On the Sensitivity of Continuous-Time Noncoherent Fading Channel Capacity

Giuseppe Durisi, *Senior Member, IEEE*, Veniamin I. Morgenshtern, and
Helmut Bölcskei, *Fellow, IEEE*

***Abstract*—The noncoherent capacity of stationary discrete-time fading channels is known to be very sensitive to the fine details of the channel model. More specifically, the measure of the support of the fading-process power spectral density (PSD) determines if noncoherent capacity grows logarithmically in SNR or slower than logarithmically. Such a result is unsatisfactory from an engineering point of view, as the support of the PSD cannot be determined through measurements. The aim of this paper is to assess whether, for general continuous-time Rayleigh-fading channels, this sensitivity has a noticeable impact on capacity at SNR values of practical interest.**

To this end, we consider the general class of band-limited continuous-time Rayleigh-fading channels that satisfy the wide-sense stationary uncorrelated-scattering (WSSUS) assumption and are, in addition, underspread. We show that, for all SNR values of practical interest, the noncoherent capacity of every channel in this class is close to the capacity of an AWGN channel with the same SNR and bandwidth, independently of the measure of the support of the scattering function (the two-dimensional channel PSD). Our result is based on a lower bound on noncoherent capacity, which is built on a discretization of the channel input-output relation induced by projecting onto Weyl-Heisenberg (WH) sets. This approach is interesting in its own right as it yields a mathematically tractable way of dealing with the mutual information between certain continuous-time random signals.

***Index Terms*—Continuous-time, ergodic capacity, fading channels, Weyl-Heisenberg sets, wide-sense stationary uncorrelated-scattering, underspread property.**

I. Introduction and Summary of Results

The capacity of fading channels in the *noncoherent setting* where neither transmitter nor receiver are aware of the realizations of the fading process, but both know its statistics,[1] is notoriously difficult to analyze, even for simple channel models. Most of the results available in the literature pertain to either low or high signal-to-noise ratio (SNR) asymptotics. While in the low-SNR regime the capacity behavior is robust with respect to the underlying channel model (see for example [1], [2]), this is not the case in the high-SNR regime, where—as we are going to argue next—capacity is very sensitive to the *fine details* of the channel model.

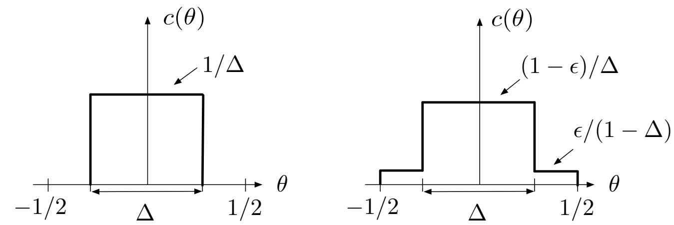


Fig. 1. Two channels with similar PSD $c(\theta)$, but drastically different high-SNR capacity behavior.

Consider, e.g., a *discrete-time* stationary frequency-flat time-selective Rayleigh-fading channel subject to additive white Gaussian noise (AWGN). Here, the channel statistics are fully specified by the fading-process power spectral density (PSD) $c(\theta)$, $\theta \in [-1/2, 1/2)$, and by the noise variance. The high-SNR capacity of this channel turns out to depend on the measure μ of the support of the PSD. More specifically, let ρ denote the SNR; if $\mu < 1$, capacity behaves as $(1-\mu)\log\rho$ in the high-SNR regime [3]. The *pre-log* factor $(1-\mu)$ quantifies the loss in signal-space dimensions (relative to coherent capacity [4], which behaves as $\log\rho$) due to the lack of channel knowledge at the receiver.[2] For $\mu \ll 1$ this loss is negligible, suggesting that, in this case, the realizations of the fading channel can be learned at the receiver (at high SNR) by sacrificing a negligible fraction of the signal-space dimensions available for communication. If $\mu = 1$ and the fading process is *regular*, i.e., $\int_{-1/2}^{1/2} \log c(\theta) d\theta > -\infty$, the high-SNR capacity behaves as $\log\log\rho$ [7]. This double-logarithmic growth behavior of capacity with SNR renders communication in the high-SNR regime extremely power inefficient.

As a consequence of the results just mentioned, we have the following: consider two discrete-time stationary Rayleigh-fading channels, the first one with PSD equal to $1/\Delta$ for $\theta \in [-\Delta/2, \Delta/2]$ and 0 else ($0 < \Delta < 1$), and the second one with PSD equal to $(1-\epsilon)/\Delta$ for $\theta \in [-\Delta/2, \Delta/2]$ and $\epsilon/(1-\Delta)$ else ($0 < \epsilon < 1$, see Fig. 1). These two channels will have completely different high-SNR capacity behavior, no matter how small ϵ is. Specifically, the capacity of the first channel behaves

G. Durisi is with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden, Email: durisi@chalmers.se

V. I. Morgenshtern is with the Department of Statistics, Stanford University, CA, USA, Email: vmorgen@stanford.edu

H. Bölcskei is with the Department of Information Technology and Electrical Engineering, ETH Zurich, Zurich, Switzerland, Email: boelcskei@nari.ee.ethz.ch

Part of the material in this paper was presented at the 2009 IEEE International Symposium on Information Theory.

[1]Capacity in the noncoherent setting is sometimes called *noncoherent* capacity; in the remainder of this paper, it will be referred to simply as capacity. We will use the adjective *coherent* to denote the setting where the channel realizations are perfectly known at the receiver but unknown at the transmitter, which is assumed to know the channel statistics only.

[2]Results of the same nature as those reported in [3] were obtained previously for the block-fading channel model (a non-stationary channel model) in [5], [6].

as $(1-\Delta)\log\rho$, whereas the capacity of the second one grows as $\log\log\rho$. A result like this is clearly unsatisfactory from an engineering point of view, as the measure of the support of a PSD cannot be determined through channel measurements. Such a sensitive dependency of the (high-SNR) capacity behavior on the fine details of the channel model (by fine details we mean details that, in the words of Slepian [8], have "…no direct meaningful counterparts in the real world …"), should make one question the usefulness of the discrete-time stationary channel model itself, at least for high-SNR analyses. In the light of this observation, an engineering-relevant problem is to assess whether this sensitivity has a noticeable impact on capacity at SNR values of practical interest. Unfortunately, this problem is still largely open. For the stationary discrete-time case, an attempt to characterize the capacity sensitivity was made in [9], where, for a first-order Gauss-Markov channel process (a regular process), the SNR beyond which capacity starts exhibiting a sub-logarithmic growth in SNR is computed as a function of the innovation variance λ of the process. More specifically, it is shown in [9] that for $\rho \gg 1$ and $\lambda \ll 1$ capacity grows as $\log\rho$ as long as $\rho < 1/\lambda$. In words, when the innovation variance is small, the high-SNR capacity grows logarithmically in SNR up to SNR values not exceeding $1/\lambda$. The main limitation of this result lies in the fact that it is based on a highly specific channel model, namely a first-order Gauss-Markov process, which is fully described by a single parameter, the innovation variance. Furthermore, it is difficult to relate this parameter to physical channel quantities such as the channel Doppler spread.

A more general approach is presented in [7], where the *fading number*, defined as the second term in the high-SNR expansion of capacity, is characterized for arbitrary discrete-time, stationary, regular fading channels. The fading number determines the rate after which the $\log\log$ regime kicks in, and communication becomes extremely power inefficient. Unfortunately, as illustrated in [10], it is, in general, not possible to relate the fading number to the SNR value at which the $\log\log$ behavior comes into effect.

The purpose of this paper is to characterize the sensitivity of capacity with respect to the channel model for the general class of *continuous-time* Rayleigh-fading linear time-varying (LTV) channels that satisfy the *wide-sense stationary* (WSS) and *uncorrelated scattering* (US) assumptions [11] and that are, in addition, *underspread* [12]. The Rayleigh-fading and the WSSUS assumptions imply that the statistics of the channel are fully characterized by its two-dimensional PSD, often referred to as the *scattering function* [11]; the underspread assumption is satisfied if the scattering function is "highly concentrated" in the delay-Doppler plane. Different definitions of the underspread property are available in the literature (e.g., in terms of the support area of the scattering function [1], [13] or in terms of its moments [14]). For the problem considered in this paper, it is crucial to adopt a novel definition of the underspread property (see Definition 1 in Section II-B), inspired by Slepian's treatment of finite-energy signals that are approximately time- and band-limited [8]. Specifically, we shall say that a WSSUS channel is underspread if its scattering function has only a fraction $\epsilon \ll 1$ of its volume outside a rectangle of area $\Delta_{\mathbb{H}} \ll 1$. This novel definition of the underspread property encompasses the underspread definitions previously proposed in the literature [13], [1], [14] and generalizes them.

When $\epsilon = 0$, i.e., when the scattering function is compactly supported, and $\Delta_{\mathbb{H}} \ll 1$ we expect—on the basis of the results obtained in [7], [3] in the context of the stationary discrete-time fading channel model—capacity to grow logarithmically in SNR. Unfortunately, it is not possible to determine through channel measurements whether a scattering function is compactly supported or not, which motivates our novel underspread definition. For the practically more relevant case $0 < \epsilon \ll 1$, we show that the sub-logarithmic growth behavior kicks in only at very large SNR. Our result is built on a lower bound on the capacity of band-limited continuous-time WSSUS underspread Rayleigh-fading channels that is explicit in the channel parameters $\Delta_{\mathbb{H}}$ and ϵ. By comparing this lower bound to a trivial capacity upper bound, namely, the capacity of a nonfading AWGN channel with the same SNR and bandwidth, we find that, for all SNR values of practical interest, the fading channel capacity is close[3] to the capacity of a nonfading AWGN channel (with the same SNR and bandwidth). As a rule of thumb, this statement is true for all SNR values in the range $\sqrt{\Delta_{\mathbb{H}}} \ll \rho \ll 1/(\Delta_{\mathbb{H}}+\epsilon)$. Hence, we conclude that the fading channel capacity essentially grows logarithmically in SNR for all SNR values of practical interest.

Information theoretic analyses of continuous-time channels are notoriously difficult. The standard approach is to discretize the continuous-time channel input-output (I/O) relation by projecting the input and output signals onto the singular functions of the *channel operator* [15], [16]. This yields a *diagonalized* discretized I/O relation consisting of countably many scalar, non-interacting I/O relations. Unfortunately, this approach is not viable in our setting because random LTV channels have random singular functions, which are not known to transmitter and receiver in the noncoherent setting [1], [2]. We will nevertheless discretize the channel by constraining the input signal to lie in the span of an orthonormal Weyl-Heisenberg (WH) set, i.e., a set of time-frequency shifted versions of a given function, and by projecting the receive signal on the same set of functions. This guarantees that the resulting discretized channel inherits the (two-dimensional) stationarity property of the underlying continuous-time channel, a fact that is essential for our analysis. This approach is interesting in its own right, as it yields a mathematically tractable way of dealing with the mutual information between certain continuous-time random signals.

In [1] a similar approach was used to obtain bounds on the capacity of continuous-time Rayleigh-fading WSSUS underspread channels at low SNR. These bounds are derived under the assumption that the off-diagonal terms in the discretized I/O relation can be neglected, which greatly simplifies the capacity analysis. Whereas this simplification was shown in [2] to be admissible at low SNR, it is unclear whether the off-diagonal terms can be neglected at high SNR. We will therefore explicitly account for the off-diagonal terms in the discretized I/O relation by treating them as (signal-dependent) additive noise, and thus obtain a firm lower bound on the capacity of the underlying continuous-time channel. This lower bound yields

[3]"Close" here means that the ratio between the capacity lower bound and the capacity of a nonfading AWGN channel (with the same SNR and bandwidth) exceeds 0.75.

an information-theoretic criterion for the design of WH sets to be used for pulse-shaped (PS) orthogonal frequency-division multiplexing (OFDM) communication systems operating over Rayleigh-fading WSSUS underspread fading channels. In particular, the lower bound suggests that the WH set should be chosen so as to optimally trade signal-space dimensions (available for communication) for minimization of the power of the off-diagonal terms in the resulting discretized I/O relation.

Notation: Uppercase boldface letters denote matrices, and lowercase boldface letters designate vectors. The Hilbert space of complex-valued finite-energy signals is denoted as $\mathcal{L}^2(\mathbb{R})$; furthermore, $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ stand for the inner product and the norm in $\mathcal{L}^2(\mathbb{R})$, respectively. The set of positive real numbers is denoted as $\mathbb{R}_+$ and the set of integers as $\mathbb{Z}$; $\mathbb{E}[\cdot]$ is the expectation operator, $\mathsf{h}(\cdot)$ denotes differential entropy, and $\mathbb{F}[\cdot]$ stands for the Fourier transform. For two vectors $\mathbf{a}$ and $\mathbf{b}$ of equal dimension, the Hadamard (element-wise) product is denoted as $\mathbf{a}\odot\mathbf{b}$. We write $\mathrm{diag}\{\mathbf{x}\}$ for the diagonal matrix that has the elements of the vector $\mathbf{x}$ on its main diagonal. The superscripts ^{T}, *, and ^{H} stand for transposition, element-wise conjugation, and Hermitian transposition, respectively. The largest eigenvalue of the Hermitian matrix $\mathbf{A}$ is denoted as $\lambda_{\max}\{\mathbf{A}\}$. For two functions $f(x)$ and $g(x)$, the notation $f(x)=\mathcal{O}(g(x))$, $x\to\infty$, means that $\limsup_{x\to\infty}|f(x)/g(x)| < \infty$. Finally, $\delta[k]$ is defined as $\delta[0]=1$ and $\delta[k]=0$ for $k\neq 0$. Throughout the paper, we shall make use of the following projection operators acting on $\mathcal{L}^2(\mathbb{R})$: the *time-limiting* operator $\mathbb{T}_D$, defined as

$$(\mathbb{T}_D x)(t) = \begin{cases} x(t), & \text{if } |t|\leq D/2 \\ 0, & \text{otherwise} \end{cases}$$

which limits $x(t)$ to the interval $[-D/2, D/2]$, and the *frequency-limiting* operator defined as

$$(\mathbb{B}_W x)(t) = \int_{t'} \frac{\sin[\pi W(t-t')]}{\pi(t-t')} x(t')dt'$$

which limits the Fourier transform of $x(t)$ to the interval $[-W/2, W/2]$.

II. System Model

A. Channel and Signal Model

The I/O relation of a continuous-time random LTV channel $\mathbb{H}$ can be written as [17]

$$\begin{aligned} y(t) &= \underbrace{(\mathbb{H}\,x)(t)}_{\triangleq\, r(t)} + w(t) \\ &= \int_{\tau} h_{\mathbb{H}}(t,\tau)x(t-\tau)d\tau + w(t). \end{aligned} \quad (1)$$

Here, $r(t)$ is the output signal in the absence of additive noise. Following [15, Model 2], we assume that the stochastic input signal $x(t)$:

i) is strictly band-limited to W Hz according to

$$X(f) = 0, \quad \text{for } |f| > W/2 \quad (2)$$

with probability one, where $X(f)\triangleq\mathbb{F}[x(t)]$;

ii) is approximately time-limited to a duration of D sec according to

$$\mathbb{E}\left[\|\mathbb{T}_D x(t)\|^2\right] \geq (1-\eta)\,\mathbb{E}\left[\|x(t)\|^2\right] \quad (3)$$

where $0<\eta\ll 1$;

iii) satisfies the average-power constraint

$$(1/D)\,\mathbb{E}\left[\|x(t)\|^2\right] \leq P. \quad (4)$$

The constraints (2) and (3) capture the fact that we are dealing with input signals that are strictly band-limited and essentially time-limited. As pointed out in [15, p. 364], time limitation is important as this allows for a physically meaningful definition of transmission rate. Note that the strict bandwidth constraint (2) implies that a nonzero $x(t)$ can be limited in time only in an approximate sense [8], a consideration that justifies the form of the constraint expressed in (3).

The signal $w(t)$ is a zero-mean proper AWGN process with double-sided PSD equal to 1. Finally, the time-varying channel impulse response $h_{\mathbb{H}}(t,\tau)$ is a zero-mean jointly proper Gaussian (JPG) process in time t and delay τ that satisfies the WSSUS assumption

$$\mathbb{E}[h_{\mathbb{H}}(t,\tau)h^*_{\mathbb{H}}(t',\tau')] = R_{\mathbb{H}}(t-t',\tau)\delta(\tau-\tau') \quad (5)$$

and is independent of $w(t)$ and $x(t)$. As a consequence of the JPG and the WSSUS assumptions, the *time-delay correlation function* $R_{\mathbb{H}}(t,\tau)$ fully characterizes the channel statistics.

