

Information Theory of Underspread WSSUS Channels

Giuseppe Durisi¹, Veniamin I. Morgenshtern², Helmut
Bölcskei², Ulrich G. Schuster³, Shlomo Shamai
(Shitz)⁴

¹Chalmers University of Technology, Gothenburg, Sweden,

²ETH Zurich, Switzerland, ³Robert Bosch GmbH, Stuttgart,
Germany, ⁴Technion–Israel Institute of Technology, Haifa, Israel

2.1 The role of a system model

2.1.1 A realistic model

In this chapter, we are interested in the *ultimate limit* on the rate of reliable communication over Rayleigh-fading channels that satisfy the wide-sense stationary (WSS) and uncorrelated scattering (US) assumptions and are underspread [Bel63, Ken69]. Therefore, the natural setting is an information-theoretic one, and the performance metric is channel *capacity* [CT91, Gal68].

The family of Rayleigh-fading underspread WSSUS channels (reviewed in Chapter 1) constitutes a good model for real-world wireless channels: their stochastic properties, like amplitude and phase distributions match channel measurement results [SB07, Sch09]. The Rayleigh-fading and the WSSUS assumptions imply that the stochastic properties of the channel are fully described by a two-dimensional power spectral density (PSD) function, often referred to as *scattering function* [Bel63]. The underspread assumption implies that the scattering function is highly concentrated in the delay-Doppler plane.

To analyze wireless channels with information-theoretic tools, a *system model*, not just a *channel model*, needs to be specified. A system model is more comprehensive than a channel model because it defines, among other parameters, the transmit-power constraints and the channel knowledge available at the transmitter and the receiver. The choice of a

realistic system model is crucial for the insights and guidelines provided by information theory to be useful for the design of practical systems. Two important aspects need to be accounted for by a model that aims at being realistic:

- i) *Neither the transmitter nor the receiver knows the realization of the channel:* In most wireless systems, channel state information (CSI) is acquired by allocating part of the available resources to channel estimation. For example, pilot symbols can be embedded into the data stream as explained in Chapters 4 and 5 to aid the receiver in the channel-estimation process. From an information-theoretic perspective, pilot-based channel estimation is just a special case of coding. Hence, the rate achievable with training-based schemes cannot exceed the capacity in the absence of CSI at transmitter and receiver.

We refer to the setting where no CSI is available at transmitter and receiver, but both know the statistics of the channel, as the *noncoherent setting* [SHN05, SWHL09, ZT02], in contrast to the *coherent setting*, where a genie provides the receiver with perfect CSI [BPS98, Section III.B]. Furthermore, we denote capacity in the noncoherent setting as *noncoherent capacity*.

- ii) *The peak power of the transmit signal is limited:* Every power amplifier has finite gain and every mobile transmitter has limited battery resources. In addition, regulatory bodies often constrain the admissible radiated power. Hence, in a realistic system model, a peak constraint should be imposed on the transmit signal.

Motivated by these two aspects, we provide an information-theoretic analysis of Rayleigh-fading underspread WSSUS channels in the noncoherent setting, under the additional assumption that the transmit signal is peak-constrained.

2.1.2 A brief literature survey

The noncoherent capacity of fading channels is notoriously difficult to characterize analytically, even for simple channel models [AFTS01, MH99]. Most of the results available in the literature pertain to either the *large-bandwidth* or the *high-SNR* regime. In the following, large-bandwidth regime refers to the case where the average power P is fixed and the bandwidth B is large, so that the SNR, which is proportional to P/B , is small. High-SNR regime refers to the case of B fixed and P large and, hence, large SNR. In this section, we briefly review the relevant literature; two elements will emerge from this review:

- i) The modeling aspects we identified in the previous section (WSSUS, noncoherent setting, peak constraint) are *fundamental* in that capacity is highly sensitive to these aspects.
- ii) In spite of the large number of results available in the literature, several questions of practical engineering relevance about the design of wireless systems operating over fading channels are still open.

Large-bandwidth (low-SNR) regime: The noncoherent capacity of fading channels has been a subject of investigation in information theory for several decades. The first contributions, which date back to the sixties [Gal68, Ken69, Pie66, Vit67] (see [BPS98] for a more complete list of references), mainly deal with the characterization of the asymptotic behavior of noncoherent capacity in the large-bandwidth limit.

The results in [Gal68, Ken69, Pie66, Vit67] illustrate the sensitivity of noncoherent capacity to the presence of a peak constraint. Specifically, the outcome of this analysis is the following rather surprising result: in the large-bandwidth limit, the noncoherent capacity of a fading channel coincides with that of an additive white Gaussian noise (AWGN) channel with the same receive power. However, the signaling schemes that achieve the noncoherent capacity of a fading channel in the wideband limit have unbounded peak power—a result recently formalized by Verdú [Ver02]. Hence, these signaling schemes are not practical.

If a peak constraint is imposed on the transmit signal, AWGN channel capacity cannot be achieved in the infinite-bandwidth limit, and the actual behavior of noncoherent capacity in the wideband regime depends on the specific form of the peak constraint [DSBS10, MG02, SH02, TT00, Vit67]. In particular, when the transmit signal is subject to a peak constraint both in time and frequency (as in most practical systems), noncoherent capacity vanishes as the bandwidth grows large [DSBS10, MG02, SH02, TT00], provided that the number of independent diversity branches of the channel scales linearly with bandwidth,¹ which is the case for WSSUS channels. Intuitively, under a peak constraint on the transmit signal, the receiver is no longer able to resolve the channel uncertainty as the bandwidth, and, hence, the number of independent diversity branches, increases. This result implies that, for a large class of fading channels, the corresponding noncoherent capacity has a global maximum at a certain finite bandwidth, commonly referred to as the *critical bandwidth*. Computing this critical bandwidth is obviously of great practical interest. Moreover, it is important to understand the role played by the spatial degrees of freedom provided by multiple antennas at the transmitter and/or the receiver: can they be used to increase capacity, or do they merely lead to a rate loss because of the increase in channel uncertainty?

High-SNR regime: Characterizing noncoherent capacity in the high-SNR regime is of great practical interest for systems that operate over narrow frequency bands. As no closed-form expressions for the noncoherent capacity are known, not even for memoryless channels, one typically resorts to analyzing the asymptotic behavior of capacity as SNR goes to infinity. Differently from the large-bandwidth regime, where capacity results are robust with respect to the underlying channel model, in the high-SNR regime the capacity behavior is highly sensitive to the *fine details* of the channel model [Lap05, LM03, LV04, MH99, Tel99, ZT02]. The following results support this claim.

In the coherent setting, capacity grows logarithmically with SNR [Tel99]; logarithmic growth also holds in the noncoherent setting for *block-fading* channels [MH99, ZT02].² An alternative—more general—approach to modeling the time variation of a wireless

¹The independent diversity branches of a fading channel are sometimes referred to as *stochastic degrees of freedom* [SB07]. We will not use this convention here. Instead, we will use the term *degrees of freedom* to refer to signal-space dimensions. Information-theoretic analyses of wireless channels for which the number of independent diversity branches scales sub-linearly with bandwidth, can be found, for example, in [PTN07, RHS07].

²The block-fading model is the simplest model that captures the time variation of a wireless channel. In this model, the channel is taken to be constant over a given time interval, called *block*, and assumed to change independently from one such block to the next. The independence assumption across blocks can be justified, e.g., for systems employing frequency hopping or if the data symbols are interleaved.

channel is to assume that the fading process is stationary.³ Surprisingly, if the fading process is stationary, the noncoherent capacity does not necessarily grow logarithmically with SNR: other scaling behaviors are possible [Lap05]. For example, consider two stationary discrete-time Rayleigh-fading channels subject to additive Gaussian noise. The fading process of the first channel has PSD equal to $1/\Delta$ on the interval $[-\Delta/2, \Delta/2]$, where $0 < \Delta < 1$, and 0 else. The fading process of the second channel has PSD equal to $(1 - \varepsilon)/\Delta$ on the interval $[-\Delta/2, \Delta/2]$ and $\varepsilon/(1 - \Delta)$ else ($0 < \varepsilon < 1$). These two channels have completely different high-SNR capacity behavior, no matter how small ε is: the noncoherent capacity of the first channel grows logarithmically in SNR, with *pre-log* factor equal to $1 - \Delta$ [Lap05]; the noncoherent capacity of the second channel grows double logarithmically in SNR [LM03].

Such a result is unsatisfactory from an engineering point of view because the support of a PSD cannot be determined through measurements (measurement noise is one of the reasons, another one is the finite time duration of any physically meaningful measurement process). In other words, capacity turns out to be highly sensitive to a parameter—the measure of the support of the PSD—that has, in the words of Slepian [Sle76], “...no direct meaningful counterparts in the real world...”. Such a dependency of the capacity behavior on fine details of the channel model suggests that the stationary model is not *robust* in the high-SNR regime. An engineering-relevant problem is then to establish the SNR value at which this lack of robustness starts to manifest itself.

2.1.3 Capacity bounds answering engineering-relevant questions

The purpose of this chapter is to present tight upper and lower bounds on the noncoherent capacity of Rayleigh-fading underspread WSSUS channels. On the basis of these bounds, answers to the following engineering-relevant questions can be given:

- i) How does the noncoherent capacity of this class of channels differ from the corresponding coherent capacity and from the capacity of an AWGN channel with the same receive power?
- ii) How much bandwidth and how many antennas should be used to maximize capacity?
- iii) How robust is the Rayleigh-fading WSSUS underspread channel model? More specifically, at which SNR values does the noncoherent capacity start being sensitive to the fine details of the channel model?

The capacity bounds presented in this chapter make use of information-theoretic tools recently developed in [GSV05, Lap05, LM03, SWHL09]. One of the difficulties we shall encounter is to adapt these tools to the continuous-time setting considered in this chapter. Harmonic analysis plays a fundamental role in this respect: it provides an effective method for converting a general continuous-time channel into a discretized channel that

³Stationarity in time, together with the assumption that scatterers corresponding to paths of different delays are uncorrelated, is the fundamental feature of the WSSUS model we focus on in this chapter.

can be analyzed using standard information-theoretic tools. The discretization is accomplished by transmitting and receiving on a highly structured set of signals, similarly to what is done in pulse-shaped (PS) orthogonal frequency-division multiplexing (OFDM) systems [KM98]. This ensures that the resulting discretized channel inherits the statistical properties of the underlying continuous-time channel (in particular, stationarity) a fact that is crucial for the ensuing analysis. As a byproduct, our results yield a novel information-theoretic criterion for the design of PS-OFDM systems (see [MSG⁺07] for a recent overview on this topic).

2.2 A discretized system model

2.2.1 The channel model

2.2.1.1 Continuous-time input-output relation

The input-output (I/O) relation of a general continuous-time stochastic linear time-varying (LTV) channel \mathbf{H} can be written as

$$r(t) = \int_{-\infty}^{\infty} h(t, \tau) s(t - \tau) d\tau + w(t). \quad (2.1)$$

Here, $s(t)$ is the (stochastic) input signal, whose realizations can be taken as elements of the Hilbert space $L^2(\mathbb{R})$ of square-integrable functions over the real line \mathbb{R} ; $r(t)$ is the output signal and $w(t)$ is a zero-mean unit-variance proper AWGN random process. Finally, the time-varying channel impulse response $h(t, \tau)$ is a zero-mean jointly proper Gaussian (JPG) random process that satisfies the WSSUS assumption [Bel63], i.e., it is stationary in time t and uncorrelated in delay τ :

$$\mathbb{E}\{h(t, \tau)h^*(t', \tau')\} \triangleq r_h(t - t', \tau)\delta(\tau - \tau').$$

As a consequence of the JPG and WSSUS assumptions, the *time-delay correlation function* $r_h(t, \tau)$ fully characterizes the channel statistics. The JPG assumption is empirically supported for narrowband channels [VB03]; even ultrawideband (UWB) channels with bandwidth up to several Gigahertz can be modeled as Gaussian-distributed [SB07]. The WSSUS assumption is widely used in wireless channel modeling [Bel63, BPS98, Ken69, MH03, Pro01, VB03]. It is in good agreement with measurements of tropospheric scattering channels [Ken69], and provides a reasonable model for many types of mobile radio channels [Cox73a, Cox73b, Jak74], at least when observed over a limited time duration and limited bandwidth [Bel63]. A more detailed review of the WSSUS channel model can be found in Chapter 1.

Often, it will be convenient to describe \mathbf{H} in domains other than the time-delay domain. The *time-varying transfer function* $L_{\mathbf{H}}(t, f) = \mathbb{E}_{\tau \rightarrow f}\{h(t, \tau)\}$ and the *delay-Doppler spreading function* $S_{\mathbf{H}}(\tau, \nu) = \mathbb{E}_{t \rightarrow \nu}\{h(t, \tau)\}$ can be used for this purpose. As a consequence of the WSSUS assumption, $L_{\mathbf{H}}(t, f)$ is WSS both in time t and in frequency f ,

and $S_{\mathbf{H}}(\tau, \nu)$ is uncorrelated in delay τ and Doppler ν :

$$\mathbb{E}\{L_{\mathbf{H}}(t, f)L_{\mathbf{H}}^*(t', f')\} \triangleq R_{\mathbf{H}}(t - t', f - f'). \quad (2.2)$$

$$\mathbb{E}\{S_{\mathbf{H}}(\tau, \nu)S_{\mathbf{H}}^*(\tau', \nu')\} \triangleq C_{\mathbf{H}}(\tau, \nu)\delta(\tau - \tau')\delta(\nu - \nu'). \quad (2.3)$$

The function $R_{\mathbf{H}}(t, f)$ is usually referred to as the channel *time-frequency correlation function*, and $C_{\mathbf{H}}(\tau, \nu)$ is called the channel *scattering function*. The two functions are related by the two-dimensional Fourier transform $C_{\mathbf{H}}(\tau, \nu) = \mathbb{F}_{t \rightarrow \nu, f \rightarrow \tau}\{R_{\mathbf{H}}(t, f)\}$. We assume throughout that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_{\mathbf{H}}(\tau, \nu) d\tau d\nu = 1 \quad (2.4)$$

for simplicity.

2.2.1.2 The underspread assumption

Almost all WSSUS channels of practical interest are *underspread*, i.e., they have a scattering function $C_{\mathbf{H}}(\tau, \nu)$ that is highly concentrated around the origin of the delay-Doppler plane [Bel63].

A mathematically precise definition of the underspread property is available for the case when $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported within a rectangle in the delay-Doppler plane. In this case, a WSSUS channel is said to be underspread if the support area of $C_{\mathbf{H}}(\tau, \nu)$ is much less than one [DSBS10, Koz97]. In practice, it is not possible to determine through channel measurements whether $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported or not. Hence, in the terminology introduced in Section 2.1.2, the measure (area) of the support of the scattering function is a fine detail of the channel model. To investigate the sensitivity of noncoherent capacity to this fine detail, and assess the robustness of the Rayleigh-fading WSSUS model, we replace the compact-support assumption by the following physically more meaningful assumption: $C_{\mathbf{H}}(\tau, \nu)$ has a small fraction of its total volume supported outside a rectangle of an area that is much smaller than 1. More precisely, we have the following definition:

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

$$\int_{-\nu_0}^{\nu_0} \int_{-\tau_0}^{\tau_0} C_{\mathbf{H}}(\tau, \nu) d\tau d\nu \geq 1 - \varepsilon.$$

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

Remark 1 Definition 1 is inspired by Slepian's treatment of $L^2(\mathbb{R})$ signals that are approximately limited in time and frequency [Sle76]. Note that $\varepsilon = 0$ in Definition 1 yields the compact-support underspread definition of [DSBS10, Koz97]. An alternative definition of the underspread property, also not requiring that $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported, was used in [MH03]. The definition in [MH03] is in terms of moments of the scattering function.

Typical wireless channels are (highly) underspread, with most of the volume of $C_{\mathbf{H}}(\tau, \nu)$ supported within a rectangle of area $\Delta_{\mathbf{H}} \approx 10^{-3}$ for typical land-mobile channels, and as low as 10^{-7} for some indoor channels with restricted mobility of the terminals [Has93, Par00, Rap02]. In the remainder of the chapter, we refer to $\Delta_{\mathbf{H}}$ as the *channel spread*.

2.2.1.3 Continuous-time channel model and capacity

The goal of this chapter is to provide a characterization of the capacity of the continuous-time channel with I/O relation (2.1), under the assumptions that:

- i) Neither the transmitter nor the receiver knows the realizations of $h(t, \tau)$, but both are aware of the channel statistics. For the Rayleigh-fading WSSUS channel model, knowledge of the channel statistics amounts to knowledge of the channel scattering function.
- ii) The input signal $s(t)$ is subject to a bandwidth constraint, an average-power constraint, and a peak-power constraint.

A formal definition of the capacity of the continuous-time channel (2.1) can be found in [DMBon]. This definition is along the lines of Wyner's treatment of the capacity of a bandlimited continuous-time AWGN channel [Gal68, Wyn66]. The key element in this definition is the precise specification of the set of constraints (approximate time duration, bandwidth, average power, ...) that are imposed on the input signal $s(t)$. For the sake of simplicity of exposition, we refrain from presenting this definition here (the interested reader is referred to [DMB09, DMBon]). We take, instead, a somewhat less rigorous approach, which has the advantage of simplifying the exposition drastically and better illustrates the harmonic analysis aspects of the problem: we first discretize the channel I/O relation, and then impose a set of constraints on the resulting discretized input signal. These constraints "mimic" the ones that are natural to impose on the underlying continuous-time input signal. We emphasize that all the results stated in the next sections can be made precise in an information-theoretic sense following the approach in [DMB09, DMBon].

2.2.2 Discretization of the continuous-time input-output relation

Different ways to discretize LTV channels have been proposed in the literature (some are reviewed in Chapter 1); not all the induced discretized I/O relations are, however, equally well suited for an information-theoretic analysis. Stationarity of the discretized system functions and statistically independent noise samples are some of the desiderata regarding the discretization step.

The most common approach to the discretization of random LTV channels is based on sampling [Bel63, Méd00, MG02], often combined with a basis expansion model (BEM) (see Chapter 1 for a detailed discussion). A different approach, particularly well suited for information-theoretic analyses [Wyn66, Gal68], is based on a channel operator eigen-decomposition or, more generally, singular-value decomposition. We briefly review these

two approaches and discuss their shortcomings when used for the problem we are dealing with in this chapter. These shortcomings will motivate us to pursue a different approach detailed in Section 2.2.2.3.

2.2.2.1 Sampling and basis expansion

Under the assumption that $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported in ν over the interval $[-\nu_0, \nu_0]$, and the input signal $s(t)$ is strictly bandlimited with bandwidth B (i.e., $S(f) \triangleq \mathbb{F}\{s(t)\} = 0$ for $|f| > B$), the I/O relation (2.1) can be discretized by means of the sampling theorem (see, for example, [AMH04] for a detailed derivation). The resulting discretized I/O relation is given by

$$r\left(\frac{n}{f_o}\right) = \frac{1}{f_i} \sum_{m=-\infty}^{\infty} h_c\left(\frac{n}{f_o}, \frac{m}{f_i}\right) s\left(\frac{n}{f_o} - \frac{m}{f_i}\right) + w\left(\frac{n}{f_o}\right) \quad (2.5)$$

where $f_o = 2(B + \nu_0)$, $f_i = 2B$, and

$$h_c(t, \tau) \triangleq \int_{-\infty}^{\infty} h(t, z) \frac{\sin[2\pi B(\tau - z)]}{\pi(\tau - z)} dz. \quad (2.6)$$

The discretized I/O relation (2.5) can be further simplified if the input signal is oversampled, i.e., if f_i is chosen to be equal to f_o . In this case, (2.5) can be rewritten as

$$r[n] = \sum_{m=-\infty}^{\infty} h_c[n, m] s[n - m] + w[n]. \quad (2.7)$$

One evident limitation of the sampling approach just discussed is the compact-support assumption on $C_{\mathbf{H}}(\tau, \nu)$ (cf. Definition 1). Furthermore, as a consequence of (2.6), $h_c[n, m]$ does not inherit the US property of $h(t, \tau)$, a fact that makes the information-theoretic analysis of (2.7) more involved. Finally, the apparently harmless oversampling step imposes an implicit constraint on the set of the input-signal samples; this constraint is hard to account for in an information-theoretic analysis. More specifically, if the input-signal samples are chosen in an arbitrary way, the resulting continuous-time input signal may have a bandwidth as large as $B + \nu_0$, rather than just B .

