = \int_\tau h(t, t-\tau)\, e^{-j2\pi f\tau} \mathrm{d}\tau.$$

The channel's scattering function $C_\mathbb{H}(\tau, \nu)$ [3] is given by

$$C_\mathbb{H}(\tau, \nu) = \int_t \int_f R_\mathbb{H}(t, f)\, e^{-j2\pi(\nu t - \tau f)} \mathrm{d}t\mathrm{d}f$$

where $R_\mathbb{H}(t, f) = \mathbb{E}\{L_\mathbb{H}(t + t', f + f')\, L^*_\mathbb{H}(t', f')\}$ denotes the channel's TF-correlation function. We invoke the common assumption [10] of a scattering function that is compactly supported within the rectangle $[-\tau_0, \tau_0] \times [-\nu_0, \nu_0]$, i.e., $C_\mathbb{H}(\tau, \nu) = 0$ for $(\tau, \nu) \notin [-\tau_0, \tau_0] \times [-\nu_0, \nu_0]$. Defining the spread of the channel as the area of this rectangle, $\Delta_\mathbb{H} = 4\tau_0\nu_0$, the channel is said to be underspread if $\Delta_\mathbb{H} \le 1$ and overspread if $\Delta_\mathbb{H} > 1$ [10]. The underspread assumption is relevant as most mobile radio channels are highly underspread.

The key idea turning (2) into a discrete problem is to recognize that underspread channels are approximately diagonalized by orthonormal Weyl-Heisenberg bases [10], which are obtained by TF-shifting of a normalized function $g(t)$ according to $g_{k,n}(t) = g(t - kT)e^{j2\pi nFt}$, where the grid parameters T and F have to satisfy $TF \ge 1$. For $T \le 1/(2\nu_0)$ and $F \le 1/(2\tau_0)$, and hence $TF \le 1/\Delta_\mathbb{H}$, the impulse response $h(t, t')$ of an underspread fading channel satisfies [16]

$$h(t, t') \approx \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} L_\mathbb{H}(kT, nF)\, g_{k,n}(t) g^*_{k,n}(t'). \qquad (3)$$

The scalar channel coefficients $L_\mathbb{H}(kT, nF)$ are circularly symmetric complex Gaussian RVs with zero mean, variance $\sigma^2_\mathbb{H} = \int_\tau \int_\nu C_\mathbb{H}(\tau, \nu)\, \mathrm{d}\tau\mathrm{d}\nu$, and correlation function

$$R_\mathbb{H}(kT, nF) = \mathbb{E}\{L_\mathbb{H}((k + k')T, (n + n')F)\, L^*_\mathbb{H}(k'T, n'F)\}.$$

Motivated by the (approximate) diagonalization (3), we write the transmit signal $x(t)$ as

$$x(t) = \sum_{k=-\infty}^{\infty} \sum_{n=0}^{N-1} x_{k,n} g_{k,n}(t) \qquad (4)$$

where the $x_{k,n}$ are the information-bearing data symbols. This modulation scheme corresponds to pulse-shaped orthogonal frequency-division multiplexing (OFDM) with OFDM symbol duration T and subcarrier spacing F. The transmit-signal bandwidth is given by $W = NF$. With the received signal $y(t) = r(t) + z(t)$ and $z(t)$ circularly symmetric additive white Gaussian noise so that[4] $\mathbb{E}\{z(t)z^*(t')\} = \delta(t - t')$, the receiver computes the inner products $y_{k,n} = \langle y, g_{k,n}\rangle$. Exploiting the orthonormality of the basis functions $g_{k,n}(t)$, we obtain the overall input-output relation

$$y_{k,n} = h_{k,n} x_{k,n} + z_{k,n} \qquad (5)$$

where $h_{k,n} = L_\mathbb{H}(kT, nF)$ and $\mathbb{E}\{z_{k,n} z^*_{k',n'}\} = \delta_{k,k'}\delta_{n,n'}$. In summary, we transmit and receive on the channel's (approximate) eigenfunctions (which are TF-translates of the prototype function $g(t)$), thereby realizing a tiling of the TF-plane. For a detailed discussion of the approximation (3) and the consequences of restricting the class of input signals to (4), the interested reader is referred to [16]. In the following, we refer to the index k as representing the "time-domain" and the index n as representing the "frequency-domain".