Often, it is convenient to describe the action of the channel $\mathbb{H}$ in domains other than the time-delay domain used in (1). Specifically, we shall frequently work with the following alternative I/O relation [cf. (1)], which is explicit in the channel *delay-Doppler spreading function* $S_{\mathbb{H}}(\tau,\nu)=\int_t h_{\mathbb{H}}(t,\tau)e^{-j2\pi\nu t}dt$ according to

$$y(t) = \underbrace{\iint_{\nu\,\tau} S_{\mathbb{H}}(\tau,\nu)x(t-\tau)e^{j2\pi\nu t}d\tau d\nu}_{=\,r(t)} + w(t).$$

This alternative I/O relation leads to the following physical interpretation: the noiseless output signal $r(t)=(\mathbb{H}\,x)(t)$ is a weighted superposition of copies of the input signal $x(t)$ that are shifted in time by the delay τ and in frequency by the Doppler shift ν. The spreading function is the corresponding weighting function. In other words, the channel operator $\mathbb{H}$ can be represented as a continuous weighted superposition of time-frequency shift operators. Note that every "reasonable" linear operator admits such a representation (see [18, Thm. 14.3.5] for a precise mathematical formulation of this statement). As a consequence of the WSSUS assumption, the spreading function $S_{\mathbb{H}}(\tau,\nu)$ is uncorrelated in τ and ν, i.e., we have

$$\mathbb{E}[S_{\mathbb{H}}(\tau,\nu)S^*_{\mathbb{H}}(\tau',\nu')] = C_{\mathbb{H}}(\tau,\nu)\delta(\tau-\tau')\delta(\nu-\nu') \quad (6)$$

where $C_{\mathbb{H}}(\tau,\nu)$ is the two-dimensional PSD of the channel process, usually referred to as *scattering function* [17]. In the remainder of the paper, we let the scattering function be normalized in volume according to

$$\iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu)d\tau d\nu = 1. \quad (7)$$

Another system function we shall need is the *time-varying transfer function*

$$L_{\mathbb{H}}(t,f) \triangleq \int_{\tau} h_{\mathbb{H}}(t,\tau) e^{-j2\pi f\tau} d\tau$$

which, as a consequence of (5), is stationary in both time and frequency:

$$\mathbb{E}[L_{\mathbb{H}}(t,f) L_{\mathbb{H}}^{*}(t',f')] = B_{\mathbb{H}}(t-t', f-f'). \tag{8}$$

Here, $B_{\mathbb{H}}(t,f)$ denotes the time-frequency correlation function of the channel process, which is related to the scattering function through a two-dimensional Fourier transform

$$B_{\mathbb{H}}(t,f) = \iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu) e^{j2\pi(\nu t - \tau f)} d\tau d\nu.$$

For a more complete description of the WSSUS channel model, the interested reader is referred to [17], [1].

B. A Robust Definition of Underspread Channels

Qualitatively speaking, WSSUS *underspread* channels are WSSUS channels with a scattering function that is highly concentrated in the delay-Doppler plane [11]. For the case where $C_{\mathbb{H}}(\tau,\nu)$ is *compactly supported*, the channel is said to be underspread if the support area of $C_{\mathbb{H}}(\tau,\nu)$ is smaller than 1 (see for example [13], [1]). The compact-support assumption on $C_{\mathbb{H}}(\tau,\nu)$, albeit mathematically convenient, is a fine detail of the channel model in the terminology introduced in Section I, because it is not possible to determine through channel measurements whether $C_{\mathbb{H}}(\tau,\nu)$ is compactly supported or not. However, the results discussed in Section I, in the context of the stationary discrete-time fading channel model, imply a high capacity sensitivity to whether the measure of the support of the PSD is smaller than 1 or not. A similar sensitivity can be expected for the continuous-time WSSUS channel model. To quantify this sensitivity, we need to work with a more general underspread definition. Specifically, we replace the underspread definition based on the compact-support assumption by the following, more robust and physically meaningful, assumption: we say that $\mathbb{H}$ is underspread if $C_{\mathbb{H}}(\tau,\nu)$ has a small fraction of its total volume outside a rectangle of area much smaller than 1. More precisely, we have the following definition.

Definition 1: Let $\tau_0, \nu_0 \in \mathbb{R}_+, \epsilon \in [0,1]$, and let $\mathcal{H}(\tau_0,\nu_0,\epsilon)$ be the set of all Rayleigh-fading WSSUS channels $\mathbb{H}$ with scattering function $C_{\mathbb{H}}(\tau,\nu)$ satisfying

$$\int_{-\nu_0}^{\nu_0} \int_{-\tau_0}^{\tau_0} C_{\mathbb{H}}(\tau,\nu) d\tau d\nu \geq 1-\epsilon. \tag{9}$$

We say that the channels in $\mathcal{H}(\tau_0,\nu_0,\epsilon)$ are *underspread* if $\Delta_{\mathbb{H}} \triangleq 4\tau_0\nu_0 \ll 1$ and $\epsilon \ll 1$.

Note that it is possible to verify, through channel measurements, whether a fading channel is underspread according to Definition 1. Typical wireless channels are (highly) underspread, with most of the volume of $C_{\mathbb{H}}(\tau,\nu)$ supported over a rectangle of area $\Delta_{\mathbb{H}} \leq 10^{-3}$ for land-mobile channels, and $\Delta_{\mathbb{H}}$ as small as 10^{-7} for certain indoor channels with restricted terminal mobility. Note that setting $\epsilon = 0$ in Definition 1 yields the compact-support underspread definition of [13], [1]. The moment-based underspread definition proposed in [14] is subsumed by Definition 1 as well.

C. Band-Limitation at the Receiver

Even though $x(t)$ has bandwidth no larger than W, the signal $r(t) = (\mathbb{H}\,x)(t)$ is, in general, not strictly band-limited, because $\mathbb{H}$ can introduce arbitrarily large frequency dispersion. However, if $\mathbb{H}$ is underspread in the sense of Definition 1, most of the energy of $r(t)$ will be supported on a frequency band of size $(W+2\nu_0)$ Hz. We therefore assume that the output signal $y(t)$ is passed through an ideal low-pass filter of bandwidth $(W + 2\nu_0)$ Hz, resulting in the filtered output signal

$$y_f(t) = (\mathbb{B}_{W+2\nu_0} y)(t). \tag{10}$$

This filtering operation yields a band-limited WSSUS fading channel.

III. Channel Capacity

A. Outline of the Information-Theoretic Analysis

We are interested in characterizing the ultimate limit on the rate of reliable communication over the continuous-time fading channel (1) in the noncoherent setting (i.e., the setting where neither the transmitter nor the receiver know the realization of $\mathbb{H}$, but both know the statistics of $\mathbb{H}$). Two main difficulties need to be overcome to obtain such a characterization. First, we need to deal with continuous-time channels and signals, which are notoriously difficult to analyze information-theoretically. Second, our focus is on the noncoherent setting, for which, even for simple discrete-time channel models, analytic capacity characterizations are not available.

To overcome these difficulties we resort to bounds on capacity. As (trivial) capacity upper bound, we take in Section III-C the capacity of a band-limited Gaussian channel [15] with the same average-power constraint as in (4) and bandwidth equal to $(W + 2\nu_0)$. A capacity lower bound is obtained in Section IV through the following two steps: first, we construct a discretized channel whose capacity is proven to be a lower bound on the capacity of the underlying continuous-time channel (1); then, we derive a lower bound on the capacity of this discretized channel that is explicit in the channel parameters $\Delta_{\mathbb{H}}$ and ϵ. In Section V, we then show that, for channels that are underspread according to Definition 1, this lower bound is close to the AWGN-channel capacity upper bound for all SNR values of practical interest, thereby sandwiching the capacity of the band-limited continuous-time fading channel tightly.

B. Mutual Information and Capacity for the Continuous-Time Channel

Dealing with continuous-time channels requires a suitable generalization of the definitions of mutual information and capacity [19] to the continuous-time case. Such a generalization can be found, e.g., in [20], [16, Ch. 8], and is reviewed here for completeness.

To define capacity of the channel (1), we represent the complex signals at the input and output of $\mathbb{H}$ in terms of projections onto complete orthonormal sets for the underlying signal spaces. More specifically, let $\{\phi_m(t)\}_{m=0}^{\infty}$ be a complete orthonormal set for the space $\mathcal{L}^2(W)$ of signals with bandwidth no larger than W. We can then describe $x(t) \in \mathcal{L}^2(W)$ uniquely in terms of the projections

$$x_m \triangleq \langle x(t), \phi_m(t)\rangle, \quad m = 0, 1, \ldots \tag{11}$$

as $x(t) = \sum_m x_m \phi_m(t)$. Similarly, let $\{\phi'_m(t)\}_{m=0}^{\infty}$ be a complete orthonormal set for $\mathcal{L}^2(W+2\nu_0)$. The low-pass filtered output signal $y_f(t) \in \mathcal{L}^2(W+2\nu_0)$ in (10) can be described uniquely in terms of the projections

$$y_m \triangleq \langle y_f(t), \phi'_m(t)\rangle, \quad m = 0, 1, \ldots \tag{12}$$

as $y_f(t) = \sum_m y_m \phi'_m(t)$. To define the mutual information between $x(t)$ and $y_f(t)$, we need to impose a probability measure on $x(t)$.[4] Concretely, let $\mathcal{Q}(W, D, \eta, P)$ be the set of probability measures on $x(t)$ that satisfy the bandwidth constraint (2), the time-limitation constraint (3), and the average-power constraint (4). Every probability measure in $\mathcal{Q}(W, D, \eta, P)$ induces a corresponding probability measure on $\{x_m\}_{m=0}^{\infty}$. For a given probability measure in $\mathcal{Q}(W, D, \eta, P)$, the mutual information between $x(t)$ and $y_f(t)$ is defined as [16, Eq. (8.151)], [20, Def. 3, Thm. 1.5]

$$I(y_f(t); x(t)) \triangleq \lim_{M\to\infty} I(\mathbf{y}^M; \mathbf{x}^M)$$

where $\mathbf{x}^M = [x_0\ x_1\ \ldots\ x_M]^T$, and, similarly, $\mathbf{y}^M = [y_0\ y_1\ \ldots\ y_M]^T$. This definition turns out to be independent of the complete orthonormal sets $\{\phi_m(t)\}_{m=0}^{\infty}$ and $\{\phi'_m(t)\}_{m=0}^{\infty}$ used [20, Thm. 1.5]. The capacity C of the channel (1) can now be defined as follows [16, Eq. (8.1.55)]:

$$C \triangleq \lim_{D\to\infty} \frac{1}{D} \sup_{\mathcal{Q}(W,D,\eta,P)} I(y_f(t); x(t)). \tag{13}$$

We conclude this section by noting that, by Fano's inequality, no rate above C is achievable [22]. However, whether the channel coding theorem applies to the general class of time-frequency selective fading channels considered in this paper is an open problem, even for the discrete-time case [23].

C. An Upper Bound on Capacity

For underspread channels in $\mathcal{H}(\tau_0, \nu_0, \epsilon)$ (see Definition 1) and input signals satisfying (2)–(4), we take as simple (yet tight, in a sense to be specified in Section V) upper bound on (13) the capacity of a (nonfading) band-limited AWGN channel with the same average-power constraint as in (4) and bandwidth $(W + 2\nu_0)$. More precisely, we show in Appendix A that $C \le C_{\text{AWGN}}$, where

$$C_{\text{AWGN}} \triangleq (W + 2\nu_0) \log\left(1 + (1-\eta)(1-\epsilon)\frac{P}{W + 2\nu_0}\right) + (\eta + \epsilon - \eta\epsilon)P. \tag{14}$$

[4] A probability measure on $x(t)$ is specified through the joint probability measure of the n-tuples $(x(t_1), \ldots, x(t_n))$ for every $n \in \mathbb{N}$ and for every choice of $t_1, \ldots, t_n \in \mathbb{R}$ [21, Sec. 25.2].

This result is based on [15, Thm. 2]. Differently from [15, Eq. (20)], the second term on the right-hand side (RHS) of (14) accounts not only for the approximate time-limitation of $x(t)$, but also for the dispersive nature of $\mathbb{H}$.

It is now appropriate to provide a preview of the nature of the results we are going to obtain. We will show that, as long as $\Delta_{\mathbb{H}} \ll 1$ and $\epsilon \ll 1$, the capacity of every channel in $\mathcal{H}(\tau_0, \nu_0, \epsilon)$, independently of whether its scattering function is compactly supported or not, is close to the AWGN-channel capacity C_{AWGN} for all SNR values typically encountered in practical wireless communication systems. To establish this result, we derive, in the next section, a lower bound on (13).

IV. A Lower Bound on Capacity

A. Outline

As the derivation of the capacity lower bound presented in this section consists of several steps, we start by providing an outline of our proof strategy. The first step entails restricting the set of input distributions in (13) to a subset of $\mathcal{Q}(W, D, \eta, P)$; this clearly yields a lower bound on C. The subset of $\mathcal{Q}(W, D, \eta, P)$ we consider is described in Section IV-B and is obtained by constraining the input signal $x(t)$ to lie in the span of an orthonormal WH set (that is not necessarily complete for $\mathcal{L}^2(W)$). The second step (see Section IV-C) consists of projecting the corresponding output signal $y_f(t)$ onto the same orthonormal WH set, an operation that further lower-bounds mutual information, as seen by application of the data-processing inequality [20, Thm. 1.4] (the orthonormal WH set is not necessarily complete for $\mathcal{L}^2(W+2\nu_0)$). As a result of these two steps, we obtain a discretization of the I/O relation. The capacity of the corresponding discretized channel, which is a lower bound on the capacity of the underlying continuous-time channel, is further lower-bounded in Section IV-E by treating the off-diagonal terms in the I/O relation as (signal-dependent) additive noise. This finally yields a lower bound on the capacity of the underlying continuous-time channel that is explicit in the channel parameters $\Delta_{\mathbb{H}}$ and ϵ.

B. A Smaller Set of Input Distributions

Let $g_{k,n}(t) \triangleq g(t - kT)e^{j2\pi nFt}$ and

$$(g, T, F) \triangleq \{g_{k,n}(t)\}_{k,n\in\mathbb{Z}}$$

be an orthonormal WH set, i.e., a set consisting of time-frequency shifts (on a rectangular lattice) of a given pulse $g(t) \in \mathcal{L}^2(\mathbb{R})$. Orthonormality of the WH set implies $TF \ge 1$, as a consequence of [18, Cor. 7.5.1, Cor. 7.3.2]. We lower-bound C by restricting the input signals to be of the form

$$x(t) = \sum_{k=-K}^{K} \sum_{n=-N}^{N} x[k,n] g_{k,n}(t) \tag{15}$$

where $\{x[k,n]\}$ are random coefficients. To guarantee that $x(t)$ in (15) satisfies (2)–(4), we impose the following constraints on (g, T, F), K, N, and $\{x[k,n]\}$.

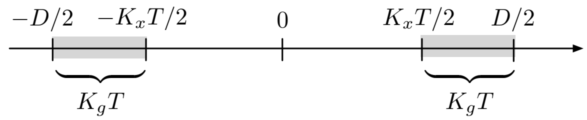


Fig. 2. Insertion of guard intervals.

1) Average-power constraint: To ensure that $x(t)$ in (15) satisfies (4), it is sufficient to choose K such that $(2K+1)T \leq D$ (further restrictions on the choice of K will be imposed in Section IV-B3), and to require that the random variables $\{x[k,n]\}$ satisfy

$$\sum_{k=-K}^{K} \sum_{n=-N}^{N} \mathbb{E}\left[|x[k,n]|^2\right] \leq (2K+1)TP. \tag{16}$$

The constraint (16), together with the orthonormality of the set (g,T,F), implies that (4) is satisfied.

2) Bandwidth limitation: To ensure that $x(t)$ in (15) satisfies (2), we require that $g(t)$ fulfill the following property.

Property 1: The function $g(t)$ is strictly band-limited with bandwidth $F \leq W$.

Furthermore, we take $N = (N_x - 1)/2$ where $N_x \triangleq W/F$. For simplicity of exposition, we shall assume, in the remainder of the paper, that N_x is an odd integer.