Often, the sampling approach is combined with a BEM. In the most common form of the BEM, the basis consists of complex exponentials and $h_c[n, m]$ is given by

$$h_c[n, m] = \sum_{i=0}^{I-1} c_i[m] e^{j2\pi \nu_i n} \quad (2.8)$$

where $\{\nu_i\}_{i=0}^{I-1}$ is a set of Doppler frequencies. For general WSSUS underspread channels, the validity of the modeling assumption underlying (2.8) is difficult to assess, and information-theoretic results obtained on the basis of (2.8) might lack generality.

2.2.2.2 Discretization through channel eigen-decomposition

As discussed in Chapter 1, the kernel $k_{\mathbf{H}}(t, t') \triangleq h(t, t - t')$ associated with a general LTV channel \mathbf{H} (under the assumption that the channel operator \mathbf{H} is normal and com-

pact [DS63, DSBS10, NS82]) can be decomposed as [NS82, Theorem 6.14.1]

$$k_{\mathbf{H}}(t, t') = \sum_{k=0}^{\infty} \lambda_k u_k(t) u_k^*(t') \quad (2.9)$$

where λ_k and $u_k(t)$ are the channel eigenvalues and eigenfunctions, respectively. The set $\{u_k(t)\}$ is orthonormal and complete in the input space and in the range space of \mathbf{H} [NS82, Theorem 6.14.1]. Hence, any input signal $s(t)$ and any output signal $r(t)$ can be expressed, without loss of generality, in terms of its projections onto the set $\{u_k(t)\}$ according to

$$s(t) = \sum_k \underbrace{\langle s(t), u_k(t) \rangle}_{\triangleq s[k]} u_k(t) \quad (2.10)$$

and

$$r(t) = \sum_k \underbrace{\langle r(t), u_k(t) \rangle}_{\triangleq r[k]} u_k(t). \quad (2.11)$$

The decomposition (2.9), together with (2.10) and (2.11), yields a particularly simple I/O relation, which we refer to as *channel diagonalization*:

$$\begin{aligned} r[k] &= \langle (\mathbf{H}s)(t) + w(t), u_k(t) \rangle = \sum_{k'} s[k'] \underbrace{\langle (\mathbf{H}u_{k'})(t), u_k(t) \rangle}_{\lambda_{k'} \delta_{k'k}} + \underbrace{\langle w(t), u_k(t) \rangle}_{\triangleq w[k]} \\ &= \lambda_k s[k] + w[k]. \end{aligned} \quad (2.12)$$

Note that in (2.12), the channel acts on the discretized input through scalar multiplications only. To summarize, it follows from (2.10) and (2.11) that channel diagonalization is achieved by transmitting and receiving on the channel eigenfunctions.

The discretization method just described is used in [Wyn66] to compute the capacity of bandlimited AWGN channels and in [Gal68] to compute the capacity of (deterministic) linear time-invariant (LTI) channels. This approach, however, is not applicable to our setting. Transmitting and receiving on the channel eigenfunctions $\{u_k(t)\}$ requires perfect knowledge of $\{u_k(t)\}$ at the transmitter and the receiver. But as \mathbf{H} is random, its eigenfunctions $u_k(t)$ are (in general) random as well, and, hence, not known at the transmitter and the receiver in the noncoherent setting.

In contrast, if the eigenfunctions of the random channel \mathbf{H} did not depend on the particular realization of \mathbf{H} , we could diagonalize \mathbf{H} without knowledge of the channel realizations. This is the case, for example, for LTI channels, where (deterministic) complex sinusoids are always eigenfunctions, independently of the realization of the channel impulse response. This observation is crucial for the ensuing analysis. Specifically, our discretization strategy is based on the following fundamental property of underspread LTV channels (see Chapter 1 for details): the eigenfunctions of a random underspread WSSUS channel can be well approximated by deterministic functions that are well localized in time and frequency.

Before illustrating the details of the discretization step, we would like to point out two difficulties that arise from the approach we are going to pursue. The discretized I/O relation induced by replacing the channel eigenfunctions in (2.10) and (2.11) with approximate eigenfunctions (or, in other words, induced by transmitting and receiving on the

approximate channel eigenfunctions) will, in general, not be as simple as (2.12), because of the presence of “off-diagonal” terms. Furthermore, the set of approximate eigenfunctions may not be complete in the input and range spaces of the channel operator \mathbf{H} . This results in a loss of dimensions in signal space, i.e., of degrees of freedom, which will need to be accounted for in our information-theoretic analysis.

2.2.2.3 Discretization through transmission and reception on a Weyl-Heisenberg set

We accomplish the discretization of the I/O relation (2.1) by transmitting and receiving on the highly structured Weyl-Heisenberg (WH) set of time-frequency shifts of a pulse $g(t)$. We denote such a WH set as $(g(t), T, F) \triangleq \{g_{k,l}(t) = g(t - kT)e^{j2\pi lFt}\}_{k,l \in \mathbb{Z}}$, where $T, F > 0$ are the *grid parameters* of the WH set [KM98]. The triple $(g(t), T, F)$ is chosen such that $g(t)$ has unit energy and that the resulting set $(g(t), T, F)$ is orthonormal.⁴ Note that we do not require that the set $(g(t), T, F)$ is complete in $L^2(\mathbb{R})$. The time-frequency localization of $g(t)$ plays an important role in our analysis, because the functions in a WH set constructed from a pulse that is well localized in time and frequency are approximate eigenfunctions of underspread WSSUS channels [Koz97, MH98, MH03].

We consider input signals of the form

$$s(t) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} s[k, l] g_{k,l}(t). \quad (2.13)$$

Whenever $g(t)$ is well localized in time and frequency, we can take $D \triangleq KT$ to be the approximate time duration of $s(t)$ and $B \triangleq LF$ to be its approximate bandwidth. Note that the lack of completeness of $(g(t), T, F)$ implies that there exist signals $s(t) \in L^2(\mathbb{R})$ that cannot be written in the form (2.13), even when $K, L \rightarrow \infty$. In other words, we may lose degrees of freedom.

The received signal $r(t)$ is projected onto the signal set $\{g_{k,l}(t)\}_{k=0, l=0}^{K-1, L-1}$ to obtain

$$\begin{aligned} \underbrace{\langle r(t), g_{k,l}(t) \rangle}_{\triangleq r[k, l]} &= \underbrace{\langle (\mathbf{H}g_{k,l})(t), g_{k,l}(t) \rangle}_{\triangleq h[k, l]} s[k, l] + \sum_{\substack{k'=0 \\ (k', l') \neq (k, l)}}^{K-1} \sum_{l'=0}^{L-1} \underbrace{\langle (\mathbf{H}g_{k',l'})(t), g_{k,l}(t) \rangle}_{\triangleq z[k', l', k, l]} s[k', l'] \\ &\quad + \underbrace{\langle w(t), g_{k,l}(t) \rangle}_{\triangleq w[k, l]} \end{aligned}$$

that is

$$r[k, l] = h[k, l] s[k, l] + \sum_{\substack{k'=0 \\ (k', l') \neq (k, l)}}^{K-1} \sum_{l'=0}^{L-1} z[k', l', k, l] s[k', l'] + w[k, l]. \quad (2.14)$$

We refer to the channel with I/O relation (2.14) as the *discretized channel induced by the WH set $(g(t), T, F)$* . The I/O relation (2.14) satisfies the desiderata we listed at

⁴A systematic approach to constructing orthonormal WH sets is described in Section 2.2.2.6.

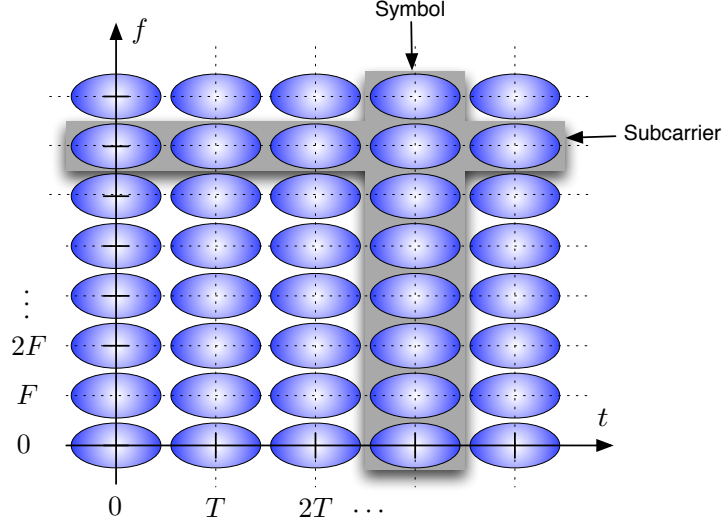


Figure 2.1: Pulse-shaped OFDM interpretation of the signaling scheme (2.13). The shaded areas represent the approximate time-frequency support of the pulses $g_{k,l}(t)$.

the beginning of Section 2.2.2. Specifically, the channel coefficients $h[k, l]$ inherit the two-dimensional stationarity property of the underlying continuous-time system function $L_{\mathbf{H}}(t, f)$ [see (2.2)]. Furthermore, the noise coefficients $w[k, l]$ are i.i.d. $\mathcal{CN}(0, 1)$ as a consequence of the orthonormality of $(g(t), T, F)$. These two properties are crucial for the ensuing analysis.

A drawback of (2.14) is the presence of (self-)interference [the second term in (2.14)], which makes the derivation of capacity bounds involved, as will be seen in Section 2.4.

The signaling scheme (2.13) can be interpreted as PS-OFDM [KM98], where the input symbols $s[k, l]$ are modulated onto a set of orthogonal signals, indexed by discrete time (symbol index) k and discrete frequency (subcarrier index) l . From this perspective, the interference term in (2.14) can be interpreted as intersymbol and intercarrier interference (ISI and ICI). Fig. 2.1 provides a qualitative representation of the PS-OFDM signaling scheme.

2.2.2.4 Outline of the information-theoretic analysis

The program pursued in the next sections is to tightly bound the noncoherent capacity of the discretized channel (2.14) under an average-power and a peak-power constraint on the input symbols $s[k, l]$. We refer the interested reader to [DMB09, DMBon] for a discussion on the relation between the capacity of the discretized channel (2.14) and the capacity of the underlying continuous-time channel (2.1).

The derivation of capacity bounds is made difficult by the presence of the interference term in (2.14). Fortunately, to establish our main results in Sections 2.4 and 2.5, it will be

sufficient to compare a trivial capacity upper bound, i.e., the capacity of an AWGN channel with the same receive power as in (2.14), with simple capacity lower bounds obtained by treating the interference term in (2.14) as noise. The corresponding results are of practical interest, as receiver algorithms that take the structure of interference explicitly into account are, in general, computationally expensive. The power of the interference term in (2.14) depends on the choice of the WH set. As it will be shown in Section 2.2.2.7, good time-frequency localization of the pulse $g(t)$ the WH set is generated from is required for the interference term to be small. But good time-frequency localization entails a loss of degrees of freedom.

Before discussing the trade-off between interference minimization and maximization of the number of degrees of freedom, we review some important theoretical results on the construction of WH sets.

2.2.2.5 Orthonormality, completeness, and localization

Orthonormality, completeness, and time-frequency localization are desirable properties of the WH set $(g(t), T, F)$. It is, therefore, sensible to ask whether complete orthonormal WH sets generated by a $g(t)$ with prescribed time-frequency localization exist. The answer is as follows:

- i) A necessary condition for the set $(g(t), T, F)$ to be orthonormal is $TF \geq 1$ [Grö01, Cor. 7.5.1, Cor. 7.3.2].
- ii) For $TF = 1$, it is possible to find orthonormal sets $(g(t), T, F)$ that are complete in $L^2(\mathbb{R})$ [Chr03, Th. 8.3.1]. These sets, however, do not exhibit good time-frequency localization, as a consequence of the Balian-Low Theorem [Chr03, Th. 4.1.1], which states that if $(g(t), T, F)$ is orthonormal and complete in $L^2(\mathbb{R})$, then

$$\left(\int_{-\infty}^{\infty} |tg(t)|^2 dt \right) \left(\int_{-\infty}^{\infty} |fG(f)|^2 df \right) = \infty$$

where $G(f) = \mathbb{F}\{g(t)\}$.

- iii) For $TF > 1$, it is possible to have orthonormality and good time-frequency localization concurrently, but the resulting set $(g(t), T, F)$ is necessarily incomplete in $L^2(\mathbb{R})$. Lack of completeness entails a loss of degrees of freedom.
- iv) For $TF < 1$, it is possible to construct WH sets generated by a well-localized $g(t)$, which are also (over)complete in $L^2(\mathbb{R})$. However, as a consequence of overcompleteness, the resulting input signal (2.13) cannot be recovered uniquely at the receiver, even in the absence of noise.

Our choice will be to privilege localization and orthonormality over completeness. The information-theoretic results in Sections 2.4 and 2.5 will show that this choice is sound. In the next two sections, we review the mathematical framework that enables the construction of (non-complete) orthonormal WH sets that are well localized in time and frequency. We furthermore review a classic criterion for the design of WH sets,

which is based on the maximization of the signal-to-interference ratio (SIR) in (2.14). The information-theoretic analysis in Section 2.5 will yield a design criterion that is more fundamental.

2.2.2.6 Construction of WH sets

We next discuss how to construct (non-complete) WH sets with prescribed time-frequency localization. Frame theory plays a fundamental role in this context.

A WH set $(g(t), T, F)$ is called a WH *frame* for $L^2(\mathbb{R})$ if there exist constants A and B (*frame bounds*) with $0 < A \leq B < \infty$ such that for all $s(t) \in L^2(\mathbb{R})$, we have

$$A\|s(t)\|^2 \leq \sum_k \sum_l |\langle s(t), g_{k,l}(t) \rangle|^2 \leq B\|s(t)\|^2.$$

When $A = B$, the frame is called *tight*. A necessary condition for $(g(t), T, F)$ to be a frame for $L^2(\mathbb{R})$ is $TF \leq 1$ [Grö01, Cor. 7.5.1]. For specific pulses, sufficient conditions for the corresponding WH set to be a frame are known. For example, for the Gaussian pulse $g(t) = 2^{1/4}e^{-\pi t^2}$, the condition $TF < 1$ is necessary and sufficient for $(g(t), T, F)$ to be a frame [Grö01, Th. 7.5.3].

A frame $(g(t), T, F)$ for $L^2(\mathbb{R})$ can be transformed into a tight frame $(g^\perp(t), T, F)$ for $L^2(\mathbb{R})$ through standard frame-theoretic methods [Chr03, Th. 5.3.4]. Furthermore, the so-obtained $g^\perp(t)$ inherits the decay properties of $g(t)$ (see [BJ00, MSG⁺07] for a mathematically precise formulation of this statement).

The key result that makes frame theory relevant for the construction of orthonormal WH sets is the so called *duality theorem* [DLL95, Jan95, RS97], which states that $(g(t), T, F)$ with $TF \geq 1$ is an orthonormal WH set if and only if $(g(t), 1/F, 1/T)$ is a tight frame for $L^2(\mathbb{R})$.

The results summarized above are used in the following example to construct an orthonormal WH set that will be used throughout this chapter.

Example 1 (Root-raised-cosine WH set) *For later use, we present an example of an orthonormal WH set for the case $T = F = \sqrt{c}$, with $1 \leq c \leq 2$. Let $G(f) \triangleq \mathbb{F}\{g(t)\}$, $\zeta = \sqrt{c}$, and $\mu = c - 1$. We choose $G(f)$ as the (positive) square root of a raised-cosine pulse:*

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

where $S(f) = \cos\left[\frac{\pi\zeta}{\mu}\left(|f| - \frac{1-\mu}{2\zeta}\right)\right]$. As $(1+\mu)/(2\zeta) = \zeta/2$, the function $G(f)$ has compact support of length $\zeta = \sqrt{c}$. Furthermore, $G(f)$ is real-valued and even, and satisfies

$$\sum_{l=-\infty}^{\infty} G(f - l/\zeta)G(f - l/\zeta - k\zeta) = \zeta \delta_{k0}.$$

By [Chr03, Th. 8.7.2], we can, therefore, conclude that the WH set $(g(t), 1/\sqrt{c}, 1/\sqrt{c})$ is a tight WH frame for $L^2(\mathbb{R})$, and, by duality, the WH set $(g(t), \sqrt{c}, \sqrt{c})$ is orthonormal. Note

that, for $c = 1$ (i.e., $TF = 1$), the pulse $G(f)$ reduces to the rectangular pulse $1_{[-1/2, 1/2]}(f)$ and, consequently, $g(t)$ reduces to a sinc function, which has poor time localization, as expected from the Balian-Low Theorem.

2.2.2.7 Diagonalization and loss of degrees of freedom

By choosing $g(t)$ to be well-localized in time and frequency and TF sufficiently large, we can make the variance of the interference term in the I/O relation (2.14) small. The drawback is a loss of degrees of freedom, as formalized next.

Let $A_g(\tau, \nu)$ denote the ambiguity function of $g(t)$, defined as

$$A_g(\tau, \nu) \triangleq \int_{-\infty}^{\infty} g(t) g^*(t - \tau) e^{-j2\pi\nu t} dt.$$

The variance of both $h[k, l] = \langle (\mathbf{H}g_{k,l})(t), g_{k,l}(t) \rangle$ and $z[k', l', k, l] = \langle (\mathbf{H}g_{k',l'})(t), g_{k,l}(t) \rangle$ can be expressed in terms of $A_g(\tau, \nu)$. In fact, as a consequence of the WSSUS property of \mathbf{H} , we have that

$$\mathbb{E}\{h[k, l]h^*[k', l']\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_{\mathbf{H}}(\tau, \nu) |A_g(\tau, \nu)|^2 e^{j2\pi[(k-k')T\nu - (l-l')F\tau]} d\tau d\nu \quad (2.15)$$

and, in particular,

$$\sigma_h^2 \triangleq \mathbb{E}\{|h[k, l]|^2\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_{\mathbf{H}}(\tau, \nu) |A_g(\tau, \nu)|^2 d\tau d\nu. \quad (2.16)$$

Because $|A_g(\tau, \nu)|^2 \leq \|g(t)\|^4 = 1$ and because the scattering function $C_{\mathbf{H}}(\tau, \nu)$ was assumed to be of unit volume [see (2.4)], we have that $\sigma_h^2 \leq 1$. The relation (2.15) implies that $R_{\mathbf{H}}[k, k', l, l'] \triangleq \mathbb{E}\{h[k, l]h^*[k', l']\} = R_{\mathbf{H}}[k - k', l - l']$, i.e., that $h[k, l]$ is WSS both in discrete time k and discrete frequency l .

The variance of the interference term in (2.14), under the assumption that the $s[k, l]$ are i.i.d. with zero mean and unit variance,⁵ can be upper-bounded as follows:

$$\begin{aligned} \mathbb{E}\left\{\left|\sum_{\substack{k'=0 \\ (k', l') \neq (k, l)}}^{K-1} \sum_{l'=0}^{L-1} z[k', l', k, l] s[k', l']\right|^2\right\} &= \sum_{\substack{k'=0 \\ (k', l') \neq (k, l)}}^{K-1} \sum_{l'=0}^{L-1} \mathbb{E}\{|z[k', l', k, l]|^2\} \\ &\leq \sum_{\substack{k'=-\infty \\ (k', l') \neq (k, l)}}^{\infty} \sum_{l'=-\infty}^{\infty} \mathbb{E}\{|z[k', l', k, l]|^2\} \\ &= \sum_{\substack{k'=-\infty \\ (k', l') \neq (k, l)}}^{\infty} \sum_{l'=-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A_g(\tau + (k' - k)T, \nu + (l' - l)F)|^2 C_{\mathbf{H}}(\tau, \nu) d\tau d\nu \quad (2.17) \\ &= \sum_{\substack{k=-\infty \\ (k, l) \neq (0, 0)}}^{\infty} \sum_{l=-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A_g(\tau, \nu)|^2 C_{\mathbf{H}}(\tau - kT, \nu - lF) d\tau d\nu \\ &\triangleq \sigma_I^2. \end{aligned}$$

⁵In Sections 2.4 and 2.5, we will use this assumption to obtain capacity lower bounds explicit in σ_I^2 .