We assume that each channel use takes place over N subcarriers and K OFDM symbols. The N-dimensional vector containing the data symbols transmitted in the kth OFDM symbol ($k = 0, 1, \ldots, K-1$) is denoted as $\mathbf{x}_k = [x_{k,0}\ x_{k,1}\ \cdots\ x_{k,N-1}]^T$. The vectors $\mathbf{y}_k$, $\mathbf{h}_k$, and $\mathbf{z}_k$ are defined correspondingly. Furthermore, we define the KN-dimensional vector containing the data symbols transmitted in one channel use as $\mathbf{x} = [\mathbf{x}_0^T\ \mathbf{x}_1^T\ \cdots\ \mathbf{x}_{K-1}^T]^T$. Again, $\mathbf{y}$, $\mathbf{h}$ and $\mathbf{z}$ are defined correspondingly. Finally, we denote the covariance matrix of the channel vector $\mathbf{h}$ by $\mathbf{C_h} = \mathbb{E}\{\mathbf{h}\mathbf{h}^H\}$. Since $h_{k,n}$ is WSS in k and n, the matrix $\mathbf{C_h}$ is block-Toeplitz.

The input-output relation corresponding to one channel use can now be written as

$$\mathbf{y} = \operatorname{diag}\{\mathbf{h}\}\, \mathbf{x} + \mathbf{z}.$$

Note that the channel in (5) will in general not be memoryless because the channel gains $h_{k,n}$ are correlated across time index k and across frequency index n. Throughout the paper, we assume an average-power constraint according to $\mathbb{E}\{\|\mathbf{x}\|^2\} \le KPT$. In addition, setting $P_\text{peak} = \beta P$, with $\beta \ge 1$, we impose a peak constraint distinguishing the following two cases:

i) *Peak constraint in time and frequency:* The data symbols

[3] In the remainder of the paper, block-Toeplitz matrices with Toeplitz blocks will simply be referred to as block-Toeplitz matrices.

[4] In the remainder of the paper (apart from Section III-E), we normalize the one-sided noise spectral density according to $N_0 = 1$.

satisfy

$$|x_{k,n}|^2 \le P_{\text{peak}} T/N \qquad \text{a.s. } \forall k, n \tag{6}$$

where a.s. stands for *almost surely*.

ii) *Peak constraint in time:* The data symbols satisfy

$$\|\mathbf{x}_k\|^2 \le P_{\text{peak}} T \qquad \text{a.s. } \forall k. \tag{7}$$

While Condition i) prohibits peakiness in time *as well as* peakiness in frequency, Condition ii) prohibits peakiness in time only, still allowing for a signaling scheme that is peaky in frequency. Note that we impose peak constraints in the channel's eigenspace (i.e., on the data symbols $x_{k,n}$) rather than on the continuous-time transmit signal $x(t)$.

III. Capacity Bounds under Peak Constraints in Time and Frequency

For a given bandwidth $W = NF$, the capacity (in nat/s) of the channel (5) is defined as

$$C(W) = \lim_{K\to\infty} \frac{1}{KT} \sup_{\mathcal{P}_{\mathbf{x}}} I(\mathbf{y};\mathbf{x}) \tag{8}$$

where the supremum is taken over the set of input distributions $\mathcal{P}_{\mathbf{x}}$ satisfying the average-power constraint $\mathbb{E}\{\|\mathbf{x}\|^2\} \le KPT$ and the peak constraint (6). We shall next derive two upper bounds and a lower bound on (8). The first upper bound and the lower bound generalize the bounds for the frequency-flat, time-selective fading case reported in [8, Prop. 3.1] and [17, Prop. 2.2], respectively, to the TF-selective underspread case. In contrast to [8] and [17], our proof techniques do not explicitly rely on the relationship between the lag-one mean-square prediction error of a WSS process and its spectral measure [18, Th. 4.3]. Instead, we use the relation between mutual information and MMSE discovered in [12], and a generalization of Szegö's theorem on the asymptotic eigenvalue distribution of Toeplitz matrices to the case of block-Toeplitz matrices [14]. The second upper bound is standard and is obtained by assuming perfect CSI at the receiver.