3) Time limitation: To ensure that $x(t)$ in (15) satisfies (3), we impose two additional constraints. First, we require that $g(t)$ satisfies the following property.

Property 2: The function $g(t)$ is even and decays faster than $1/t$, i.e.,

$$g(t) = \mathcal{O}(1/t^{1+\mu}), \quad t \to \infty \tag{17}$$

for some $\mu > 0$.

Second, we insert, in the interval $[-D/2, D/2]$, two *guard intervals*. More specifically, for a given approximate duration D of the input signal $x(t)$ [we will later take $D \to \infty$ according to (13)], the interval $[-D/2, D/2]$ is divided up into three parts (see Fig. 2): the interval $[-K_xT/2, K_xT/2]$, with $K = (K_x - 1)/2$ in (15),[5] supporting most of the energy of $x(t)$, and two guard intervals $[-D/2, -K_xT/2]$ and $[K_xT/2, D/2]$, each of length $K_gT = D/2 - K_xT/2$. This will ensure that (3) is satisfied. We will let $K_x \to \infty$ as $D \to \infty$, with K_g kept constant. This guarantees that the fraction of time allocated to the guard intervals vanishes as $D \to \infty$. For simplicity of notation, we shall assume in the remainder of the paper that K_g is an integer. For fixed η in (3), the decay property of $g(t)$ expressed in (17) implies that one can choose K_g (independent of K) so that $x(t)$ in (15) satisfies (3). This statement is proven in Appendix B.

We next show formally that our construction results in a capacity lower bound. Fix an orthonormal WH set (g,T,F) satisfying Properties 1 and 2. Furthermore, let $\mathcal{Q}_d$ be the set of probability measures on $\{x[k,n]\}$ that satisfy (16). Every probability measure in $\mathcal{Q}_d$ induces a probability measure on $x(t)$ in (15). We denote the corresponding set of probability measures on $x(t)$ by $\mathcal{Q}_{\mathrm{WH}}(W,D,\eta,P)$. As just shown, $x(t)$ satisfies (2)–(4). Hence, $\mathcal{Q}_{\mathrm{WH}}(W,D,\eta,P) \subseteq \mathcal{Q}(W,D,\eta,P)$ [recall that $\mathcal{Q}(W,D,\eta,P)$ is the set of *all* probability measures that satisfy (2)–(4)]. We can then lower-bound C in (13) as follows:

$$\begin{aligned} C &= \lim_{D\to\infty} \frac{1}{D} \sup_{\mathcal{Q}(W,D,\eta,P)} I(y_f(t); x(t)) \\ &\geq \lim_{D\to\infty} \frac{1}{D} \sup_{\mathcal{Q}_{\mathrm{WH}}(W,D,\eta,P)} I(y_f(t); x(t)). \end{aligned} \tag{18}$$

Here, the inequality follows by restricting the supremization to the smaller set $\mathcal{Q}_{\mathrm{WH}}(W,D,\eta,P)$.

C. The Discretized I/O Relation

The second step in our approach is to project the output signal $y_f(t)$ [resulting from the transmission of $x(t)$ in (15)] onto the signal set $\{g_{k,n}(t)\}$ to obtain

$$\begin{aligned} y[k,n] &\triangleq \langle y_f, g_{k,n}\rangle \\ &\overset{(a)}{=} \langle y, g_{k,n}\rangle \\ &= \underbrace{\langle \mathbb{H}\, g_{k,n}, g_{k,n}\rangle}_{\triangleq h[k,n]} x[k,n] \\ &\quad + \sum_{\substack{l=-K \\ (l,m)\neq(k,n)}}^{K} \sum_{m=-N}^{N} \underbrace{\langle \mathbb{H}\, g_{l,m}, g_{k,n}\rangle}_{\triangleq p[l,m,k,n]} x[l,m] + \underbrace{\langle w, g_{k,n}\rangle}_{\triangleq w[k,n]} \\ &= h[k,n]x[k,n] \\ &\quad + \sum_{\substack{l=-K \\ (l,m)\neq(k,n)}}^{K} \sum_{m=-N}^{N} p[l,m,k,n]x[l,m] + w[k,n] \end{aligned} \tag{19}$$

for each *time-frequency* slot (k,n), $k = -K, -K+1, \ldots, K$, $n = -N, -N+1, \ldots, N$. Here, (a) is a consequence of Property 1, which implies that the Fourier transform of $g_{k,n}(t)$ (with $k = -K, -K+1, \ldots, K$, $n = -N, -N+1, \ldots, N$) is strictly supported in the interval $[-W/2, W/2]$. We refer to the channel with I/O relation (19) as the discretized channel *induced* by the WH set (g,T,F). As we assumed that $h_{\mathbb{H}}(t,\tau)$ in (1) is a zero-mean JPG random process in t and τ, the random variables $h[k,n]$ and $p[l,m,k,n]$ are zero-mean JPG. Furthermore, the orthonormality of the WH set (g,T,F) implies that the $w[k,n]$ in (19) are i.i.d. $\mathcal{CN}(0,1)$.

For each time slot $k \in \{-K, -K+1, \ldots, K\}$, we arrange the data symbols $x[k,n]$, the output signal samples $y[k,n]$, the channel coefficients $h[k,n]$, and the noise samples $w[k,n]$ in corresponding N_x-dimensional vectors.[6] For example, the N_x-dimensional vector that contains the input symbols in the kth time slot is defined as

$$\mathbf{x}[k] \triangleq \left[x[k,-N]\ x[k,-N+1]\ \ldots\ x[k,N]\right]^T.$$

The output vector $\mathbf{y}[k]$, the channel vector $\mathbf{h}[k]$, and the noise vector $\mathbf{w}[k]$ are defined analogously. To get a compact notation, we further stack K_x contiguous input, output, channel, and noise vectors, into corresponding K_xN_x-dimensional vectors. For example, for the channel input this results in the K_xN_x-dimensional vector

$$\mathbf{x} \triangleq \left[\mathbf{x}^T[-K]\ \ \mathbf{x}^T[-K+1]\ \ \ldots\ \ \mathbf{x}^T[K]\right]^T. \tag{20}$$

[5] We assume that K_x is an odd integer.

[6] Recall that $K_x = 2K+1$ and $N_x = 2N+1$.

Again, the stacked vectors $\mathbf{y}$, $\mathbf{h}$, and $\mathbf{w}$ are defined analogously. Finally, we arrange the self-interference terms $p[l,m,k,n]$ in a $K_xN_x \times K_xN_x$ matrix $\mathbf{P}$ with entries

$$[\mathbf{P}]_{n+kN_x,m+lN_x} = \begin{cases} p[l-K,m-N,k-K,n-N], & \text{if } (l,m)\neq(k,n) \\ 0, & \text{otherwise} \end{cases}$$

for $l,k = 0,1,\dots,K_x-1$ and $m,n = 0,1,\dots,N_x-1$. With these definitions, we can now compactly express the I/O relation (19) as

$$\mathbf{y} = \mathbf{h}\odot\mathbf{x} + \mathbf{P}\mathbf{x} + \mathbf{w}. \tag{21}$$

Let now C_d be the capacity of the discretized channel (21) [induced by the WH set (g,T,F)] with $\mathbf{x}$ subject to the average-power constraint (16). We can lower-bound the RHS of (18) by C_d as follows

$$\begin{aligned} C &\overset{(a)}{\geq} \lim_{D\to\infty}\frac{1}{D}\sup_{\mathcal{Q}_{\text{WH}}(W,D,\eta,P)} I(y_f(t);x(t)) \\ &\overset{(b)}{\geq} \lim_{K_x\to\infty}\frac{1}{(K_x+2K_g)T}\sup_{\mathcal{Q}_d} I(\mathbf{y};\mathbf{x}) \\ &\triangleq C_d. \end{aligned} \tag{22}$$

Here, in (a) we used (18), and (b) is a consequence of [20, Thm. 1.4], which extends the data processing inequality to continuous-time signals. To summarize, we showed that the capacity of the discretized channel (21) induced by the WH set (g,T,F) is a lower bound on the capacity of the underlying continuous-time channel (1).

D. Why Weyl-Heisenberg Sets?

The choice of constraining $x(t)$ to lie in the span of an orthonormal WH set according to (15) results in a signaling scheme that can be interpreted as PS-OFDM [24], where the data symbols $x[k,n]$ are modulated onto a set of orthogonal signals indexed by discrete time (symbol index) k, and discrete frequency (subcarrier index) n. From this perspective, the self-interference term (the second term on the RHS of (19), which is made up of the off-diagonal terms in the I/O relation) can be interpreted as intersymbol and intercarrier interference. Discretization through WH sets is sensible for the following two reasons.

Stationarity: The structure of WH sets preserves the stationarity of the channel in the discretization. More precisely, the channel gains $h[k,n]$ in (19) inherit the two-dimensional stationarity property of the underlying continuous-time channel [see (8)], a fact that is crucial for the ensuing analysis. We prove this result in Appendix C, where we also establish properties of the statistics of $p[l,m,k,n]$ in (19) that will be needed in the remainder of the paper.

Approximate diagonalization: The presence of the self-interference term in (19) makes the computation of C_d in (22) involved. A classic approach to eliminate self-interference is to discretize the channel by projecting the input and output signals onto the channel-operator singular functions [15], [16]. This choice is convenient, as it leads to a diagonal discretized I/O relation, i.e., to countably many scalar, non-interacting I/O relations (see [2] for more details). Unfortunately, this approach is not viable in our setup, because in the LTV case the channel-operator singular functions are, in general, random and not known to transmitter and receiver (recall that we consider the noncoherent setting). Discretizing using *deterministic* orthonormal functions, as done in the previous section, yields self-interference, which we will need to take into account. This will be accomplished by treating self-interference as additive noise, which will further lower-bound capacity. The main technical difficulty in this context arises from the self-interference term being signal-dependent. Moreover, as our capacity lower bound is obtained by treating self-interference as noise, ensuring that the power in the self-interference term is small (and, hence, that the discretized I/O relation is approximately diagonal) is crucial to get a good capacity lower bound. This can be accomplished by choosing the pulse $g(t)$ to be well localized in time and frequency. In fact, it was shown in [13], [25], [14], [1] that the singular functions of random underspread operators can be well approximated by orthonormal WH sets generated by pulses that are well localized in time and frequency.

E. A Lower Bound on the Capacity of the Discretized Channel

We next derive a lower bound on C_d [and, hence, on C in (13)] by using a Gaussian input distribution, and by treating self-interference as (signal-dependent) noise. This lower bound—evaluated for an appropriately chosen WH set—will then be shown to be close (for all SNR values of practical interest) to the AWGN-channel capacity upper bound C_{AWGN} in (14), whenever the channel is underspread according to Definition 1, thereby sandwiching the capacity of the underlying continuous-time channel tightly.

Our first result is a lower bound on C_d, which we indicate as L_1, that is explicit in the power spectral density $\mathbf{C}(\theta)$ of the multivariate stationary channel process $\{\mathbf{h}[k]\}$ with autocorrelation function $\mathbf{R}[k'-k] \triangleq \mathbb{E}\left[\mathbf{h}[k']\mathbf{h}^H[k]\right]$, where

$$\mathbf{C}(\theta) \triangleq \sum_{k=-\infty}^{\infty}\mathbf{R}[k]e^{-j2\pi k\theta}, \quad |\theta|\leq\frac{1}{2}. \tag{23}$$

We then show in Corollary 3, Section IV-F that L_1 can be further lower-bounded by an expression that is explicit in the channel parameters $\Delta_{\mathbb{H}}$ and ϵ introduced in Definition 1.

Theorem 2: Let (g,T,F) be an orthonormal WH set satisfying Properties 1 and 2 in Section IV-B and consider a Rayleigh-fading WSSUS channel (not necessarily underspread) with scattering function $C_{\mathbb{H}}(\tau,\nu)$. For a given bandwidth W and a given SNR $\rho \triangleq P/W$, the capacity of the discretized channel (21) induced by (g,T,F) is lower-bounded according to $C_d(\rho)\geq L_1(\rho)$, where

$$\begin{aligned} L_1(\rho) = \frac{W}{TF}\,\mathbb{E}_h\left[\log\left(1+\frac{r[0,0]TF\rho\,|h|^2}{1+TF\rho\,\sigma_I^2}\right)\right] \\ -\inf_{0<\alpha<1}\left\{\frac{1}{T}\int_{-1/2}^{1/2}\log\det\left(\mathbf{I}+\frac{TF\rho}{\alpha}\mathbf{C}(\theta)\right)d\theta \right. \\ \left. +\frac{W}{TF}\log\left(1+\frac{TF\rho}{1-\alpha}\sigma_I^2\right)\right\}. \end{aligned} \tag{24}$$

Here,

$$h \sim \mathcal{CN}(0,1)$$

$$r[0,0] \triangleq \iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu)\,|A_g(\tau,\nu)|^2\,d\tau d\nu$$

$$\sigma_I^2 \triangleq \sum_{\substack{k=-\infty\\(k,n)\neq(0,0)}}^{\infty}\sum_{n=-\infty}^{\infty}\iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu)\,|A_g(\tau-kT,\nu-nF)|^2\,d\tau d\nu$$

where $A_g(\tau,\nu)$ denotes the ambiguity function of $g(t)$ (see Appendix C) and $\mathbf{C}(\theta)$, defined in (23), denotes the matrix-valued power spectral density of the discretized channel induced by (g,T,F).

Proof: See Appendix E. ■

F. A Lower Bound that is Explicit in the Channel Parameters $\Delta_{\mathbb{H}}$ and ϵ

For the purposes of our analysis, it is convenient to further lower-bound L_1 to get an expression that is explicit in the channel parameters $\Delta_{\mathbb{H}}$ and ϵ introduced in Definition 1. The resulting lower bound, presented in the next corollary, will allow us to assess how sensitive capacity is to whether $C_{\mathbb{H}}(\tau,\nu)$ is compactly supported or not.

Corollary 3: Let (g,T,F) be an orthonormal WH set satisfying Properties 1 and 2 in Section IV-B and consider a Rayleigh-fading WSSUS channel (not necessarily underspread) in the set $\mathcal{H}(\tau_0,\nu_0,\epsilon)$ with scattering function $C_{\mathbb{H}}(\tau,\nu)$. For a given bandwidth W and a given SNR $\rho = P/W$, and under the technical condition $\widetilde{\Delta}_{\mathbb{H}} \triangleq 2\nu_0 T < 1$, the capacity of the discretized channel (21) induced by (g,T,F) is lower-bounded as $C_d(\rho) \geq L_2(\rho)$, where

$$\begin{aligned} L_2(\rho) \triangleq \frac{W}{TF}\Bigg\{&\mathbb{E}_h\left[\log\left(1+\frac{TF\rho(1-\epsilon)m_g|h|^2}{1+TF\rho(M_g+\epsilon)}\right)\right] \\ &-\inf_{0<\alpha<1}\Bigg[\widetilde{\Delta}_{\mathbb{H}}\log\left(1+\frac{TF\rho}{\alpha\widetilde{\Delta}_{\mathbb{H}}}\right) \\ &+(1-\widetilde{\Delta}_{\mathbb{H}})\log\left(1+\frac{TF\rho\,\epsilon}{\alpha(1-\widetilde{\Delta}_{\mathbb{H}})}\right) \\ &+\log\left(1+\frac{TF\rho}{1-\alpha}(M_g+\epsilon)\right)\Bigg]\Bigg\}. \end{aligned} \tag{25}$$

Here, $h \sim \mathcal{CN}(0,1)$, $m_g \triangleq \min_{(\tau,\nu)\in\mathcal{D}} |A_g(\tau,\nu)|^2$, and

$$M_g \triangleq \max_{(\tau,\nu)\in\mathcal{D}} \sum_{\substack{k=-\infty\\(k,n)\neq(0,0)}}^{\infty}\sum_{n=-\infty}^{\infty}|A_g(\tau-kT,\nu-nF)|^2$$

with $\mathcal{D} \triangleq [-\tau_0,\tau_0]\times[-\nu_0,\nu_0]$.

Proof: See Appendix F. ■

The lower bound L_2 in (25) depends on the seven quantities $(\rho, g(t), T, F, \tau_0, \nu_0, \epsilon)$ and is therefore difficult to analyze. We show next that if T and F are chosen so that $\nu_0 T = \tau_0 F$, a condition often referred to as the *grid matching rule* [13, Eq. (2.75)], two of these seven quantities can be dropped without loss of generality.