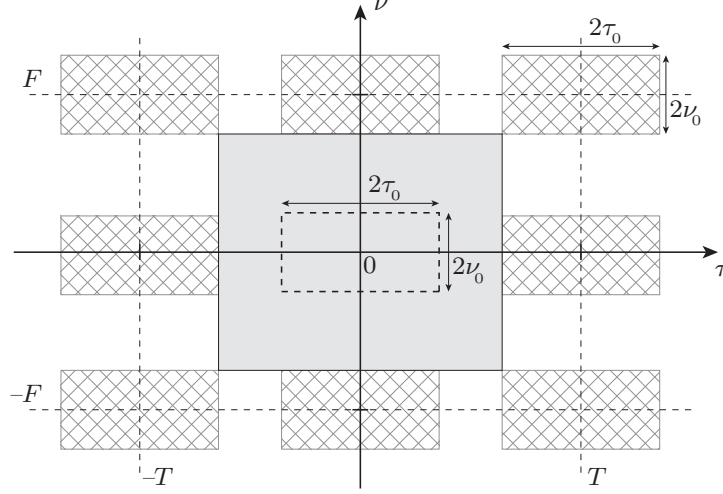


Figure 2.2: Support of $C_{\mathbf{H}}(\tau - kT, \nu - lF)$ for some (k, l) pairs. From (2.16), it follows that $\sigma_h^2 \approx 1$ if $|A_g(\tau, \nu)|^2 \approx 1$ over the support of $C_{\mathbf{H}}(\tau, \nu)$. Furthermore, from (2.17) it follows that $\sigma_l^2 \approx 0$ if $|A_g(\tau, \nu)|^2 \approx 0$ outside the area shaded in grey.

The “infinite-horizon” interference variance σ_l^2 will turn out to be of particular importance in the information-theoretic analysis in Sections 2.4 and 2.5. When $\sigma_l^2 \approx 0$, the I/O relation (2.14) can be well approximated by the following diagonalized I/O relation

$$r[k, l] = h[k, l]s[k, l] + w[k, l]. \quad (2.18)$$

This simplification eases the derivation of bounds on capacity. As the received signal power in (2.18) is proportional to σ_h^2 , it is also desirable to choose WH sets that result in $\sigma_h^2 \approx 1$ (recall that $\sigma_h^2 \leq 1$).

Next, we investigate the design criteria a WH set $(g(t), T, F)$ needs to satisfy so that $\sigma_h^2 \approx 1$ and $\sigma_l^2 \approx 0$. We assume, for simplicity, that $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported within the rectangle $[-\tau_0, \tau_0] \times [-\nu_0, \nu_0]$. Referring to (2.16), (2.17), and to Fig. 2.2, we conclude the following:

- i) $\sigma_h^2 \approx 1$ if the ambiguity function of $g(t)$ satisfies $|A_g(\tau, \nu)|^2 \approx 1$ over the support of the scattering function.
- ii) $\sigma_l^2 \approx 0$ if the ambiguity function of $g(t)$ takes on small values on the rectangles $[-\tau_0 + kT, \tau_0 + kT] \times [-\nu_0 + lF, \nu_0 + lF]$, for $(k, l) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$.

For these two conditions to be satisfied, the spread of the channel $\Delta_{\mathbf{H}} = 4\tau_0\nu_0$ needs to be small and the grid parameters need to be chosen such that the ambiguity function $A_g(\tau, \nu)$ takes on small values outside the solid grey-shaded rectangle centered at the origin in Fig. 2.2. Let D_g and B_g be the root-mean-square duration and the root-mean-square bandwidth, respectively, of the pulse $g(t)$ (see Chapter 1 for a definition of

these two quantities). Then, condition ii) above holds if $T > \tau_0 + D_g$ and $F > \nu_0 + B_g$. These two inequalities illustrate the importance of good time-frequency localization of the pulse $g(t)$ (i.e., small D_g and B_g). In fact, large D_g and B_g imply large T and F , and, consequently, a significant loss of degrees of freedom.

For fixed TF , a simple rule on how to choose the grid parameters T and F follows from the observation that for given τ_0 and ν_0 , the area of the solid rectangle centered at the origin in Fig. 2.2 is maximized if [Koz97, KM98, MSG⁺07]

$$\nu_0 T = \tau_0 F. \quad (2.19)$$

This rule is commonly referred to as *grid matching rule*.

Remark 2 Whenever T and F are chosen according to the grid matching rule, it is possible to assume, without loss of generality, that $C_{\mathbf{H}}(\tau, \nu)$ in (2.16) and (2.17) is supported on a square rather than a rectangle. A proof of this claim follows from a simple coordinate transformation.

Common approaches for the optimization of WH sets, such as the ones described in [MSG⁺07], aim at finding—for fixed TF —the pulse $g(t)$ that maximizes the ratio σ_h^2/σ_l^2 (which can be thought of as an SIR). To understand the trade-off between degrees of freedom and SIR, we proceed in a different way, and study how σ_h^2/σ_l^2 varies as a function of TF , for fixed $g(t)$. In Fig. 2.3, we plot a lower bound on σ_h^2/σ_l^2 for the root-raised-cosine WH sets of Example 1, as a function of TF and for different channel spreads $\Delta_{\mathbf{H}}$ (see Example 2 for more details). As expected, the larger TF , the larger the SIR, but the larger also the loss of degrees of freedom. A common compromise between loss of degrees of freedom and maximization of SIR is to take $TF \approx 1.2$ [MSG⁺07].

The limitation of the analysis we just outlined is that, although it sheds light on how σ_h^2/σ_l^2 depends on TF , it does not reveal the influence that σ_h^2 , σ_l^2 , and TF have on the rate achievable when interference is treated as noise. An information-theoretic analysis of the trade-off between maximization of SIR and minimization of the loss of degrees of freedom is called for. Such an analysis is provided in Section 2.5.

Example 2 (Trade-off between TF and σ_h^2/σ_l^2) Let $\mathcal{D} \triangleq [-\tau_0, \tau_0] \times [-\nu_0, \nu_0]$ and assume that $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported within \mathcal{D} and has unit volume. Then,

$$\sigma_h^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_{\mathbf{H}}(\tau, \nu) |A_g(\tau, \nu)|^2 d\tau d\nu \geq \min_{(\tau, \nu) \in \mathcal{D}} |A_g(\tau, \nu)|^2 \triangleq m_g. \quad (2.20)$$

Furthermore,

$$\begin{aligned} \sigma_l^2 &= \sum_{\substack{k=-\infty \\ (k,l) \neq (0,0)}}^{\infty} \sum_{l=-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A_g(\tau + kT, \nu + lF)|^2 C_{\mathbf{H}}(\tau, \nu) d\tau d\nu \\ &\leq \max_{(\tau, \nu) \in \mathcal{D}} \sum_{\substack{k=-\infty \\ (k,l) \neq (0,0)}}^{\infty} \sum_{l=-\infty}^{\infty} |A_g(\tau + kT, \nu + lF)|^2 \triangleq M_g. \end{aligned} \quad (2.21)$$

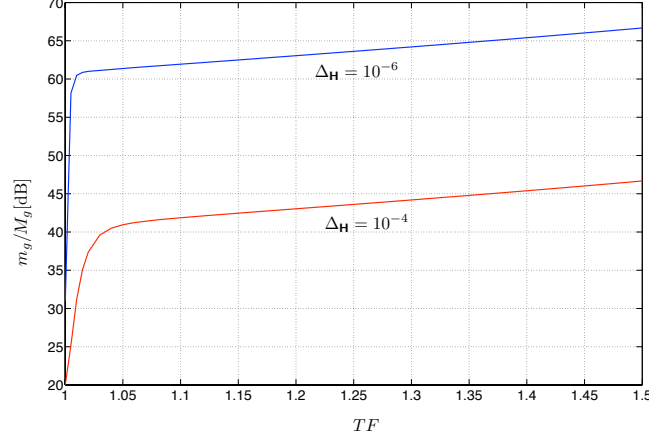


Figure 2.3: Trade-off between the product TF of the grid parameters and the signal-to-interference ratio σ_h^2/σ_I^2 for the root-raised-cosine WH sets constructed in Example 1. The two cases $\Delta_H = 10^{-4}$ and $\Delta_H = 10^{-6}$ are considered. The ratio m_g/M_g is a lower bound on σ_h^2/σ_I^2 , as detailed in Example 2.

The ratio m_g/M_g is obviously a lower bound on the SIR σ_h^2/σ_I^2 , and is easier to compute than σ_h^2/σ_I^2 , because it depends on the scattering function only through its support area and not its shape. Under the assumption that T and F are chosen according to the grid matching rule (2.19), it is sufficient to study the ratio m_g/M_g exclusively for the case when the scattering function is compactly supported within a square in the delay-Doppler plane (see Remark 2). In Fig. 2.3, the ratio m_g/M_g is plotted as a function of TF and for different values of Δ_H for the family of root-raised-cosine WH sets defined in Example 1. The curves in Fig. 2.3 can be used to determine the value of TF needed to achieve a prescribed SIR for a given channel spread. The ratio m_g/M_g increases significantly when TF is taken only slightly larger than 1. A further increase of TF produces a much less pronounced increase of the ratio m_g/M_g .

2.2.2.8 Large-bandwidth and high-SNR regimes

The effect of a loss of degrees of freedom on capacity depends on the operating regime of the system. To illustrate this point, let us consider, for simplicity, an AWGN channel and input signals subject to an average-power constraint only. Two operating regimes can be identified, where the impact of a loss of degrees of freedom is drastically different: the *large-bandwidth* (or, *power-limited*, or *low-SNR*) regime and the *high-SNR* (or *bandwidth-limited*) regime [TV05].

In the large-bandwidth regime, capacity is essentially proportional to the receive power, and only mildly sensitive to the number of degrees of freedom. Hence, a loss of degrees of freedom is irrelevant in this regime. In contrast, in the high-SNR regime, capacity grows

linearly with the number of degrees of freedom and is only mildly sensitive to the receive power. Therefore, a loss of degrees of freedom is critical in this regime.

One of the main contributions of this chapter is to illustrate that for all SNR values of practical interest, the noncoherent capacity of a Rayleigh-fading underspread WSSUS channel is close to the capacity of an AWGN channel with the same receive power. A key element to establish this result is an appropriate choice of the WH set $(g(t), T, F)$ that is used to discretize the channel. Not surprisingly, this choice turns out to depend on the operating regime of the system. In particular,

- i) In the large-bandwidth regime, where capacity is only mildly sensitive to a loss of degrees of freedom, it is sensible to choose $(g(t), T, F)$ so that $\sigma_h^2 \approx 1$ and $\sigma_f^2 \approx 0$, and replace the discretized I/O relation (2.14) by the much simpler diagonalized I/O relation (2.18). In Section 2.3, we study the noncoherent capacity of the diagonalized channel (2.18) in the large-bandwidth regime. Then, in Section 2.4, we assess how well the noncoherent capacity of (2.18) approximates that of (2.14) in this regime. The analysis in these two sections sheds light on the impact of bandwidth, number of antennas, and shape of the scattering function on capacity, and allows us to derive guidelines on the choice of the WH set.
- ii) In the high-SNR regime, where capacity is sensitive to a loss of degrees of freedom, the choice of $(g(t), T, F)$ that leads to $\sigma_h^2 \approx 1$ and $\sigma_f^2 \approx 0$ may be inadequate, as it could entail a dimension loss that is too high. We, therefore, have to work directly with the I/O relation (2.14), which accounts for ISI/ICI. In Section 2.5, we study the noncoherent capacity of (2.14) in the high-SNR regime. In particular, we address the dependency of noncoherent capacity on the support of the scattering function, and we discuss, from an information-theoretic perspective, the trade-off between maximization of the number of degrees of freedom and maximization of SIR.

2.3 The large-bandwidth regime: diagonalized I/O relation

As a first step towards the characterization of the noncoherent capacity of the channel (2.14) in the large-bandwidth regime, we present, in this section, tight bounds on the noncoherent capacity of the diagonalized channel (2.18). In Section 2.4, we will then discuss how well the noncoherent capacity of (2.18) approximates that of (2.14). We chose this two-level approach because bounds on the noncoherent capacity of (2.18) are much easier to derive than for (2.14). Furthermore, the information-theoretic tools presented in this section will also turn out useful for the analysis of the noncoherent capacity of (2.14) presented in Sections 2.4 and 2.5.

Throughout this section, we shall then focus on the diagonalized I/O relation (2.18), which we recall is obtained by discretizing the continuous-time channel I/O relation (2.1) by means of a WH set $(g(t), T, F)$ for which $\sigma_h^2 \approx 1$ and $\sigma_f^2 \approx 0$. As a consequence, the capacity bounds we shall derive in this section, will depend on $(g(t), T, F)$ only through

the grid parameters T and F . A more refined analysis, which is based on the I/O relation with ISI/ICI (2.14) and, hence, leads to bounds explicit in $g(t)$, will be presented in Section 2.4.

We shall assume throughout that the scattering function of the underlying Rayleigh-fading underspread WSSUS channel is compactly supported within $[-\tau_0, \tau_0] \times [-\nu_0, \nu_0]$, and that the grid parameters satisfy the Nyquist condition $T \leq 1/(2\nu_0)$ and $F \leq 1/(2\tau_0)$. These assumptions are not fundamental: they merely serve to simplify the analytic expressions for our capacity bounds.

It is convenient for our analysis to rewrite the I/O relation (2.18) in vector form. As discussed in Section 2.2.2.3, we let $D = KT$ be the approximate duration of the continuous-time input signal $s(t)$ in (2.13) and $B = LF$ be its approximate bandwidth. We denote by \mathbf{s} the KL -dimensional input vector that contains the input symbols $s[k, l]$. The exact way the input symbols $s[k, l]$ are organized into the vector \mathbf{s} is of no concern here, as we will only provide a glimpse of the proof of the capacity bounds. The detailed proof can be found in [DSBS10]. Similarly, the vector \mathbf{r} contains the output-signal samples $r[k, l]$, the vector \mathbf{h} contains the channel coefficients $h[k, l]$, and \mathbf{w} contains the AWGN samples $w[k, l]$. With these definitions, we can now express the I/O relation (2.18) as

$$\mathbf{r} = \mathbf{h} \odot \mathbf{s} + \mathbf{w} \quad (2.22)$$

where \odot denotes the Hadamard element-wise product.

2.3.1 Power constraints

We assume that the average power and the peak power of the input signal are constrained as follows:

- i) The average power satisfies

$$\frac{1}{T} \mathbb{E} \{ \|\mathbf{s}\|^2 \} \leq KP \quad (2.23)$$

where P denotes the admissible average power.

- ii) Among the several ways in which the peak power of the input signal in (2.18) can be constrained (see [DSBS10] for a detailed discussion), here, we exclusively analyze the case where a joint limitation in time and frequency is imposed, i.e., the case where the amplitude of the input symbols $s[k, l]$ in each time-frequency slot (k, l) is constrained:

$$\frac{1}{T} |s[k, l]|^2 \leq \beta \frac{P}{L}. \quad (2.24)$$

Here, $\beta \geq 1$ is the *nominal peak- to average-power ratio* (PAPR). This type of peak constraint is of practical relevance. It models, e.g., a limitation of the radiated peak power in a given frequency band; it also mimics regulatory peak constraints, such as those imposed on UWB systems.

Note that, according to (2.24), the admissible peak power per time-frequency slot goes to zero as the bandwidth [and, hence, L in (2.24)] goes to infinity. An important observation is that the peak constraint in (2.24) depends on the total available bandwidth, rather than the bandwidth that is effectively used by the input signal. As a consequence, it can be shown [DSBS10] that the peak constraint (2.24) enforces the use of the total available bandwidth. Consequently, the PAPR of the signal transmitted in each time-frequency slot is effectively limited, and signals with unbounded PAPR, like *flash signals* [Ver02], are ruled out. Input alphabets commonly used in current systems, like phase-shift keying (PSK) and quadrature-amplitude modulation (QAM), satisfy the constraint (2.24). In contrast, Gaussian inputs, which are often used in information-theoretic analyses, do not satisfy (2.24).

We note that the power constraints (2.23) and (2.24) are imposed on the input symbols $s[k, l]$, rather than on the continuous-time signal $s(t)$. While for the average-power constraint it is possible to make our analysis more rigorous and let the constraint on $s[k, l]$ follow from an underlying constraint on $s(t)$ [DMB09], the same does not seem to hold true for the peak-power constraint: a limit on the amplitude of $s[k, l]$ does not generally imply a limit on the amplitude of $s(t)$.

2.3.2 Definition of noncoherent capacity

Let \mathcal{Q} be the set of probability distributions on \mathbf{s} that satisfy the average-power constraint (2.23) and the peak-power constraint (2.24). The noncoherent capacity of the channel (2.18) is given by [SH05]

$$C \triangleq \lim_{K \rightarrow \infty} \frac{1}{KT} \sup_{\mathcal{Q}} I(\mathbf{r}; \mathbf{s}) \quad (2.25)$$

where $I(\mathbf{r}; \mathbf{s})$ denotes the mutual information between the KL -dimensional output vector \mathbf{r} and the KL -dimensional input vector \mathbf{s} [CT91]. The noncoherent capacity C is notoriously hard to characterize analytically. In the next subsections, we present the following bounds on C :

- i) An upper bound U_c , which we refer to as coherent-capacity upper bound, that is based on the assumption that the receiver has perfect knowledge of the channel realizations. The derivation of this bound is standard (see, e.g., [BPS98, Sec. III.C.1]).
- ii) An upper bound U_1 that is explicit in the channel scattering function and extends the upper bound [SWHL09, Prop. 2.2] on the capacity of frequency-flat time-selective channels to general underspread WSSUS channels.
- iii) A lower bound L_1 that extends the lower bound [SHN05, Prop. 2.2] to general underspread WSSUS channels. This bound is explicit in the channel scattering function only for large bandwidth.

Our focus here will be on the engineering insights that can be obtained from the bounds; we will give a flavor of the derivations and refer the reader to [DSBS10, SDBP09] for detailed derivations.

2.3.3 A coherent-capacity upper bound

The assumptions that the receiver perfectly knows the instantaneous channel realizations and that the input vector \mathbf{s} is subject only to an average-power constraint furnish the following standard capacity upper bound [BPS98, Sec. III.C.1]

$$U_c(B) \triangleq \frac{B}{TF} \mathbb{E}_h \left\{ \ln \left(1 + \frac{PTF}{B} |h|^2 \right) \right\} \quad (2.26)$$

where $h \sim \mathcal{CN}(0, 1)$. As the upper bound U_c increases monotonically with bandwidth, this bound does not reflect the noncoherent capacity behavior for large bandwidth accurately. Nevertheless, we shall see in Section 2.3.6, by means of numerical examples, that U_c is quite useful over a large range of bandwidth values of practical interest.

2.3.4 An upper bound on capacity that is explicit in $C_{\mathbf{H}}(\tau, \nu)$

In this section, we derive an upper bound on C that will help us to better understand the dependency of noncoherent capacity on bandwidth in the large-bandwidth regime. An important feature of this bound is that it is explicit in the channel scattering function. To simplify the exposition of the key steps used to obtain this bound, we first focus on a very simple channel model.