A. First Upper Bound

Theorem 1: The capacity of a TF-selective underspread Rayleigh fading channel, with scattering function $C_{\mathbb{H}}(\tau,\nu)$, under the average-power constraint $\mathbb{E}\{\|\mathbf{x}\|^2\} \le KPT$ and the peak constraint (6) with $P_{\text{peak}} = \beta P$, is upper-bounded as $C(W) \le \text{UB}_1(W)$, where

$$\text{UB}_1(W) = \frac{W}{TF} \log\left(1 + \alpha(W) P \frac{TF}{W} \sigma_{\mathbb{H}}^2\right) - \alpha(W) A(W,\beta) \tag{9}$$

with

$$\alpha(W) = \min\left\{1, \frac{W}{TF}\left(\frac{1}{A(W,\beta)} - \frac{1}{P\sigma_{\mathbb{H}}^2}\right)\right\} \tag{10}$$

and

$$A(W,\beta) = \frac{W}{\beta} \int_\tau \int_\nu \log\left(1 + \frac{\beta P}{W} C_{\mathbb{H}}(\tau,\nu)\right) \mathrm{d}\tau \mathrm{d}\nu.$$

Proof: A sketch of the proof can be found in Appendix I. ■

The upper bound (9) generalizes the upper bound [11, Eq. (2)], which holds only for constant modulus signals, i.e., for $|x_{k,n}| = \text{const.}, \forall k, n$. The bounds (9) and [11, Eq. (2)] are both explicit in the channel's scattering function, have similar structure and coincide for $\beta = 1$ when $\alpha(W) = 1$ in (10).

B. Perfect Receive CSI Upper Bound

A straightforward upper bound on (8) is obtained by assuming perfect CSI at the receiver and neglecting the peak constraint (6), but retaining the average-power constraint. The resulting bound is given by

$$\text{UB}_2(W) = \frac{W}{TF} \mathbb{E}_h\left\{\log\left(1 + \frac{PTF}{W}|h|^2\right)\right\} \tag{11}$$

with $h \sim \mathcal{CN}(0, \sigma_{\mathbb{H}}^2)$.

C. Lower Bound

Theorem 2: The capacity of a TF-selective underspread Rayleigh fading channel, with scattering function $C_{\mathbb{H}}(\tau,\nu)$, under the average-power constraint $\mathbb{E}\{\|\mathbf{x}\|^2\} \le KPT$ and the peak constraint (6) with $P_{\text{peak}} = \beta P$, is lower-bounded, for large enough W, as $C(W) \ge \text{LB}(W) = \max_{1\le\gamma\le\beta} \text{LB}(W,\gamma)$, where

$$\begin{aligned}\text{LB}(W,\gamma) = {} & \frac{W}{\gamma TF} I(y;x|h) \\ & - \frac{W}{\gamma} \int_\tau \int_\nu \log\left(1 + \frac{\gamma P}{W} C_{\mathbb{H}}(\tau,\nu)\right) \mathrm{d}\tau \mathrm{d}\nu. \end{aligned} \tag{12}$$

The first term on the RHS of (12) is the perfect-CSI mutual information of a scalar channel with input-output relation $y = hx + z$, where $h \sim \mathcal{CN}(0,\sigma_{\mathbb{H}}^2)$, $|x|^2 = \gamma PT/N$ a.s., and $z \sim \mathcal{CN}(0,1)$.

Proof: A sketch of the proof can be found in Appendix II. ■

Noting that constant modulus constellations are second-order optimal in the low-SNR regime (see [19, Th. 14]), we obtain the following explicit expression for (12):

$$\begin{aligned}\text{LB}(W,\gamma) \approx {} & P\sigma_{\mathbb{H}}^2 - \frac{\gamma (P\sigma_{\mathbb{H}}^2)^2 TF}{W} \\ & - \frac{W}{\gamma} \int_\tau \int_\nu \log\left(1 + \frac{\gamma P}{W} C_{\mathbb{H}}(\tau,\nu)\right) \mathrm{d}\tau \mathrm{d}\nu. \end{aligned} \tag{13}$$

D. Discussion of the Bounds

The upper bound (9) and the lower bound (12) have the same structure in the sense of being the difference of the mutual information of a memoryless channel and a penalty term that depends on the fading channel memory. Moreover, both bounds are explicit in the channel's scattering function. Unlike the upper bound (9), the lower bound (12) holds only for large bandwidths as the penalty term on the RHS of (12) admits a closed-form (integral) expression only for sufficiently large W.