Lemma 4: Let (g,T,F) be an orthonormal WH set satisfying Properties 1 and 2 in Section IV-B. Then, for any $\beta > 0$, we have

$$\begin{aligned} &L_2(\rho, g(t), T, F, \tau_0, \nu_0, \epsilon) \\ &\qquad = L_2\left(\rho, \sqrt{\beta}g(\beta t), \frac{T}{\beta}, \beta F, \frac{\tau_0}{\beta}, \beta\nu_0, \epsilon\right). \end{aligned}$$

In particular, assume that $\nu_0 T = \tau_0 F$ and let $\beta = \sqrt{T/F} = \sqrt{\tau_0/\nu_0}$ and $\widetilde{g}(t) = \sqrt{\beta}g(\beta t)$. Then,

$$\begin{aligned} &L_2(\rho, g(t), T, F, \tau_0, \nu_0, \epsilon) \\ &\quad = L_2\left(\rho, \widetilde{g}(t), \sqrt{TF}, \sqrt{TF}, \sqrt{\Delta_{\mathbb{H}}}/2, \sqrt{\Delta_{\mathbb{H}}}/2, \epsilon\right) \\ &\quad \triangleq L_2^{(s)}(\rho, \widetilde{g}(t), TF, \Delta_{\mathbb{H}}, \epsilon). \end{aligned} \tag{26}$$

Proof: See Appendix G. ■

In (26), the superscript (s) indicates that the scattering function is supported on a square (with sidelength $\sqrt{\Delta_{\mathbb{H}}}$). In the remainder of the paper, for the sake of simplicity of exposition, we will choose T and F such that the grid matching rule $\nu_0 T = \tau_0 F$ is satisfied. Then, as a consequence of Lemma 4, we can (and will) only consider WH sets of the form $(g, \sqrt{TF}, \sqrt{TF})$ and WSSUS channels in the set $\mathcal{H}(\sqrt{\Delta_{\mathbb{H}}}/2, \sqrt{\Delta_{\mathbb{H}}}/2, \epsilon)$.

The lower bound $L_2^{(s)}$ in (26) can be tightened by maximizing it over all WH sets $(g, \sqrt{TF}, \sqrt{TF})$ satisfying Properties 1 and 2 in Section IV-B. This maximization implicitly provides an information-theoretic criterion for choosing $g(t)$ and TF. Unfortunately, an analytic maximization of $L_2^{(s)}$ seems complicated as the dependency of m_g and M_g on $(g, \sqrt{TF}, \sqrt{TF})$ is difficult to characterize analytically. We shall therefore choose a specific $g(t)$, detailed in the next section, and numerically maximize $L_2^{(s)}$ as a function of TF.

G. A Simple WH Set

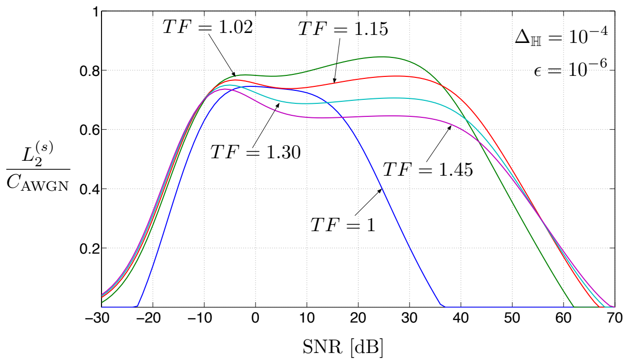
We next construct a family of WH sets $(g, \sqrt{TF}, \sqrt{TF})$ that satisfy Properties 1 and 2 in Section IV-B, and have $g(t)$ real-valued. Take $1 < TF < 2$, let $\zeta \triangleq \sqrt{TF}$, $\delta \triangleq TF - 1$, and $G(f) \triangleq \mathbb{F}\{g(t)\}$. We choose $G(f)$ as the (positive) square root of a raised-cosine pulse:

$$G(f) = \begin{cases} \sqrt{\zeta}, & \text{if} \quad |f| \leq \frac{1-\delta}{2\zeta} \\ \sqrt{\frac{\zeta}{2}(1+S(f))}, & \text{if} \quad \frac{1-\delta}{2\zeta} \leq |f| \leq \frac{1+\delta}{2\zeta} \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

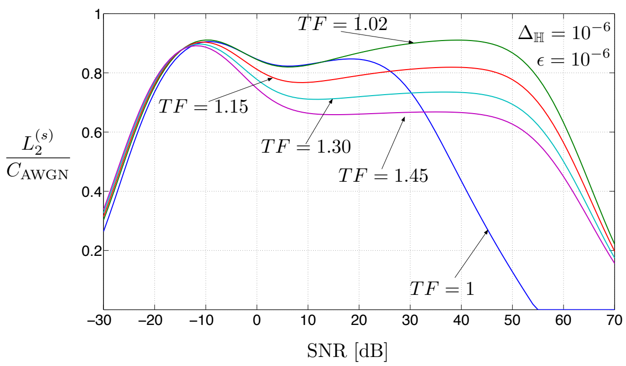
where $S(f) \triangleq \cos\left[\frac{\pi\zeta}{\delta}\left(|f| - \frac{1-\delta}{2\zeta}\right)\right]$. As $(1+\delta)/(2\zeta) = \zeta/2$, the function $G(f)$ is supported on an interval of length $\zeta = \sqrt{TF}$. Furthermore, $G(f)$ has unit norm, is real-valued and even, and satisfies

$$\sum_{n=-\infty}^{\infty} G(f-n/\zeta)G(f-n/\zeta-k\zeta) = \zeta\delta[k].$$

By [26, Thm. 8.7.2], we can therefore conclude that the WH set $(g(t), 1/\sqrt{TF}, 1/\sqrt{TF})$ is a tight WH frame for $\mathcal{L}^2(\mathbb{R})$, and, by duality [27]–[29], the WH set $(g(t), \sqrt{TF}, \sqrt{TF})$ is orthonormal. Finally, it can be shown that $g(t) = \mathcal{O}(1/t^2)$ whenever $TF > 1$.



(a)



(b)

Fig. 3. Lower bounds $L_2^{(s)}$ normalized with respect to the upper bound C_{AWGN}. The bounds are computed for WH sets based on the root-raised-cosine pulse (27), for different values of the grid-parameter product TF. $\Delta_{\mathbb{H}} = 10^{-4}$ in (a) and $\Delta_{\mathbb{H}} = 10^{-6}$ in (b). In both cases, $\epsilon = 10^{-6}$.

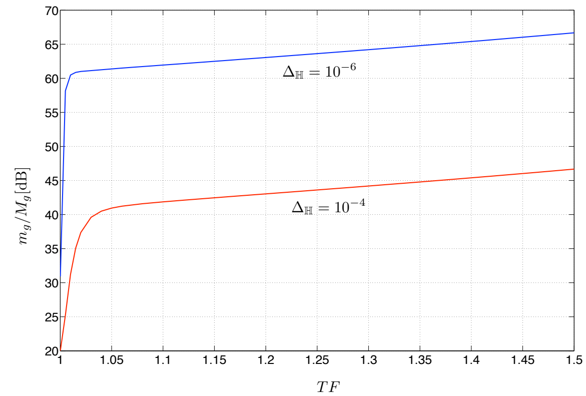


Fig. 4. Trade-off between the product TF, and the signal-to-interference ratio m_g/M_g for the root-raised-cosine WH set constructed in Section IV-G.

V. Finite-SNR Analysis of the Lower Bound $L_2^{(s)}$

We now study the behavior of the lower bound $L_2^{(s)}$ in (26) evaluated for the WH set constructed in the previous section, under the assumption that the underlying channel is underspread according to Definition 1, i.e., $\Delta_{\mathbb{H}} \ll 1$ and $\epsilon \ll 1$. Specifically, we compare $L_2^{(s)}$ to the upper bound C_{AWGN} in (14). To simplify the comparison, we assume throughout this section that $W \gg \nu_0$ (a reasonable assumption for most wireless communication systems of practical interest). Furthermore, in (3) we take $\eta \ll 1$. Under these assumptions, we have

$$C_{\text{AWGN}}(\rho) \approx W\big[\log(1 + (1-\epsilon)\rho) + \epsilon\rho\big]. \tag{28}$$

A. Trade-off between Self-Interference and Signal-Space Dimensions

In Fig. 3, we plot $L_2^{(s)}/C_{\text{AWGN}}$ for $\Delta_{\mathbb{H}} = 10^{-4}$ and for $\Delta_{\mathbb{H}} = 10^{-6}$. In both cases, we take $\epsilon = 10^{-6}$. The different curves correspond to different values of TF. We observe that the choice $TF = 1$ is highly suboptimal. The reason for this suboptimality is the poor time-frequency localization of $g(t)$ this choice entails. In fact, when $TF = 1$, the pulse $g(t)$ reduces to a $(\sin t)/t$ function, which has poor time localization. This, in turn, yields an ambiguity function $A_g(\tau, \nu)$ that is poorly localized in τ, and, hence to a small value for m_g and a large value for M_g, i.e., to small *signal-to-interference ratio* (SIR) m_g/M_g; this leads to a loose lower bound $L_2^{(s)}$ (recall that $L_2^{(s)}$ was obtained by treating self-interference as noise). A value of TF slightly larger than 1 results in a significant improvement in the SIR m_g/M_g (see Fig. 4), which is caused by the improved time localization of $g(t)$. This, in turn, yields an improved lower bound $L_2^{(s)}$ for all SNR values of practical interest, as shown in Fig. 3. A further increase of the product TF seems to be detrimental for all but very high SNR values, where the ratio $L_2^{(s)}/C_{\text{AWGN}}$ is much smaller than 1 anyways. The reason underlying this behavior is as follows: in the regime where $L_2^{(s)}$ is close to C_{AWGN}, the first term on the RHS of (25) dominates the other terms. But in this regime, the first term on the RHS of (25) is essentially linear[7] in $W/(TF)$, which can be interpreted as the number of signal-space dimensions available for communication. The loss of signal-space dimensions incurred by choosing TF much larger than 1 quickly outweighs the SIR gain resulting from improved time-frequency localization. Our numerical results suggest that a value of TF slightly larger than 1 optimally trades signal-space dimensions for SIR maximization. We hasten to add that this trade-off is a consequence of self-interference being treated as (signal-dependent) noise in deriving our lower bound.

B. Sensitivity of Capacity to the Channel Parameters $\Delta_{\mathbb{H}}$ and ϵ

The results presented in Fig. 3 suggest that, for $TF = 1.02$, the lower bound $L_2^{(s)}$ is close to the AWGN-channel capacity upper bound C_{AWGN} over a large range of SNR values. To further quantify this statement, we identify the SNR interval $[\rho_{\min}, \rho_{\max}]$ over which

$$L_2^{(s)}(\rho) \geq 0.75\, C_{\text{AWGN}}(\rho). \tag{29}$$

The corresponding interval end points $\rho_{\min}$ and $\rho_{\max}$, as a function of $\Delta_{\mathbb{H}}$ and ϵ, can easily be obtained numerically and are plotted in Figs. 5 and 6, respectively, for $TF = 1.02$. For the WH set and WSSUS underspread channels considered in this section, we have $\rho_{\min} \in [-25\,\text{dB}, -7\,\text{dB}]$ and $\rho_{\max} \in [30\,\text{dB}, 68\,\text{dB}]$. Hence, the interval $[\rho_{\min}, \rho_{\max}]$ covers all SNR values of practical interest. An analytic characterization of $\rho_{\min}$ and $\rho_{\max}$ seems

[7] Recall that $\rho = P/W$.

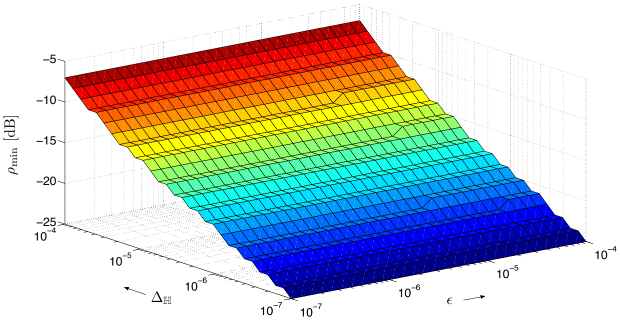


Fig. 5. Minimum SNR value ρ_{min} for which (29) holds, as a function of $\Delta_{\mathbb{H}}$ and ϵ. The lower bound $L_2^{(s)}$ is evaluated for a WH set based on the root-raised-cosine pulse (27); furthermore, $TF = 1.02$.

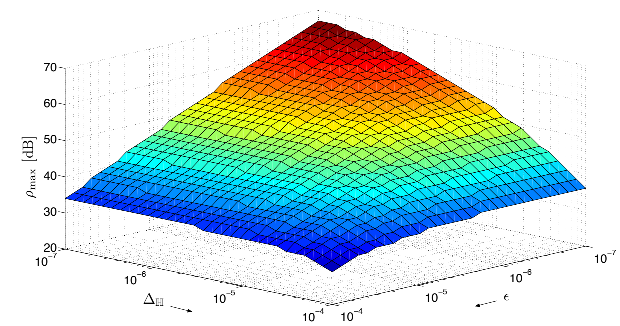


Fig. 6. Maximum SNR value ρ_{max} for which (29) holds, as a function of $\Delta_{\mathbb{H}}$ and ϵ. The lower bound $L_2^{(s)}$ is evaluated for a WH set based on the root-raised-cosine pulse (27); furthermore, $TF = 1.02$.

difficult. Insights on how these two quantities are related to the channel parameters $\Delta_{\mathbb{H}}$ and ϵ can be obtained by the following "back-of-the-envelope" analysis of $L_2^{(s)}$ (for $TF = 1.02$). We first approximate $L_2^{(s)}$ by replacing m_g and M_g (whose dependency on $\Delta_{\mathbb{H}}$ is difficult to characterize analytically) with simpler expressions that are accurate when $\Delta_{\mathbb{H}} \ll 1$. Then, we determine the SNR values for which the resulting approximate lower bound is close to (28). We start by noting that, when $\Delta_{\mathbb{H}} \ll 1$, we can approximate m_g by its first-order Taylor-series expansion around $\Delta_{\mathbb{H}} = 0$. This yields

$$\begin{aligned} m_g &= \min_{(\tau,\nu)\in\widetilde{\mathcal{D}}} |A_g(\tau,\nu)|^2 \\ &\approx 1 - c_m \Delta_{\mathbb{H}} \end{aligned} \tag{30}$$

where $\widetilde{\mathcal{D}} \triangleq [-\sqrt{\Delta_{\mathbb{H}}}/2, \sqrt{\Delta_{\mathbb{H}}}/2] \times [-\sqrt{\Delta_{\mathbb{H}}}/2, \sqrt{\Delta_{\mathbb{H}}}/2]$, and $c_m \triangleq \pi^2(T_0^2 + F_0^2)$ with

$$T_0^2 \triangleq \int t^2 \, |g(t)|^2 \, dt, \quad F_0^2 \triangleq \int f^2 \, |G(f)|^2 \, df.$$

To get (30), we used the Taylor-series expansion of $|A_g(\tau,\nu)|^2$ reported in [30, Sec. 6]. Similarly, for $\Delta_{\mathbb{H}} \ll 1$ we can approximate M_g as follows:

$$\begin{aligned} M_g &= \max_{(\tau,\nu)\in\widetilde{\mathcal{D}}} \sum_{\substack{k=-\infty \\ (k,n)\neq(0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \left| A_g(\tau - k\sqrt{TF}, \nu - n\sqrt{TF}) \right|^2 \\ &\approx c_M \Delta_{\mathbb{H}} \end{aligned} \tag{31}$$

where

$$c_M \triangleq \sum_{\substack{k=-\infty \\ (k,n)\neq(0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \left[|a_{k,n}|^2 + |b_{k,n}|^2 \right] / 4$$

with $a_{k,n}$ and $b_{k,n}$ being the first partial derivatives of $A_g(\tau,\nu)$ (with respect to ν and τ, respectively) calculated at the points $(-k\sqrt{TF}, -n\sqrt{TF})$:

$$\begin{aligned} a_{k,n} &\triangleq -j2\pi \int_t t g(t) g(t + k\sqrt{TF}) e^{j2\pi n\sqrt{TF}t} dt \\ b_{k,n} &\triangleq j2\pi \int_f f G(f - n\sqrt{TF}) G(f) e^{-j2\pi k\sqrt{TF}f} df. \end{aligned}$$