2.3.4.1 Bounding idea

Let $h \sim \mathcal{CN}(0, 1)$ denote the random gain of a memoryless flat-fading channel with input s , output r , and additive noise $w \sim \mathcal{CN}(0, 1)$, i.e., with I/O relation $r = hs + w$. Let \mathcal{Q} denote the set of all probability distributions on s that satisfy the average-power constraint $\mathbb{E}[|s|^2] \leq P$ and the peak constraint $|s|^2 \leq \beta P$. We obtain an upper bound on the capacity $\sup_{\mathcal{Q}} I(r; s)$ along the lines of [SWHL09, Prop. 2.2] (see also [BPS98, p. 2636]) as follows:

$$\begin{aligned} \sup_{\mathcal{Q}} I(r; s) &\stackrel{(a)}{=} \sup_{\mathcal{Q}} \{I(r; s, h) - I(r; h | s)\} \\ &\stackrel{(b)}{\leq} \sup_{\mathcal{Q}} \left\{ \ln(1 + \mathbb{E}\{|s|^2\}) - \mathbb{E}\{\ln(1 + |s|^2)\} \right\} \\ &= \sup_{0 \leq \alpha \leq 1} \sup_{\substack{\mathcal{Q} \\ \mathbb{E}\{|s|^2\} = \alpha P}} \left\{ \ln(1 + \mathbb{E}\{|s|^2\}) - \mathbb{E}\{\ln(1 + |s|^2)\} \right\} \\ &\stackrel{(c)}{\leq} \sup_{0 \leq \alpha \leq 1} \left\{ \ln(1 + \alpha P) - \inf_{\substack{\mathcal{Q} \\ \mathbb{E}\{|s|^2\} = \alpha P}} \mathbb{E} \left\{ \inf_{|u|^2 \leq \beta P} \frac{\ln(1 + |u|^2)}{|u|^2} |s|^2 \right\} \right\} \\ &\stackrel{(d)}{=} \sup_{0 \leq \alpha \leq 1} \left\{ \ln(1 + \alpha P) - \frac{\alpha}{\beta} \ln(1 + \beta P) \right\}. \end{aligned} \quad (2.27)$$

Here, (a) follows from the chain rule for mutual information [CT91]; the inequality (b) results because we take hs as JPG with variance $\mathbb{E}\{|hs|^2\} = \mathbb{E}\{|s|^2\}$. To obtain the in-

equality (c), we multiply and divide $\ln(1 + |s|^2)$ by $|s|^2$ and lower-bound the resulting term $\ln(1 + |s|^2)/|s|^2$ by its infimum over all inputs s that satisfy the peak constraint. Finally, (d) results because $\ln(1 + |u|^2)/|u|^2$ is monotonically decreasing in $|u|^2$, so that its infimum is achieved for $|u|^2 = \beta P$.

If the supremum in (2.27) is achieved for $\alpha = 1$, the upper bound simplifies to

$$C \leq \ln(1 + P) - \frac{1}{\beta} \ln(1 + \beta P).$$

This bound can be interpreted as the capacity of an AWGN channel with SNR equal to P minus a penalty term that quantifies the capacity loss due to channel uncertainty. The higher the allowed peakiness of the input, as measured by the PAPR β , the smaller the penalty.

In spite of its simplicity, the upper bound (2.27) is tight in the low-SNR regime we are interested in (in this section). More precisely, the Taylor-series expansion of the upper bound (2.27) around the point $P = 0$ matches that of capacity up to second order [SWHL09, Prop. 2.1]. In contrast, at high SNR the upper bound (2.27) exhibits an overly optimistic behavior. In fact, the bound scales logarithmically in P , while the high-SNR capacity scaling for memoryless channels is doubly logarithmic [TE97, LM03].

2.3.4.2 The actual bound

For the I/O relation of interest in this chapter, namely, (2.18), the derivation of a bound similar to (2.27) is complicated by the correlation that $h[k, l]$ exhibits in k and l . A key element in this derivation is the relation between mutual information and minimum mean-square error (MMSE) estimation [GSV05], which leads (through the classic formula for the infinite-horizon noncausal prediction error for stationary Gaussian processes [Poo94, Eq. (V.D.28)]) to a closed-form expression that is explicit in the channel scattering function. The resulting upper bound on C is presented in Theorem 1 below. A detailed derivation of this upper bound can be found in [DSBS10].

Theorem 1 *The noncoherent capacity of the channel (2.18), under the assumption that the input signal satisfies the average-power constraint (2.23) and the peak-power constraint (2.24), is upper-bounded as $C \leq U_1$, where*

$$U_1(B) \triangleq \frac{B}{TF} \ln \left(1 + \alpha(B) P \frac{TF}{B} \right) - \alpha(B) \psi(B) \quad (2.28a)$$

with

$$\alpha(B) \triangleq \min \left\{ 1, \frac{B}{TF} \left(\frac{1}{\psi(B)} - \frac{1}{P} \right) \right\} \quad (2.28b)$$

and

$$\psi(B) \triangleq \frac{B}{\beta} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln \left(1 + \frac{\beta P}{B} C_H(\tau, \nu) \right) d\tau d\nu. \quad (2.28c)$$

The bound U_1 vanishes for $B \rightarrow \infty$ [DSBS10, Sec. III.E], a behavior that is to be expected because the peak constraint (2.24) forces the input signal to be spread out over the total available bandwidth. This behavior is well known [MG02, SH02, TT00]. However, as the bound (2.28) is explicit in B , it allows us to characterize the capacity behavior also for finite bandwidth. In particular, through a numerical evaluation of U_1 (and of the lower bound we shall derive in Section 2.3.5), it is possible to (coarsely) identify the *critical bandwidth* for which capacity is maximized (see Section 2.3.6).

By means of a Taylor-series expansion, the bound U_1 can be shown to be tight in the large-bandwidth regime. More precisely, the Taylor-series expansion of U_1 around the point $1/B = 0$ matches that of capacity up to first order [DSBS10].

2.3.4.3 Some simplifications

Similarly to the very simple memoryless flat-fading channel analyzed in Section 2.3.4.1, if the minimum in (2.28b) were attained for $\alpha(B) = 1$, the first term of the upper bound U_1 in (2.28a) could be interpreted as the capacity of an effective AWGN channel with power P and $B/(TF)$ degrees of freedom, while the second term could be seen as a penalty term that characterizes the capacity loss due to channel uncertainty. It turns out, indeed, that $\alpha(B) = 1$ minimizes (2.28b) for virtually all wireless channels and SNR values of practical interest. In particular, a sufficient condition for $\alpha(B) = 1$ is [DSBS10]

$$\Delta_{\mathbf{H}} \leq \frac{\beta}{3TF} \quad (2.29a)$$

and

$$0 \leq \frac{P}{B} < \frac{\Delta_{\mathbf{H}}}{\beta} \left[\exp\left(\frac{\beta}{2TF\Delta_{\mathbf{H}}}\right) - 1 \right]. \quad (2.29b)$$

As virtually all wireless channels are highly underspread, as $\beta \geq 1$, and as, typically, $TF \approx 1.2$ (see Section 2.2.2.7), condition (2.29a) is satisfied in most real-world application scenarios, so that the only relevant condition is (2.29b); but even for large channel spread $\Delta_{\mathbf{H}}$, this condition holds for all SNR values⁶ P/B of practical interest. As an example, consider a system with $\beta = 1$ and spread $\Delta_{\mathbf{H}} = 10^{-2}$; for this choice, (2.29b) is satisfied for all SNR values less than 153 dB. As this value is far in excess of the SNR encountered in practical systems, we can safely claim that a capacity upper bound of practical interest results if we substitute $\alpha(B) = 1$ in (2.28a).

2.3.4.4 Impact of channel characteristics

The spread $\Delta_{\mathbf{H}}$ and the shape of the scattering function $C_{\mathbf{H}}(\tau, \nu)$ are important characteristics of wireless channels. As the upper bound (2.28) is explicit in the scattering function, we can analyze its behavior as a function of $C_{\mathbf{H}}(\tau, \nu)$. We restrict our discussion to the practically relevant case $\alpha(B) = 1$.

⁶Recall that the noise variance was normalized to 1.

Channel spread: For fixed shape of the scattering function, the upper bound U_1 decreases for increasing spread $\Delta_{\mathbf{H}}$. To see this, we define a normalized scattering function $\tilde{C}_{\mathbf{H}}(\tau, \nu)$ supported on a square with unit area, so that

$$C_{\mathbf{H}}(\tau, \nu) = \frac{1}{\Delta_{\mathbf{H}}} \tilde{C}_{\mathbf{H}}\left(\frac{\tau}{2\tau_0}, \frac{\nu}{2\nu_0}\right).$$

By a change of variables, the penalty term can now be written as

$$\begin{aligned} \psi(B) &= \frac{B}{\beta} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln\left(1 + \frac{\beta P}{B} C_{\mathbf{H}}(\tau, \nu)\right) d\tau d\nu \\ &= \frac{B\Delta_{\mathbf{H}}}{\beta} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \ln\left(1 + \frac{\beta P}{B\Delta_{\mathbf{H}}} \tilde{C}_{\mathbf{H}}(\tau, \nu)\right) d\tau d\nu. \end{aligned}$$

Because $\Delta_{\mathbf{H}} \ln(1 + \rho/\Delta_{\mathbf{H}})$ is monotonically increasing in $\Delta_{\mathbf{H}}$ for any positive constant $\rho > 0$, the penalty term $\psi(B)$ increases with increasing spread $\Delta_{\mathbf{H}}$. Consequently, as the first term in (2.28a) does not depend on $\Delta_{\mathbf{H}}$, the upper bound U_1 decreases with increasing spread. Because of the Fourier relation $C_{\mathbf{H}}(\tau, \nu) = \mathbb{E}_{t \rightarrow \nu, f \rightarrow \tau} \{R_{\mathbf{H}}(t, f)\}$, a larger spread implies less correlation in time, frequency, or both; but a channel with less correlation is harder for the receiver to learn; hence, channel uncertainty increases, which ultimately reduces capacity. In a typical system that uses pilot symbols to estimate the channel, a larger spread means that more pilots are required to reliably estimate the channel, so that fewer degrees of freedom are left to transmit data.

Shape of the scattering function: For fixed spread $\Delta_{\mathbf{H}}$, the scattering function that minimizes the upper bound U_1 is the “brick-shaped” scattering function: $C_{\mathbf{H}}(\tau, \nu) = 1/\Delta_{\mathbf{H}}$ for $(\tau, \nu) \in [-\tau_0, \tau_0] \times [-\nu_0, \nu_0]$. This observation follows from Jensen’s inequality applied to the penalty term in (2.28c), the normalization of $C_{\mathbf{H}}(\tau, \nu)$ in (2.4), and the fact that a brick-shaped scattering function achieves the resulting upper bound.

First design sketches of a communication system often rely on simple channel parameters like the maximum multipath delay τ_0 and the maximum Doppler shift ν_0 . These two parameters completely specify a brick-shaped scattering function, which we just saw to provide the minimum capacity upper bound among all WSSUS channels with a scattering function of prescribed τ_0 and ν_0 . Hence, a design on the basis of τ_0 and ν_0 alone is implicitly targeted at the worst-case channel and thus results in a robust design.

2.3.5 A lower bound on capacity

Typically, lower bounds on capacity are easier to obtain than upper bounds because it is sufficient to evaluate the mutual information in (2.25) for an input distribution that satisfies the power constraints (2.23) and (2.24). The main difficulty here is to find input distributions that lead to tight lower bounds. As we are going to show in Section 2.3.6, a good trade-off between analytical tractability and tightness of the resulting bound follows from the choice of i.i.d. zero-mean *constant modulus* input symbols. Constant modulus

input symbols are often found in practical systems in the form of PSK constellations, especially in systems designed to operate at low SNR.

2.3.5.1 Bounding idea

As with the upper bound in Theorem 1, we illustrate the main steps in the derivation of the lower bound by considering a simple memoryless channel with I/O relation $r = hu + w$. Here, we pick u to be of zero mean and constant modulus $|u|^2 = P$. A word of warning is appropriate at this point. The choice of transmitting constant-modulus signals on a memoryless fading channel (with the channel not known at the receiver) is a bad one. It is easy to show that the corresponding mutual information $I(r; u)$ (which is a lower bound on the capacity of the memoryless channel) equals zero. The underlying reason is that in constant modulus signals, the information is encoded in the phase of the signal. But for the setting we just described, the conditional probability of r given u depends on the amplitude of u only. Disturbing as this observation might seem, the bounding steps presented below lead to a perfectly sensible lower bound on the capacity of the channel (2.18) we are interested in. In fact, as the discretized channel $h[k, l]$ in (2.18) exhibits correlation both in k and l , the conditional probability of \mathbf{r} given \mathbf{s} [see (2.22)] depends on both the phase and the amplitude of the entries of \mathbf{s} . Consequently, as we shall see in Section 2.3.6, we have that $I(\mathbf{r}; \mathbf{s}) > 0$ also when constant-modulus inputs are used.

We now derive a lower bound on $I(r; u)$, which not surprisingly will turn out to be negative. We use the chain rule to split the mutual information $I(r; u)$ and obtain

$$\begin{aligned}
 I(r; u) &= I(r; u, h) - I(r; h | u) \\
 &= I(r; h) + I(r; u | h) - I(r; h | u) \\
 &\stackrel{(a)}{\geq} I(r; u | h) - I(r; h | u) \\
 &\stackrel{(b)}{=} I(r; u | h) - \mathbb{E} \left\{ \ln(1 + |u|^2) \right\} \\
 &\stackrel{(c)}{=} I(r; u | h) - \ln(1 + P). \tag{2.30}
 \end{aligned}$$

To get the inequality (a), we used the chain rule for mutual information twice, and then dropped the nonnegative term $I(r; h)$. This essentially splits the mutual information into a first component that corresponds to the case when perfect channel knowledge at the receiver is available, and a second component that [like in the upper bound (2.27)] can be interpreted as a penalty term, and quantifies the impact of channel uncertainty. Next, (b) follows because, given the input u , the output r is JPG with variance $1 + |u|^2$. Finally, (c) results as u is of constant modulus with $|u|^2 = P$. As expected, the lower bound we arrived at is negative. In fact, the first term in the lower bound (2.30), which is a “coherent” mutual information, is always less than the second term, which is the capacity of an AWGN channel with the same receive power.

If the input signal is subject to a peak-power constraint, with PAPR β strictly larger than 1, we can improve upon (2.30) by *time sharing* [SHN05, Cor. 2.1]: we let the input signal have squared magnitude γP during a fraction $1/\gamma$ of the total transmission time, where $1 \leq \gamma \leq \beta$ —that is, we set the channel input s to be $s = \sqrt{\gamma}u$ during this time; for

the remaining time, the transmitter is silent, so that the constraint on the average power is satisfied. The resulting bound is

$$\sup_{\mathcal{Q}} I(r; s) \geq \max_{1 \leq \gamma \leq \beta} \frac{1}{\gamma} \left[I(r; \sqrt{\gamma}u|h) - \ln(1 + \gamma P) \right].$$

Because a closed-form expression for the coherent mutual information $I(r; \sqrt{\gamma}u|h)$, for constant modulus u , does not exist, numerical methods are needed for evaluating the lower bound just derived.

2.3.5.2 The actual bound

Two difficulties arise when trying to derive a bound similar to (2.30) on the capacity of the channel (2.18): we need to account for the correlation that $h[k, l]$ exhibits in k and l and we need to compute the limit $K \rightarrow \infty$ in (2.25). We choose the input symbols to be i.i.d. of zero mean and constant modulus $|s[k, l]|^2 = PT/K$. The coherent mutual information $I(\mathbf{r}; \mathbf{s}|\mathbf{h})$ is then simply given by $KL \times I(r; s|h)$, i.e., KL times the coherent mutual information of the scalar memoryless Rayleigh-fading channel we analyzed previously.

In the limit $K \rightarrow \infty$, the penalty term $I(\mathbf{r}; \mathbf{h}|\mathbf{s})$ is explicit in the matrix-valued spectrum $\mathbf{P}_h(\theta)$ of the multivariate channel process $\{\mathbf{h}[k] \triangleq (h[k, 0] \ h[k, 1] \ \cdots \ h[k, L-1])^T\}$

$$\mathbf{P}_h(\theta) \triangleq \sum_{k=-\infty}^{\infty} \mathbf{R}_h[k] e^{-j2\pi k\theta}, \quad |\theta| \leq \frac{1}{2}, \quad (2.31)$$

where $\mathbf{R}_h[k] \triangleq \mathbb{E}\{\mathbf{h}[k' + k]\mathbf{h}^H[k']\}$. This result follows from a generalization of Szegő's theorem on the asymptotic distribution of Toeplitz matrices (see [DSBS10] for further details).

The final lower bound on capacity is stated in the following theorem. As before, a detailed derivation can be found in [DSBS10].

Theorem 2 *The noncoherent capacity of the channel (2.18), under the assumption that the input signal satisfies the average-power constraint (2.23) and the peak-power constraint (2.24), is lower-bounded as $C \geq L_1$, where*

$$L_1(B) = \max_{1 \leq \gamma \leq \beta} \left\{ \frac{B}{\gamma TF} I(r; \sqrt{\gamma}u|h) - \frac{1}{\gamma T} \int_{-1/2}^{1/2} \ln \det \left(\mathbf{I}_L + \frac{\gamma PTF}{B} \mathbf{P}_h(\theta) \right) d\theta \right\}. \quad (2.32)$$

2.3.5.3 Some approximations

The lower bound (2.32) differs from the upper bound in Theorem 1 in two important aspects. i) The first term inside the braces in (2.32) cannot be expressed in closed form but needs to be evaluated numerically because of the constant modulus signaling assumption. ii) The penalty term depends on the scattering function only indirectly through $\mathbf{P}_h(\theta)$, which again complicates the evaluation of the lower bound. The reason for the complicated structure are edge effects caused by the finite bandwidth $B = LF$. In the large-

bandwidth regime, however, we can approximate the penalty term by exploiting the asymptotic equivalence between Toeplitz and circulant matrices [Gra05]. This yields

$$L_1(B) \approx L_a(B) \triangleq \max_{1 \leq \gamma \leq \beta} \left\{ \frac{B}{\gamma TF} I(r; \sqrt{\gamma} u | h) - \frac{B}{\gamma} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln \left(1 + \frac{\gamma P}{B} C_{\mathbf{H}}(\tau, \nu) \right) d\tau d\nu \right\}. \quad (2.33)$$

Furthermore, we can replace the mutual information $I(r; \sqrt{\gamma} u | h)$ in (2.33) by its first-order Taylor series expansion for $B \rightarrow \infty$ [Ver02, Th. 14], to obtain the approximation

$$L_a(B) \approx L_{aa}(B) \triangleq \max_{1 \leq \gamma \leq \beta} \left\{ P - \frac{\gamma P^2 TF}{B} - \frac{B}{\gamma} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln \left(1 + \frac{\gamma P}{B} C_{\mathbf{H}}(\tau, \nu) \right) d\tau d\nu \right\}. \quad (2.34)$$

Both L_a and L_{aa} are no longer true lower bounds, yet they agree with L_1 in (2.32) for large B . More details on how well L_a and L_{aa} approximate L_1 can be found in [DSBS10].

2.3.6 A numerical example

We evaluate the bounds presented previously for the following set of practically relevant system parameters:

- i) Brick-shaped scattering function with maximum delay $\tau_0 = 0.5 \mu\text{s}$, maximum Doppler shift $\nu_0 = 5 \text{ Hz}$, and corresponding spread $\Delta_{\mathbf{H}} = 4\tau_0\nu_0 = 10^{-5}$.
- ii) Grid parameters $T = 0.35 \text{ ms}$ and $F = 3.53 \text{ kHz}$, so that $TF \approx 1.25$ and $\nu_0 T = \tau_0 F$, as suggested by the design rule (2.19).
- iii) Receive power normalized with respect to the noise spectral density $P/(1 \text{ W/Hz}) = 2.42 \cdot 10^7 \text{ s}^{-1}$.

These parameter values are representative of several different types of systems. For example:

- (a) An IEEE 802.11a system with transmit power of 200 mW, path loss of 118 dB, and receiver noise figure [Raz98] of 5 dB; the path loss is rather pessimistic for typical indoor link distances and includes the attenuation of the signal, e.g., by a concrete wall.
- (b) A UWB system with transmit power of 0.5 mW, path loss of 77 dB, and receiver noise figure of 20 dB.