E. Numerical Evaluation of the Bounds

In this section, we provide numerical results based on the upper bounds (9) and (11), and on the lower bound (12), where the first term in (12) is evaluated numerically for QPSK mod-

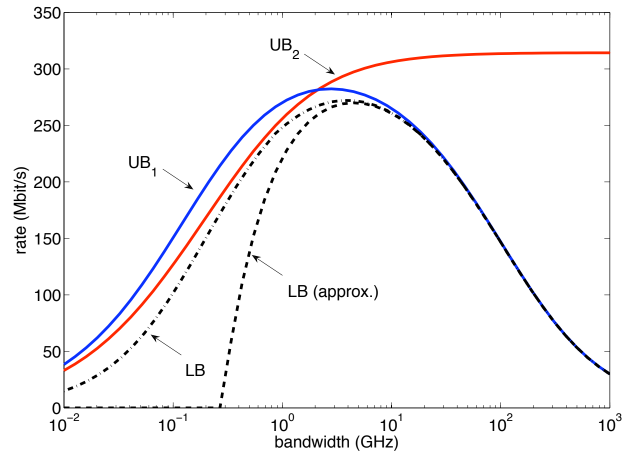

Fig. 1. UB_1, UB_2, LB, and the approximate lower bound (13) for $\beta = 1$.

ulation using the algorithm proposed in [20]. We use IEEE 802.11a related system parameters, i.e., $TF = 1.25$, transmit power equal to 1 mW and one-sided noise spectral density $N_0 = 4.14 \cdot 10^{-21}$ W/Hz. The channel's scattering function is assumed to be brick-shaped with $\Delta_{\mathbb{H}} = 10^{-3}$ (i.e., the channel is highly underspread) and path loss $\sigma_{\mathbb{H}}^2 = -90$ dB. Fig. 1 shows $\mathrm{UB}_1(W)$, $\mathrm{UB}_2(W)$, $\mathrm{LB}(W)$ and the approximation (13) for $\beta = 1$. We can see that UB_1 and LB take on a maximum at the *critical* bandwidth of approximately 1 GHz and then approach zero as the bandwidth increases. The approximation (13) is accurate for bandwidths above the critical bandwidth, and very loose otherwise. The effect of the capacity decreasing for bandwidths exceeding the critical bandwidth, and eventually going to zero, is known in the literature as *overspreading* [4], [5], [21]. Overspreading occurs since the peak constraint (6) prohibits peakiness of the signaling scheme in time as well as in frequency.

IV. Infinite-Bandwidth Capacity under Peak Constraint in Time

We shall next relax the peak constraint in time and frequency and impose a peak constraint in time only according to (7), while maintaining the average-power constraint $\mathbb{E}\{\|\mathbf{x}\|^2\} \leq KPT$. In addition, we focus on the infinite-bandwidth limit only. A capacity lower bound, which is explicit in the channel's scattering function, will allow us to conclude that the overspreading phenomenon discussed in the previous section can be eliminated by allowing the transmit signal to be peaky in frequency. We start by defining the infinite-bandwidth capacity of the channel (5) as

$$C_\infty = \lim_{K\to\infty} \lim_{N\to\infty} \sup_{\mathcal{P}_\mathbf{x}} \frac{1}{KT} I(\mathbf{y};\mathbf{x})$$

where the supremum is taken over the set of input distributions $\mathcal{P}_\mathbf{x}$ satisfying the average-power constraint $\mathbb{E}\{\|\mathbf{x}\|^2\} \leq KPT$ and the peak constraint (7). We derive a lower bound on C_∞ by evaluating $1/(KT) \cdot I(\mathbf{y};\mathbf{x})$ in the limit $K, N \to \infty$ for a specific signaling scheme. This signaling scheme, described in detail in [16] mimics the FSK scheme used in [6], and can be interpreted as a generalization (to channels with memory in time and frequency) of the on-off FSK scheme proposed in [22] for memoryless (Ricean) fading channels. The resulting lower bound is summarized as follows.