Here, (31) is obtained by performing a Taylor-series expansion of $A_g(\tau - k\sqrt{TF}, \nu - n\sqrt{TF})$ around the point $(\tau,\nu) = (0,0)$ for all k and n, and by using that $g(t)$ is real and even. For our choice of $TF = 1.02$ we have $c_m \approx 25.87$ and $c_M \approx 0.77$. Hence, (30) and (31) suggest that when $\Delta_{\mathbb{H}} \ll 1$, we can approximate m_g by 1 and M_g by $\Delta_{\mathbb{H}}$. On the basis of these two approximations, which are in good agreement with the numerical results reported in Fig. 4, and the assumption that $\epsilon \ll 1$ and $TF = 1.02 \approx 1$, we can approximate the lower bound $L_2^{(s)}$ for all SNR values satisfying $\rho(\Delta_{\mathbb{H}} + \epsilon) \ll 1$ as follows

$$\begin{aligned} L_2^{(s)}(\rho) \approx W \Bigg\{ & \mathbb{E}_h \left[\log\left(1 + \rho |h|^2\right) \right] \\ & - \sqrt{\Delta_{\mathbb{H}}} \log\left(1 + \frac{\rho}{\sqrt{\Delta_{\mathbb{H}}}}\right) \Bigg\}. \end{aligned} \tag{32}$$

The RHS of (32) is close to the AWGN-channel capacity upper bound (apart from the Jensen penalty in the first term) for all SNR values that satisfy $\rho \gg \sqrt{\Delta_{\mathbb{H}}}$. In fact, when $\rho \gg \sqrt{\Delta_{\mathbb{H}}}$ (and $\Delta_{\mathbb{H}} \ll 1$), the second term on the RHS of (32) can be approximated as

$$\begin{aligned} \sqrt{\Delta_{\mathbb{H}}} \log\left(1 + \frac{\rho}{\sqrt{\Delta_{\mathbb{H}}}}\right) &\approx \sqrt{\Delta_{\mathbb{H}}} \log \rho - \sqrt{\Delta_{\mathbb{H}}} \log \sqrt{\Delta_{\mathbb{H}}} \\ &\approx \sqrt{\Delta_{\mathbb{H}}} \log \rho \\ &\ll \log \rho \end{aligned}$$

which implies that, when $\rho \gg \sqrt{\Delta_{\mathbb{H}}}$ (and $\Delta_{\mathbb{H}} \ll 1$), the first term on the RHS of (32) dominates the second term on the RHS of (32).

We can therefore summarize our findings in the following rule of thumb: the capacity of a Rayleigh-fading WSSUS underspread channel with scattering function $C_{\mathbb{H}}(\tau,\nu)$ and parameters $\Delta_{\mathbb{H}}$ and ϵ in Definition 1, is close to C_{AWGN} for all ρ that satisfy $\sqrt{\Delta_{\mathbb{H}}} \ll \rho \ll 1/(\Delta_{\mathbb{H}} + \epsilon)$, independently of whether $C_{\mathbb{H}}(\tau,\nu)$ is compactly supported or not, and independently of its shape. In particular, this implies that capacity essentially grows logarithmically with SNR up to SNR values $\rho \ll 1/(\Delta_{\mathbb{H}} + \epsilon)$. We conclude by noting that the condition $\sqrt{\Delta_{\mathbb{H}}} \ll \rho \ll 1/(\Delta_{\mathbb{H}} + \epsilon)$ holds for all channels and SNR values of practical interest.

VI. Conclusions

We studied the noncoherent capacity of *continuous-time* Rayleigh-fading channels that satisfy the WSSUS and the underspread assumptions. Our main result is a capacity lower bound obtained by (i) discretizing the continuous-time I/O relation and (ii) treating the (signal-dependent) self-interference term in the resulting discretized I/O relation as noise. Discretization is performed by constraining the input signal to lie in the span of an orthonormal WH set and by projecting the output signal onto the same orthonormal set. The resulting lower bound was shown to be close to the AWGN-channel capacity upper bound C_{AWGN} for all SNR values of practical interest, as long as the underlying channel is underspread according to Definition 1. In particular, this result implies that—for all SNR values typically encountered in real-world systems—the capacity of Rayleigh-fading underspread WSSUS channels is *not sensitive* to whether the channel scattering function is compactly supported or not. It also shows that—for all SNR values of practical interest—lack of channel knowledge at the receiver has little impact on the capacity of this class of channels. From a practical point of view, the underspread assumption is not restrictive as the fading channels commonly encountered in wireless communications are, in fact, highly underspread.

On the basis of our capacity lower bound, we derived an information-theoretic criterion for the design of capacity-approaching WH sets to be used in PS-OFDM schemes. This criterion is more fundamental than criteria based on SIR maximization (see [31] and references therein), because it sheds light on the trade-off between self-interference reduction and maximization of the number of signal-space dimensions available for communication. Unfortunately, the corresponding optimization problem is hard to solve, analytically as well as numerically. It turns out, however, that the simple choice of taking $g(t)$ to be a root-raised-cosine pulse and letting the grid-parameter product TF be close to 1 (but strictly larger than 1) yields a lower bound that is close to C_{AWGN} for all SNR values of practical interest. In particular, this result suggests that—when self-interference is treated as (signal-dependent) noise—the maximization of the number of signal-space dimensions available for communication should be privileged over SIR maximization.

An interesting open problem, the solution of which would strengthen our results, is to compute an upper bound on the capacity of (1) by assuming perfect channel state information at the receiver. The main difficulty here lies in dealing with self-interference. In particular, we expect that nonstandard tools from large random matrix theory will be needed for this analysis. Recent results along these lines, for a specific channel model, can be found in [32].

Appendix A
AWGN Capacity Upper Bound

Let $\mathbb{H} \in \mathcal{H}(\tau_0, \nu_0, \epsilon)$. To establish that $C \leq C_{\text{AWGN}}$, where C_{AWGN} is defined in (14), we start by upper-bounding the mutual information on the RHS of (13) as follows:

$$I(y_f(t); x(t)) \leq I(y_f(t); r_f(t)). \tag{33}$$

Here, $r_f(t) \triangleq (\mathbb{B}_{W+2\nu_0} r)(t)$, and the inequality follows by noting that $x(t)$ and $y_f(t)$ are conditionally independent given $r_f(t)$ and by using the data-processing inequality for continuous-time random signals [20, Thm. 1.4]. If we now substitute (33) into (13), we obtain

$$C \leq \lim_{D\to\infty} \frac{1}{D} \sup_{\mathcal{Q}(W,D,\eta,P)} I(y_f(t); r_f(t)). \tag{34}$$

The mutual information in (34) is between the input and the output of a continuous-time band-limited AWGN channel. Hence, we can establish an upper bound on the RHS of (34) by invoking [15, Thm. 2], provided that an inequality, in the spirit of (3), on the energy of the restriction of $r_f(t)$ to a certain time interval can be established. More specifically, we shall show next that the energy of the restriction of $r_f(t)$ to the interval $[-D/2 - \tau_0, D/2 + \tau_0]$, i.e., the energy of $(\mathbb{T}_{D+2\tau_0} r_f)(t)$, is bounded from below by $(1-\eta)(1-\epsilon)\,\mathbb{E}\left[\|x(t)\|^2\right]$. Let

$$x_f^{(\tau,\nu)}(t) \triangleq \mathbb{B}_{W+2\nu_0}\big(x(t-\tau)e^{j2\pi t\nu}\big).$$

Using

$$(\mathbb{T}_{D+2\tau_0} r_f)(t) = \begin{cases} \displaystyle\iint_{\nu\,\tau} S_{\mathbb{H}}(\tau,\nu) x_f^{(\tau,\nu)}(t) d\tau d\nu, & \text{if } |t| \leq D/2+\tau_0 \\ 0, & \text{otherwise} \end{cases}$$

we get

$$\begin{aligned}
&\mathbb{E}\left[\|(\mathbb{T}_{D+2\tau_0} r_f)(t)\|^2\right] \\
&\overset{(a)}{=} \iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu)\,\mathbb{E}\left[\int_{-D/2-\tau_0}^{D/2+\tau_0} \left|x_f^{(\tau,\nu)}(t)\right|^2 dt\right] d\tau d\nu \\
&\overset{(b)}{\geq} \int_{-\nu_0}^{\nu_0}\int_{-\tau_0}^{\tau_0} C_{\mathbb{H}}(\tau,\nu)\,\mathbb{E}\left[\int_{-D/2-\tau_0}^{D/2+\tau_0} \left|x_f^{(\tau,\nu)}(t)\right|^2 dt\right] d\tau d\nu
\end{aligned} \tag{35}$$

where (a) follows from the WSSUS property of $\mathbb{H}$ [see (6)], and (b) follows from the non-negativity of the integrand. Because $x(t)$ is subject to the bandwidth constraint (2) and to the time-concentration constraint (3), we have that, for every $(\tau,\nu) \in [-\tau_0, \tau_0] \times [-\nu_0, \nu_0]$,

$$\mathbb{E}\left[\int_{-D/2-\tau_0}^{D/2+\tau_0} \left|x_f^{(\tau,\nu)}(t)\right|^2 dt\right] \geq (1-\eta)\,\mathbb{E}\left[\|x(t)\|^2\right]. \tag{36}$$

Substituting (36) into (35), we get

$$\begin{aligned}
&\mathbb{E}\left[\|(\mathbb{T}_{D+2\tau_0} r_f)(t)\|^2\right] \\
&\quad\geq (1-\eta)\,\mathbb{E}\left[\|x(t)\|^2\right] \int_{-\nu_0}^{\nu_0}\int_{-\tau_0}^{\tau_0} C_{\mathbb{H}}(\tau,\nu) d\tau d\nu \\
&\quad\geq (1-\eta)(1-\epsilon)\,\mathbb{E}\left[\|x(t)\|^2\right]
\end{aligned} \tag{37}$$

where the last step follows from Definition 1. We now observe that

$$\begin{aligned}\mathbb{E}\left[\|r_f(t)\|^2\right] &\leq \mathbb{E}\left[\|r(t)\|^2\right] \\ &= \iint\limits_{\nu\ \tau} C_{\mathbb{H}}(\tau,\nu)\,\mathbb{E}\left[\|x(t-\tau)e^{j2\pi\nu t}\|^2\right] d\tau d\nu \\ &= \iint\limits_{\nu\ \tau} C_{\mathbb{H}}(\tau,\nu)\,\mathbb{E}\left[\|x(t)\|^2\right] d\tau d\nu \\ &= \mathbb{E}\left[\|x(t)\|^2\right]. \end{aligned} \tag{38}$$

Here, the last step follows from the normalization (7). The inequality (38), combined with (37), yields the following time-concentration inequality for $r_f(t)$ [cf. (3)]

$$\mathbb{E}\left[\|\mathbb{T}_{D+2\tau_0} r_f(t)\|^2\right] \geq (1-\eta)(1-\epsilon)\,\mathbb{E}\left[\|r_f(t)\|^2\right]. \tag{39}$$

To obtain the desired upper bound (14), we now note that every probability measure on $x(t)$ in the set $\mathcal{Q}(W,D,\eta,P)$ induces a probability measure on $r_f(t)$ (through the map $r_f(t) = (\mathbb{B}_{W+2\nu_0}\,\mathbb{H}\,x)(t)$) that satisfies the following constraints [cf. (2)–(4)]:

i) the bandwidth of $r_f(t)$ is no larger than $(W+2\nu_0)$,
ii) $\mathbb{E}\left[\|r_f(t)\|^2\right] \leq DP$, which follows from (38) and (4), and
iii) (39) holds.

Let $\widetilde{\mathcal{Q}}$ be the set of *all* probability measures on $r_f(t)$ satisfying i)–iii). Note that the set of probability measures on $r_f(t)$ induced by probability measures on $x(t)$ in $\mathcal{Q}(W,D,\eta,P)$ through the map $r_f(t) = (\mathbb{B}_{W+2\nu_0}\,\mathbb{H}\,x)(t)$ is contained in $\widetilde{\mathcal{Q}}$, as shown above. This property can be used to upper-bound the RHS of (34) according to

$$\begin{aligned}&\lim_{D\to\infty}\frac{1}{D}\sup_{\mathcal{Q}(W,D,\eta,P)} I(y_f(t);r_f(t)) \\ &\qquad\qquad \leq \lim_{D\to\infty}\frac{1}{D}\sup_{\widetilde{\mathcal{Q}}} I(y_f(t);r_f(t)).\end{aligned}$$

A direct application of [15, Thm. 2] yields (14).

Appendix B
The Input Signal (15) Satisfies (3)

We show that for every orthonormal WH set satisfying Properties 1 and 2 in Section IV-B and for every $\eta > 0$, and $D > 0$, one can find a $K_g > 0$ such that the corresponding $x(t)$ in (15) (with K chosen as specified in Section IV-B3) satisfies (3). To this end, it will turn out convenient to reformulate (3) as follows:

$$\mathbb{E}\left[\|(\mathbb{I}-\mathbb{T}_D)x(t)\|^2\right] \leq \eta\,\mathbb{E}\left[\|x(t)\|^2\right] \tag{40}$$

where $\mathbb{I}$ denotes the identity operator. Let $\mathbf{x}$ be the vector of dimension K_xN_x obtained by stacking the data symbols $x[k,n]$ as in (20). Furthermore, let

$$d[k,n,l,m] \triangleq \int\limits_{|t|>D/2} g_{k,n}(t)g^*_{l,m}(t)dt$$

and define $\mathbf{D}$ to be the square matrix of dimension $K_xN_x \times K_xN_x$ with entries

$$[\mathbf{D}]_{\tilde{n}+\tilde{k}N_x,\tilde{m}+\tilde{l}N_x} \triangleq d[\tilde{k}-K,\tilde{n}-N,\tilde{l}-K,\tilde{m}-N]$$

for $\tilde{k},\tilde{l} = 0,1,\ldots,K_x-1$ and $\tilde{n},\tilde{m} = 0,1,\ldots,N_x-1$. Note that $\mathbf{D}$ is Hermitian, by construction. We have that

$$\mathbb{E}\left[\|(\mathbb{I}-\mathbb{T}_D)x(t)\|^2\right] = \mathbb{E}\left[\mathbf{x}^H\mathbf{D}\mathbf{x}\right] \leq \lambda_{\max}\{\mathbf{D}\}\,\mathbb{E}\left[\|\mathbf{x}\|^2\right].$$

Here, the first equality follows by definition, and the inequality follows by application of the Rayleigh-Ritz theorem [33, Thm. 4.2.2].[8] We next use the Geršgorin disc theorem [33, Cor. 6.1.5] to derive an upper bound on $\lambda_{\max}\{\mathbf{D}\}$ that is explicit in the entries of $\mathbf{D}$:

$$\lambda_{\max}\{\mathbf{D}\} \leq \max_{k\in[-K,K],n\in[-N,N]}\left[\sum_{l=-K}^{K}\sum_{m=-N}^{N}|d[k,n,l,m]|\right]. \tag{41}$$

Each term on the RHS of (41) can be bounded as follows

$$\begin{aligned}|d[k,n,l,m]| &= \left|\int\limits_{|t|>D/2} g_{k,n}(t)g^*_{l,m}(t)dt\right| \\ &\leq \int\limits_{|t|>D/2}\left|g_{k,n}(t)g^*_{l,m}(t)\right| dt \\ &= \int\limits_{|t|>D/2}\left|g(t-kT)g^*(t-lT)\right| dt.\end{aligned}$$

Recall that $D/2 = (K+K_g+1/2)T$, by construction. As, by assumption, $g(t)$ is even and satisfies $g(t) = \mathcal{O}(1/t^{1+\mu})$, there exist constants $\gamma > 0$, and $t_0 > 0$ such that $|g(t)| < \gamma/|t|^{1+\mu}$ for $|t| \geq t_0$. Hence, if we choose K_g such that $K_gT > t_0$, we get[9]

$$\begin{aligned}&|d[k,n,l,m]| \\ &\leq \gamma^2\int\limits_{|t|>(K+K_g+1/2)T}\frac{1}{|t-kT|^{1+\mu}}\frac{1}{|t-lT|^{1+\mu}}dt \\ &= \gamma^2\int\limits_{(K+K_g+1/2)T}^{\infty}\frac{1}{|t-kT|^{1+\mu}}\frac{1}{|t-lT|^{1+\mu}}dt \\ &\quad + \gamma^2\int\limits_{-\infty}^{-(K+K_g+1/2)T}\frac{1}{|t-kT|^{1+\mu}}\frac{1}{|t-lT|^{1+\mu}}dt \\ &\stackrel{(a)}{\leq} \gamma^2\int\limits_{(K+K_g+1/2)T}^{\infty}\frac{1}{|t-KT|^{1+\mu}}\frac{1}{|t-lT|^{1+\mu}}dt \\ &\quad + \gamma^2\int\limits_{-\infty}^{-(K+K_g+1/2)T}\frac{1}{|t+KT|^{1+\mu}}\frac{1}{|t-lT|^{1+\mu}}dt \\ &\stackrel{(b)}{=} \gamma^2\int\limits_{(K_g+1/2)T}^{\infty}\frac{1}{|t|^{1+\mu}}\frac{1}{|t-(l-K)T|^{1+\mu}}dt\end{aligned}$$

[8] With slight abuse of notation, we used $\|\cdot\|$, a symbol which we reserved for the norm in $\mathcal{L}^2(\mathbb{R})$, to denote the Euclidean norm in a finite-dimensional vector space.