As illustrated in Fig. 2.4, the upper bound U_1 and the lower bound L_1 (which is evaluated for a QPSK input alphabet) take on their maximum at a large but finite bandwidth; above

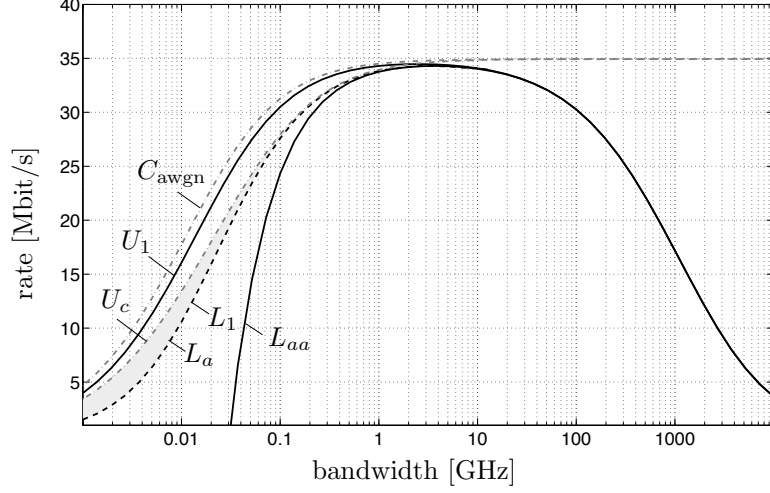


Figure 2.4: The coherent-capacity upper bound U_c in (2.26), the upper bound U_1 in (2.28), as well as the lower bound L_1 in (2.32) (evaluated for a QPSK input alphabet) and its large-bandwidth approximations L_a and L_{aa} in (2.33) and (2.34), respectively, for $\beta = 1$ and a brick-shaped scattering function with spread $\Delta_{\mathbf{H}} = 10^{-5}$. The noncoherent capacity is confined to the area shaded in grey. The AWGN-capacity upper bound (2.35) is also plotted for reference.

this *critical bandwidth*, additional bandwidth is detrimental and the capacity approaches zero as the bandwidth increases further. In this regime, the rate gain resulting from the additional degrees of freedom is offset by the resources required to resolve channel uncertainty. In particular, we can see from Fig. 2.4 that many current wireless systems operate well below the critical bandwidth.

For bandwidth values smaller than the critical bandwidth, L_1 comes quite close to the coherent-capacity upper bound U_c ; we show in Section 2.4.2 that this gap can be further reduced by a more sophisticated choice of the input distribution.

We finally note that, for a large range of bandwidth values of practical interest, both U_1 and L_1 [and, hence, the noncoherent capacity of (2.18)] are close to the capacity of an AWGN channel with the same receive power [a trivial upper bound on (2.25)], given by⁷

$$C_{\text{awgn}}(B) = B \ln \left(1 + \frac{P}{B} \right). \quad (2.35)$$

This observation will be helpful in Section 2.4, where we quantify how well the noncoherent capacity of the diagonalized channel (2.18) approximates that of the channel with ISI/ICI (2.14) in the large-bandwidth regime.

⁷In the remainder of the chapter, we will refer to C_{awgn} as AWGN-capacity upper bound.

2.3.7 Extension to the multi-antenna setting

Bandwidth is the main source of degrees of freedom in wireless communications but it also results in channel uncertainty. Multiple antennas at the transmitter and the receiver can be used to increase the number of degrees of freedom even further. Hence, questions of engineering relevance are (i) how channel uncertainty grows with the number of antennas, (ii) if there is a regime where the additional *spatial* degrees of freedom are detrimental just as the degrees of freedom associated with bandwidth are, and (iii) what is the impact on capacity of spatial correlation across antennas.

2.3.7.1 Modeling multi-antenna channels—a formal extension

To answer the questions above, we need an appropriate model for multi-antenna under-spread WSSUS channels. As the accurate modeling of multi-antenna LTV channels goes beyond the scope of this chapter, we limit our analysis to a multi-antenna channel model that is a *formal* extension of the single-input single-output (SISO) model used so far. More precisely, we extend the diagonalized SISO I/O relation (2.18) to a multiple-input multiple-output (MIMO) I/O relation with M_T transmit antennas, indexed by q , and M_R receive antennas, indexed by r . We assume that all *component channels*, which we denote by $h_{r,q}[k, l]$, are identically distributed, though not necessarily statistically independent.

For a given slot (k, l) , we arrange the corresponding component channels $h_{r,q}[k, l]$ in an $M_R \times M_T$ matrix $\mathbf{H}[k, l]$ with entries $[\mathbf{H}[k, l]]_{r,q} = h_{r,q}[k, l]$. The diagonalized I/O relation of the MIMO channel is then given by

$$\mathbf{r}[k, l] = \mathbf{H}[k, l]\mathbf{s}[k, l] + \mathbf{w}[k, l] \quad (2.36)$$

where, for each slot (k, l) , $\mathbf{w}[k, l]$ is the M_R -dimensional noise vector, $\mathbf{s}[k, l]$ is the M_T -dimensional input vector, and $\mathbf{r}[k, l]$ is the output vector of dimension M_R . We allow for spatial correlation according to the separable (Kronecker) correlation model [CTKV02, KSP⁺02], i.e.,

$$\mathbb{E}[h_{r,q}[k' + k, l' + l]h_{r',q'}^*[k', l']] = B[r, r']A[q, q']R_{\mathbf{H}}[k, l].$$

The $M_T \times M_T$ matrix \mathbf{A} with entries $[\mathbf{A}]_{q,q'} = A[q, q']$ is called the *transmit correlation matrix*, and the $M_R \times M_R$ matrix \mathbf{B} with entries $[\mathbf{B}]_{r,r'} = B[r, r']$ is the *receive correlation matrix*. We normalize \mathbf{A} and \mathbf{B} so that $\text{tr}(\mathbf{A}) = M_T$ and $\text{tr}(\mathbf{B}) = M_R$. Finally, we let $\sigma_0 \geq \sigma_1 \geq \dots \geq \sigma_{M_T-1}$ be the eigenvalues of \mathbf{A} , and $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{M_R-1}$ the eigenvalues of \mathbf{B} .

A detailed discussion and formal description of the MIMO extension just outlined can be found in [Sch09, SDBP09].

2.3.7.2 Capacity bounds for MIMO channels in the large-bandwidth regime

Upper and lower bounds on the capacity of the MIMO channel (2.36), under the average-power constraint

$$\frac{1}{T} \mathbb{E} \left\{ \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} \|\mathbf{s}[k, l]\|^2 \right\} \leq KP \quad (2.37)$$

and the peak-power constraint

$$\frac{1}{T} \|\mathbf{s}[k, l]\|^2 \leq \beta \frac{P}{L} \quad (2.38)$$

can be obtained using the same techniques as in the SISO case. The resulting upper and lower bounds are presented in Theorems 3 and 4 below. A detailed derivation of these bounds can be found in [Sch09, SDBP09].

2.3.7.3 The upper bound

Theorem 3 *The noncoherent capacity of the channel (2.36), under the assumption that the input signal satisfies the average-power constraint (2.37) and the peak-power constraint (2.38), is upper-bounded as $C \leq U_1^{\text{mimo}}$, where*

$$U_1^{\text{mimo}}(B) \triangleq \sup_{0 \leq \alpha \leq \sigma_0} \sum_{r=0}^{M_R-1} \left[\frac{B}{TF} \ln \left(1 + \alpha \lambda_r \frac{PTF}{B} \right) - \alpha \psi_r(B) \right] \quad (2.39a)$$

with

$$\psi_r(B) \triangleq \frac{B}{\sigma_0 \beta} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln \left(1 + \frac{\sigma_0 \lambda_r \beta P}{B} C_{\mathbf{H}}(\tau, \nu) \right) d\tau d\nu. \quad (2.39b)$$

Similarly to the single-antenna case (see Section 2.3.4.3), it can be shown that for virtually all SNR values of practical interest the supremum in (2.39a) is attained for $\alpha = \sigma_0$. Hence, the upper bound can be interpreted as the capacity of a set of M_R parallel AWGN channels with $B/(TF)$ degrees of freedom and received power $\sigma_0 \lambda_r P$, minus a penalty term that quantifies the capacity loss because of channel uncertainty. The observations on the impact of the shape and spread of the scattering function made in Section 2.3.4.4 remain valid.

2.3.7.4 The lower bound

Theorem 4 *The noncoherent capacity of the channel (2.36), under the assumption that the input signal satisfies the average-power constraint (2.37) and the peak-power constraint (2.38), is lower-bounded as*

$$C(B) \geq \max_{1 \leq Q \leq M_T} L_1^{\text{mimo}}(B, Q)$$

where

$$L_1^{\text{mimo}}(B, Q) = \max_{1 \leq \gamma \leq \beta} \left\{ \frac{B}{\gamma TF} I(\tilde{\mathbf{r}}; \sqrt{\gamma} \tilde{\mathbf{s}} | \tilde{\mathbf{H}}) - \frac{1}{\gamma T} \sum_{q=0}^{Q-1} \sum_{r=0}^{M_R-1} \int_{-1/2}^{1/2} \ln \det \left(\mathbf{I}_L + \sigma_q \lambda_r \frac{\gamma PTF}{QB} \mathbf{P}_h(\theta) \right) d\theta \right\}$$

with

$$\tilde{\mathbf{r}} = \mathbf{\Lambda}^{1/2} \tilde{\mathbf{H}} \mathbf{\Sigma}^{1/2} \tilde{\mathbf{s}} + \tilde{\mathbf{w}}.$$

Here, $\tilde{\mathbf{s}}$ is a M_T -dimensional vector whose first Q elements are i.i.d. of zero mean and constant modulus $|\tilde{s}_q|^2 = PT/(QL)$, and the remaining $M_T - Q$ elements are equal to zero. Both the $M_R \times M_T$ matrix $\tilde{\mathbf{H}}$ and the M_R -dimensional vector $\tilde{\mathbf{w}}$ have i.i.d. JPG entries of zero mean and unit variance. Finally, $\mathbf{\Sigma} = \text{diag}\{(\sigma_0 \ \sigma_1 \ \cdots \ \sigma_{M_T-1})^T\}$ and $\mathbf{\Lambda} = \text{diag}\{(\lambda_0 \ \lambda_1 \ \cdots \ \lambda_{M_R-1})^T\}$.

The lower bound L_1^{mimo} needs to be optimized with respect to the number of active transmit eigenmodes Q (i.e., the number of eigenmodes of the transmit correlation matrix \mathbf{A} being signaled over, see [SDBP09] for more details). Note that when the channel is spatially uncorrelated at the transmitter, Q simply denotes the number of active transmit antennas.

It can be shown that, at very large bandwidth, it is optimal to signal along the strongest eigenmode only [SDBP09], a scheme often referred to as *rank-one statistical beamforming* or *eigen-beamforming* [Ver02]. In particular, when the channel is spatially uncorrelated at the transmitter, at very large bandwidth it is optimal to use a single transmit antenna, an observation previously made in [SWHL09] for frequency-flat time-selective channels. At intermediate bandwidth values, the number of transmit antennas to use (for uncorrelated component channels at the transmitter side), or the number of transmit eigenmodes to signal over (if correlation is present), can be determined by numerical evaluation of the bounds, as shown in Section 2.3.8.

2.3.8 Numerical examples

For a 3×3 MIMO system, we show in Figs 2.5–2.7 plots of the upper bound U_1^{mimo} , and—for a number of active transmit eigenmodes Q ranging from 1 to 3—plots of the lower bound L_1^{mimo} and of a large-bandwidth approximation of L_1^{mimo} denoted as L_{aa}^{mimo} .⁸

Parameter settings: All plots are obtained for receive power normalized with respect to the noise spectral density $P/(1 \text{ W/Hz}) = 1.26 \cdot 10^8 \text{ s}^{-1}$; this corresponds, e.g., to a transmit power of 0.5 mW, thermal noise level at the receiver of -174 dBm/Hz , free-space path loss corresponding to a distance of 10 m, and a rather conservative receiver noise figure of 20 dB. Furthermore, we assume that the scattering function is brick-shaped with $\tau_0 = 5 \mu\text{s}$, $\nu_0 = 50 \text{ Hz}$, and corresponding spread $\Delta_{\mathbf{H}} = 10^{-3}$. We also set $\beta = 1$. For this set of parameter values, we analyze three different scenarios: a spatially uncorrelated channel, spatial correlation at the receiver only, and spatial correlation at the transmitter only.

Spatially uncorrelated channel: Fig. 2.5 shows the upper bound U_1^{mimo} and the lower bound L_1^{mimo} for the spatially uncorrelated case. For comparison, we also plot a standard

⁸The analytic expression for L_{aa}^{mimo} , which is similar to (2.34), can be found in [SDBP09, Eq. (19)].

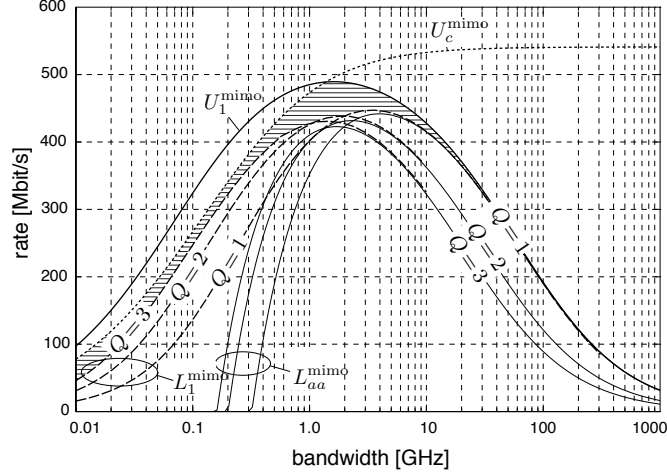


Figure 2.5: Upper and lower bounds on the noncoherent capacity of a spatially uncorrelated MIMO underspread WSSUS channel with $M_T = M_R = 3$, $\beta = 1$, and $\Delta_{\mathbf{H}} = 10^{-3}$. The bounds confine the noncoherent capacity to the hatched area.

capacity upper bound U_c^{mimo} obtained for the coherent setting and with input subject to an average-power constraint only [TV05]. We can observe that U_c^{mimo} is tighter than our upper bound U_1^{mimo} for small bandwidth; this holds true in general, as for small bandwidth and spatially uncorrelated channels, $U_1^{\text{mimo}}(B) \approx [(M_R B)/(TF)] \ln(1 + PTF/B)$, which is the Jensen upper bound on U_c^{mimo} .

For small and medium bandwidth, the lower bound L_1^{mimo} increases with Q and comes surprisingly close to the coherent capacity upper bound U_c^{mimo} for $Q = 3$.

As for the SISO case, when the bandwidth exceeds a certain critical bandwidth, both U_1^{mimo} and L_1^{mimo} start to decrease, because the rate gain due to the additional degrees of freedom is offset by the increase in channel uncertainty. The same argument holds in the wideband regime for the degrees of freedom provided by multiple transmit antennas: in this regime, using a single transmit antenna is optimal [SDBP09].

Impact of receive correlation: Fig. 2.6 shows the same bounds as before, but evaluated for spatial correlation at the receiver according to a correlation matrix with eigenvalues $\{2.6, 0.3, 0.1\}$ and a spatially uncorrelated channel at the transmitter. The curves in Fig. 2.6 are very similar to the ones shown in Fig. 2.5 for the spatially uncorrelated case, yet they are shifted toward higher bandwidth and the maximum is lower. In general, receive correlation reduces capacity at small bandwidth, but is beneficial at large bandwidth [SDBP09].

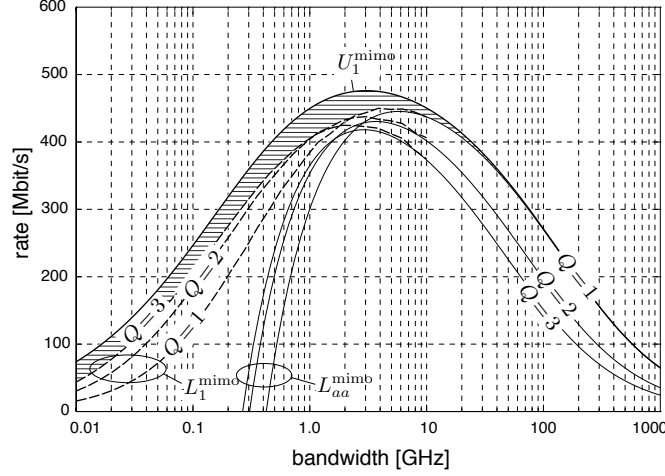


Figure 2.6: Upper and lower bounds on the noncoherent capacity of a MIMO under-spread WSSUS channel that is spatially uncorrelated at the transmitter but correlated at the receiver with eigenvalues of the receive correlation matrix given by $\{2.6, 0.3, 0.1\}$; $M_T = M_R = 3$, $\beta = 1$, and $\Delta_{\mathbf{H}} = 10^{-3}$. The bounds confine the noncoherent capacity to the hatched area.

Impact of transmit correlation: We evaluate the same bounds once more, but this time for spatial correlation at the transmitter according to a correlation matrix with eigenvalues $\{1.7, 1.0, 0.3\}$ and a spatially uncorrelated channel at the receiver. The corresponding curves are shown in Fig. 2.7. The maximum of both the upper and the lower bound is higher than the corresponding maxima in the previous two examples. This rate increase at large bandwidth is caused by the power gain due to statistical beamforming [SDBP09]. The impact of transmit correlation at small bandwidth is more difficult to characterize, because the distance between upper and lower bound is larger than for the spatially uncorrelated case.

An observation of practical importance is that both U_1^{mimo} and L_1^{mimo} are rather flat over a large range of bandwidth values around their respective maxima. Further numerical results (not presented here) point at the following: (i) for smaller values of the channel spread $\Delta_{\mathbf{H}}$, these maxima broaden and extend toward higher bandwidth; (ii) an increase in β increases the gap between upper and lower bounds.

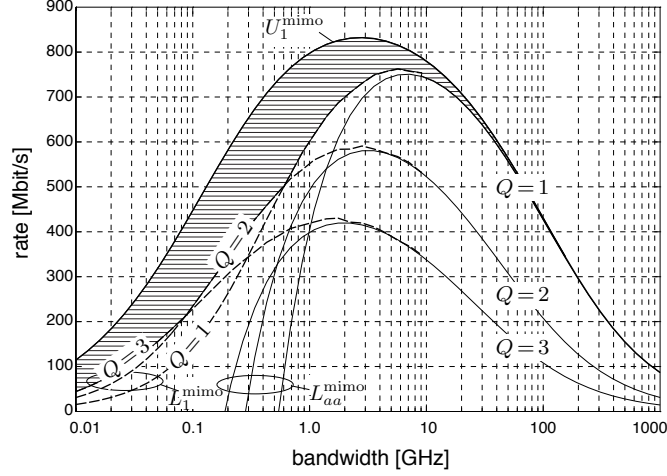


Figure 2.7: Upper and lower bounds on the noncoherent capacity of a MIMO underspread WSSUS channel that is spatially correlated at the transmitter with eigenvalues of the transmit correlation matrix $\{1.7, 1.0, 0.3\}$ and uncorrelated at the receiver; $M_T = M_R = 3$, $\beta = 1$, and $\Delta_H = 10^{-3}$. The bounds confine the noncoherent capacity to the hatched area.

2.4 The large-bandwidth regime: I/O relation with interference

So far, we based our analysis on the diagonalized I/O relation (2.18). The goal of this section is to determine how well the noncoherent capacity of (2.18) approximates that of the channel with ISI and ICI (2.14) in the large-bandwidth regime. The presence of interference makes the derivation of tight capacity bounds (in particular, upper bounds) technically challenging. Nevertheless, our analysis will be sufficient to establish that for a large range of bandwidth values of practical interest, the presence of interference does not change the noncoherent capacity behavior significantly, whenever the channel is underspread.

To establish this result, we derive a lower bound on the capacity of (2.14) by treating interference as noise. We then show that whenever the channel is underspread, this lower bound, evaluated for an appropriately chosen root-raised-cosine WH set (see Example 1), is close for a large range of bandwidth values of practical interest both to the lower bound L_1 we derived in the previous section [for the diagonalized I/O relation (2.18)] and to the AWGN-capacity upper bound C_{awgn} .

To get an I/O relation in vector form, similar to the one we worked with in the previous section, we arrange the intersymbol and intercarrier interference terms $\{z[k', l', k, l]\}$

in (2.14) in a $KL \times KL$ matrix \mathbf{Z} with entries

$$[\mathbf{Z}]_{(l'+k'L)(l+kL)} = \begin{cases} z[k', l', k, l], & \text{if } (k', l') \neq (k, l) \\ 0, & \text{otherwise.} \end{cases}$$

This definition, together with the definitions in Section 2.3, allows us to compactly express (2.14) as

$$\mathbf{r} = \mathbf{h} \odot \mathbf{s} + \mathbf{Z}\mathbf{s} + \mathbf{w}.$$

For a given WH set, the noncoherent capacity C of the induced discretized channel (2.14) is defined as in (2.25). We next derive a lower bound on C . As in Section 2.3, we first illustrate the main steps in the derivation of the lower bound using a simplified setting.