Theorem 3: The infinite-bandwidth capacity of a TF-selective underspread Rayleigh fading channel, with scattering function $C_{\mathbb{H}}(\tau,\nu)$, under the average-power constraint $\mathbb{E}\{\|\mathbf{x}\|^2\} \leq KPT$ and the peak constraint (7) with $P_{\text{peak}} = \beta P$, is lower-bounded as $C_\infty \geq \mathrm{LB}(\beta)$, where

$$\mathrm{LB}(\beta) = P\sigma_{\mathbb{H}}^2 - \frac{1}{\beta}\int_\nu \log\Big(1 + \beta P \widetilde{C}_{\mathbb{H}}(\nu)\Big)\,\mathrm{d}\nu \qquad (14)$$

with $\widetilde{C}_{\mathbb{H}}(\nu) = \int_\tau C_{\mathbb{H}}(\tau,\nu)\,\mathrm{d}\tau$ denoting the power-Doppler profile [1] of the channel.

Proof: The proof of this result can be found in [16]. ∎

The lower bound in (14) coincides with Viterbi's lower bound (1) when $\beta = 1$. The proof technique used to obtain Theorem 3 is, however, conceptually different from that in [6]. Specifically, in [6] an appropriate choice of the codebook, namely FSK, reduces a WSSUS TF-selective fading channel with scattering function $C_{\mathbb{H}}(\tau,\nu)$ to an effective frequency-flat, time-selective fading channel with power-Doppler profile $\widetilde{C}_{\mathbb{H}}(\nu)$. A Karhunen-Loève decomposition of the resulting effective time-selective fading channel then leads to a closed-form expression for the error exponent, which finally yields the capacity expression (1). The proof of Theorem 3, on the other hand, starts from an eigenfunction decomposition of the WSSUS TF-selective channel's impulse response $h(t,t')$, and establishes an infinite-bandwidth capacity lower bound by computing the rate achievable for a specific signaling scheme (which mimics FSK). The main tool used in the proof of Theorem 3 is a property of the information divergence of FSK constellations, first presented in a different context in [15].

Appendix I

Sketch of the proof of Theorem 1: Fix K and N. Following [8], we use the chain rule for mutual information to write the supremum in (8) as

$$\begin{aligned}\sup_{\mathcal{P}_\mathbf{x}} I(\mathbf{y};\mathbf{x}) &= \sup_{\mathcal{P}_\mathbf{x}}\{I(\mathbf{y};\mathbf{x},\mathbf{h}) - I(\mathbf{y};\mathbf{h}|\mathbf{x})\}\\ &\leq \sup_{0\leq\alpha\leq 1}\left\{\sup_{\widetilde{\mathcal{P}}_\mathbf{x}} I(\mathbf{y};\mathbf{x},\mathbf{h}) - \inf_{\widetilde{\mathcal{P}}_\mathbf{x}} I(\mathbf{y};\mathbf{h}|\mathbf{x})\right\} \qquad (15)\end{aligned}$$

where (15) follows by rewriting the supremum over $\mathcal{P}_\mathbf{x}$ as a double supremum over $\alpha \in [0,1]$ and over the set of input distributions $\widetilde{\mathcal{P}}_\mathbf{x}$ satisfying $\mathbb{E}\{\|\mathbf{x}\|^2\} = \alpha KPT$ and the peak constraint (6). Define $\mathbf{X} = \mathrm{diag}\{\mathbf{x}\}$. The first term on the RHS of (15) can be upper-bounded as follows