[9] If $t_0 > D/2$, we let the guard-interval cover the whole transmission time $[-D/2, D/2]$. In this case (40) is trivially satisfied.

$$+\gamma^2 \int_{-\infty}^{-(K_g+1/2)T} \frac{1}{|t|^{1+\mu}} \frac{1}{|t-(l+K)T|^{1+\mu}} dt. \quad (42)$$

Here, (a) follows by replacing k by K in the first term of the sum and k by $-K$ in the second term of the sum; these substitutions lead to an upper bound; (b) follows by a simple change of variables. Note now that, for $t \geq K_g T$, we have

$$\begin{aligned}\sum_{l=-K}^{K} \frac{1}{|t-(l-K)T|^{1+\mu}} &= \sum_{l=0}^{2K} \frac{1}{|t+lT|^{1+\mu}} \\ &\leq \sum_{l=0}^{2K} \frac{1}{[(K_g+l)T]^{1+\mu}} \\ &\leq \sum_{l=1}^{\infty} \frac{1}{(lT)^{1+\mu}} \\ &\triangleq \gamma' < \infty \end{aligned} \quad (43)$$

where in the last step we used that $\mu > 0$ and, hence, the series converges. Similarly, for $t \leq -K_g T$, we have

$$\begin{aligned}\sum_{l=-K}^{K} \frac{1}{|t-(l+K)T|^{1+\mu}} &= \sum_{l=0}^{2K} \frac{1}{|t-lT|^{1+\mu}} \\ &\leq \sum_{l=0}^{2K} \frac{1}{[(K_g+l)T]^{1+\mu}} \\ &\leq \gamma'. \end{aligned} \quad (44)$$

Inserting (42) into (41) and using (43) and (44), we get

$$\begin{aligned}&\sum_{l=-K}^{K} \sum_{m=-N}^{N} |d[k,n,l,m]| \\ &\leq \sum_{l=-K}^{K} \sum_{m=-N}^{N} \Bigg[\gamma^2 \int_{(K_g+1/2)T}^{\infty} \frac{1}{|t|^{1+\mu}} \frac{1}{|t-(l-K)T|^{1+\mu}} dt \\ &\quad +\gamma^2 \int_{-\infty}^{-(K_g+1/2)T} \frac{1}{|t|^{1+\mu}} \frac{1}{|t-(l+K)T|^{1+\mu}} dt\Bigg] \\ &\leq 2(2N+1)\gamma^2\gamma' \int_{(K_g+1/2)T}^{\infty} \frac{1}{t^{1+\mu}} dt.\end{aligned}$$

To summarize, we have the following upper bound on the RHS of (41):

$$\lambda_{\max}\{\mathbf{D}\} \leq 2(2N+1)\gamma^2\gamma' \int_{(K_g+1/2)T}^{\infty} \frac{1}{t^{1+\mu}} dt. \quad (45)$$

The RHS of (45) can be made arbitrarily small by choosing K_g sufficiently large. In other words, we can find a finite K_g for which the RHS of (45) is smaller than η. This concludes the proof.

Appendix C
Statistical Properties of the Channel Coefficients in (19)

We establish basic properties of the statistics of $h[k,n]$ and $p[l,m,k,n]$ in (19) that will be needed in the proof of the capacity lower bound in Theorem 2. The first property concerns the autocorrelation function of $h[k,n]$. Let the *cross-ambiguity function* of two signals $f(t)$ and $g(t)$ be defined as [34]

$$A_{f,g}(\tau,\nu) \triangleq \int_t f(t)g^*(t-\tau)e^{-j2\pi\nu t}dt \quad (46)$$

and let the ambiguity function of $g(t)$ be defined as $A_g(\tau,\nu) \triangleq A_{g,g}(\tau,\nu)$.[10] The autocorrelation function of $h[k,n]$ turns out to be explicit in the ambiguity function of $g(t)$, as the following calculation reveals:

$$\begin{aligned}&\mathbb{E}[h[k,n]h^*[l,m]] \\ &\quad = \mathbb{E}[\langle \mathbb{H}\, g_{k,n}, g_{k,n}\rangle \langle \mathbb{H}\, g_{l,m}, g_{l,m}\rangle^*] \\ &\quad \stackrel{(a)}{=} \int_\nu\int_\tau C_{\mathbb{H}}(\tau,\nu) A^*_{g_{k,n}}(\tau,\nu) A_{g_{l,m}}(\tau,\nu) d\tau d\nu \\ &\quad \stackrel{(b)}{=} \int_\nu\int_\tau C_{\mathbb{H}}(\tau,\nu)\, |A_g(\tau,\nu)|^2\, e^{j2\pi[(k-l)T\nu-(n-m)F\tau]} d\tau d\nu \\ &\quad \triangleq r[k-l,n-m]. \end{aligned} \quad (47)$$

Here, (a) follows from Property 6 in Appendix D and because $\mathbb{H}$ is WSSUS [see (6)], while (b) follows from Property 5 in Appendix D [see in particular (52)]. As a consequence of (47), we have that $\{h[k,n]\}$ is stationary both in discrete time k and in discrete frequency n. The corresponding power spectral density function is given by

$$c(\varphi,\theta) \triangleq \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} r[k,n] e^{-j2\pi(k\theta-n\varphi)}, \quad |\varphi|, |\theta| \leq 1/2. \quad (48)$$

The Fourier transform relation (48) together with the Poisson summation formula allow us to relate $c(\varphi,\theta)$ to the channel scattering function $C_{\mathbb{H}}(\tau,\nu)$ as follows

$$\begin{aligned} c(\varphi,\theta) &= \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} e^{-j2\pi(k\theta-n\varphi)} \\ &\quad \times \int_\nu\int_\tau C_{\mathbb{H}}(\tau,\nu)\, |A_g(\tau,\nu)|^2\, e^{j2\pi(kT\nu-nF\tau)} d\tau d\nu \\ &= \frac{1}{TF} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{\mathbb{H}}\left(\frac{\varphi-n}{F}, \frac{\theta-k}{T}\right) \\ &\quad \times \left|A_g\left(\frac{\varphi-n}{F}, \frac{\theta-k}{T}\right)\right|^2. \end{aligned} \quad (49)$$

Another property we shall often use is

$$\begin{aligned} r[0,0] &= \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} c(\varphi,\theta) d\varphi d\theta \\ &= \int_\nu\int_\tau C_{\mathbb{H}}(\tau,\nu)\, |A_g(\tau,\nu)|^2\, d\tau d\nu \\ &\leq 1 \end{aligned}$$

[10] Basic results on the ambiguity function that will be needed in our analysis are reviewed in Appendix D.

where the last step follows from Property 3 in Appendix D, from the assumption that $g(t)$ has unit norm, and from the normalization (7).

A characterization of the autocorrelation function of $p[l,m,k,n]$ is possible, but not particularly insightful. For our purposes, it will be sufficient to study the variance of $p[l,m,k,n]$. As $p[l,m,k,n]$ has zero mean (see Section IV-C), its variance is given by

$$\begin{aligned}
&\mathbb{E}\Big[|p[l,m,k,n]|^2\Big] \\
&= \mathbb{E}\Big[|\langle \mathbb{H}\, g_{l,m}, g_{k,n}\rangle|^2\Big] \\
&\overset{(a)}{=} \iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu)\left|A_{g_{k,n},g_{l,m}}(\tau,\nu)\right|^2 d\tau d\nu \\
&\overset{(b)}{=} \iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu)\left|A_g(\tau+(l-k)T,\nu+(m-n)F)\right|^2 d\tau d\nu \\
&\triangleq \sigma_p^2[l-k,m-n] \qquad (50)
\end{aligned}$$

where in (a) we used Property 6 in Appendix D together with the WSSUS property of $\mathbb{H}$, and (b) follows from Property 5 in Appendix D.

Appendix D
Properties of the Ambiguity Function

We summarize properties of the (cross-)ambiguity function defined in (46) that are needed for our analysis.

Property 3: For every function $g(t) \in \mathcal{L}^2(\mathbb{R})$, the ambiguity surface $|A_g(\tau,\nu)|^2$ attains its maximum at the origin, i.e., $|A_g(\tau,\nu)|^2 \le [A_g(0,0)]^2 = \|g(t)\|^4$, for all τ and ν. This property, as shown in [18, Lem. 4.2.1], follows directly from the Cauchy-Schwarz inequality.

Property 4: Let $g(t) \in \mathcal{L}^2(\mathbb{R})$ and $e(t) = \sqrt{\beta} g(\beta t)$. Then

$$\begin{aligned}
A_e(\tau,\nu) &= \int_t e(t)e^*(t-\tau)e^{-j2\pi\nu t}dt \\
&= \beta\int_t g(\beta t)g^*(\beta(t-\tau))e^{-j2\pi\nu t}dt \\
&\overset{(a)}{=} \int_z g(z)g^*(z-\beta\tau)e^{-j2\pi\nu z/\beta}dz \\
&= A_g\left(\beta\tau,\frac{\nu}{\beta}\right)
\end{aligned}$$

where (a) follows from the change of variables $z = \beta t$.

Property 5: The cross-ambiguity function between the two time- and frequency-shifted versions $g_{(\alpha,\beta)}(t) \triangleq g(t-\alpha)e^{j2\pi\beta t}$ and $g_{(\alpha',\beta')}(t) \triangleq g(t-\alpha')e^{j2\pi\beta' t}$ of $g(t) \in \mathcal{L}^2(\mathbb{R})$ is given by

$$\begin{aligned}
&A_{g_{(\alpha,\beta)},g_{(\alpha',\beta')}}(\tau,\nu) \\
&= \int_t g(t-\alpha)e^{j2\pi\beta t}g^*(t-\alpha'-\tau)e^{-j2\pi\beta'(t-\tau)}e^{-j2\pi\nu t}dt \\
&\overset{(a)}{=} e^{j2\pi\beta'\tau}e^{-j2\pi(\nu+\beta'-\beta)\alpha} \\
&\quad\times\int_{t'} g(t')g^*(t'-(\alpha'-\alpha)-\tau)e^{-j2\pi(\nu+\beta'-\beta)t'}dt' \\
&= A_g(\tau+\alpha'-\alpha,\nu+\beta'-\beta) \\
&\quad\times e^{-j2\pi(\nu\alpha-\tau\beta')}e^{-j2\pi(\beta'-\beta)\alpha}. \qquad (51)
\end{aligned}$$

Here, (a) follows from the change of variables $t' = t-\alpha$. As a direct consequence of (51), we have that

$$A_{g_{(\alpha,\beta)}}(\tau,\nu) = A_g(\tau,\nu)e^{-j2\pi(\nu\alpha-\tau\beta)}. \qquad (52)$$

Property 6: Let $S_{\mathbb{H}}(\tau,\nu)$ be the delay-Doppler spreading function of the channel $\mathbb{H}$. Then, for $g(t) \in \mathcal{L}^2(\mathbb{R})$, and $f(t) \in \mathcal{L}^2(\mathbb{R})$, we have

$$\begin{aligned}
\langle \mathbb{H}\, g, f\rangle &= \iiint_{t\,\nu\,\tau} S_{\mathbb{H}}(\tau,\nu)g(t-\tau)e^{j2\pi t\nu}f^*(t)d\tau d\nu dt \\
&= \iint_{\nu\,\tau} S_{\mathbb{H}}(\tau,\nu)\left[\int_t f(t)g^*(t-\tau)e^{-j2\pi t\nu}dt\right]^* d\tau d\nu \\
&= \iint_{\nu\,\tau} S_{\mathbb{H}}(\tau,\nu)A^*_{f,g}(\tau,\nu)d\tau d\nu.
\end{aligned}$$

Appendix E
Proof of Theorem 2

We obtain a lower bound on C_d in (22) by evaluating the mutual information $I(\mathbf{y};\mathbf{x})$ for a specific input distribution. In particular, we take $x[k,n]$ to be i.i.d. JPG with zero mean and variance $TF\rho$ for all k,n, so that the average-power constraint (16) is satisfied. The corresponding input vector $\mathbf{x}$ is independent of $\mathbf{h}$, $\mathbf{P}$, and $\mathbf{w}$. We use the chain rule for mutual information and the fact that mutual information is nonnegative to obtain the following standard lower bound:

$$\begin{aligned}
I(\mathbf{y};\mathbf{x}) &= I(\mathbf{y};\mathbf{x},\mathbf{h}) - I(\mathbf{y};\mathbf{h}\,|\,\mathbf{x}) \\
&= I(\mathbf{y};\mathbf{h}) + I(\mathbf{y};\mathbf{x}\,|\,\mathbf{h}) - I(\mathbf{y};\mathbf{h}\,|\,\mathbf{x}) \\
&\ge I(\mathbf{y};\mathbf{x}\,|\,\mathbf{h}) - I(\mathbf{y};\mathbf{h}\,|\,\mathbf{x}). \qquad (53)
\end{aligned}$$

The first term on the RHS of (53) can be interpreted as a "coherent" mutual information term (i.e., the mutual information between $\mathbf{x}$ and $\mathbf{y}$ under perfect knowledge of the channel realization at the receiver), while the second term can be interpreted as quantifying the rate penalty due to the lack of channel knowledge [1].

1) The "Coherent" Term: The first term can be further lower-bounded as follows

$$\begin{aligned}
&I(\mathbf{y};\mathbf{x}\,|\,\mathbf{h}) \\
&= \mathsf{h}(\mathbf{x}\,|\,\mathbf{h}) - \mathsf{h}(\mathbf{x}\,|\,\mathbf{h},\mathbf{y}) \\
&\overset{(a)}{=} \mathsf{h}(\mathbf{x}) - \mathsf{h}(\mathbf{x}\,|\,\mathbf{h},\mathbf{y}) \\
&\overset{(b)}{=} \sum_{k=-K}^{K}\sum_{n=-N}^{N}\Big[\mathsf{h}\Big(x[k,n]\,|\,\mathbf{x}_{\text{prec}}^{(k,n)}\Big) \\
&\qquad -\mathsf{h}\Big(x[k,n]\,|\,\mathbf{h},\mathbf{y},\mathbf{x}_{\text{prec}}^{(k,n)}\Big)\Big] \\
&\overset{(c)}{=} \sum_{k=-K}^{K}\sum_{n=-N}^{N}\Big[\mathsf{h}(x[k,n]) - \mathsf{h}\Big(x[k,n]\,|\,\mathbf{h},\mathbf{y},\mathbf{x}_{\text{prec}}^{(k,n)}\Big)\Big] \\
&\overset{(d)}{\ge} \sum_{k=-K}^{K}\sum_{n=-N}^{N}\left[\mathsf{h}(x[k,n]) - \mathsf{h}(x[k,n]\,|\,h[k,n],y[k,n])\right] \\
&= \sum_{k=-K}^{K}\sum_{n=-N}^{N} I(y[k,n];x[k,n]\,|\,h[k,n]).
\end{aligned}$$

Here, (a) follows because $\mathbf{x}$ and $\mathbf{h}$ are independent; (b) is a consequence of the chain rule for differential entropy [$\mathbf{x}_{\text{prec}}^{(k,n)}$ denotes the vector containing all entries of $\mathbf{x}$ up to and including the one before $x[k,n]$]. Next, (c) holds because $\mathbf{x}$ has i.i.d. entries, and (d) follows because conditioning reduces entropy.