2.4.1 A lower bound on capacity

2.4.1.1 The bounding idea

We consider, for simplicity, a block-fading channel with block length 2 and I/O relation

$$\begin{aligned} r[1] &= hs[1] + zs[2] + w[1] \\ r[2] &= hs[2] + zs[1] + w[2] \end{aligned} \tag{2.40}$$

where $h, w[1], w[2] \sim \mathcal{CN}(0, 1)$ are mutually independent, while $z \sim \mathcal{CN}(0, \sigma_z^2)$ is independent of the noise samples $w[1]$ and $w[2]$, but not necessarily independent of h . For this setting, the interference matrix \mathbf{Z} reduces to

$$\mathbf{Z} = \begin{bmatrix} 0 & z \\ z & 0 \end{bmatrix}$$

and the I/O relation (2.40) can be recast as:

$$\mathbf{r} = h\mathbf{s} + \mathbf{Z}\mathbf{s} + \mathbf{w}.$$

Let \mathcal{Q} be the set of probability distributions on \mathbf{s} satisfying the average-power constraint $\mathbb{E}\{|s[i]|^2\} \leq P$ and the peak constraint $|s[i]|^2 \leq \beta P$, for $i = 1, 2$. To obtain a lower bound on noncoherent capacity, which is given by⁹ $(1/2) \sup_{\mathcal{Q}} I(\mathbf{r}; \mathbf{s})$, we compute the mutual information for a specific probability distribution in \mathcal{Q} . In particular, we take a probability distribution for which $s[1]$ and $s[2]$ are i.i.d. and $h(s[1]) > -\infty$, where $h(\cdot)$ denotes differential entropy. We denote by ρ_s the average power of $s[1]$ and $s[2]$, i.e., $\rho_s \triangleq \mathbb{E}\{|s[1]|^2\} = \mathbb{E}\{|s[2]|^2\} \leq P$. The presence of the interference term in (2.40) makes the derivation of a capacity lower bound involved even for this simple setting; bounding steps more sophisticated than the ones we used in Section 2.3.5.1 are

⁹The factor 1/2 arises because of the normalization with respect to the block length (2 in this case).

needed. As in (2.30), we first use the chain rule and the non-negativity of mutual information to split $I(\mathbf{r}; \mathbf{s})$ as

$$\begin{aligned} I(\mathbf{r}; \mathbf{s}) &= I(\mathbf{r}; h) + I(\mathbf{r}; \mathbf{s} | h) - I(\mathbf{r}; h | \mathbf{s}) \\ &\geq I(\mathbf{r}; \mathbf{s} | h) - I(\mathbf{r}; h | \mathbf{s}). \end{aligned} \quad (2.41)$$

We next bound each of the two terms on the right-hand side (RHS) of (2.41).

A lower bound on the first term: To lower-bound the first term on the RHS of (2.41), we use the chain rule for differential entropy, the independence of $s[1]$ and $s[2]$ and the fact that conditioning reduces entropy [CT91, Sec. 9.6]:

$$\begin{aligned} I(\mathbf{r}; \mathbf{s} | h) &= h(\mathbf{s} | h) - h(\mathbf{s} | \mathbf{r}, h) \\ &= h(s[1]) + h(s[2]) - h(s[1] | \mathbf{r}, h) - h(s[2] | \mathbf{r}, h, s[1]) \\ &\geq h(s[1]) + h(s[2]) - h(s[1] | r[1], h) - h(s[2] | r[2], h) \\ &= I(r[1]; s[1] | h) + I(r[2]; s[2] | h) \\ &= 2I(r[1]; s[1] | h). \end{aligned} \quad (2.42)$$

The mutual information $I(r[1]; s[1] | h)$ is still difficult to evaluate because of the presence of the interference term $zs[2]$. A simple way to deal with this issue is to treat the interference as noise, as shown below [Lap05, App. I]:

$$\begin{aligned} I(r[1]; s[1] | h) &= h(s[1]) - h(s[1] | r[1], h) \\ &\stackrel{(a)}{=} h(s[1]) - \min_{\delta} h(s[1] - \delta r[1] | r[1], h) \\ &\stackrel{(b)}{\geq} h(s[1]) - \min_{\delta} h(s[1] - \delta r[1] | h) \\ &\stackrel{(c)}{\geq} h(s[1]) - \min_{\delta} \mathbb{E}_h \left\{ \ln \left[\pi e \mathbb{E} \left\{ |s[1] - \delta r[1]|^2 | h \right\} \right] \right\} \\ &\stackrel{(d)}{=} h(s[1]) - \mathbb{E}_h \left\{ \ln \left[\pi e \frac{\rho_s (\sigma_z^2 \rho_s + 1)}{|h|^2 \rho_s + \sigma_z^2 \rho_s + 1} \right] \right\} \\ &= \mathbb{E}_h \left\{ \ln \left(1 + \frac{|h|^2 \rho_s}{\sigma_z^2 \rho_s + 1} \right) \right\} - [\ln(\pi e \rho_s) - h(s[1])]. \end{aligned} \quad (2.43)$$

In (a) we used the fact that the differential entropy is invariant to translations, in (b) that conditioning reduces entropy, (c) follows because the Gaussian distribution maximizes the differential entropy, and (d) follows by choosing δ to be the coefficient of the linear estimator of $s[1]$ (from $r[1]$) that minimizes the mean square error. Note that the term $\ln(\pi e \rho_s) - h(s[1])$ is always positive, and can be interpreted as a penalty term that quantifies the rate loss due to the peak constraint. In fact, for any distribution for which $\mathbb{E}\{|s[1]|^2\} = \rho_s$, we have $h(s[1]) \leq \ln(\pi e \rho_s)$, with equality if and only if $s[1] \sim \mathcal{CN}(0, \rho_s)$. But the complex Gaussian distribution does not belong to \mathcal{Q} , because it violates the peak constraint.

An upper bound on the second term: We next upper-bound the second term on the RHS of (2.41). Let $\mathbf{w}^{(1)} \sim \mathcal{CN}(\mathbf{0}, \alpha \mathbf{I}_2)$ and $\mathbf{w}^{(2)} \sim \mathcal{CN}(\mathbf{0}, (1 - \alpha) \mathbf{I}_2)$, where $0 < \alpha < 1$, be two independent JPG vectors. Furthermore, let $\mathbf{r}^{(1)} = h\mathbf{s} + \mathbf{w}^{(1)}$ and $\mathbf{r}^{(2)} = \mathbf{Z}\mathbf{s} + \mathbf{w}^{(2)}$. By the data-processing inequality and the chain rule for mutual information, we have that

$$\begin{aligned} I(\mathbf{r}; h | \mathbf{s}) &\leq I(\mathbf{r}^{(1)}, \mathbf{r}^{(2)}; h | \mathbf{s}) \\ &= I(\mathbf{r}^{(1)}; h | \mathbf{s}) + I(\mathbf{r}^{(2)}; h | \mathbf{s}, \mathbf{r}^{(1)}). \end{aligned} \quad (2.44)$$

The first term on the RHS of (2.44) can be upper-bounded according to

$$I(\mathbf{r}^{(1)}; h | \mathbf{s}) = \mathbb{E} \left\{ \ln \left(1 + \frac{\|\mathbf{s}\|^2}{\alpha} \right) \right\} \leq \ln \left(1 + \frac{\rho_s}{\alpha} \right).$$

For the second term on the RHS of (2.44), we proceed as follows:

$$\begin{aligned} I(\mathbf{r}^{(2)}; h | \mathbf{s}, \mathbf{r}^{(1)}) &= h(\mathbf{r}^{(2)} | \mathbf{s}, \mathbf{r}^{(1)}) - h(\mathbf{r}^{(2)} | \mathbf{s}, \mathbf{r}^{(1)}, h) \\ &\stackrel{(a)}{=} h(\mathbf{r}^{(2)} | \mathbf{s}, \mathbf{r}^{(1)}) - h(\mathbf{r}^{(2)} | \mathbf{s}, h) \\ &\stackrel{(b)}{\leq} h(\mathbf{r}^{(2)} | \mathbf{s}) - h(\mathbf{r}^{(2)} | \mathbf{s}, h, \mathbf{Z}) \\ &\stackrel{(c)}{=} h(\mathbf{r}^{(2)} | \mathbf{s}) - h(\mathbf{r}^{(2)} | \mathbf{s}, \mathbf{Z}) \\ &= I(\mathbf{r}^{(2)}; \mathbf{Z} | \mathbf{s}) \\ &= \mathbb{E} \left\{ \ln \left(1 + \frac{\sigma_z^2}{1 - \alpha} \|\mathbf{s}\|^2 \right) \right\} \\ &\leq \ln \left(1 + \frac{\sigma_z^2 \rho_s}{1 - \alpha} \right). \end{aligned}$$

Here, (a) holds because $\mathbf{r}^{(1)}$ and $\mathbf{r}^{(2)}$ are conditionally independent given \mathbf{s} and h , in (b) we used that conditioning reduces entropy, and (c) follows because $\mathbf{r}^{(2)}$ and h are conditionally independent given \mathbf{Z} . Combining the two bounds, we get

$$I(\mathbf{r}; h | \mathbf{s}) \leq \ln \left(1 + \frac{\rho_s}{\alpha} \right) + \ln \left(1 + \frac{\rho_s \sigma_z^2}{1 - \alpha} \right).$$

Furthermore, as the bound holds for any $\alpha \in (0, 1)$, we have

$$I(\mathbf{r}; h | \mathbf{s}) \leq \inf_{0 < \alpha < 1} \left\{ \ln \left(1 + \frac{\rho_s}{\alpha} \right) + \ln \left(1 + \frac{\rho_s \sigma_z^2}{1 - \alpha} \right) \right\}. \quad (2.45)$$

Completing the bound: To get the final bound, we first insert (2.43) into (2.42), and then (2.42) and (2.45) into (2.41):

$$\begin{aligned} \sup_{\mathcal{Q}} I(\mathbf{r}; \mathbf{s}) &\geq 2 \left[\mathbb{E}_h \left\{ \ln \left(1 + \frac{|h|^2 \rho_s}{\sigma_z^2 \rho_s + 1} \right) \right\} - \left(\ln(\pi e \rho_s) - h(s[1]) \right) \right] \\ &\quad - \inf_{0 < \alpha < 1} \left\{ \ln \left(1 + \frac{\rho_s}{\alpha} \right) + \ln \left(1 + \frac{\rho_s \sigma_z^2}{1 - \alpha} \right) \right\}. \end{aligned} \quad (2.46)$$

The bound just obtained can be tightened by maximizing it over the set of probability distributions on $s[1]$ that satisfy the average-power constraint $\mathbb{E}\{|s[1]|^2\} \leq P$ and the peak-power constraint $|s[1]|^2 \leq \beta P$.

2.4.1.2 The actual bound

The application of the bounding steps just illustrated to the channel (2.14), is made difficult by the correlation exhibited by $h[k, l]$ and $z[k', l', k, l]$. We deal with this difficulty as follows. As in Section 2.3.5.2, the stationarity of $h[k, l]$ in k allows us to use Szegő's theorem on the eigenvalue distribution of Toeplitz matrices to obtain a bound explicit in the matrix-valued spectral density $\mathbf{P}_h(\theta)$ defined in (2.31). The statistical properties of $z[k', l', k, l]$ are captured through the interference variance σ_I^2 defined in (2.17), which replaces σ_z^2 in (2.46). More details on these bounding steps can be found in [DMBon].

As in Section 2.3, we consider, for simplicity of exposition, scattering functions that are compactly supported within the rectangle $[-\tau_0, \tau_0] \times [-\nu_0, \nu_0]$ and grid parameters satisfying the Nyquist condition $T \leq 1/(2\nu_0)$ and $F \leq 1/(2\tau_0)$. The resulting lower bound on capacity is presented in the following theorem, whose proof can be found in [DMBon].

Theorem 5 *Let $(g(t), T, F)$ be a WH set and consider a Rayleigh-fading WSSUS channel with scattering function $C_H(\tau, \nu)$. Then, for a given bandwidth B , the noncoherent capacity of the discretized channel (2.14) induced by $(g(t), T, F)$, with input subject to the average-power constraint (2.23) and the peak-power constraint (2.24), is lower-bounded as $C \geq L_1^{\text{int}}$, where*

$$L_1^{\text{int}}(B) \triangleq \frac{B}{TF} \left[\mathbb{E}_h \left\{ \ln \left(1 + \frac{TF |h|^2 \rho_s \sigma_h^2}{1 + TF \rho_s \sigma_I^2} \right) \right\} - \left(\ln(\pi e \rho_s) - h(s) \right) \right] \\ - \inf_{0 < \alpha < 1} \left\{ \frac{1}{T} \int_{-1/2}^{1/2} \ln \det \left(\mathbf{I}_L + \frac{TF \rho_s}{\alpha} \mathbf{P}_h(\theta) \right) d\theta + \frac{B}{TF} \ln \left(1 + \frac{TF \rho_s \sigma_I^2}{1 - \alpha} \right) \right\}. \quad (2.47)$$

Here, $h \sim \mathcal{CN}(0, 1)$, and s is a complex random variable that satisfies

$$\rho_s \triangleq \mathbb{E}\{|s|^2\} \leq \frac{P}{B} \quad \text{and} \quad |s|^2 \leq \beta \frac{P}{B}.$$

For WH sets that result in $\sigma_h^2 \approx 1$ and $\sigma_I^2 \approx 0$ (see Section 2.2.2.7), the lower bound L_1^{int} is close—for a large range of bandwidth values of practical interest—to the lower bound L_1 in (2.32). A qualitative justification of this statement is provided below, followed by numerical results in Section 2.4.2. When $\sigma_h^2 \approx 1$ and $\sigma_I^2 \approx 0$, we can approximate L_1^{int} as

$$L_1^{\text{int}}(B) \approx \frac{B}{TF} \left[\mathbb{E}_h \left\{ \ln \left(1 + TF |h|^2 \rho_s \right) \right\} - \left(\ln(\pi e \rho_s) - h(s) \right) \right] \\ - \frac{1}{T} \int_{-1/2}^{1/2} \ln \det(\mathbf{I}_L + TF \rho_s \mathbf{P}_h(\theta)) d\theta. \quad (2.48)$$

If, for simplicity, we neglect the peak constraint and take $s \sim \mathcal{CN}(0, \rho_s)$, we have that $\ln(\pi e \rho_s) - h(s) = 0$. Furthermore, in the large-bandwidth regime of interest in this section, we can approximate the first term on the RHS of (2.48) by its first-order Taylor series expansion, and obtain (for $\rho_s = P/B$)

$$\mathbb{E}_h \left\{ \ln \left(1 + TF |h|^2 \rho_s \right) \right\} \approx P - \frac{P^2 TF}{B}.$$

This expansion coincides with that of the mutual information $I(r; \sqrt{\gamma}u|h)$ in (2.32) [see also (2.34)]. From this approximate analysis, it follows that, at least for Gaussian inputs, the difference between L_1^{int} and L_1 is small in the large-bandwidth regime (when $\sigma_h^2 \approx 1$ and $\sigma_f^2 \approx 0$). In Section 2.4.2, we present numerical results that support this statement. Specifically, these numerical results reveal that:

- i) When the channel \mathbf{H} is underspread, there exist WH sets for which (for Gaussian inputs) $L_1^{\text{int}} \approx L_1$ in the large-bandwidth regime; one such set is the root-raised-cosine WH set of Example 1, with T and F chosen so that the grid matching rule (2.19) is satisfied. Furthermore, both L_1^{int} and L_1 are close to the AWGN-capacity upper bound C_{awgn} for a large range of bandwidth values of practical interest.
- ii) The difference between L_1^{int} and L_1 can be made small for a large range of bandwidth values of practical interest also when the Gaussian input distribution is replaced by an appropriate distribution that satisfies the peak constraint $|s|^2 \leq \beta P$.

Before providing the corresponding numerical results, we make three remarks on L_1^{int} , which simplify its numerical evaluation.

2.4.1.3 Large-bandwidth approximation

To ease the numerical evaluation of L_1^{int} , we proceed as in Section 2.3.5.3. Specifically, we use the asymptotic equivalence between Toeplitz and circulant matrices [Gra05], to obtain the following large-bandwidth approximation of the first term inside the braces in (2.47):

$$\begin{aligned} \frac{1}{T} \int_{-1/2}^{1/2} \ln \det \left(\mathbf{I}_L + \frac{TF \rho_s}{\alpha} \mathbf{P}_h(\theta) \right) d\theta \\ \approx B \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln \left(1 + \frac{\rho_s}{\alpha} C_{\mathbf{H}}(\tau, \nu) |A_g(\tau, \nu)|^2 \right) d\tau d\nu \\ \leq B \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln \left(1 + \frac{\rho_s}{\alpha} C_{\mathbf{H}}(\tau, \nu) \right) d\tau d\nu. \end{aligned} \quad (2.49)$$

In the first step, we used the fact that $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported and that T and F are chosen so as to satisfy the Nyquist condition $T \leq 1/(2\nu_0)$ and $F \leq 1/(2\tau_0)$. The last step follows because $|A_g(\tau, \nu)| \leq 1$. If we further use that $\sigma_h^2 \geq m_g$ and $\sigma_f^2 \leq M_g$ [see (2.20) and (2.21), respectively], we get the following large-bandwidth approximation

of L_1^{int} , which is easier to evaluate numerically than L_1^{int} :

$$L_1^{\text{int}}(B) \gtrsim L_a^{\text{int}}(B) \triangleq \frac{B}{TF} \left[\mathbb{E}_h \left\{ \ln \left(1 + \frac{TF|h|^2 \rho_s m_g}{1 + TF \rho_s M_g} \right) \right\} - \left(\ln(\pi e \rho_s) - h(s) \right) \right. \\ \left. - \inf_{0 < \alpha < 1} \left\{ TF \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln \left(1 + \frac{\rho_s}{\alpha} C_{\mathbf{H}}(\tau, \nu) \right) d\tau d\nu + \ln \left(1 + \frac{TF \rho_s M_g}{1 - \alpha} \right) \right\} \right]. \quad (2.50)$$

2.4.1.4 Reduction to a square setting

The evaluation of the lower bound is simplified if T and F are chosen according to the grid matching rule (2.19). In this case, a simple coordinate transformation yields the following result.

Lemma 6 *Let $(g(t), T, F)$ be an orthonormal WH set, and assume that the channel scattering function $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported within the rectangle $[-\tau_0, \tau_0] \times [-\nu_0, \nu_0]$. Then, for any $\zeta > 0$, we have*

$$L_a^{\text{int}}(B, g(t), T, F, C_{\mathbf{H}}(\tau, \nu)) = L_a^{\text{int}} \left(B, \sqrt{\zeta} g(\zeta t), \frac{T}{\zeta}, \zeta F, C_{\mathbf{H}} \left(\frac{\tau_0}{\zeta}, \zeta \nu_0 \right) \right).$$

In particular, assume that $\nu_0 T = \tau_0 F$ and let $\zeta = \sqrt{T/F} = \sqrt{\tau_0/\nu_0}$, $\tilde{g}(t) \triangleq \sqrt{\zeta} g(\zeta t)$, and $\tilde{C}_{\mathbf{H}}(\tau, \nu) = C_{\mathbf{H}}(\tau_0/\zeta, \zeta \nu_0)$, where $\tilde{C}_{\mathbf{H}}(\tau, \nu)$ is compactly supported within the square $[-\sqrt{\Delta_{\mathbf{H}}}/2, \sqrt{\Delta_{\mathbf{H}}}/2] \times [-\sqrt{\Delta_{\mathbf{H}}}/2, \sqrt{\Delta_{\mathbf{H}}}/2]$, with $\Delta_{\mathbf{H}} = 4\tau_0\nu_0$. Then,

$$L_a^{\text{int}}(B, g(t), T, F, C_{\mathbf{H}}(\tau, \nu)) = L_a^{\text{int}}(B, \tilde{g}(t), \sqrt{TF}, \sqrt{TF}, \tilde{C}_{\mathbf{H}}(\tau, \nu)).$$

In what follows, for the sake of simplicity of exposition, we choose T and F such that the grid matching rule $\nu_0 T = \tau_0 F$ is satisfied. Then, as a consequence of Lemma 6, we will, without loss of generality, only consider WH sets of the form $(g(t), \sqrt{TF}, \sqrt{TF})$ and WSSUS channels with scattering function compactly supported within a square.