$$\begin{aligned}\sup_{\widetilde{\mathcal{P}}_\mathbf{x}} I(\mathbf{y};\mathbf{x},\mathbf{h}) &\overset{(a)}{\leq} \sup_{\mathbb{E}\{\|\mathbf{x}\|^2\}=\alpha KPT} \log\det\left(\mathbf{I} + \mathbb{E}\{\mathbf{X}\mathbf{C}_\mathbf{h}\mathbf{X}^H\}\right)\\ &\overset{(b)}{\leq} \sup_{\mathbb{E}\{\|\mathbf{x}\|^2\}=\alpha KPT} \sum_{k=0}^{K-1}\sum_{n=0}^{N-1} \log\left(1 + \mathbb{E}\{|x_{k,n}|^2\}\,\sigma_{\mathbb{H}}^2\right)\\ &\overset{(c)}{\leq} KN\log\left(1 + \frac{\alpha PT}{N}\sigma_{\mathbb{H}}^2\right) \qquad (16)\end{aligned}$$

where (a) follows by ignoring the peak constraint and upper-bounding $I(\mathbf{y};\mathbf{x},\mathbf{h})$ by the capacity of an AWGN channel with input $\mathbf{Xh}$, (b) follows from Hadamard's inequality, and (c) follows from the concavity of the log function. The second term on the RHS of (15) can be lower-bounded as

$$\inf_{\widetilde{\mathcal{P}}_{\mathbf{x}}} I(\mathbf{y};\mathbf{h}|\mathbf{x}) \geq c \int_{\tau}\int_{\nu} \log\left(1+\frac{\beta P}{NF}C_{\mathbb{H}}(\tau,\nu)\right)\mathrm{d}\tau\mathrm{d}\nu \quad (17)$$

with $c = \alpha KNTF/\beta$. The proof of (17), detailed in [16], is based on the relation between mutual information and MMSE discovered in [12] and on the closed-form expression for the noncausal MMSE of a two-dimensional stationary random process obtained in [23]. Inserting (16) and (17) in (15), dividing by KT, and using $W = NF$, we get

$$\mathrm{UB}_1(W) = \sup_{0\leq\alpha\leq 1}\left\{\frac{W}{TF}\log\left(1+\frac{\alpha PTF}{W}\sigma_{\mathbb{H}}^2\right)\right. \\ \left. -\frac{\alpha W}{\beta}\int_{\tau}\int_{\nu}\log\left(1+\frac{\beta P}{W}C_{\mathbb{H}}(\tau,\nu)\right)\mathrm{d}\tau\mathrm{d}\nu\right\}.$$

As the function the supremum is taken over is concave in α, the maximizing value given in (10) is unique and can be obtained by differentiating with respect to α.

Appendix II

Sketch of the proof of Theorem 2: For the sake of simplicity of exposition, we outline the proof for $\beta = 1$ only. The result for general β follows from a simple time-sharing argument [17, Corollary 2.1]. We first lower-bound $C(W)$ in (8) by assuming a specific input distribution, namely, by letting $\mathbf{x}$ have independent and identically distributed (i.i.d.), constant modulus entries, each of which satisfies $|x|^2 = PT/N$. In the following, vectors $\mathbf{x}$ with this property will be denoted as $\mathbf{x}_{\text{i.i.d.}}$. Next, we use the well-known inequality

$$I(\mathbf{y};\mathbf{x}_{\text{i.i.d.}}) = I(\mathbf{y};\mathbf{x}_{\text{i.i.d.}},\mathbf{h}) - I(\mathbf{y};\mathbf{h}|\mathbf{x}_{\text{i.i.d.}}) \\ \geq I(\mathbf{y};\mathbf{x}_{\text{i.i.d.}}|\mathbf{h}) - I(\mathbf{y};\mathbf{h}|\mathbf{x}_{\text{i.i.d.}}). \quad (18)$$

The i.i.d. and the constant modulus assumptions imply that

$$I(\mathbf{y};\mathbf{x}_{\text{i.i.d.}}|\mathbf{h}) = KNI(y;x_{\text{i.i.d.}}|h) \quad (19)$$

and

$$I(\mathbf{y};\mathbf{h}|\mathbf{x}_{\text{i.i.d.}}) = \log\det\left(\mathbf{I}+\frac{PT}{N}\mathbf{C_h}\right). \quad (20)$$