We next seek a lower bound on $I(y[k,n]; x[k,n] \,|\, h[k,n])$ that does not depend on $[k,n]$. Let $\widetilde{w}[k,n]$ be the sum of the self-interference and noise terms in $y[k,n]$ [see (19)], i.e.,

$$\widetilde{w}[k,n] \triangleq \sum_{\substack{l=-K \\ (l,m)\neq(k,n)}}^{K} \sum_{m=-N}^{N} p[l,m,k,n]x[l,m] + w[k,n].$$

Furthermore, let $\widetilde{w}_G[k,n]$ be a proper Gaussian random variable that has the same variance as $\widetilde{w}[k,n]$. It follows from [35, Lem. II.2] that $I(y[k,n]; x[k,n] \,|\, h[k,n])$ does not increase if we replace $w[k,n]$ by $\widetilde{w}_G[k,n]$. Hence,

$$\begin{aligned}
&I(y[k,n]; x[k,n] \,|\, h[k,n]) \\
&\quad = I(h[k,n]x[k,n] + \widetilde{w}[k,n]; x[k,n] \,|\, h[k,n]) \\
&\quad \geq I(h[k,n]x[k,n] + \widetilde{w}_G[k,n]; x[k,n] \,|\, h[k,n]) \\
&\quad \overset{(a)}{=} \mathbb{E}_{h[k,n]}\left[\log\left(1 + TF\rho\frac{|h[k,n]|^2}{\mathbb{E}\left[|\widetilde{w}_G[k,n]|^2\right]}\right)\right] \\
&\quad \overset{(b)}{=} \mathbb{E}_h\left[\log\left(1 + \frac{r[0,0]TF\rho\,|h|^2}{\mathbb{E}\left[|\widetilde{w}_G[k,n]|^2\right]}\right)\right] \qquad (54)
\end{aligned}$$

where (a) follows because $x[k,n] \sim \mathcal{CN}(0, TF\rho)$, and (b) follows because $h[k,n] \sim \mathcal{CN}(0, r[0,0])$ [see (47)], so that we can replace $h[k,n]$ by $r[0,0]h$, where $h \sim \mathcal{CN}(0,1)$. As the input symbols $x[k,n]$ are independent, and as $\mathbb{E}\left[|p[l,m,k,n]|^2\right] = \sigma_p^2[l-k, m-n]$ [see (50)], we have that

$$\begin{aligned}
\mathbb{E}\left[|\widetilde{w}_G[k,n]|^2\right] &= \mathbb{E}\left[|\widetilde{w}[k,n]|^2\right] \\
&= 1 + TF\rho \sum_{\substack{l=-K \\ (l,m)\neq(k,n)}}^{K} \sum_{m=-N}^{N} \sigma_p^2[l-k, m-n].
\end{aligned} \qquad (55)$$

The nonnegativity of $\sigma_p^2[k,n]$ allows us to upper-bound (55) as follows

$$\begin{aligned}
\mathbb{E}\left[|\widetilde{w}_G[k,n]|^2\right] &\leq 1 + TF\rho \sum_{\substack{l=-\infty \\ (l,m)\neq(k,n)}}^{\infty} \sum_{m=-\infty}^{\infty} \sigma_p^2[l-k, m-n] \\
&= 1 + TF\rho \sum_{\substack{l=-\infty \\ (l,m)\neq(0,0)}}^{\infty} \sum_{m=-\infty}^{\infty} \sigma_p^2[l,m] \\
&= 1 + TF\rho\,\sigma_I^2 \qquad (56)
\end{aligned}$$

where we set

$$\sigma_I^2 \triangleq \sum_{\substack{l=-\infty \\ (l,m)\neq(0,0)}}^{\infty} \sum_{m=-\infty}^{\infty} \sigma_p^2[l,m]. \qquad (57)$$

If we now substitute (56) into (54), we obtain

$$I(y[k,n]; x[k,n] \,|\, h[k,n]) \geq \mathbb{E}_h\left[\log\left(1 + \frac{r[0,0]TF\rho\,|h|^2}{1 + TF\rho\,\sigma_I^2}\right)\right]$$

and, consequently,

$$I(\mathbf{y}; \mathbf{x} \,|\, \mathbf{h}) \geq K_x N_x\, \mathbb{E}_h\left[\log\left(1 + \frac{r[0,0]TF\rho\,|h|^2}{1 + TF\rho\,\sigma_I^2}\right)\right]. \qquad (58)$$

2) The Penalty Term: We next seek an upper bound on the penalty term $I(\mathbf{y}; \mathbf{h} \,|\, \mathbf{x})$ in (53). The main difficulty lies in the self-interference term being signal-dependent. Our approach is to split $\mathbf{y}$ into a self-interference-free part and a self-interference-only part. Specifically, let $\mathbf{w}_1 \sim \mathcal{CN}(\mathbf{0}, \alpha\mathbf{I})$ and $\mathbf{w}_2 \sim \mathcal{CN}(\mathbf{0}, (1-\alpha)\mathbf{I})$, where $0 < \alpha < 1$, be two $K_x N_x$-dimensional independent JPG vectors.[11] Then,

$$\begin{aligned}
\mathbf{y} &= \mathbf{h} \odot \mathbf{x} + \mathbf{P}\mathbf{x} + \mathbf{w} \\
&= \underbrace{\mathbf{h} \odot \mathbf{x} + \mathbf{w}_1}_{\triangleq \mathbf{y}_1} + \underbrace{\mathbf{P}\mathbf{x} + \mathbf{w}_2}_{\triangleq \mathbf{y}_2}.
\end{aligned}$$

By the data-processing inequality [19, Thm. 2.8.1] and the chain rule for mutual information, we have that

$$\begin{aligned}
I(\mathbf{y}; \mathbf{h} \,|\, \mathbf{x}) &\leq I(\mathbf{y}_1, \mathbf{y}_2; \mathbf{h} \,|\, \mathbf{x}) \\
&= I(\mathbf{y}_1; \mathbf{h} \,|\, \mathbf{x}) + I(\mathbf{y}_2; \mathbf{h} \,|\, \mathbf{x}, \mathbf{y}_1). \qquad (59)
\end{aligned}$$

As $\mathbf{h}$ is JPG, the first term on the RHS of (59) can be bounded as follows:

$$\begin{aligned}
&I(\mathbf{y}_1; \mathbf{h} \,|\, \mathbf{x}) \\
&\quad = I(\mathbf{h} \odot \mathbf{x} + \mathbf{w}_1; \mathbf{h} \,|\, \mathbf{x}) \\
&\quad = \mathbb{E}_{\mathbf{x}}\left[\log\det\left(\mathbf{I} + \frac{1}{\alpha}\operatorname{diag}\{\mathbf{x}\}\,\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]\operatorname{diag}\{\mathbf{x}^H\}\right)\right] \\
&\quad \overset{(a)}{=} \mathbb{E}_{\mathbf{x}}\left[\log\det\left(\mathbf{I} + \frac{1}{\alpha}\operatorname{diag}\{\mathbf{x}^H\}\operatorname{diag}\{\mathbf{x}\}\,\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]\right)\right] \\
&\quad \overset{(b)}{\leq} \log\det\left(\mathbf{I} + \frac{TF\rho}{\alpha}\,\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]\right). \qquad (60)
\end{aligned}$$

Here, (a) follows from the identity $\det\left(\mathbf{I} + \mathbf{A}\mathbf{B}^H\right) = \det\left(\mathbf{I} + \mathbf{B}^H\mathbf{A}\right)$ for any pair of matrices $\mathbf{A}$ and $\mathbf{B}$ of appropriate dimensions [33, Thm. 1.3.20] and (b) is a consequence of Jensen's inequality.

For the second term on the RHS of (59) we note that

$$\begin{aligned}
I(\mathbf{y}_2; \mathbf{h} \,|\, \mathbf{x}, \mathbf{y}_1) &= \mathsf{h}(\mathbf{y}_2 \,|\, \mathbf{x}, \mathbf{y}_1) - \mathsf{h}(\mathbf{y}_2 \,|\, \mathbf{x}, \mathbf{y}_1, \mathbf{h}) \\
&\overset{(a)}{=} \mathsf{h}(\mathbf{y}_2 \,|\, \mathbf{x}, \mathbf{y}_1) - \mathsf{h}(\mathbf{y}_2 \,|\, \mathbf{x}, \mathbf{h}) \\
&\overset{(b)}{\leq} \mathsf{h}(\mathbf{y}_2 \,|\, \mathbf{x}) - \mathsf{h}(\mathbf{y}_2 \,|\, \mathbf{x}, \mathbf{h}, \mathbf{P}) \\
&\overset{(c)}{=} \mathsf{h}(\mathbf{y}_2 \,|\, \mathbf{x}) - \mathsf{h}(\mathbf{y}_2 \,|\, \mathbf{x}, \mathbf{P}) \\
&= I(\mathbf{y}_2; \mathbf{P} \,|\, \mathbf{x}).
\end{aligned}$$

Here, (a) holds because $\mathbf{y}_1$ and $\mathbf{y}_2$ are conditionally independent given $\mathbf{x}$ and $\mathbf{h}$, in (b) we used twice that conditioning reduces entropy, and (c) follows because $\mathbf{y}_2$ and $\mathbf{h}$ are conditionally independent given $\mathbf{P}$.

[11] The role of α will become clear later.

Let $\mathbf{K}(\mathbf{x}) \triangleq \mathbb{E}_{\mathbf{P}}\left[\mathbf{P}\mathbf{x}\mathbf{x}^H\mathbf{P}^H\right]$ be the $K_xN_x \times K_xN_x$ conditional covariance matrix of the vector $\mathbf{Px}$ given $\mathbf{x}$. We next upper-bound $I(\mathbf{y}_2;\mathbf{P}\,|\,\mathbf{x})$ as follows:

$$\begin{aligned}
&I(\mathbf{y}_2;\mathbf{P}\,|\,\mathbf{x})\\
&= I(\mathbf{Px}+\mathbf{w}_2;\mathbf{P}\,|\,\mathbf{x})\\
&\overset{(a)}{=} \mathbb{E}_{\mathbf{x}}\left[\log\det\left(\mathbf{I}+\frac{1}{1-\alpha}\mathbf{K}(\mathbf{x})\right)\right]\\
&\overset{(b)}{\leq} \sum_{\tilde{k}=0}^{K_x-1}\sum_{\tilde{n}=0}^{N_x-1}\mathbb{E}_{\mathbf{x}}\left[\log\left(1+\frac{1}{1-\alpha}\left[\mathbf{K}(\mathbf{x})\right]_{(\tilde{n}+\tilde{k}N_x,\tilde{n}+\tilde{k}N_x)}\right)\right]\\
&\overset{(c)}{\leq} \sum_{\tilde{k}=0}^{K_x-1}\sum_{\tilde{n}=0}^{N_x-1}\log\left(1+\frac{1}{1-\alpha}\,\mathbb{E}_{\mathbf{x}}\left[\left[\mathbf{K}(\mathbf{x})\right]_{(\tilde{n}+\tilde{k}N_x,\tilde{n}+\tilde{k}N_x)}\right]\right)
\end{aligned}$$

where (a) follows because, given $\mathbf{x}$, the vector $\mathbf{Px}$ is JPG, in (b) we used Hadamard's inequality, and (c) follows from Jensen's inequality. As the entries of $\mathbf{x}$ are i.i.d. with zero mean, we have that

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x}}\left[\left[\mathbf{K}(\mathbf{x})\right]_{(\tilde{n}+\tilde{k}N_x,\tilde{n}+\tilde{k}N_x)}\right]\\
&\quad= \mathbb{E}_{\mathbf{x}}\left[\mathbb{E}_{\mathbf{P}}\left[\left[\mathbf{P}\mathbf{x}\mathbf{x}^H\mathbf{P}^H\right]_{(\tilde{n}+\tilde{k}N_x,\tilde{n}+\tilde{k}N_x)}\right]\right]\\
&\quad= TF\rho \sum_{\substack{l=-K \\ (l,m)\neq(\tilde{k}-K,\tilde{n}-N)}}^{K}\sum_{m=-N}^{N}\sigma_p^2[l-\tilde{k}+K, m-\tilde{n}+N]\\
&\quad\leq TF\rho\,\sigma_I^2
\end{aligned}$$

where σ_I^2 was defined in (57). Hence,

$$I(\mathbf{y}_2;\mathbf{P}\,|\,\mathbf{x}) \leq K_xN_x\log\left(1+\frac{TF\rho}{1-\alpha}\sigma_I^2\right). \tag{61}$$

If we now substitute (60) and (61) into (59), we obtain

$$\begin{aligned}
I(\mathbf{y};\mathbf{h}\,|\,\mathbf{x}) \leq{}& \log\det\left(\mathbf{I}+\frac{TF\rho}{\alpha}\,\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]\right)\\
&+K_xN_x\log\left(1+\frac{TF\rho}{1-\alpha}\sigma_I^2\right).
\end{aligned} \tag{62}$$

3) Putting the Pieces Together: We substitute (58) and (62) into (53) and then (53) into (22) to get the following lower bound on capacity:

$$\begin{aligned}
C(\rho) \geq{}& \frac{N_x}{T}\,\mathbb{E}_h\left[\log\left(1+\frac{r[0,0]TF\rho\,|h|^2}{1+TF\rho\,\sigma_I^2}\right)\right]\\
&-\left\{\lim_{K_x\to\infty}\frac{1}{(K_x+2K_g)T}\log\det\left(\mathbf{I}+\frac{TF\rho}{\alpha}\,\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]\right)\right.\\
&\left.+\frac{N_x}{T}\log\left(1+\frac{TF\rho}{1-\alpha}\sigma_I^2\right)\right\}.
\end{aligned}$$

Furthermore, as the bound holds for all $\alpha\in(0,1)$, we can tighten it according to

$$\begin{aligned}
C(\rho) \geq{}& \frac{N_x}{T}\,\mathbb{E}_h\left[\log\left(1+\frac{r[0,0]TF\rho\,|h|^2}{1+TF\rho\,\sigma_I^2}\right)\right]\\
&-\inf_{0<\alpha<1}\left\{\lim_{K_x\to\infty}\frac{1}{(K_x+2K_g)T}\log\det\left(\mathbf{I}+\frac{TF\rho}{\alpha}\,\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]\right)\right.\\
&\left.+\frac{N_x}{T}\log\left(1+\frac{TF\rho}{1-\alpha}\sigma_I^2\right)\right\}.
\end{aligned} \tag{63}$$

By direct application of [36, Thm. 3.4], an extension of Szegő's theorem (on the asymptotic eigenvalue distribution of Toeplitz matrices) to two-level Toeplitz matrices, we obtain

$$\begin{aligned}
&\lim_{K_x\to\infty}\frac{1}{(K_x+2K_g)T}\log\det\left(\mathbf{I}+\frac{TF\rho}{\alpha}\,\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]\right)\\
&\qquad= \frac{1}{T}\int_{-1/2}^{1/2}\log\det\left(\mathbf{I}+\frac{TF\rho}{\alpha}\mathbf{C}(\theta)\right)d\theta.
\end{aligned}$$

Substituting this expression into (63) and noting that [see (50)]

$$\begin{aligned}
\sigma_I^2 &= \sum_{\substack{k=-\infty \\ (k,n)\neq(0,0)}}^{\infty}\sum_{n=-\infty}^{\infty}\sigma_p^2[k,n]\\
&= \sum_{\substack{k=-\infty \\ (k,n)\neq(0,0)}}^{\infty}\sum_{n=-\infty}^{\infty}\iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu)\,|A_g(\tau-kT,\nu-nF)|^2\,d\tau d\nu
\end{aligned} \tag{64}$$

completes the proof.