2.4.1.5 Maximization of the lower bound

The lower bound L_a^{int} in (2.50) can be tightened by maximizing it over the orthonormal WH set and the probability distribution on s . The maximization over all orthonormal WH sets is difficult to carry out, because the dependency of m_g and M_g on the WH set $(g(t), T, F)$ is, in general, hard to characterize analytically. This problem can be partially overcome when $\Delta_{\mathbf{H}} \ll 1$. In this case, a first-order Taylor-series expansion of m_g and M_g around $\Delta_{\mathbf{H}} = 0$ yields an accurate picture, as discussed in [DMB09]. Here, we simplify the maximization problem by considering only WH sets that are based on a root-raised-cosine pulse (see Example 1).

For simplicity, we also avoid the maximization over the probability distribution on s . Instead, we consider a simple distribution, obtained by truncating a complex Gaussian distribution so that s satisfies the peak constraint $|s|^2 \leq \beta P/B$. More specifically, we

take $s = (\sqrt{P/B})\tilde{s}$, where the phase of \tilde{s} is uniformly distributed on $[0, 2\pi)$; furthermore, $|\tilde{s}|^2 = z$ is distributed according to

$$f_z(z) = c_1 e^{-z/c_2} 1_{[0, \beta)}(z)$$

where $c_1 = c_2/(1 - e^{-\beta/c_2})$, and c_2 is chosen so that $\mathbb{E}\{|\tilde{s}|^2\} = 1$. A straightforward calculation reveals that for this input distribution

$$h(\tilde{s}) = \log c_1 + \frac{1}{c_2} + \ln \pi.$$

Hence, the penalty term $\ln(\pi e \rho_s) - h(s)$ in (2.50) reduces to

$$\begin{aligned} \ln(\pi e \rho_s) - h(s) &\stackrel{(a)}{=} \ln(\pi e) - h(\tilde{s}) \\ &= 1 - \log c_1 - \frac{1}{c_2} \end{aligned}$$

where in (a) we used that $\rho_s = P/B$, by construction. Note that, for the input distribution under consideration, the penalty term $\ln(\pi e \rho_s) - h(s)$ is independent of the SNR P/B , and depends only on the PAPR β through c_1 and c_2 (more precisely, the penalty term is monotonically decreasing in β , and vanishes in the limit $\beta \rightarrow \infty$). In the large-bandwidth regime, the lower bound L_a^{int} turns out to be highly sensitive to the value of this penalty term.

2.4.2 Numerical examples

In this section, we evaluate the bound L_a^{int} for a set of parameters similar to the one considered in Section 2.3.6. In particular, we take $P/(1 \text{ W/Hz}) = 2.42 \cdot 10^7 \text{ s}^{-1}$ and $TF = 1.25$, and consider a brick-shaped scattering function with $\Delta_{\mathbf{H}} = 10^{-5}$ (see Section 2.3.6 for a discussion on the practical relevance of this set of parameters). Furthermore, we focus on WH sets based on a root-raised-cosine pulse (see Example 1), and assume $\beta = 10$. Finally, we take as probability distribution on s , the truncated complex Gaussian distribution discussed in Section 2.4.1.5.

Fig. 2.8 shows the AWGN-capacity upper bound C_{awgn} in (2.35) (same curve as in Fig. 2.4), the lower bound L_1 in (2.32), which, differently from Fig. 2.4 is evaluated for $\beta = 10$, and the lower bound L_a^{int} , evaluated for $\beta = 6, 8$, and 10 , to illustrate the impact of the penalty term $\ln(\pi e \rho_s) - h(s)$ on L_a^{int} for different values of β . As mentioned above, L_a^{int} is extremely sensitive to β in the large-bandwidth regime. Fortunately, this sensitivity manifests itself only for bandwidth values that are significantly above those encountered in practical systems. We can observe that for $\beta = 10$ the difference between L_1 and L_a^{int} is negligible over a large range of bandwidth values of practical interest. Furthermore, also the difference between L_a^{int} and C_{awgn} is small. We can, therefore, conclude that the presence of interference in (2.14) does not change the capacity behavior significantly for a large range of bandwidth values of practical interest.

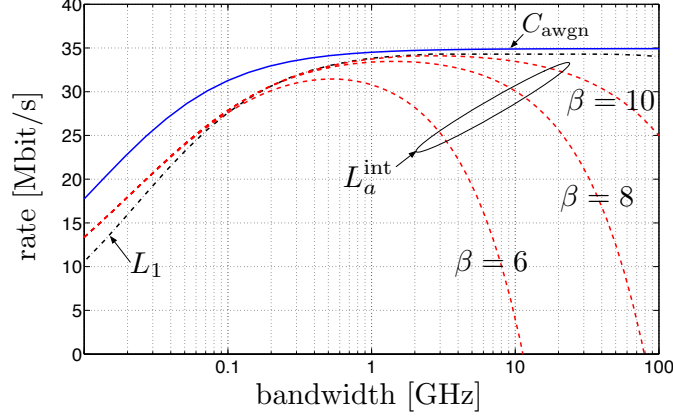


Figure 2.8: The AWGN-capacity upper bound C_{awgn} in (2.35), the lower bound L_1 in (2.32) for $\beta = 10$ and the lower bound L_a^{int} in (2.50) for $\beta = 6, 8$, and 10 . The channel scattering function is brick-shaped, with $\Delta_{\mathbf{H}} = 10^{-5}$.

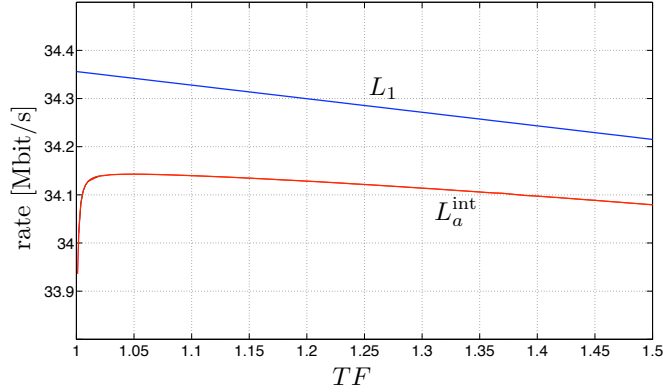


Figure 2.9: The lower bounds L_1 in (2.32) and L_a^{int} in (2.50), for $B = 3\text{GHz}$. The channel scattering function is brick-shaped with $\Delta_{\mathbf{H}} = 10^{-5}$. Furthermore, $\beta = 10$.

We next analyze the dependency of L_1 and L_a^{int} on TF . Fig. 2.9 shows the lower bounds L_1 and L_a^{int} as a function of TF for $B = 3\text{GHz}$, which roughly corresponds to the capacity-maximizing bandwidth for the setting considered in Fig. 2.8. Even though the lower bound L_1 decreases monotonically with TF , the rate loss experienced when TF is increased from 1 to 1.5 is extremely small (about 0.1%). This is in good agreement with our claim that, in the large-bandwidth regime, capacity is only mildly sensitive to a loss of degrees of freedom. The lower bound L_a^{int} is surprisingly close to L_1 over almost the entire range of TF values considered. Only for $TF \approx 1$, the gap widens significantly. As

remarked in Example 1, when $TF = 1$, the pulse $g(t)$ reduces to a sinc function, which has poor time localization. As a consequence, the variance of the interference term increases (as illustrated in Fig. 2.3), and so does the gap between L_1 and L_a^{int} .

Compared to the characterization of the noncoherent capacity of the diagonalized channel (2.18), the characterization we provided here on the basis of L_a^{int} and C_{awgn} is less accurate. At moderate bandwidth values, our analysis could be strengthened by replacing C_{awgn} with a tighter upper bound based on perfect CSI at the receiver. Finding such a bound is an interesting open problem. In the very large bandwidth regime, the difference between C_{awgn} and L_a^{int} is large. In fact C_{awgn} approaches P as $B \rightarrow \infty$ [see (2.35)], while L_a^{int} approaches 0. To obtain a more accurate capacity characterization in the very large bandwidth regime, both upper and lower bound would need to be tightened. From an engineering point of view, however, our analysis is sufficient to (coarsely) determine the optimal (i.e., critical) bandwidth at which to operate.

2.5 The high-SNR regime

Next, we consider the high-SNR regime. As argued in Section 2.2.2.8, in this regime the choice of the WH set $(g(t), T, F)$ that leads to $\sigma_h^2 \approx 1$ and $\sigma_f^2 \approx 0$ might be suboptimal because of a loss of degrees of freedom. Hence, we work with the I/O relation (2.14) directly. In the previous section, we considered, for simplicity of exposition, only scattering functions that are compactly supported in the delay-Doppler plane. Here, we drop this assumption and consider the larger class of WSSUS channels that satisfy the more general underspread notion in Definition 1. We recall that the crucial parameters in Definition 1 are the channel spread $\Delta_{\mathbf{H}} = 4\tau_0\nu_0$, which is the area of the rectangle in the delay-Doppler plane that supports most of the volume of the scattering function, and ε , which is the maximal fraction of volume of $C_{\mathbf{H}}(\tau, \nu)$ that lies outside the rectangle of area $\Delta_{\mathbf{H}}$. The compact-support assumption on $C_{\mathbf{H}}(\tau, \nu)$ is dropped because the results summarized in Section 2.1.2 hint at a high sensitivity of capacity in the high-SNR regime to whether $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported or not, a property of $C_{\mathbf{H}}(\tau, \nu)$ that, however, cannot be verified through measurements (see Section 2.2.1.2).

The goal of the high-SNR analysis presented in this section is twofold: i) to shed light on the trade-off between interference reduction and maximization of number of degrees of freedom (see also Section 2.2.2.7), and ii) to assess the robustness of the Rayleigh-fading WSSUS underspread model by determining the SNR values at which capacity becomes sensitive to whether $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported or not.

To this end, in the next section, we present a capacity lower bound that is explicit in the parameters $\Delta_{\mathbf{H}}$ and ε and in the WH set $(g(t), T, F)$. We then compare this lower bound to the AWGN-capacity upper bound C_{awgn} . Our main result is the following: we show that for an appropriate choice of the grid-parameter product TF , the lower bound obtained using WH sets based on a root-raised-cosine pulse (see Example 1) is close to C_{awgn} for all SNR values of interest in practical systems, and for all Rayleigh-fading WSSUS channels

that are underspread according to Definition 1, independently of whether the scattering function $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported or not.

2.5.1 A lower bound on capacity

The lower bound we analyze in this section is obtained from the bound L_1^{int} in (2.47) through some additional bounding steps to make the dependency on $\Delta_{\mathbf{H}}$ and ε explicit. More specifically, we use the fact that for any Rayleigh-fading WSSUS channel in the set $\mathcal{H}(\tau_0, \nu_0, \varepsilon)$ (see Definition 1) the following inequalities hold [DMBon]:

$$\begin{aligned}\sigma_h^2 &\geq m_g(1 - \varepsilon) \\ \sigma_l^2 &\leq M_g + \varepsilon.\end{aligned}\tag{2.51}$$

The quantities m_g and M_g are defined in (2.20) and (2.21), respectively. Furthermore, whenever $\tilde{\Delta}_{\mathbf{H}} \triangleq 2\nu_0 T < 1$, we also have [DMBon]

$$\begin{aligned}\frac{1}{T} \int_{-1/2}^{1/2} \ln \det \left(\mathbf{I}_L + \frac{TF\rho_s}{\alpha} \mathbf{P}_h(\theta) \right) d\theta \\ \leq \tilde{\Delta}_{\mathbf{H}} \ln \left(1 + \frac{TF\rho_s}{\alpha \tilde{\Delta}_{\mathbf{H}}} \right) + (1 - \tilde{\Delta}_{\mathbf{H}}) \ln \left(1 + \frac{TF\rho_s \varepsilon}{\alpha(1 - \tilde{\Delta}_{\mathbf{H}})} \right).\end{aligned}$$

The condition $\tilde{\Delta}_{\mathbf{H}} < 1$ is not restrictive for underspread channels if T and F are chosen according to the grid matching rule (2.19). In this case, $\tilde{\Delta}_{\mathbf{H}} = \sqrt{\Delta_{\mathbf{H}} T F} \ll 1$ for all values of TF of practical interest. By using (2.51) in (2.47), we get the following result.

Theorem 7 *Let $(g(t), T, F)$ be a WH set and consider any Rayleigh-fading WSSUS channel in the set $\mathcal{H}(\tau_0, \nu_0, \varepsilon)$ (see Definition 1). Then, for a given bandwidth B and a given SNR $\rho \triangleq P/B$, and under the technical condition $\tilde{\Delta}_{\mathbf{H}} = 2\nu_0 T < 1$, the capacity of the discretized channel (2.14) induced by $(g(t), T, F)$, with input subject to the average-power constraint (2.23) and the peak-power constraint (2.24), is lower-bounded as $C \geq L_2^{\text{int}}$, where*

$$\begin{aligned}L_2^{\text{int}}(\rho) \triangleq \frac{B}{TF} \left[\mathbb{E} \left\{ \ln \left(1 + \frac{TF\rho_s(1 - \varepsilon)m_g|h|^2}{1 + TF\rho_s(M_g + \varepsilon)} \right) \right\} - \left(\ln(\pi e \rho_s) - h(s) \right) \right. \\ \left. - \inf_{0 < \alpha < 1} \left\{ \tilde{\Delta}_{\mathbf{H}} \ln \left(1 + \frac{TF\rho_s}{\alpha \tilde{\Delta}_{\mathbf{H}}} \right) + (1 - \tilde{\Delta}_{\mathbf{H}}) \ln \left(1 + \frac{TF\rho_s \varepsilon}{\alpha(1 - \tilde{\Delta}_{\mathbf{H}})} \right) \right. \right. \\ \left. \left. + \ln \left(1 + \frac{TF\rho_s}{1 - \alpha} (M_g + \varepsilon) \right) \right\} \right].\end{aligned}\tag{2.52}$$

Here, $h \sim \mathcal{CN}(0, 1)$, and s is a complex random variable that satisfies

$$\rho_s \triangleq \mathbb{E}\{|s|^2\} \leq \rho \quad \text{and} \quad |s|^2 \leq \beta \rho.$$

2.5.1.1 Some remarks on the lower bound

The remarks we listed in Sections 2.4.1.4 and 2.4.1.5 about the reduction to a square setting, and about the maximization of L_a^{int} with respect to the choice of the WH set and the probability distribution on s , apply to the bound L_2^{int} as well. As in Section 2.4, we consider only WH sets based on a root-raised-cosine pulse and use for s the truncated complex-Gaussian distribution described in Section 2.4.1.5. Differently from the behavior of L_a^{int} in the large-bandwidth regime, the bound L_2^{int} is hardly sensitive to the value of the penalty term $\ln(\pi e \rho_s) - h(s)$.

For fixed $g(t)$ (in our case, a root-raised-cosine pulse), the lower bound L_2^{int} can be tightened by maximizing it over the product TF of the grid parameters. This provides an information-theoretic criterion for the design of the WH set $(g(t), T, F)$. Compared to the design criterion based on SIR maximization discussed in Section 2.2.2.7, the maximization of the lower bound L_2^{int} yields a more complete picture, because it reveals the influence of the number of degrees of freedom (reflected through the grid-parameter product TF), and of the SIR (reflected through the quantities m_g and M_g) on capacity.

The lower bound L_2^{int} is not useful in the asymptotic regime $\rho \rightarrow \infty$. In fact, the bound even turns negative when ρ is sufficiently large. Nevertheless, as shown in Section 2.5.2 below, when the channel is underspread, the bound L_2^{int} , evaluated for the TF value that maximizes it, is close to the AWGN-capacity upper bound for all SNR values of practical interest.

2.5.2 Numerical examples

In this section, we evaluate the lower bound L_2^{int} in (2.52) for the WH set based on a root-raised-cosine pulse considered in Example 1, under the assumption that the underlying WSSUS channel is underspread according to Definition 1, i.e., $\Delta_{\mathbf{H}} \ll 1$ and $\varepsilon \ll 1$. More precisely, we assume $\Delta_{\mathbf{H}} \leq 10^{-4}$ and $\varepsilon \leq 10^{-4}$, respectively. We also assume that s follows the truncated complex-Gaussian distribution described in Section 2.4.1.5.

Trade-off between interference reduction and maximization of the number of degrees of freedom: In Fig. 2.10, we plot the ratio $L_2^{\text{int}}/C_{\text{awgn}}$ for $\beta = 10$, $\varepsilon = 10^{-6}$, and $\Delta_{\mathbf{H}} = 10^{-4}$ and $\Delta_{\mathbf{H}} = 10^{-6}$. The different curves correspond to different values of TF . We can observe that the choice $TF = 1$ is highly suboptimal. The reason for this suboptimality is the significant reduction in SIR this choice entails (see Fig. 2.3). In fact (as already mentioned), when $TF = 1$, the pulse $g(t)$ reduces to a sinc function, which has poor time localization. A value of TF slightly above 1 leads to a significant improvement in the time localization of $g(t)$ and to a corresponding increase in the lower bound L_2^{int} for all SNR values of practical interest, despite the (small) loss of degrees of freedom. A further increase of the product TF seems to be detrimental for all but very high SNR values, where the ratio $L_2^{\text{int}}/C_{\text{awgn}}$ is much smaller than 1: the rate loss due to the reduction of the number of degrees of freedom is more significant than the rate increase due to the resulting SIR improvement.

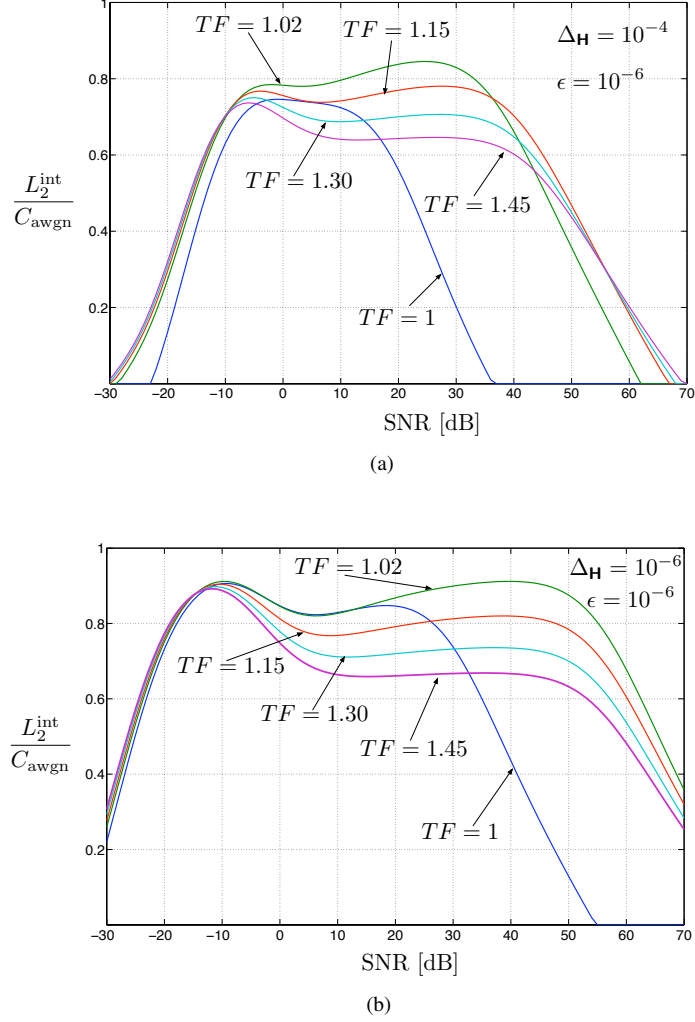


Figure 2.10: Lower bounds L_2^{int} on the capacity of the channel (2.14) normalized with respect to the AWGN capacity. The bounds are computed for WH sets based on a root-raised-cosine pulse (see Example 1), for different values of the grid-parameter product TF . The channel spread $\Delta_{\mathbf{H}}$ (see Definition 1) is equal to 10^{-4} in (a) and equal to 10^{-6} in (b); furthermore, in both cases $\epsilon = 10^{-6}$ and $\beta = 10$.

Sensitivity to the PAPR β : Fig. 2.11 shows the ratio $L_2^{\text{int}}/C_{\text{awgn}}$ for $TF = 1.02$, $\Delta_{\mathbf{H}} = \epsilon = 10^{-6}$, and different values of β . We can observe that in the high-SNR regime, L_2^{int} is only mildly sensitive to the value of the penalty term $\ln(\pi e \rho_s) - h(s)$ (recall that the value of the penalty term decreases as β increases).