Next, by inserting (19) and (20) in (18), dividing by KT, and taking the limit $K\to\infty$, we obtain

$$C(NF) \geq \mathrm{LB}(NF) = \frac{N}{T}I(y;x_{\text{i.i.d.}}|h) \\ -\lim_{K\to\infty}\frac{1}{KT}\log\det\left(\mathbf{I}+\frac{PT}{N}\mathbf{C_h}\right). \quad (21)$$

Finally, we use the generalization of Szegö's theorem to block-Toeplitz matrices provided in [14] to show that for large enough N the second term on the RHS of (21) admits the following closed-form approximate integral expression

$$\lim_{K\to\infty}\frac{1}{KT}\log\det\left(\mathbf{I}+\frac{PT}{N}\mathbf{C_h}\right) \\ \approx NF\int_{\tau}\int_{\nu}\log\left(1+\frac{P}{NF}C_{\mathbb{H}}(\tau,\nu)\right)\mathrm{d}\tau\mathrm{d}\nu. \quad (22)$$

The proof is complete by replacing (22) in (21) and substituting $W = NF$.

References

[1] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun.*, vol. 11, no. 4, pp. 360–393, Dec. 1963.
[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, U.S.A.: Wiley, 1968.
[3] R. S. Kennedy, *Fading Dispersive Communication Channels*. New York, NY, U.S.A.: Wiley, 1969.
[4] M. Médard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 840–852, Apr. 2002.
[5] V. G. Subramanian and B. Hajek, "Broad-band fading channels: Signal burstiness and capacity," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 809–827, Apr. 2002.
[6] A. J. Viterbi, "Performance of an M-ary orthogonal communication system using stationary stochastic signals," *IEEE Trans. Inf. Theory*, vol. 13, no. 3, pp. 414 – 422, July 1967.
[7] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3102–3120, Sept. 2005.
[8] ——, "Low SNR capacity of fading channels with peak and average power constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seattle, WA, U.S.A., July 2006, pp. 689–693.
[9] W. Zhang and J. N. Laneman, "How good is phase-shift keying for peak-limited Rayleigh fading channels in the low-SNR regime?" *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 236–251, Jan. 2007.
[10] W. Kozek, "Matched Weyl-Heisenberg expansions of nonstationary environments," Ph.D. dissertation, Vienna University of Technology, Vienna, Austria, Mar. 1997.
[11] D. Schafhuber, H. Bölcskei, and G. Matz, "System capacity of wide-band OFDM communications over fading channels without channel knowledge," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Chicago, IL, U.S.A., June / July 2004, p. 389 (with corrections available at http://www.nari.ee.ethz.ch/commth/pubs/p/ofdm04).
[12] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[13] U. Grenander and G. Szegö, *Toeplitz forms and their applications*. New York, NY, U.S.A.: Chelsea Publications, 1984.
[14] E. E. Tyrtyshnikov and N. L. Zamarashkin, "Spectra of multilevel Toeplitz matrices: Advanced theory via simple matrix relationships," *Lin. Algeb. Appl.*, vol. 270, pp. 15–27, Feb. 1998.
[15] S. Butman and M. J. Klass, "Capacity of noncoherent channels," Jet Propulsion Lab., Pasadena, CA, U.S.A., Tech. Rep. 32-1526, Sept. 1973.
[16] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz), "Capacity of underspread WSSUS fading channels in the wideband regime under peak constraints," in preparation.
[17] V. Sethuraman, B. Hajek, and K. Narayanan, "Capacity bounds for noncoherent fading channels with a peak constraint," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Adelaide, Australia, Sept. 2005, pp. 515–519.
[18] J. Doob, *Stochastic processes*. New York, NY, U.S.A.: Wiley, 1953.
[19] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.
[20] W. He and C. N. Georghiades, "Computing the capacity of a MIMO fading channel under PSK signaling," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1794–1803, May 2005.
[21] I. E. Telatar and D. N. C. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1384–1400, July 2000.
[22] M. C. Gursoy, H. V. Poor, and S. Verdú, "On-off frequency-shift keying for wideband fading channels," *EURASIP J. Wireless Commun. Netw.*, vol. 6, pp. 95–109, Mar. 2006.
[23] C. W. Helstrom, "Image restoration by the method of least squares," *J. Opt. Soc. Amer.*, vol. 57, pp. 297–303, Mar. 1967.