APPENDIX F
PROOF OF COROLLARY 3

To prove the corollary we further bound each term in (24) separately.

a) The "$\log\det$" *term:* We start with an upper bound on the "$\log\det$" term on the RHS of (24). The matrix $\mathbf{C}(\theta)$ is Toeplitz [see (23)]. Hence, the entries on the main diagonal of $\mathbf{C}(\theta)$ are all equal. Let $c_0(\theta)$ denote one such entry; then

$$\begin{aligned}
c_0(\theta) &\overset{(a)}{=} \sum_{k=-\infty}^{\infty} r[k,0]e^{-j2\pi k\theta}\\
&\overset{(b)}{=} \int_{-1/2}^{1/2} c(\varphi,\theta)d\varphi, \quad |\theta|\leq 1/2.
\end{aligned} \tag{65}$$

Here, (a) follows from (23) and (47); (b) follows from (48) and by applying the Poisson summation formula. By Hadamard's inequality, we can upper-bound the "$\log\det$" term on the RHS of (24) as follows:

$$\frac{1}{T}\int_{-1/2}^{1/2}\log\det\left(\mathbf{I}+\frac{TF\rho}{\alpha}\mathbf{C}(\theta)\right)d\theta \le \frac{N_x}{T}\int_{-1/2}^{1/2}\log\left(1+\frac{TF\rho}{\alpha}c_0(\theta)\right)d\theta. \quad (66)$$

Let $\mathcal{B} \triangleq \{\theta : |\theta| < \nu_0 T\}$ and $\bar{\mathcal{B}} \triangleq \{\theta : \nu_0 T < |\theta| < 1/2\}$. We next use that $\nu_0 T < 1/2$, by assumption, to first split the integral into two parts and then use Jensen's inequality on both terms to obtain

$$\begin{aligned}
&\frac{N_x}{T}\int_{-1/2}^{1/2}\log\left(1+\frac{TF\rho}{\alpha}c_0(\theta)\right)d\theta \\
&= \frac{N_x}{T}\int_{\theta\in\mathcal{B}}\log\left(1+\frac{TF\rho}{\alpha}c_0(\theta)\right)d\theta \\
&\quad + \frac{N_x}{T}\int_{\theta\in\bar{\mathcal{B}}}\log\left(1+\frac{TF\rho}{\alpha}c_0(\theta)\right)d\theta \\
&\le 2\nu_0 N_x \log\left(1+\frac{F\rho}{2\nu_0\alpha}\int_{\theta\in\mathcal{B}}c_0(\theta)d\theta\right) + \frac{N_x}{T}(1-2\nu_0 T) \\
&\quad \times \log\left(1+\frac{TF\rho}{(1-2\nu_0 T)\alpha}\int_{\theta\in\bar{\mathcal{B}}}c_0(\theta)d\theta\right). \qquad (67)
\end{aligned}$$

Let $F(\tau,\nu) \triangleq C_{\mathbb{H}}(\tau,\nu)\,|A_g(\tau,\nu)|^2$. Note that

$$\begin{aligned}
&\int_{\theta\in\mathcal{B}} c_0(\theta)d\theta \\
&\overset{(a)}{=} \int_{-1/2}^{1/2}\int_{\theta\in\mathcal{B}} c(\varphi,\theta)d\theta d\varphi \\
&\overset{(b)}{=} \int_{-1/2}^{1/2}\int_{\theta\in\mathcal{B}} \frac{1}{TF}\sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} F\left(\frac{\varphi-n}{F},\frac{\theta-k}{T}\right)d\theta d\varphi \\
&\overset{(c)}{\le} \frac{1}{TF}\int_{-1/2}^{1/2}\int_{\theta\in\mathcal{B}} \sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} C_{\mathbb{H}}\left(\frac{\varphi-n}{F},\frac{\theta-k}{T}\right)d\theta d\varphi \\
&\le \frac{1}{TF}\int_{-1/2}^{1/2}\int_{-1/2}^{1/2} \sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} C_{\mathbb{H}}\left(\frac{\varphi-n}{F},\frac{\theta-k}{T}\right)d\theta d\varphi \\
&= \iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu)d\tau d\nu = 1 \qquad (68)
\end{aligned}$$

where (a) follows from (65), (b) follows from (49), and (c) follows from Property 3 in Appendix D. Similar steps lead to

$$\begin{aligned}
&\int_{\theta\in\bar{\mathcal{B}}} c_0(\theta)d\theta \\
&\le \frac{1}{TF}\int_{-1/2}^{1/2}\int_{\theta\in\bar{\mathcal{B}}} \sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} C_{\mathbb{H}}\left(\frac{\varphi-n}{F},\frac{\theta-k}{T}\right)d\theta d\varphi \\
&\le \int_{|\nu|\ge\nu_0}\int_{\tau} C_{\mathbb{H}}(\tau,\nu)d\tau d\nu \le \epsilon \qquad (69)
\end{aligned}$$

where the last step follows from (9). If we now substitute (68) and (69) into (67), insert the result into (66), set $\widetilde{\Delta}_{\mathbb{H}} = 2\nu_0 T$, and use $W = N_x F$, we get

$$\begin{aligned}
&\frac{1}{T}\int_{-1/2}^{1/2}\log\det\left(\mathbf{I}+\frac{TF\rho}{\alpha}\mathbf{C}(\theta)\right)d\theta \\
&\le \frac{W\widetilde{\Delta}_{\mathbb{H}}}{TF}\log\left(1+\frac{TF\rho}{\alpha\widetilde{\Delta}_{\mathbb{H}}}\right) \\
&\quad + \frac{W}{TF}(1-\widetilde{\Delta}_{\mathbb{H}})\log\left(1+\frac{TF\rho\epsilon}{\alpha(1-\widetilde{\Delta}_{\mathbb{H}})}\right). \qquad (70)
\end{aligned}$$

b) Bounds on $r[0,0]$ and on σ_I^2: To further lower-bound the RHS of (24), we next derive a lower bound on $r[0,0]$ and an upper bound on σ_I^2; the resulting bounds are explicit in $\Delta_{\mathbb{H}}$ and ϵ, and in the ambiguity function of $g(t)$.

Let $\mathcal{D} = \{(\tau,\nu) \in [-\tau_0,\tau_0]\times[-\nu_0,\nu_0]\}$ be the rectangular area in the delay-Doppler plane that supports at least $1-\epsilon$ of the volume of $C_{\mathbb{H}}(\tau,\nu)$ according to (9). The following chain of inequalities holds:

$$\begin{aligned}
r[0,0] &= \iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu)\,|A_g(\tau,\nu)|^2\, d\tau d\nu \\
&\ge \iint_{\mathcal{D}} C_{\mathbb{H}}(\tau,\nu)\,|A_g(\tau,\nu)|^2\, d\tau d\nu \\
&\ge \min_{(\tau,\nu)\in\mathcal{D}}\left\{|A_g(\tau,\nu)|^2\right\}\iint_{\mathcal{D}} C_{\mathbb{H}}(\tau,\nu)d\tau d\nu \\
&\ge \min_{(\tau,\nu)\in\mathcal{D}}\left\{|A_g(\tau,\nu)|^2\right\}(1-\epsilon). \qquad (71)
\end{aligned}$$

We now seek an upper bound on σ_I^2. Let

$$M(\tau,\nu) = \sum_{\substack{k=-\infty \\ (k,n)\ne(0,0)}}^{\infty}\sum_{n=-\infty}^{\infty} |A_g(\tau-kT,\nu-nF)|^2$$

and note that

$$\begin{aligned}
M(\tau,\nu) &\le \sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} |A_g(\tau-kT,\nu-nF)|^2 \\
&= \sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \left|\langle g(t+\tau)e^{-j2\pi\nu t}, g_{k,n}(t)\rangle\right|^2 \\
&\overset{(a)}{\le} \|g(t)\|^2 = 1 \qquad (72)
\end{aligned}$$

where (a) follows from Bessel's inequality [37, Thm. 3.4-6]. The following chain of inequalities holds:

$$\begin{aligned}
\sigma_I^2 &= \iint_{\nu\,\tau} C_{\mathbb{H}}(\tau,\nu) M(\tau,\nu) d\tau d\nu \\
&= \iint_{\mathcal{D}} C_{\mathbb{H}}(\tau,\nu) M(\tau,\nu) d\tau d\nu + \iint_{\mathbb{R}^2\setminus\mathcal{D}} C_{\mathbb{H}}(\tau,\nu) M(\tau,\nu) d\tau d\nu \\
&\leq \max_{(\tau,\nu)\in\mathcal{D}} \Big\{ M(\tau,\nu) \Big\} \iint_{\mathcal{D}} C_{\mathbb{H}}(\tau,\nu) d\tau d\nu \\
&\quad + \max_{(\tau,\nu)\in\mathbb{R}^2\setminus\mathcal{D}} \Big\{ M(\tau,\nu) \Big\} \iint_{\mathbb{R}^2\setminus\mathcal{D}} C_{\mathbb{H}}(\tau,\nu) d\tau d\nu \\
&\overset{(a)}{\leq} \max_{(\tau,\nu)\in\mathcal{D}} \Big\{ M(\tau,\nu) \Big\} + \epsilon
\end{aligned} \tag{73}$$

where (a) follows from (72), (7), and (9).

The proof is completed by substituting (70) into (24), and using (71) and (73) in (24).

Appendix G
Proof of Lemma 4

To prove the lemma, we verify that after the substitutions

$$\begin{aligned}
e(t) &= \sqrt{\beta} g(\beta t) \\
\widetilde{T} &= T/\beta \\
\widetilde{F} &= \beta F \\
\widetilde{\tau}_0 &= \tau_0/\beta \\
\widetilde{\nu}_0 &= \beta \nu_0
\end{aligned}$$

the lower bound L_2 in (25) does not change. Note first that $\widetilde{T}\widetilde{F} = TF$ and $\nu_0 T = \widetilde{\nu}_0 \widetilde{T}$. Furthermore, $\|e(t)\| = \|g(t)\| = 1$ and, by Property 4 in Appendix D, the orthonormality of $\{g_{k,n}(t)\}$ implies the orthonormality of $\{e(t-k\widetilde{T})e^{j2\pi n\widetilde{F}t}\}$. Let now $\mathcal{E} = [-\widetilde{\tau}_0, \widetilde{\tau}_0] \times [-\widetilde{\nu}_0, \widetilde{\nu}_0]$; we have that

$$\begin{aligned}
m_g &= \min_{(\tau,\nu)\in\mathcal{D}} |A_g(\tau,\nu)|^2 \\
&= \min_{(\tau,\nu)\in\mathcal{D}} \left| A_e\left(\frac{\tau}{\beta}, \beta\nu\right) \right|^2 \\
&= \min_{(\tau,\nu)\in\mathcal{E}} |A_e(\tau,\nu)|^2 .
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
M_g &= \max_{(\tau,\nu)\in\mathcal{D}} \sum_{\substack{k=-\infty \\ (k,n)\neq(0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} |A_g(\tau-kT, \nu-nF)|^2 \\
&= \max_{(\tau,\nu)\in\mathcal{D}} \sum_{\substack{k=-\infty \\ (k,n)\neq(0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \left| A_e\left(\frac{\tau-kT}{\beta}, \beta(\nu-nF)\right) \right|^2 \\
&= \max_{(\tau,\nu)\in\mathcal{D}} \sum_{\substack{k=-\infty \\ (k,n)\neq(0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \left| A_e\left(\frac{\tau}{\beta} - k\widetilde{T}, \beta\nu - n\widetilde{F}\right) \right|^2 \\
&= \max_{(\tau,\nu)\in\mathcal{E}} \sum_{\substack{k=-\infty \\ (k,n)\neq(0,0)}}^{\infty} \sum_{n=-\infty}^{\infty} \left| A_e\left(\tau - k\widetilde{T}, \nu - n\widetilde{F}\right) \right|^2 .
\end{aligned}$$

To conclude, we note that for $\beta = \sqrt{T/F}$ and under the assumption $\nu_0 T = \tau_0 F$, we get $\widetilde{T} = \widetilde{F} = \sqrt{TF}$, and $\widetilde{\tau}_0 = \widetilde{\nu}_0 = \sqrt{\Delta_{\mathbb{H}}}/2$, which implies (26).

References

[1] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz), "Noncoherent capacity of underspread fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 367–395, Jan. 2010.

[2] G. Durisi, V. I. Morgenshtern, H. Bölcskei, U. G. Schuster, and S. Shamai (Shitz), "Information theory of underspread WSSUS channels," in *Wireless Communications over Rapidly Time-Varying Channels*, F. Hlawatsch and G. Matz, Eds. Academic Press, Mar. 2011, ch. 2, pp. 65–115.

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

[4] E. Biglieri, J. G. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

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

[6] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

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

[8] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, no. 3, pp. 292–300, Mar. 1976.

[9] R. H. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, Apr. 2006.

[10] T. Koch and A. Lapidoth, "Degrees of freedom in non-coherent stationary MIMO fading channels," in *Winter-School on Coding and Information Theory*, Bratislava, Slovakia, Feb. 2005.

[11] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun.*, vol. 11, no. 4, pp. 360–393, Dec. 1963.

[12] R. S. Kennedy, *Fading Dispersive Communication Channels*. New York, NY, U.S.A.: Wiley, 1969.

[13] W. Kozek, "Matched Weyl-Heisenberg expansions of nonstationary environments," Ph.D. dissertation, Vienna University of Technology, Department of Electrical Engineering, Vienna, Austria, Mar. 1997.

[14] G. Matz and F. Hlawatsch, "Time-frequency characterization of randomly time-varying channels," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, B. Boashash, Ed. Oxford, U.K.: Elsevier, 2003, ch. 9.5, pp. 410–419.

[15] A. D. Wyner, "The capacity of the band-limited Gaussian channel," *Bell Syst. Tech. J.*, vol. 45, no. 3, pp. 359–395, Mar. 1966.

[16] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, U.S.A.: Wiley, 1968.

[17] G. Matz and F. Hlawatsch, "Fundamentals of time-varying communication channels," in *Wireless Communications over Rapidly Time-Varying Channels*, F. Hlawatsch and G. Matz, Eds. Academic Press, Mar. 2011, ch. 1, pp. 1–63.

[18] K. Gröchenig, *Foundations of Time-Frequency Analysis*. Boston, MA, U.S.A.: Birkhäuser, 2001.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, U.S.A.: Wiley, 2006.

[20] I. M. Gel'fand and A. M. Yaglom, "Calculation of the amount of information about a random function contained in another such function," *Uspekhi Mat. Nauk*, vol. 12, pp. 3–52, 1957, English transl., Amer. Math. Soc. Transl., Ser. 2, Vol. 12, 1959, pp. 199-246.

[21] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[22] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[23] T. Koch and A. Lapidoth, "On multipath fading channels at high SNR," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 5945–5957, Dec. 2010.

[24] W. Kozek and A. F. Molisch, "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1579–1589, Oct. 1998.

[25] G. Matz and F. Hlawatsch, "Time-frequency transfer function calculus (symbolic calculus) of linear time-varying systems (linear operators) based on a generalized underspread theory," *J. Math. Phys.*, vol. 39, no. 8, pp. 4041–4070, Aug. 1998.

[26] O. Christensen, *An Introduction to Frames and Riesz Bases*. Boston, MA, U.S.A.: Birkhäuser, 2003.

[27] I. Daubechies, H. Landau, and Z. Landau, "Gabor time-frequency lattices and the Wexler-Raz identity," *J. Fourier Anal. Appl.*, vol. 1, no. 4, pp. 437–478, 1995.

[28] A. J. E. M. Janssen, "Duality and biorthogonality for Weyl-Heisenberg frames," *J. Fourier Anal. Appl.*, vol. 1, no. 4, pp. 403–437, 1995.

[29] A. Ron and Z. Shen, "Weyl-Heisenberg frames and Riesz bases in $L_2(\mathbb{R}^d)$," *Duke Math J.*, vol. 89, no. 2, pp. 237–282, 1997.

[30] C. H. Wilcox, "The synthesis problem for radar ambiguity functions," in *Radar and Sonar*, R. E. Blahut, W. Miller, Jr., and C. H. Wilcox, Eds. New York, NY, U.S.A.: Springer Verlag, 1991, vol. 1, pp. 229–260.

[31] G. Matz, D. Schafhuber, K. Gröchenig, M. Hartmann, and F. Hlawatsch, "Analysis, optimization, and implementation of low-interference wireless multicarrier systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1921–1931, May 2007.

[32] A. Tulino, G. Caire, S. Shamai, and S. Verdú, "Capacity of channels with frequency-selective and time-selective fading," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1187–1215, Mar. 2010.

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

[34] P. M. Woodward, *Probability and Information Theory, with Applications to Radar*. London, U.K.: Pergamon Press, 1953.

[35] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.

[36] M. Miranda and P. Tilli, "Asymptotic spectra of Hermitian block Toeplitz matrices and preconditioning results," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 3, pp. 867–881, Feb. 2000.

[37] E. Kreyszig, *Introduction to Functional Analysis with Applications*. New York, NY, U.S.A.: Wiley, 1989.