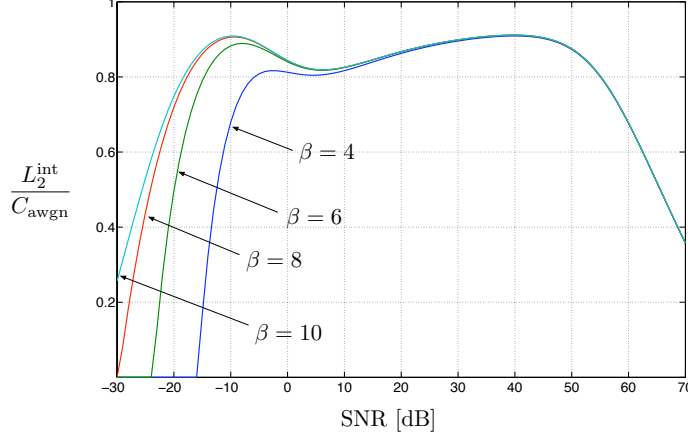


Figure 2.11: Lower bounds L_2^{int} on the capacity of the channel (2.14) normalized with respect to the AWGN capacity. The lower bounds are computed for WH sets based on a root-raised-cosine pulse (see Example 1), for different values of the PAPR β . The channel spread $\Delta_{\mathbf{H}}$ (see Definition 1) is equal to 10^{-6} ; furthermore, $\varepsilon = 10^{-6}$ and $TF = 1.02$.

Sensitivity of capacity to the measure of the support of $C_{\mathbf{H}}(\tau, \nu)$: The results presented in Fig. 2.10 suggest that, for $TF = 1.02$, the lower bound L_2^{int} is close to the AWGN-capacity upper bound C_{awgn} over a quite large range of SNR values. To make this statement precise, we compute the SNR interval $[\rho_{\min}, \rho_{\max}]$ over which

$$L_2^{\text{int}} \geq 0.75 C_{\text{awgn}}. \quad (2.53)$$

The interval end points ρ_{\min} and ρ_{\max} can easily be computed numerically; the corresponding values for ρ_{\min} and ρ_{\max} are illustrated in Figs. 2.12 and 2.13, respectively, for different $(\Delta_{\mathbf{H}}, \varepsilon)$ pairs. For the WH set and WSSUS underspread channels considered in this section, we have $\rho_{\min} \in [-25 \text{ dB}, -7 \text{ dB}]$ and $\rho_{\max} \in [32 \text{ dB}, 68 \text{ dB}]$.

Insights into how ρ_{\min} and ρ_{\max} are related to the channel parameters $\Delta_{\mathbf{H}}$ and ε can be obtained by replacing both sides of the inequality (2.53) by corresponding low-SNR approximations (to get ρ_{\min}) and high-SNR approximations (to get ρ_{\max}). Under the assumption that $\Delta_{\mathbf{H}} \leq 10^{-4}$ and $\varepsilon \leq 10^{-4}$, this analysis, detailed in [DMB0n], yields $\rho_{\min} \approx 13\sqrt{\Delta_{\mathbf{H}}}$ and $\rho_{\max} \approx 0.22/(\Delta_{\mathbf{H}} + \varepsilon)$ for the WH set considered in this section. The following rule of thumb then holds: the capacity of all WSSUS underspread channels with scattering function $C_{\mathbf{H}}(\tau, \nu)$ having no more than ε of its volume outside a rectangle (in the delay-Doppler plane) of area $\Delta_{\mathbf{H}}$, is close to C_{awgn} for all SNR values ρ that satisfy $\sqrt{\Delta_{\mathbf{H}}} \ll \rho \ll 1/(\Delta_{\mathbf{H}} + \varepsilon)$, independently of whether $C_{\mathbf{H}}(\tau, \nu)$ is compactly supported or not, and independently of the shape of $C_{\mathbf{H}}(\tau, \nu)$. To conclude, we stress that the condition $\sqrt{\Delta_{\mathbf{H}}} \ll \rho \ll 1/(\Delta_{\mathbf{H}} + \varepsilon)$ holds for all channels and SNR values of practical interest.

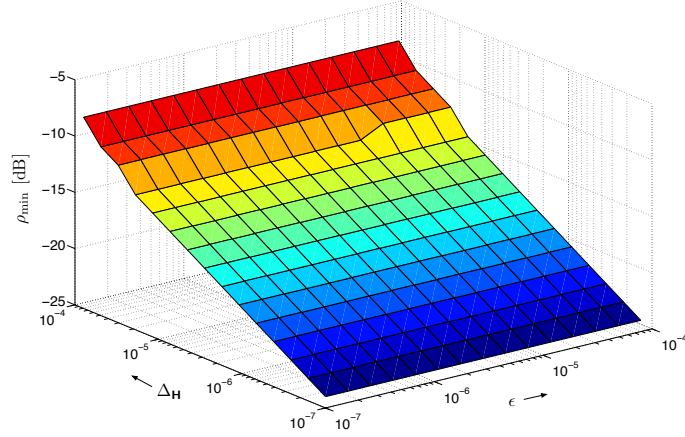


Figure 2.12: Minimum SNR value ρ_{\min} for which (2.53) holds, as a function of $\Delta_{\mathbf{H}}$ and ϵ . The lower bound L_2^{int} is computed for a WH set based on a root-raised-cosine pulse (see Example 1); furthermore, $TF = 1.02$, and $\beta = 10$.

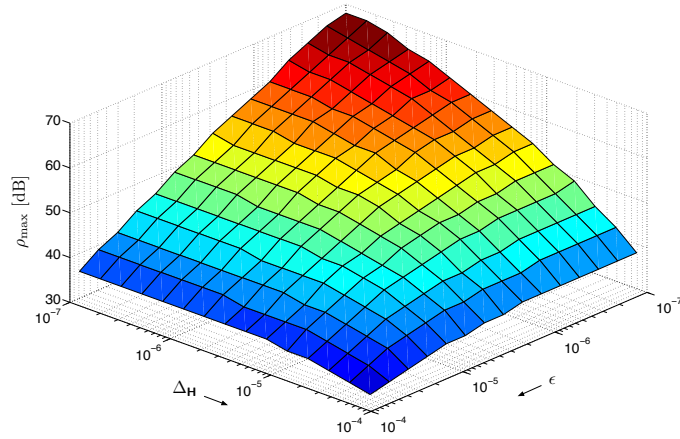


Figure 2.13: Maximum SNR value ρ_{\max} for which (2.53) holds, as a function of $\Delta_{\mathbf{H}}$ and ϵ . The lower bound L_2^{int} is evaluated for a WH set based on a root-raised-cosine pulse (see Example 1); furthermore, $TF = 1.02$, and $\beta = 10$.

2.6 Conclusions

In this chapter, we provided an information-theoretic characterization of Rayleigh-fading channels that satisfy the WSSUS and the underspread assumptions. The information-

theoretic analysis is built upon a discretization of WSSUS underspread channels that takes the underspread property explicitly into account to minimize the ISI/ICI in the discretized I/O relation. The channel discretization is accomplished by transmitting and receiving on a WH set generated by a pulse that is well localized in time and frequency.

We obtained bounds on the noncoherent capacity under both an average and a peak-power constraint, and used these bounds to study engineering-relevant questions.

For the large-bandwidth regime, we provided upper and lower bounds on the noncoherent capacity on the basis of a diagonalized I/O relation that closely approximates the underlying continuous-time I/O relation. These upper and lower bounds provide important guidelines for the design of wireless communication systems operating over a Rayleigh-fading WSSUS underspread channel with known scattering function. Even if the scattering function is not known completely and the channel is only characterized coarsely by its maximum delay τ_0 and maximum Doppler shift ν_0 , the bounds may serve as an efficient design tool. The maximum delay and the maximum Doppler shift are sufficient to characterize a brick-shaped scattering function, for which our upper and lower bounds are easy to evaluate. Furthermore, a brick-shaped scattering function was shown to minimize the capacity upper bound for a given τ_0 and ν_0 . Therefore, the widely accepted practice to characterize wireless channels simply by means of these two parameters leads to robust design guidelines.

The numerical results in Section 2.3.6 indicate that the upper and lower bounds are surprisingly close over a large range of bandwidth values. In particular, our lower bound is close to the coherent-capacity upper bound in Fig. 2.4 for bandwidth values of up to 1 GHz. It is exactly this regime that is of most interest for current wideband and UWB communication systems; this is also the regime for which we can expect the WSSUS model to be most accurate.

The advent of UWB communication systems spurred the current interest in wireless communications over channels with very large bandwidth. Our bounds make it possible to assess whether multiple antennas at the transmitter are beneficial for UWB systems. The system parameters used to numerically evaluate the bounds in Section 2.3.8 are compatible with a UWB system that operates over a bandwidth of 7 GHz and transmits at -41.3 dBm/MHz. Figs. 2.5, 2.6, and 2.7 show that the maximum rate increase that can be expected from the use of multiple antennas at the transmitter does not exceed 7%. For channels with smaller spreads than the one in Section 2.3.6, the possible rate increase is even smaller.

For the high-SNR regime, we provide a capacity lower bound that is obtained by treating interference as noise. This lower bound yields valuable insights into the capacity of Rayleigh-fading underspread WSSUS channels over a large range of SNR values of practical interest. On the basis of this lower bound, we derived an information-theoretic criterion for the design of capacity-optimal WH sets. This criterion is more fundamental than criteria based on SIR maximization (see Section 2.2.2.7), because it sheds light on the trade-off between number of degrees of freedom and time-frequency localization of the pulse $g(t)$. Unfortunately, the corresponding optimization problem is hard to solve. We simplified the problem by fixing $g(t)$ to be a root-raised-cosine pulse and performing an optimization over the grid parameters T and F . Our analysis shows that the optimal

value of the grid-parameter product TF is close to 1 (but strictly larger than 1) for a large range of SNR values of practical interest. This result suggests that the maximization of the number of degrees of freedom should be privileged over the SIR maximization in the design of capacity-maximizing WH sets.

Even though our analysis was confined to a specific pulse shape (i.e., root-raised-cosine), we were able to show that for all Rayleigh-fading WSSUS channels that are underspread according to Definition 1, the corresponding capacity lower bound is close to the AWGN-capacity upper bound for all SNR values of practical interest, independently of whether the scattering function is compactly supported or not (a fine detail of the channel model). In other words, the capacity of Rayleigh-fading underspread WSSUS channels starts being sensitive to this fine detail of the channel model only for SNR values that lie outside the SNR range typically encountered in real-world systems. Hence, the Rayleigh-fading WSSUS underspread model is a robust model.

References

- [AFTS01] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz). The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE Trans. Inf. Theory*, 47(4):1290–1301, May 2001.
- [AMH04] H. Artés, G. Matz, and F. Hlawatsch. Unbiased scattering function estimators for underspread channels and extension to data-driven operation. *IEEE Trans. Signal Process.*, 52(5):1387–1402, May 2004.
- [Bel63] P. Bello. Characterization of randomly time-variant linear channels. *IEEE Trans. Commun.*, 11(4):360–393, December 1963.
- [BJ00] H. Bölcskei and A. J. E. M. Janssen. Gabor frames, unimodularity, and window decay. *J. Fourier Anal. Appl.*, 6(3):255–276, May 2000.
- [BPS98] E. Biglieri, J. Proakis, and S. Shamai (Shitz). Fading channels: Information-theoretic and communications aspects. *IEEE Trans. Inf. Theory*, 44(6):2619–2692, October 1998.
- [Chr03] O. Christensen. *An Introduction to Frames and Riesz Bases*. Birkhäuser, Boston, MA, U.S.A., 2003.
- [Cox73a] D. C. Cox. 910 MHz urban mobile radio propagation: Multipath characteristics in New York City. *IEEE Trans. Commun.*, 21(11):1188–1194, November 1973.
- [Cox73b] D. C. Cox. A measured delay-Doppler scattering function for multipath propagation at 910 MHz in an urban mobile radio environment. *Proc. IEEE*, 61(4):479–480, April 1973.
- [CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, NY, U.S.A., 1991.
- [CTKV02] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela. Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Trans. Inf. Theory*, 48(3):637–650, March 2002.
- [DLL95] I. Daubechies, H. Landau, and Z. Landau. Gabor time-frequency lattices and the Wexler-Raz identity. *J. Fourier Anal. Appl.*, 1(4):437–478, 1995.
- [DMB09] G. Durisi, V. I. Morgenshtern, and H. Bölcskei. On the sensitivity of noncoherent capacity to the channel model. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 2174–2178, Seoul, Korea, June 2009.

- [DMBon] G. Durisi, V. I. Morgenshtern, and H. Bölcskei. On the information-theoretic robustness of the WSSUS underspread channel model. In preparation.
- [DS63] N. Dunford and J. T. Schwarz. *Linear Operators*, volume 2. Wiley, New York, NY, U.S.A., 1963.
- [DSBS10] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz). Noncoherent capacity of underspread fading channels. *IEEE Trans. Inf. Theory*, 56(1):367–395, January 2010.
- [Gal68] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, NY, U.S.A., 1968.
- [Gra05] R. M. Gray. Toeplitz and circulant matrices: A review. In *Foundations and Trends in Communications and Information Theory*, volume 2. now Publishers, Delft, The Netherlands, 2005.
- [Grö01] K. Gröchenig. *Foundations of Time-Frequency Analysis*. Birkhäuser, Boston, MA, U.S.A., 2001.
- [GSV05] D. Guo, S. Shamai (Shitz), and S. Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inf. Theory*, 51(4):1261–1282, April 2005.
- [Has93] H. Hashemi. The indoor radio propagation channel. *Proc. IEEE*, 81(7):943–968, July 1993.
- [Jak74] W. C. Jakes, editor. *Microwave Mobile Communications*. Wiley, New York, NY, U.S.A., 1974.
- [Jan95] A. J. E. M. Janssen. Duality and biorthogonality for Weyl-Heisenberg frames. *J. Fourier Anal. Appl.*, 1(4):403–437, 1995.
- [Ken69] R. S. Kennedy. *Fading Dispersive Communication Channels*. Wiley, New York, NY, U.S.A., 1969.
- [KM98] W. Kozek and A. F. Molisch. Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels. *IEEE J. Sel. Areas Commun.*, 16(8):1579–1589, October 1998.
- [Koz97] W. Kozek. *Matched Weyl-Heisenberg Expansions of Nonstationary Environments*. PhD thesis, Vienna University of Technology, Department of Electrical Engineering, Vienna, Austria, March 1997.
- [KSP⁺02] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen. A stochastic MIMO radio channel model with experimental validation. *IEEE J. Sel. Areas Commun.*, 20(6):1211–1226, August 2002.
- [Lap05] A. Lapidoth. On the asymptotic capacity of stationary Gaussian fading channels. *IEEE Trans. Inf. Theory*, 51(2):437–446, February 2005.

- [LM03] A. Lapidoth and S. M. Moser. Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels. *IEEE Trans. Inf. Theory*, 49(10):2426–2467, October 2003.
- [LV04] Y. Liang and V. V. Veeravalli. Capacity of noncoherent time-selective Rayleigh-fading channels. *IEEE Trans. Inf. Theory*, 50(12):3096–3110, December 2004.
- [Méd00] M. Médard. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Trans. Inf. Theory*, 46(3):933–946, May 2000.
- [MG02] M. Médard and R. G. Gallager. Bandwidth scaling for fading multipath channels. *IEEE Trans. Inf. Theory*, 48(4):840–852, April 2002.
- [MH98] 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.*, 39(8):4041–4070, August 1998.
- [MH99] T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE Trans. Inf. Theory*, 45(1):139–157, January 1999.
- [MH03] G. Matz and F. Hlawatsch. Time-frequency characterization of randomly time-varying channels. In B. Boashash, editor, *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, chapter 9.5, pages 410–419. Elsevier, Oxford, U.K., 2003.
- [MSG⁺07] 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.*, 6(5):1921–1931, May 2007.
- [NS82] A. W. Naylor and G. R. Sell. *Linear Operator Theory in Engineering and Science*. Springer, New York, NY, U.S.A., 1982.
- [Par00] J. D. Parsons. *The Mobile Radio Propagation Channel*. Wiley, Chichester, U.K., second edition, 2000.
- [Pie66] J. R. Pierce. Ultimate performance of M -ary transmission on fading channels. *IEEE Trans. Inf. Theory*, 12(1):2–5, January 1966.
- [Poo94] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer, New York, NY, U.S.A., second edition, 1994.
- [Pro01] J. G. Proakis. *Digital Communications*. McGraw-Hill, New York, NY, U.S.A., fourth edition, 2001.

- [PTN07] D. Porrat, D. N. C. Tse, and Ş. Nacu. Channel uncertainty in ultra-wideband communication systems. *IEEE Trans. Inf. Theory*, 53(1):194–208, January 2007.
- [Rap02] T. S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, Upper Saddle River, NJ, U.S.A., second edition, 2002.
- [Raz98] B. Razavi. *RF Microelectronics*. Prentice Hall, Upper Saddle River, NJ, U.S.A., 1998.
- [RHS07] V. Raghavan, G. Hariharan, and A. M. Sayeed. Capacity of sparse multipath channels in the ultra-wideband regime. *IEEE J. Sel. Topics Signal Process.*, 1(3):357–371, October 2007.
- [RS97] A. Ron and Z. Shen. Weyl-Heisenberg frames and Riesz bases in $L_2(\mathbb{R}^d)$. *Duke Math J.*, 89(2):237–282, 1997.
- [SB07] U. G. Schuster and H. Bölcskei. Ultrawideband channel modeling on the basis of information-theoretic criteria. *IEEE Trans. Wireless Commun.*, 6(7):2464–2475, July 2007.
- [Sch09] U. G. Schuster. *Wireless Communication over Wideband Channels*, volume 4 of *Series in Communication Theory*. Hartung-Gorre Verlag, Konstanz, Germany, 2009.
- [SDBP09] U. G. Schuster, G. Durisi, H. Bölcskei, and H. V. Poor. Capacity bounds for peak-constrained multiantenna wideband channels. *IEEE Trans. Commun.*, 57(9):2686–2696, September 2009.
- [SH02] V. G. Subramanian and B. Hajek. Broad-band fading channels: Signal burstiness and capacity. *IEEE Trans. Inf. Theory*, 48(4):809–827, April 2002.
- [SH05] V. Sethuraman and B. Hajek. Capacity per unit energy of fading channels with peak constraint. *IEEE Trans. Inf. Theory*, 51(9):3102–3120, September 2005.
- [SHN05] V. Sethuraman, B. Hajek, and K. Narayanan. Capacity bounds for noncoherent fading channels with a peak constraint. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 515–519, Adelaide, Australia, September 2005.
- [Sle76] D. Slepian. On bandwidth. *Proc. IEEE*, 64(3):292–300, March 1976.
- [SWHL09] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidath. Low SNR capacity of noncoherent fading channels. *IEEE Trans. Inf. Theory*, 55(4):1555–1574, April 2009.
- [TE97] G. Taricco and M. Elia. Capacity of fading channel with no side information. *Electron. Lett.*, 33(16):1368–1370, July 1997.

- [Tel99] Í. E. Telatar. Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecommun.*, 10:585–595, November 1999.
- [TT00] Í. E. Telatar and D. N. C. Tse. Capacity and mutual information of wideband multipath fading channels. *IEEE Trans. Inf. Theory*, 46(4):1384–1400, July 2000.
- [TV05] D. N. C. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge Univ. Press, Cambridge, U.K., 2005.
- [VB03] R. Vaughan and J. Bach Andersen. *Channels, Propagation and Antennas for Mobile Communications*. The Institution of Electrical Engineers, London, U.K., 2003.
- [Ver02] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Trans. Inf. Theory*, 48(6):1319–1343, June 2002.
- [Vit67] A. J. Viterbi. Performance of an M-ary orthogonal communication system using stationary stochastic signals. *IEEE Trans. Inf. Theory*, 13(3):414–422, July 1967.
- [Wyn66] A. D. Wyner. The capacity of the band-limited Gaussian channel. *Bell Syst. Tech. J.*, 45(3):359–395, March 1966.
- [ZT02] 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*, 48(2):359–383, February 2002.